The Review *of* Economics *and* Statistics

VOL. LXXXIV NOVEMBER 2002 NUMBER 4

THE UNRELIABILITY OF OUTPUT-GAP ESTIMATES IN REAL TIME

Athanasios Orphanides and Simon van Norden*

Abstract—We examine the reliability of alternative output detrending methods, with special attention to the accuracy of real-time estimates of the output gap. We show that ex post revisions of the estimated gap are of the same order of magnitude as the estimated gap itself and that these revisions are highly persistent. Although important, the revision of published data is not the primary source of revisions in measured output gaps; the bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the trend in output. Multivariate methods that incorporate information from inflation to estimate the output gap are not more reliable than their univariate counterparts.

I. Introduction

UNDERSTANDING macroeconomic fluctuations entails the study of an economy's output relative to its trend or potential level. The difference between the two is commonly referred to as the business cycle or the output gap. Although macroeconomic analysis often takes measurement of the output gap for granted, its construction is subject to considerable uncertainty. As a practical matter, empirical estimates of the output gap for any given method may not be particularly reliable. This may pose an acute difficulty for economic stabilization policy that requires reliable estimates of the output gap in real time when policy decisions are made.

Three distinct issues complicate measurement of the output gap in real time. First, output data may be revised, implying that output gaps estimated from real-time data may differ from those estimated from data for the same period published later. Second, as data on output in subsequent quarters become available, hindsight may clarify our position in the business cycle even in the absence of data revision. Third, the arrival of new data may instead make us revise our model of the economy, which in turn revises our estimated output gaps.

Received for publication August 11, 1999. Revision accepted for publication July 26, 2001.

* Board of Governors of the Federal Reserve System, and HEC (Montréal) and CIRANO, respectively.

We would like to thank Bryan Campbell, Andy Filardo and Tiff Macklem, as well as discussants and seminar participants at the Atelier économetrique de Montréal, the Centre for Growth and Business Cycle Research at the University of Manchester, the University of Ottawa, les Journées d'optimization de Montréal, the World Congress of the Econometric Society in Seattle, the meetings of the American Economic Association and Canadian Economics Association, the editor, and our referees for their comments. Professor van Norden would also like to thank the SSHRC and the HEC for their financial 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.

This paper investigates the relevance of these issues for the measurement of the output gap in the United States since the 1960s, using several well-known detrending methods.[1] For each method, we examine the behavior of end-of-sample output-gap estimates and of the revisions of these estimates over time. We also decompose the revisions into their various sources, including that due to revisions of the underlying output data and that due to reestimation of the process generating potential output.

Presuming that revisions improve our estimates, the total amount of revision gives us a lower bound on the measurement error thought to be associated with real-time output gaps. This is informative when and if we find that revision errors are relatively large, because we can conclude that the total error of these estimators must be larger still. Furthermore, our results are quite general; they apply regardless of whether output gaps are used to cyclically adjust budget balances, to forecast inflation, or for other purposes, and do not require a priori assumptions on the true structure of the economy or on the true time series model generating observed output.

II. Alternative Detrending Methods

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

$$q_t = \mu_t + z_t. \quad (1)$$

Some methods use the data to estimate the trend μ_t, and define the cyclical component as the residual. Others specify

[1] An early exposition of issues pertaining to estimating trends appeared in the inaugural issue of this Review (Persons, 1919). The potential quantitative relevance of the issues we investigate has been pointed out before. 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. Orphanides (1998, 2000) 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 (1999), Orphanides (1998, 2001), and Smets (1998), have elaborated on the policy implications of this issue. This study is the first to assess and decompose the measurement errors associated with several techniques and is the first to assess these techniques with real-time data. This issue also closely relates to investigations of uncertainty regarding estimation of the *unemployment gap*, that is, the difference between the actual rate of unemployment and estimates of the natural rate of unemployment. Staiger, Stock, and Watson (1997a,b) document that these estimates are very imprecise, which parallels the unreliability of the output gaps we discuss here.

a dynamic structure for both the trend and cycle components and estimate them jointly. We examine detrending methods that fall into both categories.

A. Deterministic Trends

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

Because of the noticeable downturn in GDP growth after 1973, another simple deterministic technique is a breaking linear trend that allows for the slowdown in that year. Our implementation of the breaking-trend method 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 (for example, 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.[2]

B. Unobserved-Components Models and the Hodrick-Prescott Filter

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

This framework can also nest smoothing splines, such the popular filter proposed by Hodrick and Prescott (1997) (the HP filter).[3] We implement the HP filter, following Harvey and Jaeger (1993) and King and Rebelo (1993), by writing it in its unobserved-components form. Assuming that the trend in (1) follows

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

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

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

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

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

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

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

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

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

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

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

C. Unobserved-Components Models with a Phillips Curve

Multivariate formulations of UC models attempt to refine estimates of the output gap by incorporating information from other variables linked to the gap. However, they also introduce additional sources of misspecification and parameter uncertainty, which may offset potential improvements. To examine this issue, we consider two models which add a Phillips curve to the univariate formulations described above: those of Kuttner (1994) and Gerlach and Smets (1997).

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

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

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

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

In each case the shock e_t is assumed i.i.d., mean-zero, and Gaussian. In the Gerlach-Smets model, e_t is also assumed uncorrelated with shocks driving the dynamics of the trend and cycle components of output in the model. Thus, by adding the Phillips curve, the Gerlach-Smets model introduces an additional six parameters that require estimation ($\{\phi_1, \ldots, \phi_5\}$ and the variance of e_t). The Kuttner model

[2] We also investigated alternatives, including ones with a break of unknown location and multiple breaks. Qualitatively, the results were similar for the other alternatives. We also used Bai-Perron tests to determine when an econometrician would have been able to detect the change in trend and obtained similar conclusions.

