The Reliability of Output Gap Estimates in Real Time

Athanasios Orphanides*
Board of Governors of the Federal Reserve System

Simon van Norden
École des Hautes Études Commerciales, Montréal

August 1999

Abstract

Compared to its central role in policy discussions in the United States and most other developed countries, the reliability of the measurement of the output gap has attracted relatively little academic study. Furthermore, both the academic literature and the debate among practitioners have tended to neglect a key factor. Although in a policy setting, it is necessary to estimate the current (i.e. end-of-sample) output gap without the benefit of knowing the future, most studies concentrate on measurement that employs data that only become available later. In this paper we examine the reliability of alternative output detrending methods, with special attention to the accuracy of real-time estimates. We show that ex post revisions of the output gap are of the same order of magnitude as the output gap itself, that these ex post revisions are highly persistent and that real-time estimates tend to be severely biased around business cycle turning points, when the cost of policy induced errors due to incorrect measurement is at its greatest. We investigate the reasons for these ex post revisions, and find that, although important, the ex post revision of published data is not the primary source of revisions in output gap measurements. The bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the trend in output.

KEYWORDS: Real-time data, output gap, business cycle measurement.

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

* We would like to thank Bryan Campbell and Andy Filardo for their comments, as well as seminar participants at the Atelier économetrique de Montréal, the University of Ottawa, Optimization Days and the meetings of the Canadian Economics Association. Prof. van Norden would also like to thank the SSHRC for their support. The opinions expressed are those of the authors and do not necessarily reflect views of the Board of Governors of the Federal Reserve System.

1 Introduction

One of the fundamental issues in macroeconomics is understanding macroeconomic fluctuations. At the most aggregate level this entails the study of an economy's output relative to its potential level. Understanding whether the economy is operating at its full potential, however, presupposes accurate measurement of both actual output as well as potential output. The difference between the two is commonly referred to as the business cycle or the output gap. Although macroeconomic analysis often takes the availability of such measures for granted, considerable uncertainty surrounds them in practice.

The issue is of some importance for empirical macroeconomics since testing and comparison of alternative models can be easily obscured by inaccurate measurements. Bluntly, to evaluate whether a specific theory or model can provide an adequate accounting of macroeconomic fluctuations we must first measure the fluctuations that are to be accounted for.

The problem is especially acute for economic policy. While academic investigations can afford the luxury of waiting for the accumulation of accurate historical data before estimates of past actual and potential output need to be constructed, policy decisions require such estimates in real-time and policy actions based on incorrect real-time estimates may inadvertently contribute to undesirable macroeconomic outcomes.

For fiscal policy, it is often useful to abstract from cyclical influences to assess whether policy is expansionary or contractionary and also to evaluate the path of government expenditures and finance. The resulting "full employment" budget estimates, however, squarely rest on accurate assessments of the economy's performance relative to potential.

Uncertainty regarding the measurement of the business cycle arguably presents a bigger problem for monetary policy. A central bank can influence credit conditions and consequently aggregate demand via its monetary policy instrument. This potentially allows monetary policy to dampen aggregate demand fluctuations and, when necessary, counteract inflationary pressures. However, since such policy actions affect aggregate demand and inflation with a lag, timely measures and forecasts of the output gap are essential. Obviously, unless the economy's potential can be reliably measured, policy choices may fail to react to the true underlying economic conditions and may instead partially reflect measurement error.

Three distinct issues complicate assessment of the economy's performance relative to its potential in real-time. First, output data (and other officially published macroeconomic time series) are continually revised in response to more complete reporting, adjustment of seasonal factors, refinements in concept or methodology, etc. This implies that measures of the output gap available in real-time may differ from those constructed from data published many years later. Second, most methods for estimating potential provide different estimates of potential output for a given quarter if data on actual output in years following the relevant quarter are made available. This may be because hindsight makes clearer which part of the business cycle the economy was in at a particular point in time, even if our beliefs about the processes driving output growth do not change. In this way, the passage of time may allow better estimates of a specific quarter's output gap to be made ex post, even if no revisions are made to actual output data. Third, the subsequent evolution of output may indicate that the economy has undergone a structural change. This in turn may lead to a change in our beliefs about the economy and the expected evolution of potential output. It may also cause us to revise our beliefs about potential output and the output gap in the period prior to our becoming convinced that a structural change had taken place.

This paper investigates the quantitative relevance of these issues for the measurement of the output gap in the United States over the last thirty years. We investigate several well-known methods for estimating the output gap. For each method, we examine the behaviour of end-of-sample output gap estimates and of the revisions of these estimates over time. Specifically, we calculate the statistical propoerties of the revisions and decompose them into their various sources, including the component due to revisions of the underlying output data and that due to re-estimation of the process generating potential output. We then compare the revision behaviour of the alternative methods.

In the present paper, we restrict our attention to univariate methods for estimation of the output gap. To conduct a thorough analysis based on multivariate techniques would require compilation of unrevised data series for all variables involved and would also introduce additional conceptual issues. Briefly, by utilizing information from additional sources, multivariate techniques may reduce the errors associated with the end-of-sample estimates of the output trend from univariate methods. However, multivariate techniques also introduce additional sources of mispecification and parameter estimation problems which may more than offset the potential improvement these methods offer. By concentrating on the univariate methods we provide a benchmark against which these additional issues can be examined. To help assess the pertinence of our results, we also provide a brief comparison of our real time univariate estimates of the output gap to "official" real time output gap estimates as constructed in the United States from the mid 1960s.

The potential quantitative relevance of the issues we investigate has been pointed out before. Using final data, Kuttner (1994) and St-Amant and van Norden (1998) pointed out that differences between end-sample and mid-sample estimates of the output gap can differ substantially for some commonly used methods for estimating the output gap, such as unobserved component and smoothing spline methods. Orphanides (1997,1999) documented that the errors in "official" estimates of the output gap available to policymakers have indeed been substantial and several authors, including Kuttner (1992), McCallum and Nelson (1998), Orphanides (1998) and Smets (1998) have elaborated on the policy implications of this issue. As far as we know this study is the first attempt at comprehensive measurement and evaluation of the measurement errors associated with various techniques based on real-time data for the past thirty years.

2 Data Sources, Revisions and Concepts

2.1 How to measure the reliability of measured output gaps

Our aim in this study is to understand better the reliability and statistical accuracy of commonly used estimates of output gaps. While there are many approaches to measuring their reliability and accuracy, none is without limitations.

One way would be to generate artificial output data from an economic model which would then be detrended by the various different methods under study. The different estimates of the true output gap could then be compared to the known output gaps from the economic model. The problem with such an approach is that results will in general depend on the specification of the economic model and a wide range of specifications could reasonably be considered plausible.[1] Furthermore, it would ignore the uncertainty introduced by the ongoing revision of published data.

Another way would be to simply use the statistical uncertainty associated with our estimate of potential or trend output to put confidence intervals around these estimates and therefore around our calculated output gaps. Unfortunately, some popular methods (such as the HP filter) do not give statistical confidence intervals. Furthermore, this method implicitly assumes that the statistical model is not misspecified, an assumption which often appears to be at odds with the evidence. Finally, this too would ignore the effects of data revision.

A third way would be to specify a particular measure of the value of output gap measures. For example, if the goal of measuring output gaps is to aid policy makers in controlling inflation, we might seek to measure the marginal forecasting power of output gaps for subsequent inflation. Again, ambiguity about the goal of measuring output gaps implies that different criteria might reasonably be used and could give varying results for any particular method.[2] More seriously, such a methodology would have to address the special

[1] For an example of such sensitivity analysis, see Guay and St-Amant (1996).

[2] This is not a criticism of the methodology, of course. This simply reflects the fact that some measures may be better suited for some purposes than for others.

problems posed by the Lucas Critique.[3]

The alternative approach which we use in this paper allows us to capture the effects of errors due both to data revision and to misspecification of statistical models used to estimate output gaps. At the same time, it is simple to implement and does not require a priori assumptions on the true structure of the economy or on the time-series model generating observed output. We explain this method in detail below. Briefly, it consists of measuring the degree to which estimates of the output gap at any point in time vary as data are revised and as data about the subsequent evolution of output becomes available.

To be sure, this method is not without its own limitations. We measure only revisions in estimates of the output gap. However, it is reasonable to assume that some uncertainty remains with long-past historical estimates of the output gap. Since the total amount of uncertainty at the end of a sample is presumably the sum of the uncertainty from these two sources, this approach gives us an overestimate of the precision and accuracy associated with any detrending method. This has implications for the way we can interpret our results. A finding that revision errors are small might not be very meaningful, since it would not necessarily imply that the remaining "unrevised" errors are small. Similarly, it would be naive to attempt to rank different methods on the basis of the size of their revisions. It is nonetheless informative when and if we find that revision errors are relatively large, since we can conclude that the total error of these estimators must be larger still.

2.2 Data

Most of our data is taken from the real-time data set compiled by Croushore and Stark (1999). From their database we use the real-time variables for real output from 1965 to 1997. In each quarter, these time series reflect real output as published during that quarter by the Department of Commerce. The latest observation is always the one corresponding

[3] To see this, consider the case where we directly observe the true output gap. If the gap is unknown to the monetary policy authority, it will presumably have some forecasting power for inflation. However, if this information is available to the policy maker and is used efficiently in setting policy, then it will appear to have no forecasting power if there have been offsetting adjustments in monetary policy.

to the previous quarter.[4] The data are seasonally adjusted, and therefore alternative data vintages reflect, among other changes, re-estimation of seasonal factors. The concept of real output has also evolved over time. In the U.S. the benchmark series was GNP until the end of 1991 and GDP since then. In addition, changes also reflect the choice of deflator. Until the end of 1995 real output was measured in constant dollars with the benchmark year changing once or twice in every decade. Since then a chain-weighted deflator is being used.

We use 1999:Q1 data as "final data" recognizing, of course, that "final" is very much an ephemeral concept in the measurement of output.

Even when the output concept and deflator are same, first released output data differ significantly from subsequent releases. The biggest revisions are in the first few quarters after the release. However, once a year a major revision is made and seasonals adjusted with changes that are, at times, substantial for the few most recent years.

2.3 Measuring the revision of output gaps

We use the data set mentioned above with a variety of detrending methods (described in the next section) to produce many different estimated output gap series. However, we also apply each of these detrending methods in a number of different ways in order to estimate and decompose the extent of the revisions in the estimated gap series. To understand how the extent of the revisions is measured, we define several conceptually different ways in which any existing detrending method may be applied. In the remainder of this section, we describe how these methods were applied and their corresponding interpretations.[5]

[4]The Croushore and Stark database samples information in the middle of every quarter. As a result on a few occasions when the data were released later than usual the real output data for the previous quarter are not available. To avoid missing observations we supplemented the data with information published towards the end of the quarter on those occasions using the first *Survey of Current Business* issue where information for the previous quarter was reported.