[3] The development of smoothing splines dates back to the work of Whittaker (1923) and Henderson (1924), and discussion of its use for measuring business cycles may be found in Orphanides and van Norden (1999).

also allows for a nonzero correlation between e_t and the shock to the cycle, η_t. Thus, it introduces eight additional parameters that require estimation ($\{\xi_1, \ldots, \xi_6\}$, the variance of e_t, and its covariance with η_t).

III. Data Sources and Revision Concepts

A. Data

Most of our data are taken from the real-time data set compiled by Croushore and Stark (2001); we use the quarterly real-time variables for real output from 1965:1 to 1997:4. Construction of the series and its revision over time is further described in Orphanides and van Norden (1999). We use 2000:1 data as final data, recognizing, of course, that "final" is very much an ephemeral concept in the measurement of output. To implement the bivariate models, we also use the quarterly rate of inflation in the consumer price index (CPI) as available in 2000:1. CPI data do not generally undergo a revision similar to that for output data. We therefore use this vintage of CPI data for all the analysis, allowing us to focus our attention on the effects of revisions in the output data.

B. Measuring the Revision of Output Gaps

We use our data with each of the detrending methods described earlier to produce estimated output-gap series. We apply each detrending method in a number of different ways in order to estimate and decompose the extent of the revisions in the estimated gap series.

The first of these estimates for each method simply takes the last available vintage of data (2000:1) and detrends it. The resulting series of deviations from trend constitutes our *final* estimate of the output gap corresponding to that method.

The *real-time* 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. That is, in every quarter we apply the detrending method with data as available during that quarter. Next, we use these different vintages to construct a new series, which consists of the latest available estimate of the output gap for each point in time. The resulting real-time estimate represents the most timely estimate of the output gap which could be constructed in real time using the method employed.

The difference between the real-time and the final estimate gives us the total revision in the estimated output gap at each point in time. This revision may 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 *quasi-real* estimate. The quasi-real estimate of the output gap is simply the rolling estimate based on the final data series. That is, the gap at period t is calculated using only observations 1 through t to estimate the long-run trend and the deviations around it. The difference between the real-time and the quasi-real 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 period.

For UC models, we further decompose the revision in the estimated gap by defining a *quasi-final* estimate. 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 to construct filtered and smoothed estimates of the output gap. For this class of models, smoothed estimates of the output gap are used to construct the final series, whereas filtered estimates are used for the quasi-final series.[4]

The difference between the quasi-final and the quasi-real series reflects the use of different parameter estimates (full-sample ones versus partial-sample ones) to filter the data. The extent of the difference will reflect the importance of parameter instability in the underlying UC model. The difference between the quasi-final and the final series reflects the importance of ex post information in estimating the output gap given the parameter values of the process generating output.[5]

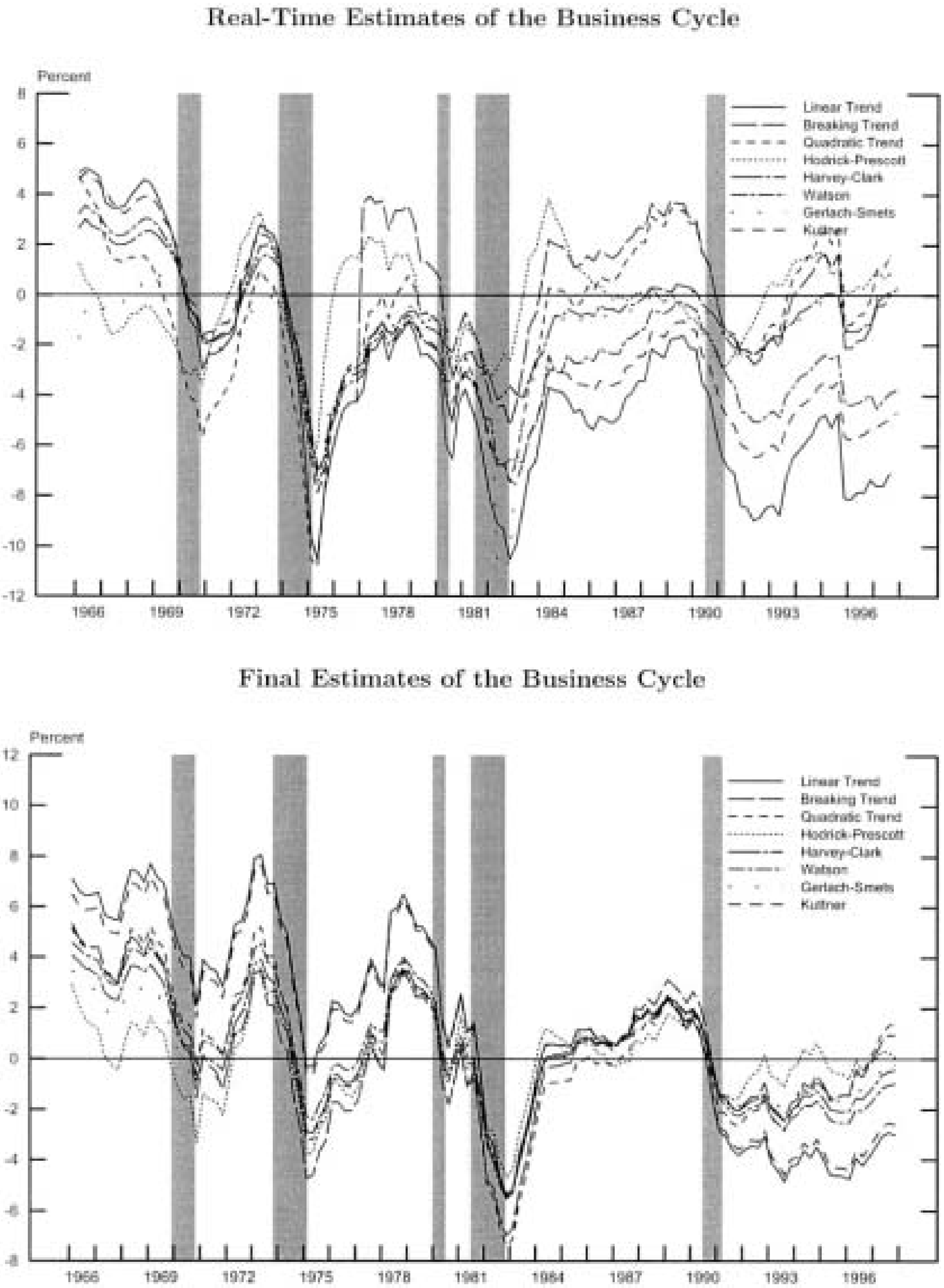
C. Standard Errors and Confidence Intervals