[5]A more technical description of the methods we used may be found in the Appendix.

2.3.1 Final Estimates

The first of these methods gives rise to a "Final" estimate of the output gap. This simply takes the last available vintage of data we have available (in our case, this is the series as published in 1999Q1) and detrends it. The resulting series of deviations from trend constitutes the "Final" estimate of the output gap. This is the typical way in which such detrending methods are employed.

2.3.2 RealTime Estimates

The "RealTime" estimate of the output gap is constructed in two stages. First, we detrend each and every vintage of data available to construct an ensemble of output gap series. Of course, earlier vintage output gap series are shorter than later vintages since the output series on which they are based end earlier. Next, we use these different vintages to construct a new series which consists entirely of the first available estimate of the output gap for each point in time.

This new series is the "RealTime" estimate of the output gap. It represents the most timely estimate of the output gap which policy makers could have constructed at any point in time. The difference between the RealTime and the Final estimate give us the total revision in the estimated output gap at each point in time. We use the statistical properties of these revisions as our guide to reliability and accuracy of estimated output gaps recalling, of course, that this is an overestimate of the true reliability of the RealTime estimates since it ignores the estimation error in the final series.

2.3.3 QuasiReal Estimates

The differences between the RealTime and the Final estimates have several sources, one of which is the ongoing revision of published data. To isolate the importance of this factor, we define a third output gap measure, the "QuasiReal" estimate. Like the RealTime estimate, the QuasiReal estimate is constructed in two steps.

The first step is to construct an ensemble of "rolling" estimates of the output gap. That is, we begin by taking the Final vintage of the output series but use only the observations up to and including 1966:Q1 in order to compute the QuasiReal estimate for 1966:Q1. Next, we extend the sample period by one observation and repeat the detrending. We continue in this way until we have used the full sample period for the Final output series and we have a full set of corresponding output gap series.

The second step is the same as that used to construct the RealTime series; we collect the first available estimate of the output gap at each point in time from the various series we constructed in step one. This sequence of output gaps is the QuasiReal series. The difference between the RealTime and the QuasiReal series is entirely due to the effects of data revision, since estimates in the two series at any particular point in time are based on data samples covering exactly the same time period.

2.3.4 QuasiFinal Estimates

For unobserved component (UC) models, we are able to further decompose the revision in the estimated gap by defining another estimate of the output gap. This QuasiFinal estimate uses more information than the QuasiReal estimate (which uses subsamples of Final data) but less than the Final estimate (which uses the full sample of Final data.) This is relevant because UC models use the data in two distinct phases. First, they use the available data sample to estimate the parameters of a time-series model of output. Next, they use these estimated parameters in the Kalman filter to arrive at estimates of the output gap. However, they distinguish between "filtered" and "smoothed" estimates of the output gap. The smoothed estimate uses the full sample parameter estimates and data from 1 to T to form an optimal estimate of the gap in quarter t ($1 \leq t \leq T$). However, the filtered estimate uses only data from 1 to t with the full sample parameter estimates to make an optimal estimate of the output gap at t.

For this class of models, smoothed estimates of the output gap are used to construct the Final series, while filtered estimates are used for the QuasiFinal series.[6] The difference

[6]In both cases, the UC model's parameters are estimated using the full sample of the Final vintage data,

between the QuasiFinal and the QuasiReal series then reflect solely the effects of using different parameter estimates for the model to filter the data (i.e. the full-sample ones versus the partial sample ones). The extent of the difference will reflect the importance of parameter instability in the underlying UC model. The difference between the QuasiReal and the RealTime series reflects the importance of ex post information in estimating the output gap given the parameter values of the process generating output.[7]

3 Alternative Detrending Methods

Having explained how we will measure the precision and reliability of different detrending methods, we now briefly review a variety of detrending methods.

We consider four types of methods. They are:

1. Deterministic Trends.
2. The Hodrick Prescott Filter
3. The Beveridge Nelson Decomposition
4. Unobserved Component Models.

Next we briefly discuss each of these four groups and the variants of these methods which we apply. Readers familiar with these detrending methods may wish to just skim this section and pass rapidly onto section 4, where we discuss our results.

3.1 Deterministic Trends

The first set of detrending methods we consider assume that the trend in (the logarithm of) output is well approximated as a simple deterministic function of time. We consider three such functions; linear, quadratic, and piece-wise linear functions.

---

and the same data is then used for filtering and smoothing.

[7] St-Amant and van Norden (1998) argue that the degree to which the subsequent behaviour of output is informative about the output gap is linked to presence or absence of hysteresis in output.

The linear trend is the oldest and simplest of these models. It assumes that output may be decomposed into a cyclical component and a linear function of time

$$y_t = \alpha + \beta \cdot t + c_t \tag{1}$$

where c_t is the business cycle and y_t is our chosen measure of output (in logarithms). The quadratic trend adds a second term in the deterministic component:

$$y_t = \alpha + \beta \cdot t + \gamma \cdot t^2 + c_t \tag{2}$$

This allows the flexibility to detect a slowly changing trend in a simple way. Because of the noticeable downturn in GDP growth after 1973, another simple deterministic technique is a breaking linear trend that allows for the slowdown in that year. In general, the breaking trend model can be written as:

$$y_t = \alpha + \beta \cdot t + c_t \quad \text{for } t \leq t_1$$

$$y_t = \alpha + \beta \cdot t + \gamma \cdot (t - t_1) + c_t \quad \text{for } t > t_1 \tag{3}$$

Breaking trends were first formally studied by Perron (1989), who allowed also for multiple breaks in the trend.

Our implementation of the breaking trend method will incorporate the assumption that the location of the break is fixed and known. Specifically we assume that a break in the trend at the end of 1973 would have been incorporated in real time from 1977 on. This conforms with the debate regarding the productivity slowdown during the 1970s and evidence (e.g. Council of Economic Advisers, 1977) that it would not have been reasonable to introduce the 1973 break earlier but would be appropriate to do so as early as 1977.[8]

Due to their simplicity, deterministic trends remain appealing. Some authors use deterministic trend methods, particularly when simplicity is greatly valued as in some applications regarding monetary policy evaluation. For example, Taylor (1993) relied on deviations

[8]We also investigated alternatives, including ones with a break of unknown location and also the possibility of multiple breaks. For compactness we only report the fixed break in 1973 case since this method is more common for practical applications, especially ones relating to productivity and output. Qualitatively, the results were similar for the other alternatives.

from a linear trend to measure the cycle, and Clarida, Gali and Gertler (1998) employed a quadratic trend. The use of deterministic trends, however, remains a matter of controversy. Nelson and Plosser's (1982) seminal critique of the adequacy of deterministic trend model, has sparked fully two decades of research and debate. To briefly summarize a vast and still unsettled literature, there is still no consensus on the adequacy of the model, with at least some recent papers disputing Nelson and Plosser's claim that output was better modeled as containing a stochastic rather than a deterministic trend.[9] However, the possibility that output contained a unit root (and possibly more than one) suggested a variety of other detrending methods which we consider next.

3.2 The Hodrick Prescott Filter and Smoothing Splines

In recent years, smoothing splines have frequently been used to detrend output and other time series. The most popular of these is that proposed by Hodrick and Prescott (1997) and is commonly called the HP filter.[10] The HP filter decomposes a time series y_t into an additive cyclical component, y_t^c, and a growth component y_t^g,

$$y_t = y_t^c + y_t^g \tag{4}$$

and then chooses the series $\{y_t^g\}$ to minimize the variance of the cyclical component y_t^c subject to a penalty for the variation in the second difference of the growth component y_t^g. Formally, the HP-filtered trend is given by

$$\{y_t^g\}_{t=0}^{T+1} = \text{argmin} \sum_{t=1}^{T} \{(y_t - y_t^g)^2 + \lambda[(y_{t+1}^g - y_t^g) - (y_t^g - y_{t-1}^g)]^2\} \tag{5}$$

and y_t^c is the resulting measure of the output gap. λ is called the "smoothness parameter" and penalizes the variability in the growth component. The larger the value of λ, the smoother the growth component and the greater the variability of the output gap. As λ approaches infinity, the growth component corresponds to a linear time trend. For quarterly data, Hodrick and Prescott propose setting λ equal to 1600.

[9] For example, see Rudebusch (1993), Rothman (1997), Cheung and Chinn (1997).

[10] The method was proposed by Hodrick and Prescott in their influential 1981 working paper. The development of smoothing splines dates back to the work of Whittaker (1923) and Henderson (1924).

King and Rebelo (1993) show that under some conditions the HP filter will be the optimal filter for identifying the cyclical component of a series. Harvey and Jaeger (1993) compare it to a structural time-series model and conclude "...that the HP filter is tailor-made for extracting the business cycle component from US GNP" (p. 236). Baxter and King (1995) show that the HP filter "...can, in some cases, produce reasonable approximations to an ideal business cycle filter" (p. 21-22). However, use of the HP filter remains controversial. King and Rebelo note that the conditions for optimality are unlikely to be satisfied and Harvey and Jaeger find the HP filter performs less well on other series. Cogley and Nason (1995) discuss the dangers of spurious cyclicality induced by the HP filter while Guay and St-Amant (1996) argue that the HP filter does a poor job of extracting business cycle frequencies from macroeconomic time series.[11] Despite this, the HP filter remains popular in applied work (e.g. Taylor, 1998). Multivariate applications of the filter have also been developed (e.g. Laxton and Tetlow, 1992 and Kozicki, 1998).

3.3 The Beveridge-Nelson Decomposition

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

$$\Delta y_t = \varepsilon_t + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \cdots = e_t \tag{6}$$

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

$$y_{t+s} - y_t = \sum_{j=1}^{s} \Delta y_{t+j} = \sum_{j=1}^{s} e_{t+j} \tag{7}$$

[11]A summary of these critiques and others may be found in St-Amant and van Norden (1997). See also Christiano and Fitzgerald (1999) for comparisons of the HP filter with the band pass filter.

The trend is defined to be

$$\lim_{s\to\infty} E_t(y_{t+s}) = y_t + \lim_{s\to\infty} E_t(\sum_{j=1}^{s} e_{t+j}) \tag{8}$$

From equation 6, we can see that

$$E_t(e_{t+j}) = E_t(\varepsilon_{t+j} + \beta_1 \cdot \varepsilon_{t+j-1} + \beta_2 \cdot \varepsilon_{t+j-2} + \cdots) = \sum_{i=0}^{\infty} \beta_{j+i} \cdot \varepsilon_{t-i} \tag{9}$$

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

$$y_t = \tau_t + c_t \tag{10}$$

where the trend is

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

and e_t is white noise.

To use the Beveridge-Nelson decomposition we must therefore: (1) Identify p and q in our ARIMA(p,1,q) model. (2) Identify the $\{\beta_j\}$ in equation 6. (3) Choose some large enough but finite value of s to approximate the limit in equation 8.[12] (4) For all t and for $j = 1, \cdots, s$, calculate $E_t(e_{t+j})$ from equation 9. (5) Calculate the trend at time t as $y_t + E_t(\sum_{j=1}^{s} e_{t+j})$ and the cycle as y_t minus the trend.

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

When applied to GDP, the Beveridge-Nelson decomposition typically implies relatively small and not very persistent output gaps.[13] The Beveridge-Nelson decomposition was influential in the 1980s when the small variance of its cycles in output was interpreted as implying that real rather than nominal shocks dominated output fluctuations. This reasoning has been discredited by the work of Watson (1986) and Quah(1992), who stressed

[12]This need not be very large since changes in the detrended log of output may not be very persistent. For example, Blanchard and Fischer (1989) argue that changes in the detrended log of U.S. GDP are well approximated by an MA(2), implying that the correct model for log output is an ARIMA(0,1,2) and that $s = 2$ is sufficient.

[13]This reflects the fact that ARMA models have little ability to forecast changes in output.

that other decompositions could lead to other conclusions, and Lippi and Reichlin (1994) who noted that the random walk assumption imposed on the trend does not match the implications of business cycle models.[14] Perhaps as a result, multivariate extensions of this method have been much more influential in recent years. (See e.g. Rotemberg and Woodford, 1996, for such an application for business cycle analysis.) Such methods currently form the basis of the OECD's measures of the output gap and their work on cyclical adjustment of government deficits and surpluses. (Giorno et al., 1995.)

3.4 Unobserved Component Models

Unobserved component (UC) models attempt to specify the time-series properties of output and use the resulting model to identify cyclic and trend components. Surveys of its use in business cycle estimation may be found in Enders (1994) and Maravall (1996). Among the simplest UC models are the Local Level models,

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{11}$$

and the Local Linear Trend models,

$$y_t = \mu_t + \varepsilon_t,$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \tag{12}$$

$$\beta = \beta_{t-1} + \zeta_t.$$

In the former (equation 11), the observed output series y_t is composed of a random walk component μ_t and white noise ε_t. ε_t and the increments of the random walk are assumed to be mutually uncorrelated and follow independent Gaussian distributions. This implies that y_t follows an IMA(1,1), with the size of the MA term determined by the relative variances of ε and μ. The local linear trend modifies the local level model by assuming that the

[14]Quah (1992) notes that of all possible decompositions, the Beveridge-Nelson decomposition minimizes the variance of the cyclical component.

increments to the trend component, μ_t, are not i.i.d, but themselves follow a local level model.[15] This implies that y_t must be I(2) rather than I(1).

Popular models of quarterly output are typically based on one of these two basic models, adding only richer short-term dynamics. The first of these to be applied was that of Watson (1986), who modified the linear level model by replacing the white noise error term ε_t with an AR(2) process to allow for more business cycle persistence.

$$
\begin{aligned}
y_t &= \mu_t + c_t \\
\mu_t &= \delta + \mu_{t-1} + \eta_t \\
c_t &= \rho_1 \cdot c_{t-1} + \rho_2 \cdot c_{t-2} + \varepsilon_t
\end{aligned}
\tag{13}
$$

Next was Clark (1987), who similarly modified the local linear trend model to allow for an AR(2) cycle.

$$
\begin{aligned}
y_t &= n_t + x_t \\
n_t &= g_{t-1} + n_{t-1} + \nu_t \\
g_t &= g_{t-1} + w_t \\
x_t &= \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + e_t
\end{aligned}
\tag{14}
$$

where ν_t, w_t and e_t are i.i.d mean-zero gaussian processes.

Finally, Harvey and Jaeger (1993) offered a different modification of the local linear trend model in which Clark's AR(2) cycle is replaced by a sinusoidal stochastic process, ψ_t.

$$
\begin{aligned}
y_t &= \mu_t + \psi_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\
\beta_t &= \beta_{t-1} + \zeta_t \\
\psi_t &= \rho \cdot \cos(\lambda_c \cdot \psi_{t-1}) + \rho \cdot \sin(\lambda_c \cdot \psi^*_{t-1}) + \chi_t
\end{aligned}
\tag{15}
$$

[15] Again, all the error terms are assumed to be normally distributed and mutually independent at all leads and lags.

$$\psi_t^* = -\rho \cdot \sin(\lambda_c \cdot \psi_{t-1}) + \rho \cdot \sin(\lambda_c \cdot \psi_{t-1}^*) + \chi_t^*$$

where $\{\varepsilon, \eta, \zeta, \chi, \chi^*\}$ are all mean-zero gaussian i.i.d. errors and are uncorrelated at all leads and lags.

All three of the above-mentioned papers suggested using the cycle-trend decompositions implied by these models as a measure of the business cycle.[16] These univariate models have led to a series of multivariate extensions which are currently used extensively in output gap measurement.[17]

We examine the simpler univariate models in this paper for a variety of reasons. First, there are some indications that the multivariate versions are not always much more precise than their simpler univariate counterparts. In those cases, our analysis of the revision errors should help us understand the reliability of the resulting estimates. Second, the inclusion of the UC models allows us to further decompose the difference between the QuasiReal and the Final estimates and thereby better understand the importance of parameter instability in causing revisions to output gap estimates. Finally, since UC models also allow us to calculate the confidence intervals around our estimated output gaps, the revision errors serve as a useful check on the accuracy of these standard errors in the face of possible misspecification.

4 Results

Figure 1 compares the estimated business cycles for the eight different methods mentioned in Section 3. RealTime estimates are shown in the top half of the figure while Final estimates are shown in the lower half. Several features are readily apparent.

First, the different methods have strong short-term comovements. Most appear to be moving upwards or downwards at roughly the same time, although the amount of these moves vary from one method to another.

[16] Clark (1987) also considered a bivariate model of output growth.

[17] For example, see Kuttner (1992, 1994), Amato (1997), Gerlach and Smets (1997) and Kichian (1999).

Second, despite having similar short-term movements, the different methods typically give rise to a wide range of different estimates of the output gap. The difference between the highest and lowest estimate is frequently over 4 percent of output and is the same order of magnitude as the size of the business cycle itself. The dispersion of estimates is sufficiently great that estimates of both signs can usually be found and exceptions to this rule tend to be short-lived. Curiously, both the RealTime and the Final estimates show a period during which all the estimates tended to be tightly clustered. However, these periods are quite different for the two kinds of estimates; around 1973 for the RealTime estimates and 1984-1990 for the Final estimates.

To provide a first impression of the variation and size of the revisions implied by the real-time and final estimates shown in figure 1, we plot the difference of the two series for each method in figure 2. As with the estimates themselves, the dispersion of revisions is great, especially in the mid 1970s, suggesting that interpreting the accuracy of estimates during that period might have been especially difficult. The mid 1970s also coincides with the period when the "official" estimates of the output gap (which were prepared at the time by the Council of Economic Advisers) were most inaccurate. At the time, those estimates were based on a segmented trend method for estimating potential output which proved particularly misleading for assessing the productive capacity of the economy following the productivity slowdown of the late 1960s and early 1970s. By 1975, these estimates suggested that output was more than 10 percentage points below potential—similar to what is shown in figure 1 for our linear and quadratic trend method estimates.

4.1 Revision size and persistence

To better understand the differences between the RealTime and the Final estimates, Table 1 provides descriptive statistics on the various RealTime, QuasiReal, QuasiFinal and Final estimates, while Table 2 provides similar statistics for the total revision (i.e. Final estimate - RealTime estimate). Comparing the two tables, we see that the revisions are of the same order of magnitude as the estimated output gaps, although this varies somewhat across methods. The last column of table 2 reports the estimated first order autocorrelation coefficients for the revisions, showing that they are highly persistent. Aside from the Beveridge-Nelson model, the persistence ranges from 0.80 for the Breaking Trend to 0.96 for the Linear Trend and the Watson model.

It is worth noting that the statistical properties of these revisions are broadly in line with those of the revisions of "official" output gap estimates for the U.S. One such series is examined in Orphanides (1999), who has compiled the real-time output gap estimates available at the Federal Reserve from 1965 to 1993. These were based on the Council of Economic Advisers estimates during the 1960s and 1970s and Federal Reserve staff estimates during the 1980s and 1990s. The standard deviation of these real-time estimates from 1966Q1 to 1993Q4 is 3.8 percent. Comparison of these real-time estimates with the historical Federal Reserve staff estimates available in 1994Q4 suggests large and highly persistent revisions. The standard deviation of these revisions is 2.6 percent and their first order serial correlation is about 0.9.

Because the various methods have substantial variation in the size of the cyclical component they produce, it is easier to compare their reliability in real-time by looking at comparably scaled measures of the revisions. Table 3 presents some such measures. In column 1 we present the correlation between the Final and RealTime series for each method. (This would be 1 in the ideal case where no revisions to the RealTime estimates were ever required.) As can be seen these correlations range from a low of 0.53 for the Hodrick-Prescott filter and 0.56 for the Harvey-Jaeger model to a high of 0.87 for the Breaking Trend and 0.81 for the Linear Trend.

The remaining three indicators in Table 3 measure in different ways the relative importance of the revisions. (In the ideal case of no revisions, each of these indicators would equal 0.) The first of these indicators, NS, reports the ratio of the standard deviation of the total revision to the standard deviation of the final estimate of the gap; this gives us a proxy for the "noise-to-signal ratio" in the RealTime estimates. For example, looking at the Hodrick Prescott method, we see that this ratio is 1.03 (i.e. the revision has a slightly larger variance than the final estimate of the output gap itself). This is the worst ratio for the eight methods, although it is not far from the 0.93 and 0.92 for the Quadratic Trend and Harvey-Jaeger models, respectively. By this criterion, even the best models have rather large ratios, between one-half and two-thirds.

The last two columns provide the frequencies with which the RealTime estimate is "bad." The OPSIGN column shows the frequency with which the RealTime and Final gaps were of opposite signs. For the Watson and Linear trend models, this frequency exceeds 50 percent. Not all methods do as badly by this criterion with the Breaking Trend model misclassifying only 12 percent and the Beveridge Nelson only 21 percent. The XSIZE column shows the frequency with which the absolute value of the revision exceeds the absolute value of the Final series. The different detrending methods give more similar results in this respect. In five of the eight models this frequency exceeds 50 percent and in two others it exceeds 40 percent. The Breaking Trend again stands out as the best with revisions larger than Final gaps only 30 percent of the time.