For the UC models, we compute standard errors and the corresponding Gaussian confidence intervals for the estimates of the output gaps and revisions. The Kalman filter and smoother provide estimates of the mean squared error associated with the quasi-final (filtered) and final (smoothed) estimates of the output gap. We use these to construct 95% confidence intervals for these estimates of the gap and for their revision. The Kalman-filter standard errors are appropriate for gauging the size of the final–quasi-final revisions, in that both estimates are conditioned on a given parameter vector. Because these standard-error estimates ignore the effect of parameter uncertainty on estimation of the gap, we also employ the approximation suggested by Ansley and Kohn (1986) to compute a comparable set of confidence intervals that capture this uncertainty. We use the Ansley-Kohn errors and confidence intervals to gauge the size of the total revisions. The Ansley-Kohn standard errors approximate the uncertainty associated with the final parameter estimates. In this respect, they are typical of the reliability calculations found previously in the output-gap literature. We stress, however, that these capture neither the effects of data revision nor the presumably greater parameter uncertainty found in the shorter samples available for estimation in real time. A test statistic can also be constructed, in the spirit of Diebold and Mariano (1995), of the

[4] In both cases, the UC model's parameters are estimated using the full sample of the same data, which is then used for filtering and smoothing. The sole exception is the HP filter, for which no parameters are estimated.

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

FIGURE 1.—ESTIMATES OF THE BUSINESS CYCLE



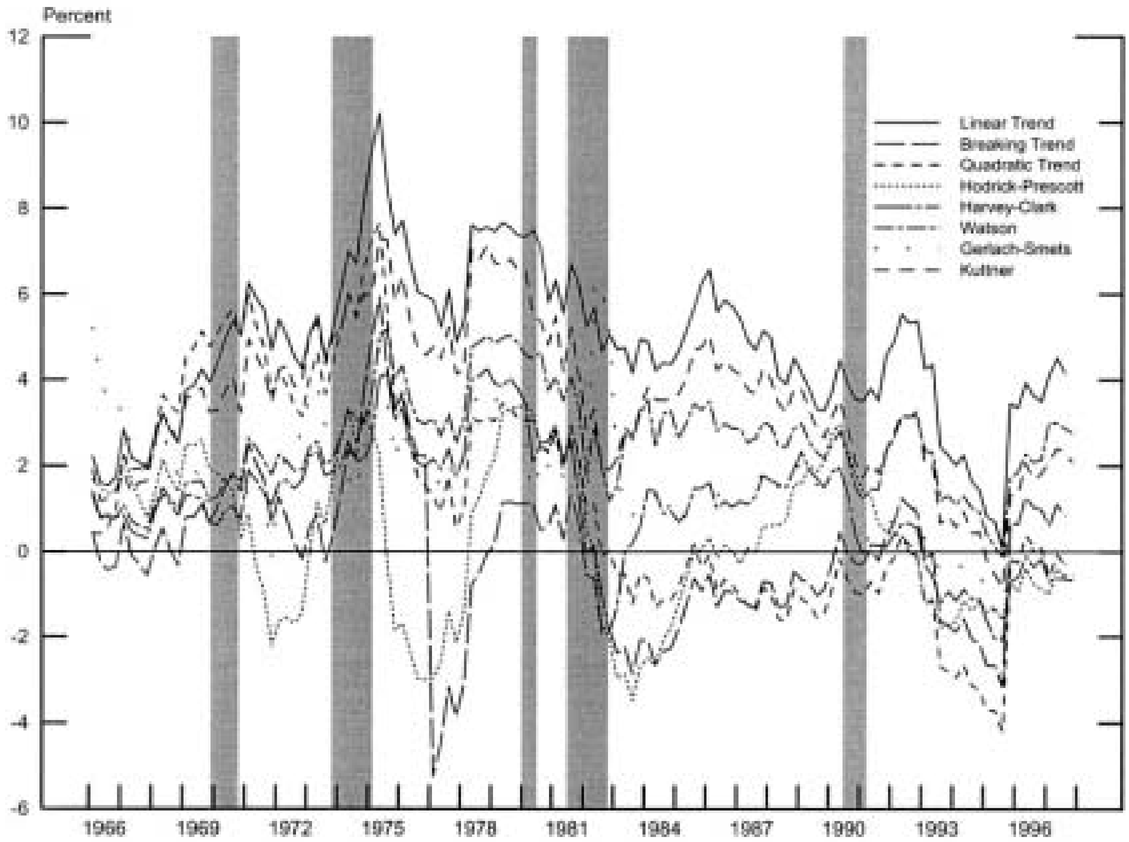
null hypothesis that the size of the revisions is consistent with the estimated confidence intervals. Details on these calculations may be found in the Appendix.

IV. Results

Figure 1 compares the estimated business cycles for the eight different methods mentioned in section 2. Real-time

FIGURE 2.

Total Revision in Business Cycle Estimates



estimates are shown in the upper half of the figure, and final estimates are shown in the lower half. The shaded regions reflect recessions as dated by the National Bureau of Economic Research (NBER). Several features are readily apparent. The different methods have strong short-term comovements; most appear to be moving upwards or downwards at roughly the same time. Further, the different methods typically give rise to a wide range of estimates for the output gap, though the range of final estimates is not as wide as the range of real-time estimates.

A. *Revision Size and Persistence*

Figure 2 shows the total revision in the output gap for each method, that is, the difference between the final and real-time estimates. Table 1 provides descriptive statistics on the various real-time, quasi-real, quasi-final, and final estimates, and Table 2 provides similar statistics for the total revision. 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. All the revision series are highly persistent, with coefficients ranging from 0.80 for the Gerlach-Smets model to 0.96 for the quadratic trend.

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 United States. For example, the revisions of Federal Reserve staff estimates of the output gap for the 1980s and early 1990s reported in Orphanides (1998) have a root mean square of 2.84%, compared to a standard deviation of 2.44% for historical estimates available at the end of 1994. The autocorrelation of those revisions also exceeds 0.8.[6]

[6] During the 1960s and 1970s, Federal Reserve staff relied on the Council of Economic Advisers estimates of potential output to construct estimates of the output gap. As shown in Orphanides (2000), the official estimates for the 1960s and 1970s, produced and published by the Council of Economic Advisors and Commerce Department, were subject to even greater revision errors. Of course, such comparisons should be interpreted with caution, as official estimates have been based on statistical methodologies that have evolved over time—presumably reflecting changes in beliefs about how best to estimate the output gap—and also incorporate judgemental considerations that cannot be fully captured with statistical methods.

TABLE 1.—OUTPUT-GAP SUMMARY STATISTICS

Method	*MEAN*	*SD*	*MIN*	*MAX*	*COR*
Hodrick-Prescott					
Final	0.04	1.65	−4.67	3.60	1.00
Quasi-real	−0.12	1.70	−3.96	3.79	0.55
Real-time	−0.27	1.90	−6.63	3.84	0.49
Breaking trend					
Final	0.18	2.58	−6.98	5.31	1.00
Quasi-real	0.56	2.79	−6.55	7.02	0.85
Real-time	0.21	3.15	−10.52	5.02	0.82
Quadratic trend					
Final	0.30	2.72	−7.39	5.20	1.00
Quasi-real	−0.70	2.71	−7.23	6.19	0.60
Real-time	−0.96	3.03	−10.83	4.70	0.58
Linear trend					
Final	1.30	3.87	−5.44	8.06	1.00
Quasi-real	−2.65	3.49	−10.32	7.02	0.88
Real-time	−3.45	3.98	−10.52	5.02	0.89
Watson					
Final	0.45	2.37	−5.34	4.56	1.00
Quasi-final	−0.26	2.19	−5.07	5.06	0.95
Quasi-real	−1.71	2.37	−7.31	4.42	0.83
Real-time	−2.08	2.61	−7.43	3.56	0.89
Kuttner					
Final	1.20	3.63	−5.52	7.69	1.00
Quasi-final	0.78	3.51	−5.61	6.92	0.99
Quasi-real	−1.63	2.79	−6.81	6.23	0.87
Real-time	−2.37	3.16	−7.91	4.86	0.88
Harvey-Clark					
Final	0.25	2.17	−5.51	4.06	1.00
Quasi-final	−0.71	1.53	−4.62	3.21	0.89
Quasi-real	−0.66	1.60	−4.14	3.41	0.81
Real-time	−0.93	1.91	−6.99	3.02	0.77
Gerlach-Smets					
Final	0.08	1.95	−5.37	3.51	1.00
Quasi-final	−0.57	1.55	−4.85	3.30	0.92
Quasi-real	−0.89	2.57	−13.17	1.95	0.56
Real-time	−1.57	2.08	−11.05	0.90	0.75

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. All statistics are for 1966:1–1997:4.

Table 3 presents some measures of the relative importance of the revision in each series. Column 1 presents the correlation between the final and real-time series for each method, which ranges from a low of 0.49 for the Hodrick-Prescott filter to a high of 0.89 for the linear-trend and Watson models. The next two columns, *NS* and *NSR*, provide two proxies for the noise-to-signal ratio in the real-time estimates. *NS* (*NSR*) is the ratio of the standard deviation (the root mean square) of the total revision to the standard deviation of the final estimate of the gap. *NSR* therefore captures the effects of persistent upward or downward revisions and exceeds 1 for six out of the eight methods reported.[7] Even the best methods have rather large ratios by these criteria.[8] The last column provides the frequency with which the real-time and final gaps were of opposite signs. For five methods this frequency exceeds 40%, and for the Kuttner and linear-trend models it is almost 50%. These results show that the errors associated with real-time estimates of the output gap are substantial. The ex post revisions are of the same order of magnitude as the ex post estimates of the gap, the estimation errors appear to contain a highly persistent component of substantial size, and the real-time estimates frequently misclassify the sign of the gap.

TABLE 2.—SUMMARY REVISION STATISTICS: FINAL VERSUS REAL-TIME ESTIMATES

Method	*MEAN*	*SD*	*RMS*	*MIN*	*MAX*	*AR*
Hodrick-Prescott	0.30	1.81	1.83	−3.48	3.44	0.93
Breaking trend	−0.04	1.78	1.78	−5.24	5.93	0.85
Quadratic trend	1.25	2.64	2.91	−4.20	7.65	0.96
Linear trend	4.78	1.82	5.12	0.09	10.21	0.91
Watson	2.53	1.17	2.78	−0.11	5.18	0.89
Kuttner	3.57	1.75	3.97	−0.83	7.29	0.92
Harvey-Clark	1.17	1.39	1.82	−2.07	4.25	0.92
Gerlach-Smets	1.64	1.43	2.17	−1.42	6.33	0.80

The detrending method and statistics are as described in the notes to Table 1. *RMS* denotes the root mean square of the revision series shown, and *AR* the first-order serial correlation of the series.