We reiterate that the revision errors we measure here are underestimates of the total estimate error; we are measuring only the estimation errors which we subsequently correct. This also means that we must be particularly cautious in trying to compare the reliability of the different methods. With this caveat firmly in mind, we may note that some methods appear on the surface to be less desirable than others. For example, the Hodrick Prescott filter combines the lowest correlation (0.53) between the Final and RealTime estimates and the worst noise-signal ratio with a higher than average persistence of revisions (0.93). The Quadratic Trend does not fair much better, with the second-worst noise-signal ratio and the third-highest persistence (0.95.) In contrast, the Breaking Trend combines the highest correlation (0.87) with the second-lowest persistence (0.80) and by far the best frequency of correctly signing the output gap.

4.2 Decomposition of Revisions

Figure 3 through Figure 6 help us understand the importance of different factors in accounting for the total revision in the estimated output gap as we move from RealTime to Final estimates. Table 4 presents detailed related summary statistics for the various methods.

Figure 3 shows results for the Linear Trend method in the upper panel and the Watson model in the lower panel. (The reason for this grouping will become clear shortly.) In each graph, we see the RealTime estimate of the output gap together with the total subsequent revision (Final - RealTime) of that estimate. The fact that the revision is roughly equal to the RealTime estimate at the trough of the 1975 recession tells us that our final estimate of the output gap is roughly zero. In other words, despite the extreme evidence of recession in the RealTime estimate, ex post we would judge that the economy was operating roughly at potential at that time. The size of these revisions (about 8 to 10 percentage points in this period) underline the lack of precision of these methods' RealTime estimates.

To understand the source of these revisions, both graphs also show the effects of data revision. (This is constructed as the RealTime estimate minus the QuasiReal estimate.) This is simply the component of the overall revision which is due to subsequent changes in the published data (as opposed to the addition of new data points to the sample.) For example, since we see that the total revision and data revision are roughly equal in both graphs in late 1995, this means that nearly all of the revision in our estimated output gap for those quarters was due to subsequent revisions in the published data.

Looking at the whole sample period, the data revision is typically less than $\pm$ 2 percent of output and its variability tends to be small compared to that of the total revision. This in turn means that most of the revision is due to the addition of new points to our data sample. However, data revisions still play a role as can be confirmed by looking at the summary statistics of the difference between the QuasiReal and RealTime estimates of the output gap shown in Table 4.

In the case of the Watson model, we can further identify the source of the revisions by identifying the effects of parameter revisions (calculated as QuasiReal - QuasiFinal). The lower panel of Figure 3 shows that these parameter revisions account for much of the revisions of our estimates of the output gap.[18]

Considering the evidence presented so far on the Linear Trend and Watson models, we are led to the conclusion that they are not well suited to the estimation of business cycles due to their assumption of a constant long-term trend in output growth. This assumption leads to parameter instability as samples are lengthened and the trend rate of growth is revised downwards. It gives us output gap estimates which seem to contain a downward trend (see Figure 2), output gaps which are furthest from zero and the largest standard deviation of revision.

Figure 4 considers the two other deterministic trend models, the Quadratic Trend and the Breaking Trend. The two give visually similar RealTime estimates, the main difference coming in 1977, when the Breaking Trend estimates undergo a discrete shift as the trend break is introduced in 1973. The total revision is again often close to the size of the RealTime output gap (particularly in the mid-1990s.) We note that although the data revisions seem to play a secondary role in explaining the total revision of the RealTime estimates, a major exception appears during 1974 and 1975 when substantial data revisions eventually helped to moderate initial perceptions of a disastrous recession.

Figure 5 again presents visually similar results from two conceptually different methods, this time from the Hodrick-Prescott filter and the Harvey-Jaeger unobserved component model.[19] In both cases we find revisions that are fully as large as the RealTime estimates and that cannot be attributed to the effects of data revisions (particularly once we exclude the 1974-75 revisions.) Results for the Harvey-Jaeger model further indicate that the effects of parameter revision are similarly small, unlike the first case we considered above. The

[18]The parameter instability was evident when performing the rolling estimation of the Watson model; parameter estimates tended to fluctuate between two different sets of parameters with quite different implications for the estimated business cycle.

[19]The similarity in the Final estimates from these two methods was noted in the original article by Harvey and Jaeger (1993).

large revision of our estimates must therefore be due almost entirely to the addition of subsequent observations to our sample.

A further striking feature of these two methods is that the revision seems to systematically lead the RealTime estimate by about one year. This does not imply that the RealTime estimates use the available data inefficiently, since the revisions can obviously only be calculated with Final data. These results appear to contrast with those of St-Amant and van Norden (1997), who examined the spectral properties of HP filters at the end of sample (similar to our QuasiReal estimates.) They found that while there was a phase lag of about 2 quarters at most business cycle frequencies, the overall phase shift was effectively zero due to the effects of spectral leakage from lower frequencies.[20]

In Figure 6, we consider the results from the last pair of models, the Beveridge-Nelson and Clark models. The upper frame shows that results for the Beveridge-Nelson decomposition are atypical in almost every way. The estimated output gap is much smaller and much less persistent that produced by any other method, facts which were also evident in Table 1. However, we now also see that the RealTime estimates look very little like the output gaps we would associate with U.S. business cycles. For example, the recessions of 1982 and 1991 are difficult to distinguish from the background "noise" and appear to be very brief and mild (with the gap never exceeding 1.5 percent of output in absolute value.) By far the largest output gap in the sample, that of 1975, is largely accounted for by data revisions and becomes unremarkable in Final estimates. Indeed, the total revisions for this method are dominated by the effects of data revision; the two series are highly correlated and their plots are often difficult to separate visually.

The lower frame shows that the results for Clark are much more typical of those for the other unobserved components models. Revisions are almost as large as the RealTime gaps and are persistent. Both parameter revision and data revision effects are relatively minor. Perhaps the most striking feature of the RealTime estimates are that after 1973 they are

[20]St-Amant and van Norden (1997), p. 32.

almost never strongly positive; that is, in real-time the economy appears to be virtually continuously at or below potential for twenty-five years with this method.

4.3 Turning Points

It is particularly interesting to know how the different business cycle measures do around business cycle turning points, since these are presumably periods where accurate and timely estimate of the output gap (and its changes) would be of particular interest to policy makers. To help assess this, we calculated a number of descriptive statistics regarding the size or the revision in RealTime estimates in the three quarters centered about each of the NBER business cycle peaks from 1966 to 1997. Results are shown in Table 5.

We see that all methods seem to underestimate the output gap in RealTime at cyclical peaks, although the degree to which this is true varies considerably from one method to another. The Linear Trend and Watson methods have by far the most severe underestimates while the Beveridge-Nelson has the smallest.

4.4 Revisions and Confidence Intervals

As noted previously, our revision errors overestimate the overall reliability of the output gap series since they neglect the estimation error which remains in the Final estimates. Alternatively, we can also use standard statistical methods to calculate the reliability of some of the output gap measures. These too will overestimate the reliability of the gap since they ignore the effects of data revision and model misspecification. Of course, if these two are relatively small, statistical methods may be a useful guide to the reliability of RealTime output gap estimates.

To investigate this question, we focused on the three UC models and calculated 95% confidence intervals about the RealTime estimates of the output gap.[21] The results are shown in figures 7 through 9, which compare these confidence intervals to the final estimates

[21] These were calculated using the usual formulas for the standard errors surrounding estimates produced by the Kalman filter. Note that in addition to the effects of data revision and model misspecification mentioned above, these also do not take account of the uncertainty in the model's estimated parameters.

of the output gap. If the statistical confidence intervals are reliable, we should find that our Final estimates fall outside the 95% RealTime confidence interval very infrequently.

The figures show that the reliability of calculated confidence intervals varies. Final estimates from the Watson model are often outside the real-time confidence intervals. This happens only rarely (and then briefly) for the Harvey-Jaeger model, and not at all for the Clark model. This finding suggests that the assumption of a constant drift rate for trend output growth embedded in the Watson model is at odds with the data and implies that the calculated confidence intervals for this model omit an important source of error.

5 Conclusions

We have examined the reliability of univariate detrending methods for estimating the output gap in real time. In doing so, we have focused on the internal consistency of output gap estimates over time as more information arrives and data are revised. This gives us results which are robust to alternative assumptions about the structure of the economy, but may tend to overestimate the reliability of the estimated output gaps from any given method.

Our results suggest that the reliability of output gap estimates in real time tends to be quite low. Different methods give widely different estimates of the output gap in real time and often do not even agree on the sign of the gap. The standard error of the revisions is of the same order of magnitude as the standard error of the output gap for all the methods. The measurement error problem is compounded by a high degree of persistence of the revisions and further by a systematic bias around business cycle turning points. These findings suggest that measurement error would pose a serious policy problem for any of these measures of the output gap. The relative size and persistence of the revision errors we report are also similar to those associated with "official" real-time output gap estimates, such as those reported in Orphanides (1999).

Some important differences between the alternative methods also emerged. The Beveridge-Nelson method does not give reasonably sized or persistent gaps. Methods which assume a cycle around a constant growth trend (Linear Trend and Watson models) have particularly large revisions due to parameter instability in the estimated trend rate of growth. This confirms that models with time-varying trend rates of growth should be preferred.

We also found that, although important, the revision of published data does not appear to be the primary source of revisions for any of the methods we examined. Rather, the subsequent evolution of the economy seems to be very informative for estimation of the current position in the business cycle. Thus, even if the reliability of the underlying real-time data were to improve, real-time estimates of the output gap would remain unreliable.

References

Amato, Jeffrey D. (1997), *Empirical Models for Monetary Policy Making*, unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Baxter, Marianne and Robert G. King (1995), "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," NBER Working Paper, No. 5022.

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

Blanchard, Olivier Jean and Stanley Fischer (1989), *Lectures on Macroeconomics* MIT Press, Cambridge, MA.

Cheung, Yin-Wong and Menzie Chinn (1997), "Further Investigation of the Uncertain Unit Root in GNP" *Journal of Business and Economic Statistics*, 15(1), 68-73.

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

Clarida, Richard, Jordi Gali, and Mark Gertler (1998), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," NBER Working Paper, No, 6442, March.

Clark, Peter K. (1987), "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics* v102 n4, pp. 797-814.

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

Council of Economic Advisers (1977), *Economic Report of the President,* U.S. Printing Office, Washington, D.C.

Croushore, Dean and Tom Stark, (1999) "A Real-Time Data Set for Macroeconomists," Manuscript, May.

Enders, Walter (1994), *Applied Econometric Time Series*, Wiley.

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

Giorno, Claude, Pete Richardson, Deborah Roseveare, and Paul van den Noord (1995), "Potential Output, Output Gaps and Structural Budget Balances," *OECD Economic*

Studies No. 24, 1, 167-209.

Guay, Alain and Pierre St-Amant (1996) "Do Mechanical Filters Provide a Good Approximation of Business Cycles?" Bank of Canada Technical Report No. 78, Ottawa.

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

Henderson, R. (1924). "A New Method of Graduation." *Actuarial Society of America Transactions*, 25, 29-40.