TABLE 3.—SUMMARY RELIABILITY INDICATORS

Method	*COR*	*NS*	*NSR*	*OPSIGN*
Hodrick-Prescott	0.49	1.10	1.11	0.41
Breaking trend	0.82	0.69	0.69	0.22
Quadratic trend	0.58	0.97	1.07	0.35
Linear trend	0.89	0.47	1.32	0.49
Watson	0.89	0.49	1.17	0.42
Kuttner	0.88	0.48	1.09	0.49
Harvey-Clark	0.77	0.64	0.84	0.34
Gerlach-Smets	0.75	0.73	1.11	0.41

The table shows measures evaluating the size, sign, and variability of the revisions for alternative methods. *COR* denotes the correlation of the real-time and final estimates (from Table 1). *NS* denotes the ratio of the standard deviation of the revision to that of the final estimate of the gap. *NSR* denotes the ratio of the root mean square of the revision to the standard deviation of the final estimate of the gap. *OPSIGN* denotes the frequency with which the real-time and final gap estimates have opposite signs.

B. *Decomposition of Revisions*

To help us understand the importance of different factors in accounting for the total revision for each method, in Figures 3 through 8 we plot the real-time estimate of the output gap together with its total subsequent revision and the components of that revision. Table 4 presents related summary statistics.

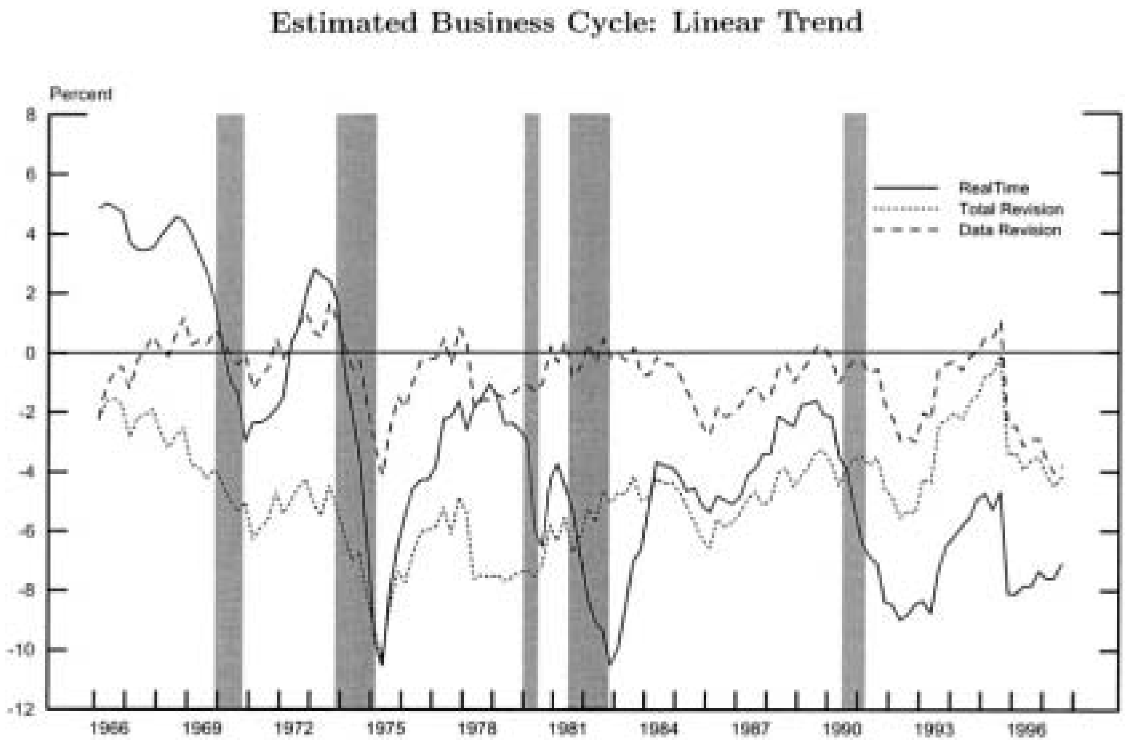
Figure 3 shows results for the linear trend. As a guide to subsequent figures for the other methods, we discuss this figure in some detail. First, compare the total revision with the real-time estimate. The fact that the revision is roughly equal to the real-time estimate at the trough of the 1975 recession tells us that our final estimate of the output gap is roughly zero. In other words, despite the extreme evidence of recession in the real-time estimate, ex post we would judge that the economy was operating roughly at potential at that time, by this method.

To understand the source of these revisions, the graph also shows the effects of data revision (measured as the real-time estimate minus the quasi-real estimate). For example, the total revision and data revision are roughly the same in both graphs in late 1995, which means that nearly

[7] The *NSR* value for the Federal Reserve staff estimates mentioned earlier is 1.16.

[8] Using the root mean square of the output gap as the benchmark for comparison yields similar conclusions. These alternative ratios can be constructed from Tables 1 and 2.

FIGURE 3.



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 $\pm 2\%$ of output, and its variability tends to be small compared to that of the total revision.

FIGURE 4.