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

Kichian, Maral (1999), "Measuring Potential Output within a State-Space Framework" Bank of Canada Working Paper 99-9.

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

Kuttner, Kenneth N. (1992), "Monetary Policy with Uncertain Estimates of Potential Output," *Economic Perspectives*, Federal Reserve Bank of Chicago, 16, 2-15.

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

Kozicki, Sharon (1998), "Multivariate Detrending Under Common Trend Restrictions: Implications for Business Cycle Research *Journal of Economic Dynamics and Control*, (forthcoming).

Laxton, Doug and Robert Tetlow (1992), *A Simple Multivariate Filter for the Measurement of Potential Output.* Technical Report No. 59, Ottawa: Bank of Canada

Lippi, M. and L. Reichlin (1994), "Diffusion of Technical Change and the Decomposition of Output into Trend and Cycle." *Review of Economic Studies*, 61(1), 19-30.

Maravall, Agustin (1996) "Unobserved components in economic time series." Banco de Espana documento de trabajo 9609.

McCallum, Bennett and Edward Nelson (1998), "Performance of Operational Policy Rules in An Estimated Semi-Classical Structural Model," NBER Working Paper, No. 6599, June.

Nelson, Charles R. and Charles I. Plosser (1982), "Trends and Random Walks in macroeconomic time series: some evidence and implications." Journal of Monetary Economics, 10(2), 139-62.

Orphanides, Athanasios (1997), "Monetary Policy Rules Based on Real-Time Data," Finance and Economics Discussion Series, 1998-03, Federal Reserve Board, December.

Orphanides, Athanasios (1998), "Monetary Policy Evaluation With Noisy Information," Finance and Economics Discussion Series, 1998-50, Federal Reserve Board, October.

Orphanides, Athanasios (1999), "The Quest for Prosperity Without Inflation," Manuscript, May.

Perron, Pierre (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica* 10(2), 130-62.

Quah, Danny (1992), "The relative importance of permanent and transitory compoenents: Identification and some theoretical bounds." Econometrica 60(1), 107-18.

Rotemberg, Julio and Michael Woodford (1996), "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption," *American Economic Review*, 86(1), 71-89, March.

Rothman, Philip (1997), "More uncertainty about the unit root in U.S. real GNP." *Journal of Macroeconomics* 19(4), 771-780.

Rudebusch, Glenn (1993), "The uncertain unit root in GNP." *American Economic Review* 83, 264-72.

Smets, Frank (1998), "Output Gap Uncertainty: Does it Matter for the Taylor Rule?" BIS Working Paper No. 60, November.

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

Taylor, John B. (1993), "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, December, 195-214.

Taylor, John B. (1998), "An Historical Analysis of Monetary Policy Rules," NBER Working Paper 6768, October.

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

Whittaker, E. T. (1923) "On a New Method of Graduation." *Proceedings of the Edinburgh Mathematical Society*, 41, 63-75.

Appendix: Alternative Measures of the Output Gap

Let y_t^v be the value of output published at time v for an observation at time t. Due to publication lags, we require $t \leq v - 1$. The full series published at any point in time v can be written as the vector $Y^v \equiv [y_1^v, y_2^v, \ldots, y_{v-1}^v]$. We can also refer to the subvector $Y_N^v \equiv [y_1^v, y_2^v, \ldots, y_N^v]$.

Now suppose Z is an $N \times M$ matrix consisting of real and non-real numbers. We restrict all its non-real entries (which represent unavailable observations) to lie below the main diagonal. We construct matrices of this form by placing series of different length in the columns, with each series starting in row 1. The remaining entries in each column (after the end of each series) are then filled with some non-real constant. We denote this as

$$Z \equiv z(A, B, \ldots, M) \tag{A.1}$$

where the arguments $A, B, \ldots, M$ are simply column vectors of (possibly) unequal length. We also define the **last-value function** $l(Z) : R^{N \times M} \to R^M$, which simply selects the last real observation (i.e. the one with the highest row number) in each column of Z. Combining the z and l functions into one gives us

$$\ell(A, B, \ldots, M) \equiv l(z(A, B, \ldots, M)) \tag{A.2}$$

Suppose that we also have an arbitrary detrending function $f(X) : R^N \to R^N$. The **Final** estimate of the output gap for this detrending function is just

$$\hat{Y}_{Final} \equiv f(Y^M) \tag{A.3}$$

where M is the "final" vintage of data available (in our case, 1999Q1.)

The **RealTime** estimate of the output gap is

$$\hat{Y}_{RealTime} \equiv \ell(f(Y^1), f(Y^2), \ldots, f(Y^M)) \tag{A.4}$$

The **QuasiReal** estimate of the output gap is given by

$$\hat{Y}_{QuasiReal} \equiv \ell(f(Y^M_{N-M+1}), f(Y^M_{N-M+2}), \ldots, f(Y^M_N)) \tag{A.5}$$

The **QuasiFinal** estimate of the output gap only exists for detrending functions of the form $f(X, \theta)$ where θ is a set of parameters. Typically, these parameters describe the data-generating process for X and the maximum-likelihood estimate of the parameters may be denoted $\hat{\theta}(X)$. When the samples which we detrend are the same as those used to estimate the parameters, then we may define a new detrending function $g(X) \equiv f(X, \hat{\theta}(X))$ which can be used to construct the conventional RealTime, QuasiReal and Final output gap estimates. In the case of the **QuasiFinal** estimate, however, we compute

$$\hat{Y}_{QuasiFinal} \equiv \ell(f(Y^M_{N-M+1}, \hat{\theta}(Y^M)), f(Y^M_{N-M+2}, \hat{\theta}(Y^M), \ldots, f(Y^M_N, \hat{\theta}(Y^M)) \tag{A.6}$$

Table 1

Output Gap Summary Statistics:
1966:1 – 1997:4

Method	MEAN	SD	MIN	MAX	COR
Hodrick-Prescott					
Final	0.06	1.71	−4.58	3.70	1.00
Quasi-Real	−0.15	1.75	−4.30	3.84	0.56
Real-Time	−0.27	1.90	−6.63	3.84	0.53
Breaking Trend					
Final	0.33	2.51	−6.24	4.84	1.00
Quasi-Real	0.25	2.86	−6.90	6.94	0.91
Real-Time	0.21	3.15	−10.52	5.02	0.87
Quadratic Trend					
Final	0.55	2.54	−6.93	5.35	1.00
Quasi-Real	−1.02	2.72	−7.57	6.16	0.72
Real-Time	−0.96	3.03	−10.83	4.70	0.65
Linear Trend					
Final	1.47	4.96	−7.15	9.68	1.00
Quasi-Real	−3.74	4.17	−11.32	6.94	0.88
Real-Time	−3.45	3.98	−10.52	5.02	0.81
Beveridge-Nelson					
Final	−0.06	0.53	−1.80	1.66	1.00
Quasi-Real	−0.10	0.51	−1.81	1.54	0.99
Real-Time	−0.20	0.75	−4.14	1.98	0.79

(Continued next page)

Table 1 (continued)

Method	MEAN	SD	MIN	MAX	COR
Clark					
Final	0.24	2.11	−5.38	3.84	1.00
Quasi-Final	−0.61	1.45	−4.15	3.11	0.87
Quasi-Real	−0.69	1.63	−4.34	3.41	0.79
Real-Time	−0.93	1.91	−6.99	3.02	0.77
Harvey-Jaeger					
Final	0.03	1.55	−3.89	3.91	1.00
Quasi-Final	−0.07	1.22	−2.68	2.91	0.63
Quasi-Real	−0.04	1.35	−3.12	3.21	0.58
Real-Time	−0.10	1.48	−5.04	3.01	0.56
Watson					
Final	1.32	3.44	−4.37	7.19	1.00
Quasi-Final	0.16	3.35	−4.73	6.37	0.96
Quasi-Real	−2.38	2.65	−7.75	4.41	0.81
Real-Time	−2.08	2.61	−7.43	3.56	0.78

Notes: The alternative detrending methods are as described in the text. The statistics shown for each variable are: MEAN, the mean; SD, the standard deviation; and MIN and MAX, the minimum and maximum values. COR, denotes the correlation with the final estimate of the gap for that method.

Table 2

Summary Revision Statistics
Final vs Real-Time Estimates
1966:1 – 1997:4

Method	MEAN	SD	MIN	MAX	AR
Hodrick-Prescott	0.32	1.77	−3.41	3.42	0.93
Breaking Trend	0.12	1.54	−4.85	5.40	0.80
Quadratic Trend	1.49	2.36	−3.40	7.56	0.95
Linear Trend	4.97	2.83	−2.33	11.51	0.96
Beveridge-Nelson	0.14	0.46	−1.11	2.66	0.29
Clark	1.17	1.37	−1.90	4.35	0.92
Harvey-Jaeger	0.12	1.43	−2.93	3.67	0.85
Watson	3.40	2.16	−1.93	7.53	0.96

Notes: The detrending method and statistics are as described in the notes to Table 1. AR denotes the first order serial correlation of the revision series shown.

Table 3

Summary Reliability Indicators
1966:1 – 1997:4

Method	COR	NS	OPSIGN	XSIZE
Hodrick-Prescott	0.53	1.03	0.40	0.60
Breaking Trend	0.87	0.62	0.12	0.30
Quadratic Trend	0.65	0.93	0.34	0.52
Linear Trend	0.81	0.57	0.52	0.59
Beveridge-Nelson	0.79	0.87	0.21	0.43
Clark	0.77	0.65	0.31	0.49
Harvey-Jaeger	0.56	0.92	0.41	0.58
Watson	0.78	0.63	0.51	0.57

Notes: The table shows measures evaluating the size, sign and variability of the revision between the final and the real-time estimates for alternative methods. COR, denotes the correlation of the real-time and final estimates (from table 1). NS indicates the ratio of the standard deviation of the revision and the standard deviation of the final estimate of the gap. OPSIGN indicates the frequency with which the real-time and final gap estimates have opposite signs. XSIZE indicates the frequency with which the absolute value of the revision exceeds the absolute value of the final gap.

Table 4

Detailed Breakdown of Revision Statistics

1966:1 – 1997:4

Method	MEAN	SD	MIN	MAX	AR
Hodrick-Prescott					
Final/Real-Time	0.32	1.77	−3.41	3.42	0.93
Final/Quasi-Real	0.21	1.62	−3.52	3.27	0.97
Quasi-Real/Real-Time	0.11	0.59	−0.97	2.71	0.60
Breaking Trend					
Final/Real-Time	0.12	1.54	−4.85	5.40	0.80
Final/Quasi-Real	0.08	1.18	−3.76	2.24	0.87
Quasi-Real/Real-Time	0.03	1.06	−2.98	3.84	0.77
Quadratic Trend					
Final/Real-Time	1.49	2.36	−3.40	7.56	0.95
Final/Quasi-Real	1.56	1.97	−1.80	5.14	0.98
Quasi-Real/Real-Time	−0.09	1.04	−2.89	3.80	0.76
Linear Trend					
Final/Real-Time	4.97	2.83	−2.33	11.51	0.96
Final/Quasi-Real	5.20	2.35	0.00	8.25	0.96
Quasi-Real/Real-Time	−0.27	1.20	−3.62	3.84	0.81
Beveridge-Nelson					
Final/Real-Time	0.14	0.46	−1.11	2.66	0.29
Final/Quasi-Real	0.04	0.06	−0.25	0.26	0.59
Quasi-Real/Real-Time	0.10	0.46	−1.19	2.62	0.31

(Continued next page)

Table 4 (continued)

Method	MEAN	SD	MIN	MAX	AR
Clark					
Final/Real-Time	1.17	1.37	−1.90	4.35	0.92
Final/Quasi-Final	0.85	1.11	−1.23	3.24	0.93
Quasi-Final/Quasi-Real	0.07	0.47	−0.86	1.20	0.93
Quasi-Real/Real-Time	0.24	0.59	−0.75	2.65	0.84
Harvey-Jaeger					
Final/Real-Time	0.12	1.43	−2.93	3.67	0.85
Final/Quasi-Final	0.10	1.22	−2.47	3.30	0.87
Quasi-Final/Quasi-Real	−0.03	0.24	−0.48	0.60	0.95
Quasi-Real/Real-Time	0.06	0.39	−0.60	1.91	0.82
Watson					
Final/Real-Time	3.40	2.16	−1.93	7.53	0.96
Final/Quasi-Final	1.16	0.92	−0.56	2.81	0.95
Quasi-Final/Quasi-Real	2.53	1.59	−0.45	4.59	0.98
Quasi-Real/Real-Time	−0.29	0.95	−2.45	2.35	0.85

Notes: See notes to tables 1 and 2.

Table 5

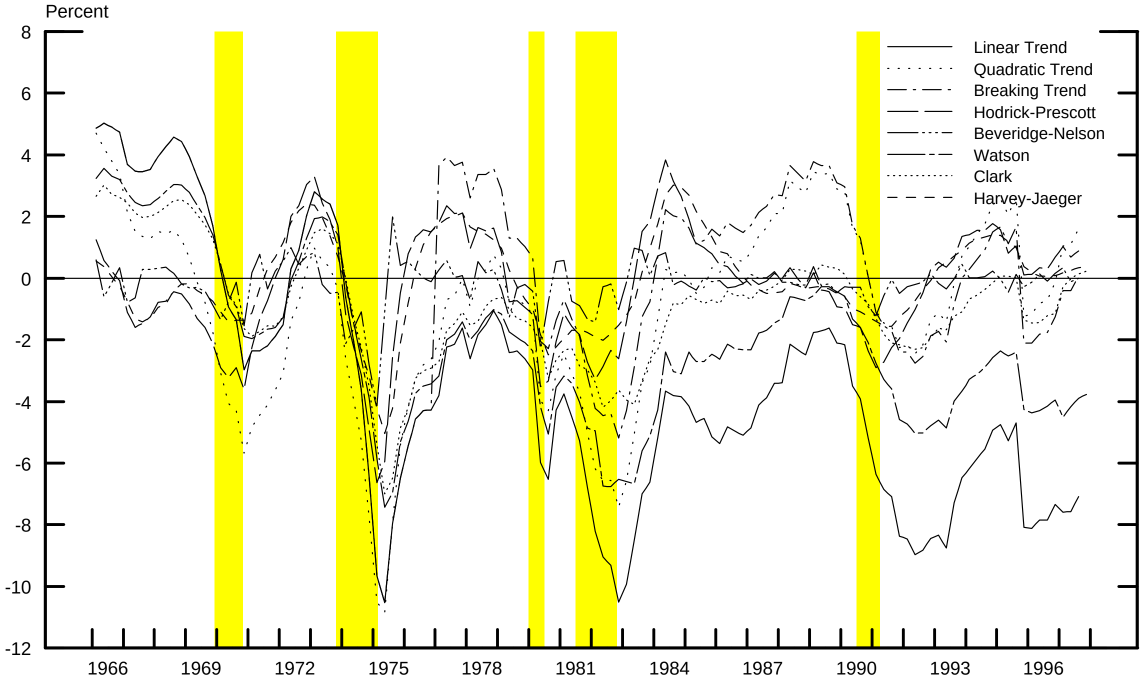
Revision Statistics at NBER Peaks
Final vs Real-Time Estimates

Name	MEAN	SD	MIN	MAX
Hodrick-Prescott	2.33	0.73	0.77	3.42
Breaking Trend	0.71	0.73	−0.79	1.81
Quadratic Trend	2.95	1.73	−0.51	5.17
Linear Trend	6.11	1.98	2.49	8.60
Beveridge-Nelson	0.31	0.49	−0.44	1.26
Clark	1.71	1.09	0.01	3.74
Harvey-Jaeger	1.68	0.94	−0.16	2.96
Watson	4.29	1.64	0.94	6.65

Notes: The revision is defined as the difference between the final and the real-time estimates. For each method, the sample used to compute the revision statistics is limited to the three quarters centered around each of the NBER peaks from 1966 to 1997. See also notes to Table 1.