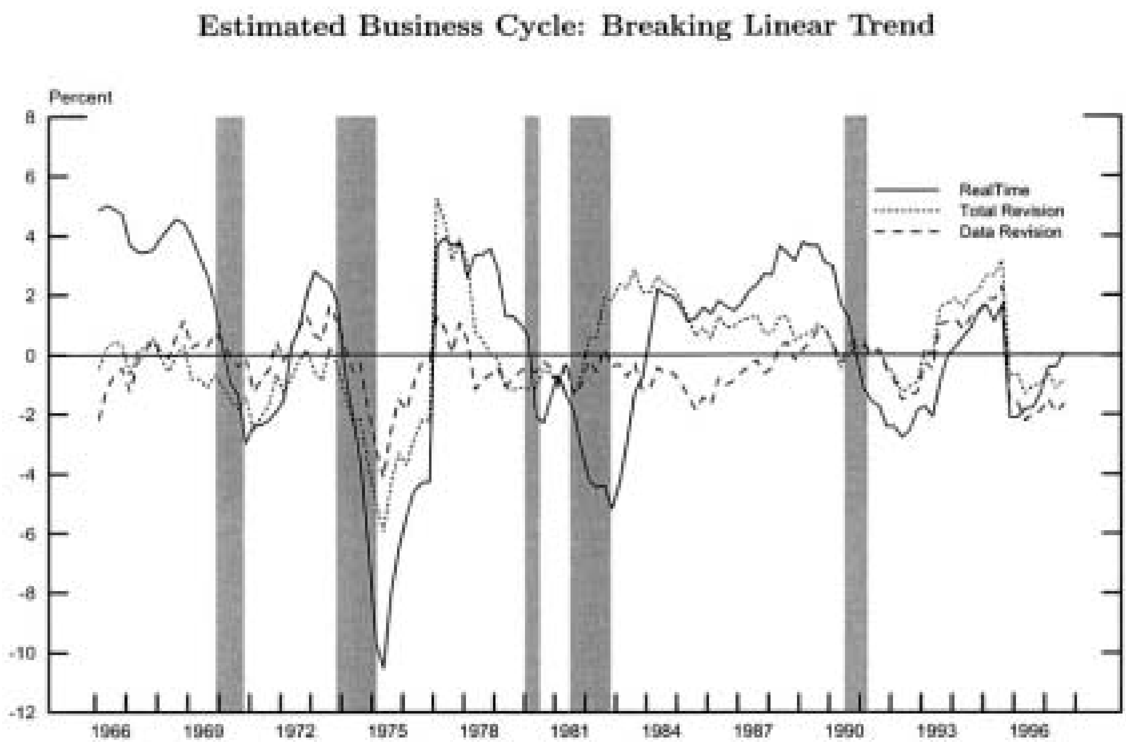
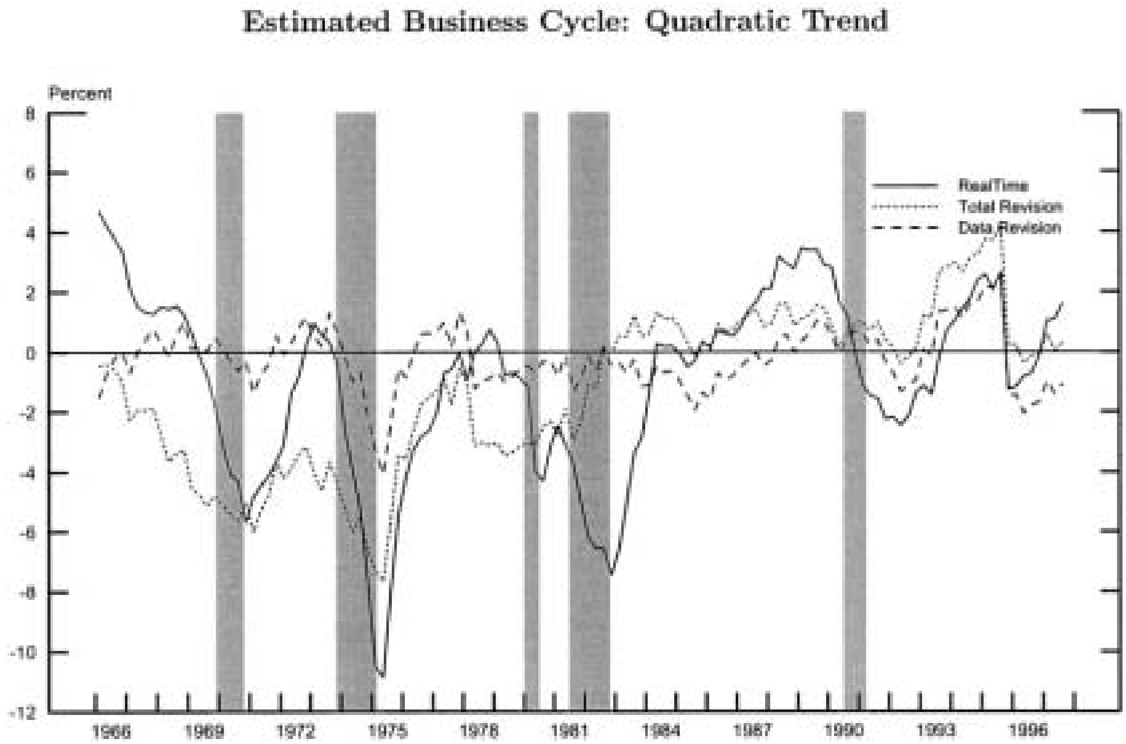


FIGURE 5.



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 quasi-real and real-time estimates of the output gap shown in Table 4.

Figures 4, 5, and 6 show results for the breaking-trend, quadratic-trend, and HP filter models. Again we note that

FIGURE 6.