Figure 1

Real-Time Estimates of the Business Cycle



Final Estimates of the Business Cycle

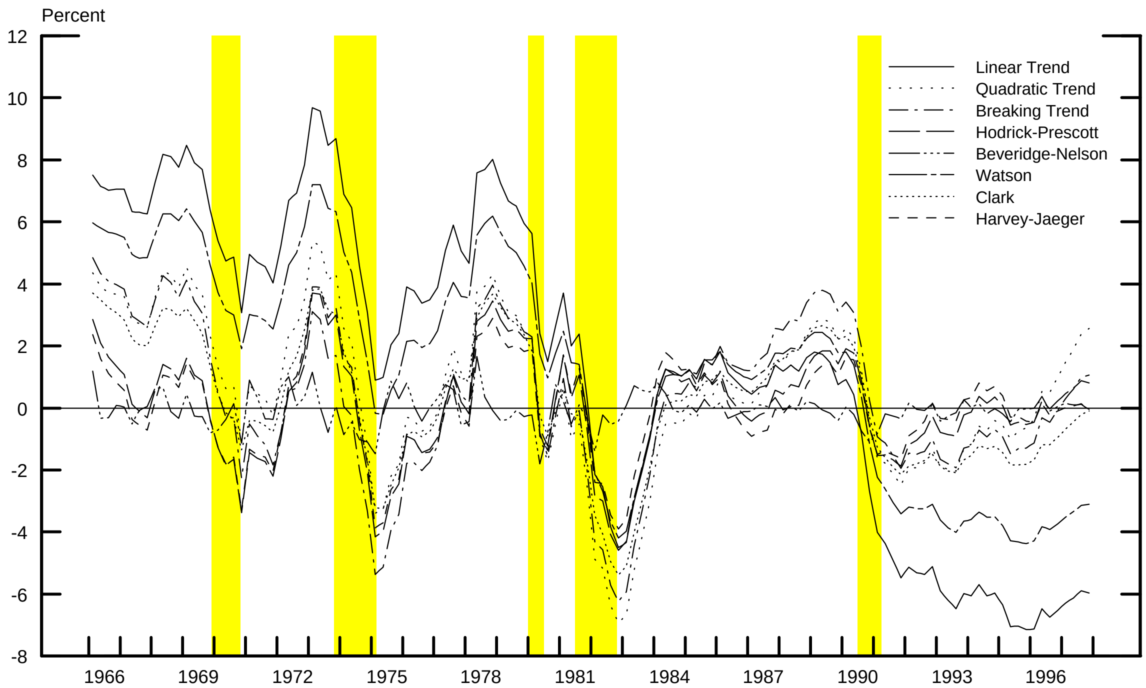


Figure 2

Total Revision in Business Cycle Estimates

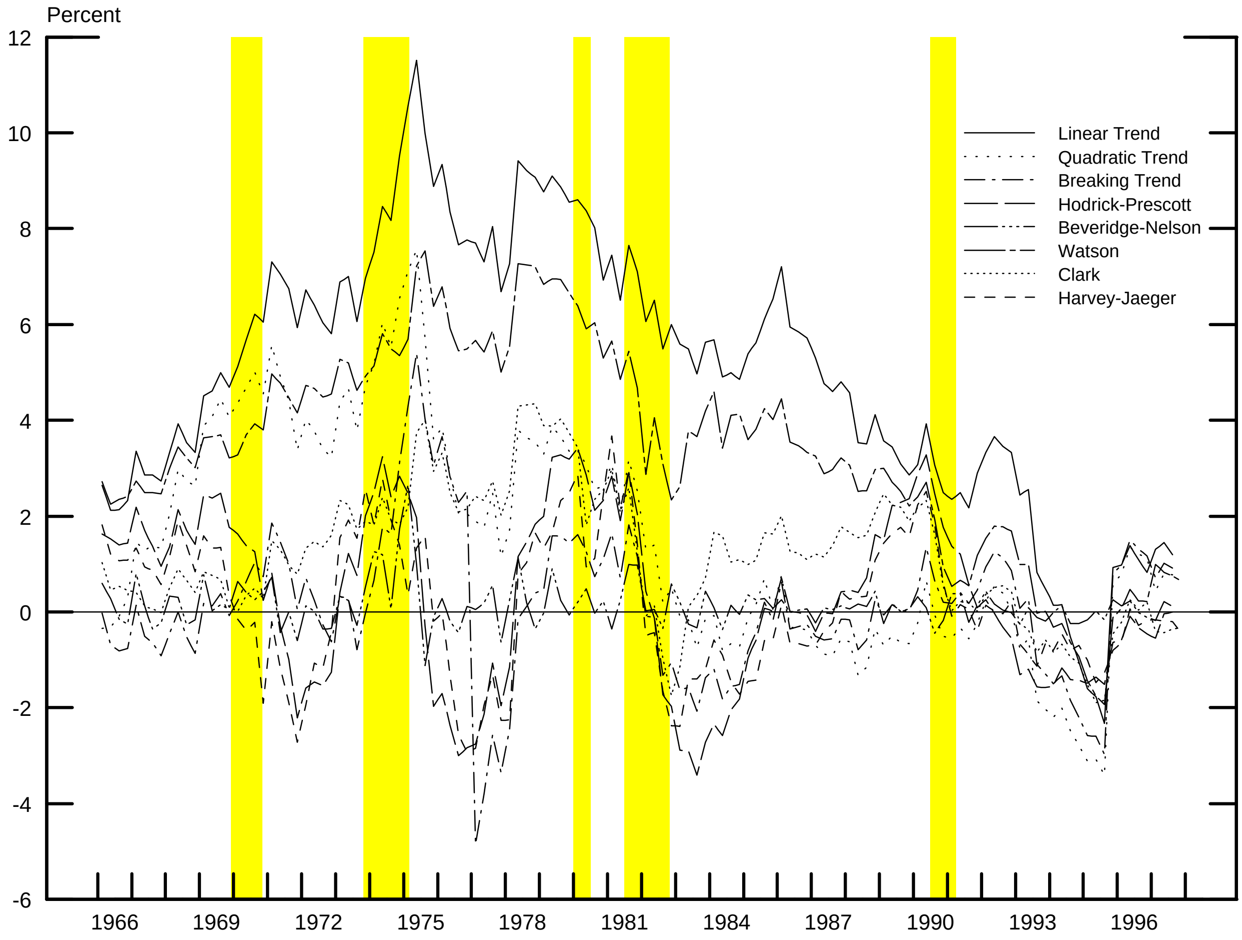
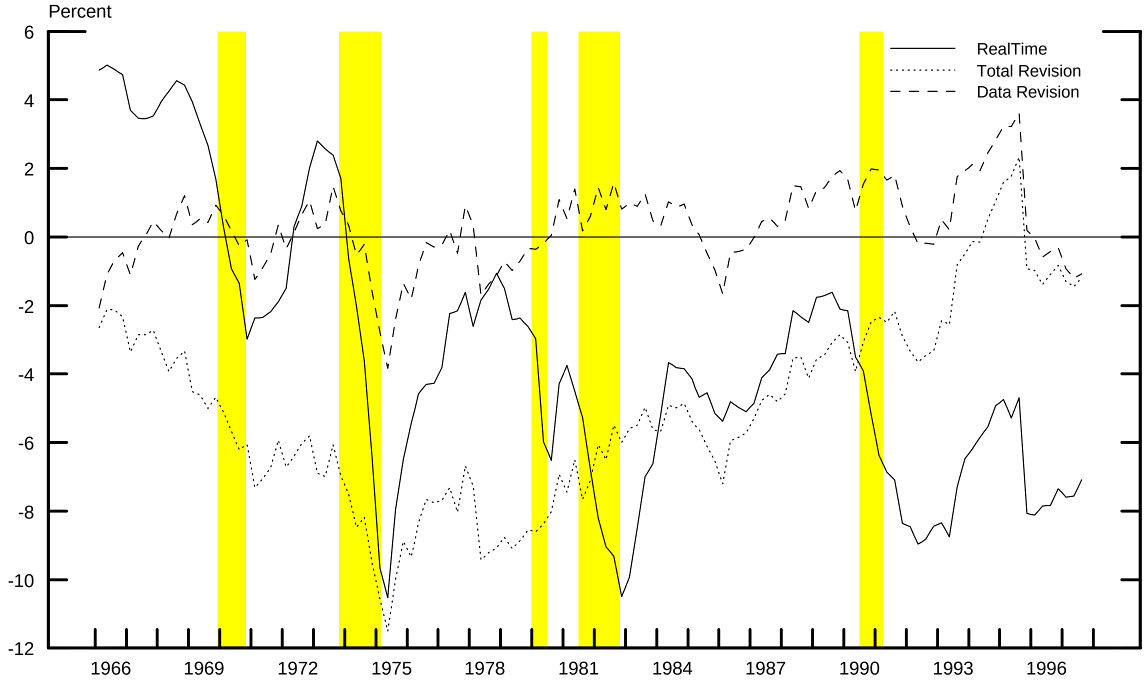


Figure 3

Estimated Business Cycle: Linear Trend



Estimated Business Cycle: Watson

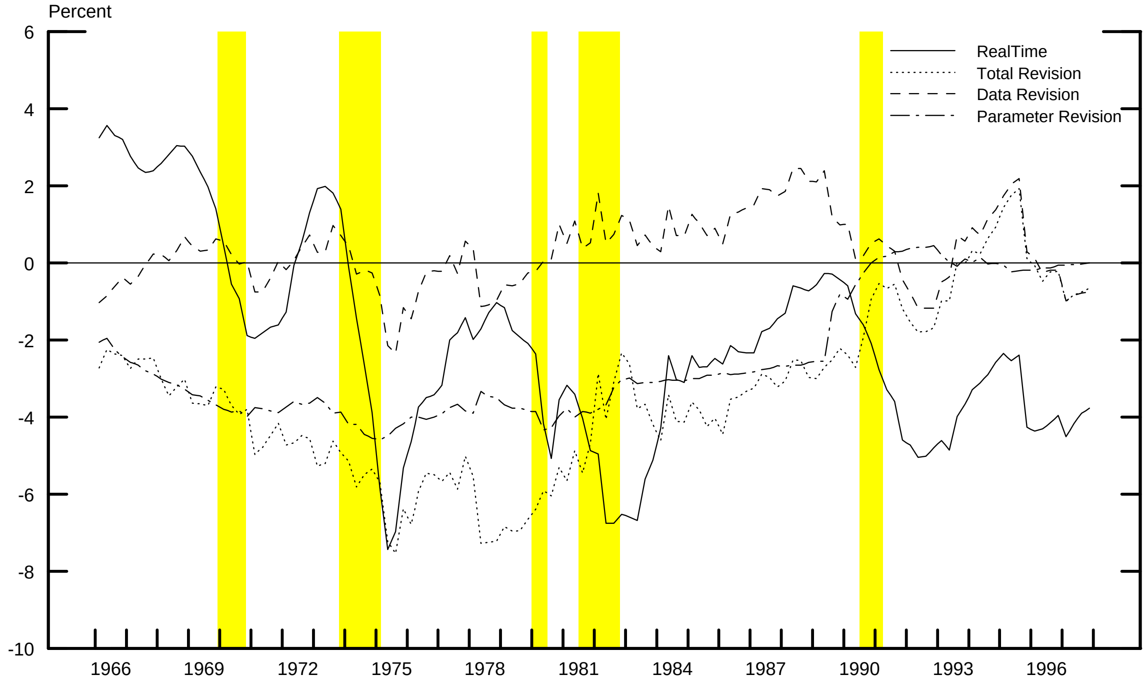
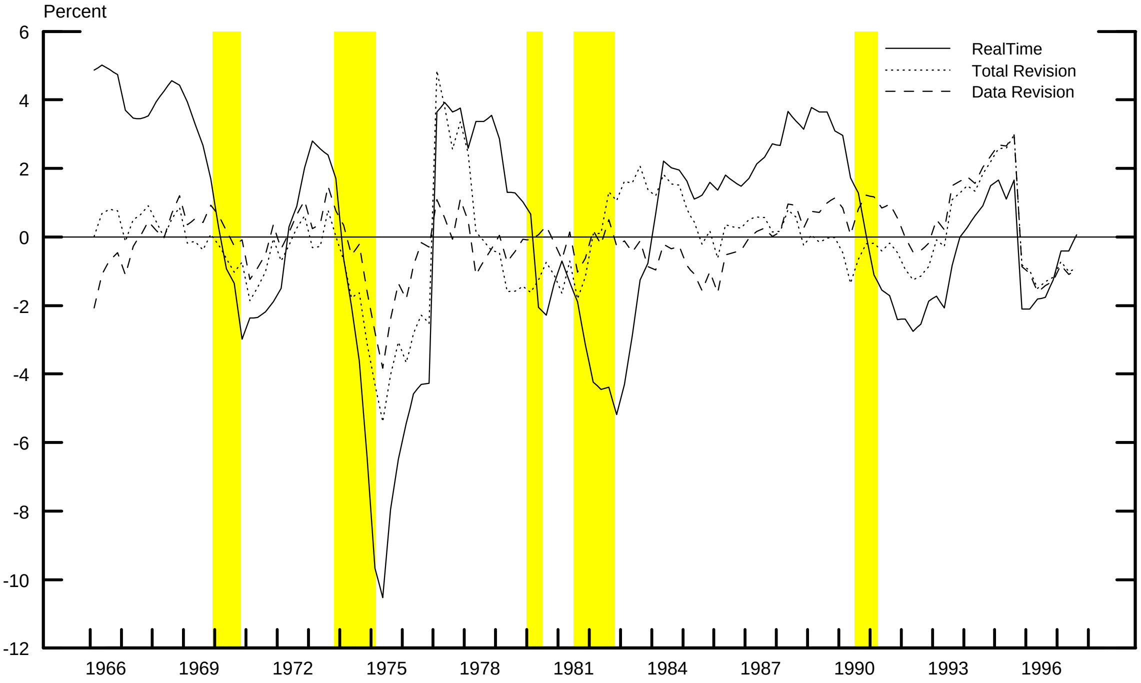


Figure 4

Estimated Business Cycle: Breaking Linear Trend



Estimated Business Cycle: Quadratic Trend

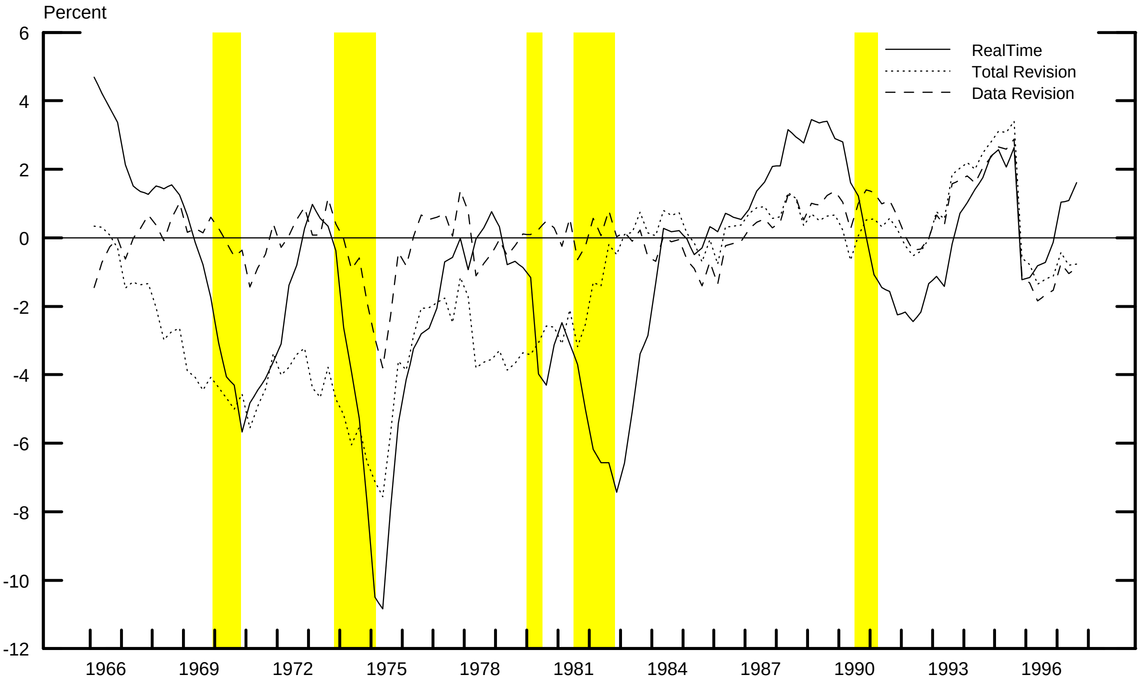
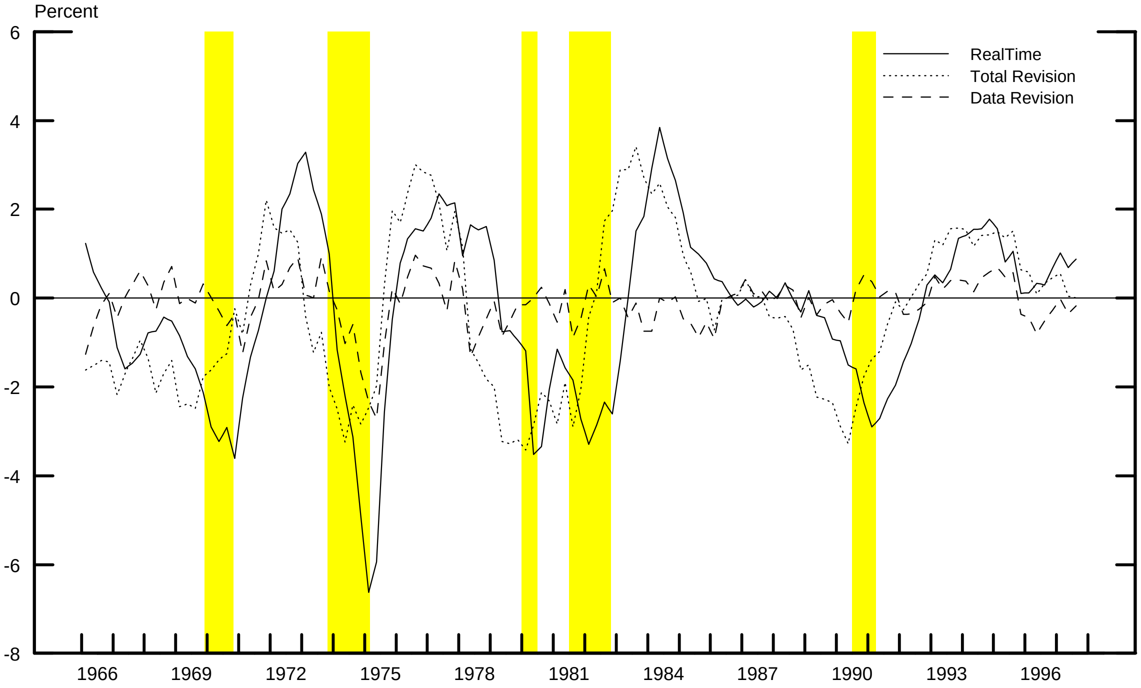


Figure 5

Estimated Business Cycle: Hodrick-Prescott



Estimated Business Cycle: Harvey-Jaeger

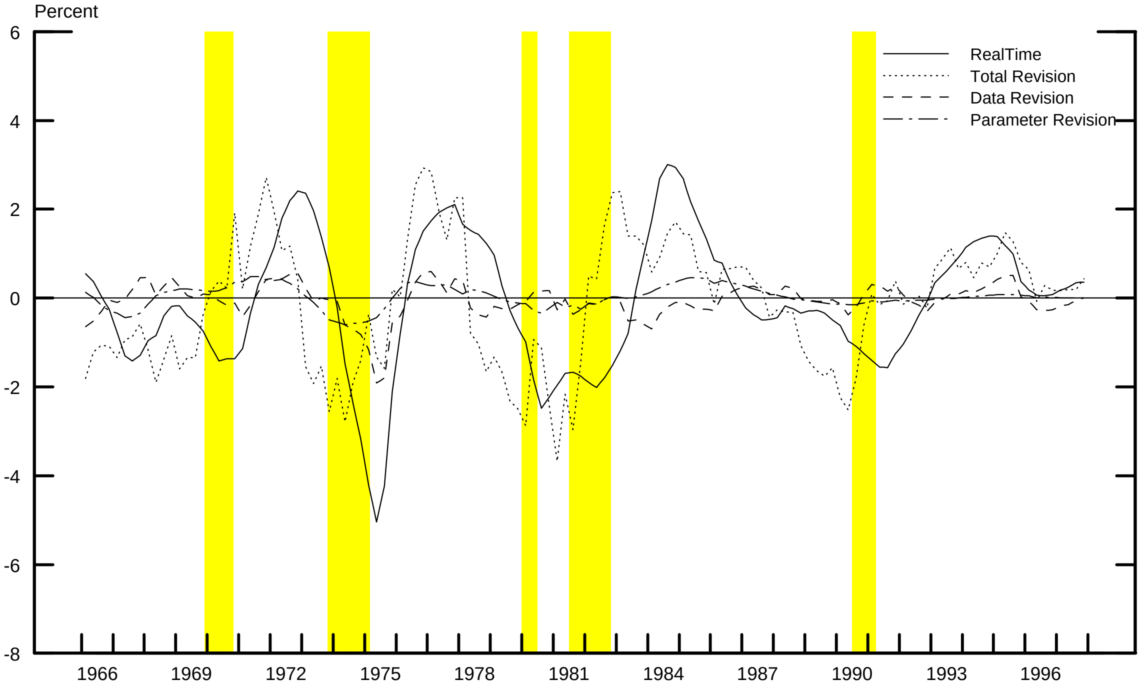
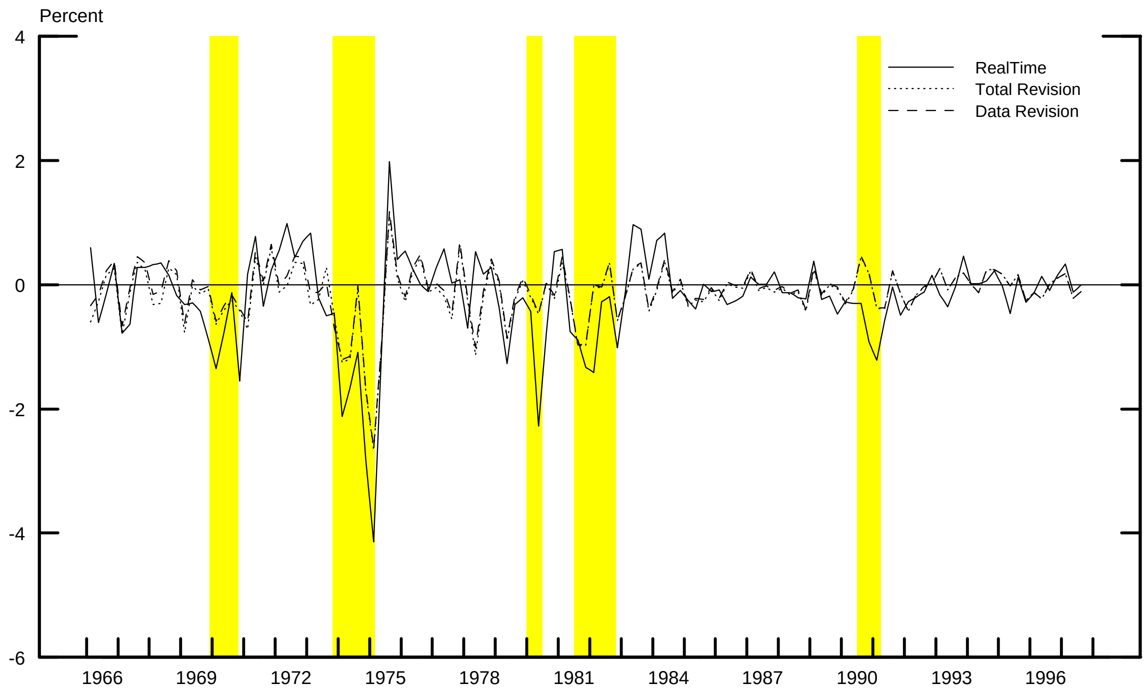


Figure 6

Estimated Business Cycle: Beveridge-Nelson



Estimated Business Cycle: Clark

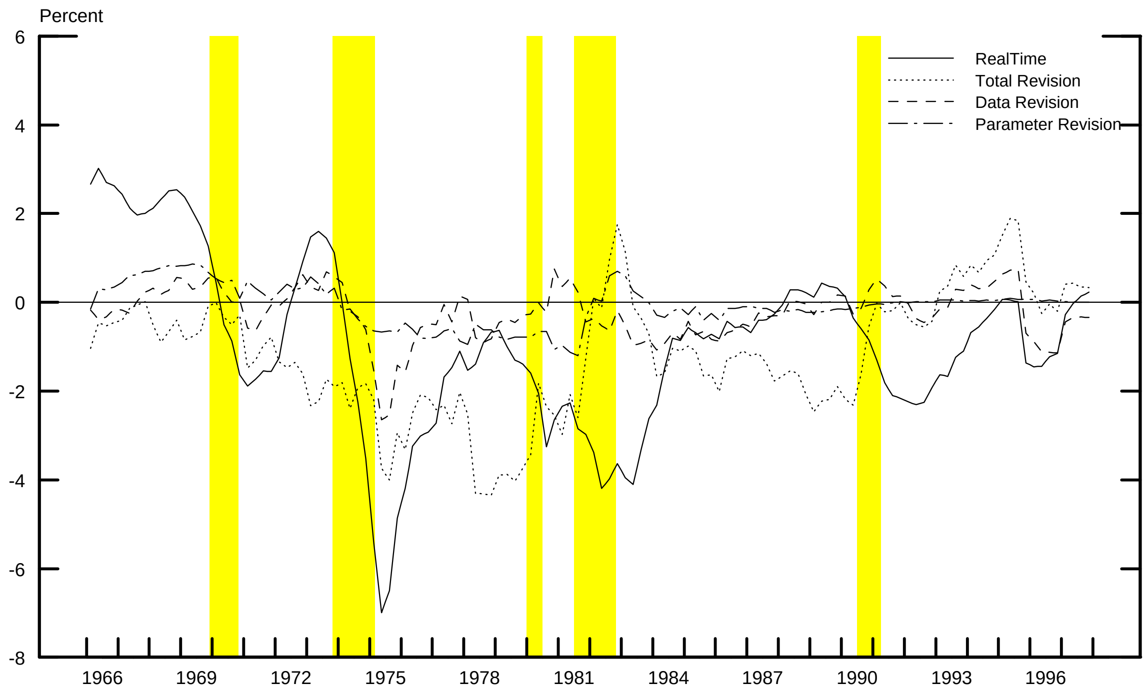


Figure 7

Real-Time 95% Confidence Interval and Final Estimates
Unobserved Component Models