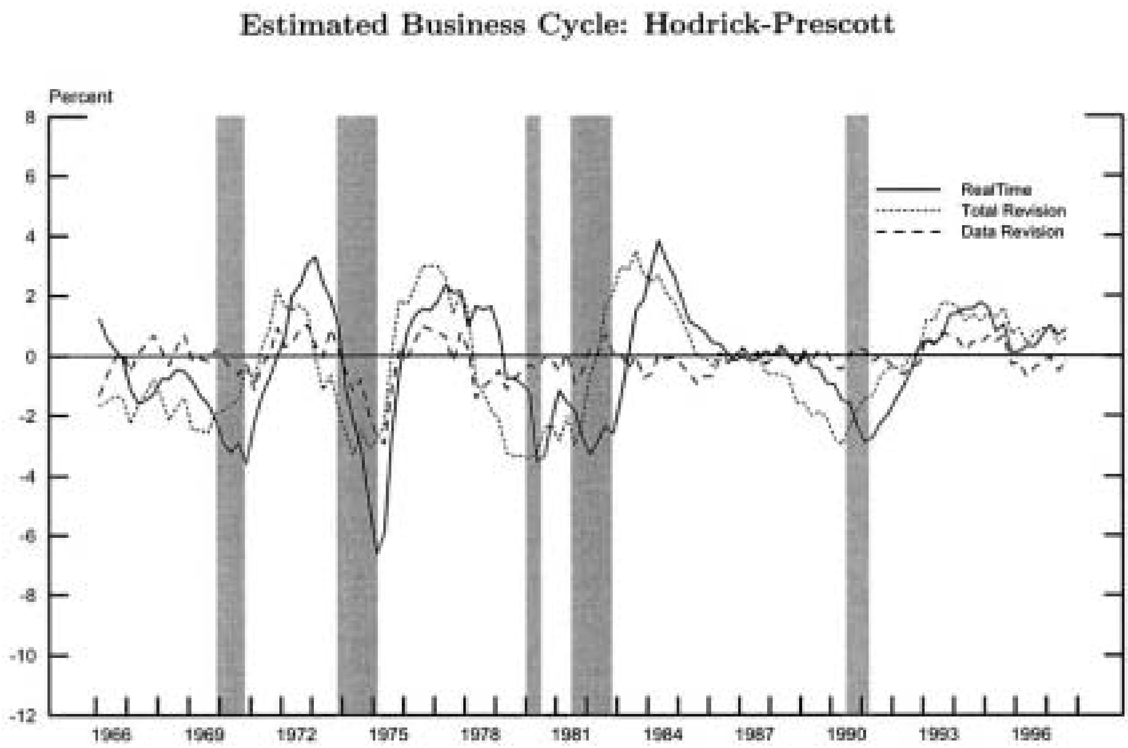
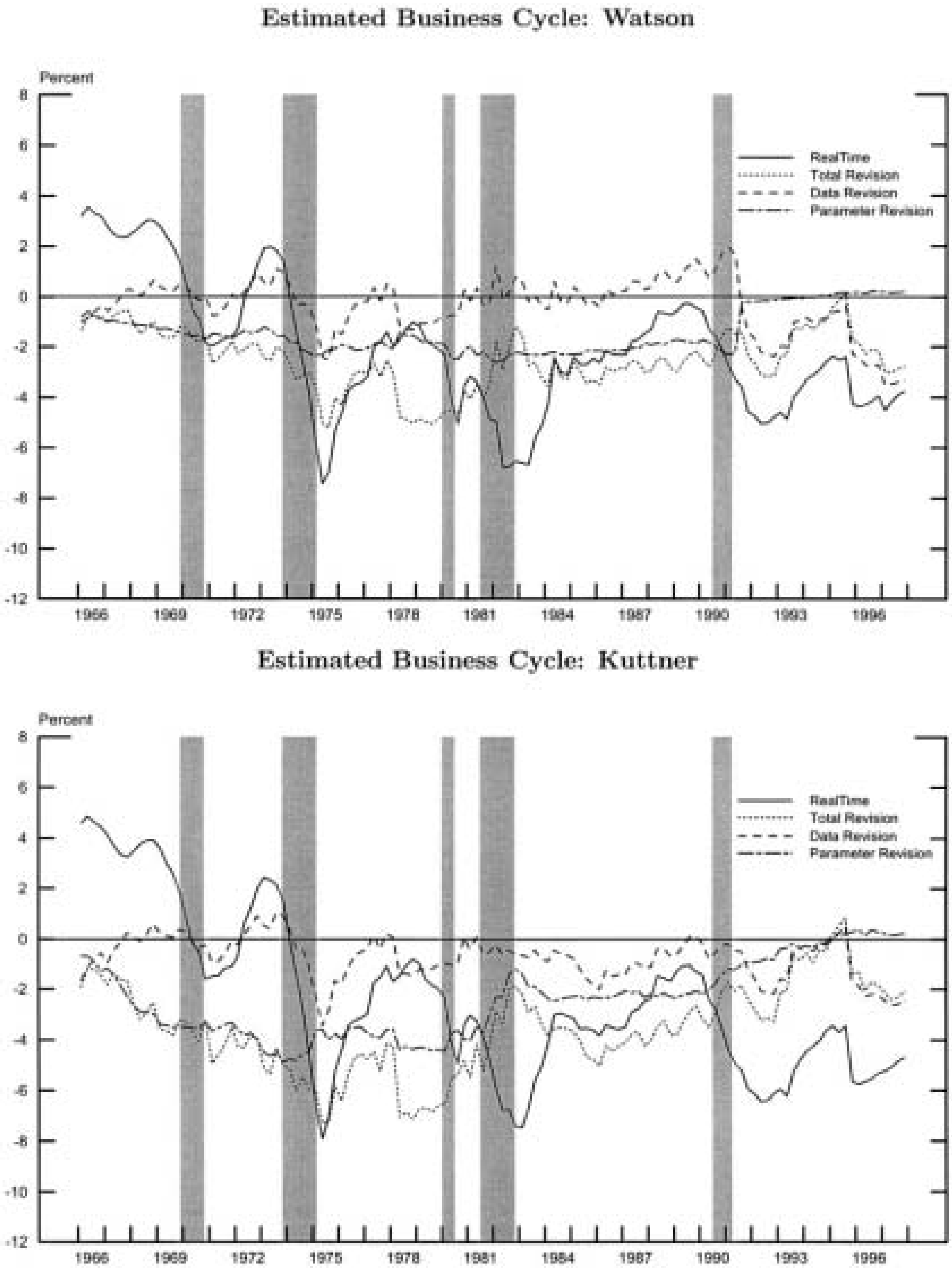


FIGURE 7.



the total revision is often close to the size of the real-time output gap. Further, although the data revisions seem to play a secondary role in explaining the total revision of the real-time estimates, some exceptions are notable.

Figures 7 and 8 show results from the four estimated UC models. The models are paired so that each figure shows results from a univariate model (upper panel) and its multivariate counterpart that incorporates information from a

TABLE 4.—DETAILED BREAKDOWN OF REVISION STATISTICS

Method	*MEAN*	*SD*	*RMS*	*MIN*	*MAX*	*AR*
Hodrick-Prescott						
Final–real-time	0.30	1.81	1.83	−3.48	3.44	0.93
Final–quasi-real	0.16	1.59	1.60	−3.49	3.12	0.97
Quasi-real–real-time	0.14	0.65	0.66	−1.05	2.95	0.66
Breaking trend						
Final–real-time	−0.04	1.78	1.78	−5.24	5.93	0.85
Final–quasi-real	−0.38	1.47	1.51	−3.96	1.99	0.92
Quasi-real–real-time	0.34	1.05	1.10	−2.30	4.14	0.76
Quadratic trend						
Final–real-time	1.25	2.64	2.91	−4.20	7.65	0.96
Final–quasi-real	1.00	2.44	2.63	−1.80	5.27	0.99
Quasi-real–real-time	0.23	1.04	1.06	−2.57	4.08	0.76
Linear trend						
Final–real-time	4.78	1.82	5.12	0.09	10.21	0.91
Final–quasi-real	3.95	1.81	4.34	0.07	6.43	0.96
Quasi-real–real-time	0.80	1.21	1.44	−1.67	4.14	0.79
Watson						
Final–real-time	2.53	1.17	2.78	−0.11	5.18	0.89
Final–quasi-final	0.71	0.75	1.03	−0.68	2.17	0.94
Quasi-final–quasi-real	1.45	0.85	1.68	−0.23	2.62	0.95
Quasi-real–real-time	0.37	1.13	1.19	−1.96	3.54	0.86
Kuttner						
Final–real-time	3.57	1.75	3.97	−0.83	7.29	0.92
Final–quasi-final	0.42	0.43	0.60	−0.63	1.29	0.91
Quasi-final–quasi-real	2.40	1.49	2.82	−0.39	4.86	0.97
Quasi-real–real-time	0.74	0.86	1.14	−1.06	3.45	0.83
Harvey-Clark						
Final–real-time	1.17	1.39	1.82	−2.07	4.25	0.92
Final–quasi-final	0.96	1.08	1.44	−1.06	3.23	0.94
Quasi-final–quasi-real	−0.05	0.37	0.37	−1.08	0.93	0.91
Quasi-real–real-time	0.27	0.61	0.66	−0.81	2.85	0.84
Gerlach-Smets						
Final–real-time	1.64	1.43	2.17	−1.42	6.33	0.80
Final–quasi-final	0.65	0.79	1.02	−0.88	2.57	0.93
Quasi-final–quasi-real	0.32	2.08	2.09	−3.48	8.73	0.69
Quasi-real–real-time	0.68	1.94	2.05	−7.88	5.67	0.61

See notes to Tables 1 and 2.

Phillips curve (lower panel). Figure 7 presents the Watson and Kuttner models. The two models provide somewhat similar real-time estimates of the gap. As with the models discussed earlier, the total revision is frequently close to the size of the real-time output gap, and the data revision only accounts for a small part of the total. Instead, changing parameter estimates play a large role and systematically revise potential output downwards.

The revisions of the Watson and Kuttner models resemble those of the linear-trend model seen in figure 3. This suggests that these models' performance suffers from their common assumption of a constant long-term trend in output growth. Given the secular decline in output growth over our sample, this assumption leads to persistent downward revisions in estimates of the "constant" trend rate of growth.

Note that the addition of the Phillips curve in the Kuttner model does not enhance the reliability of the output-gap estimates relative to the Watson model. The figure and Table 4 show that the total revision is both more biased and more variable for the Kuttner model than for the Watson model. Comparison of the standard deviation of the quasi-final–quasi-real revisions for the two models indicates that the error introduced by the estimation of the additional parameters required for the Kuttner model is substantial.

In Figure 8 we consider the results from the Harvey-Clark and Gerlach-Smets models. For the Harvey-Clark model, both parameter-revision and data-revision effects are relatively minor. In contrast, the Gerlach-Smets model exhibits much larger parameter revision, due in part to particularly severe parameter instability in the quasi-real estimates of the output gap. Again, the addition of the Phillips curve does not appear to enhance the reliability of the resulting output-gap estimates.

C. *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 when an accurate and timely estimate of the output gap (and its changes) would be of particular interest to policymakers. To help assess this, we calculated a number of descriptive statistics regarding the size or the revision in real-time 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 the real-time estimates at cyclical peaks, although the degree to which this is true varies considerably from one method to another. The linear-trend, Watson, and Kuttner methods make the most severe underestimates, and all but the breaking-trend method underestimate the gap by more than 1.5% on average.

D. *Revisions and Confidence Intervals*

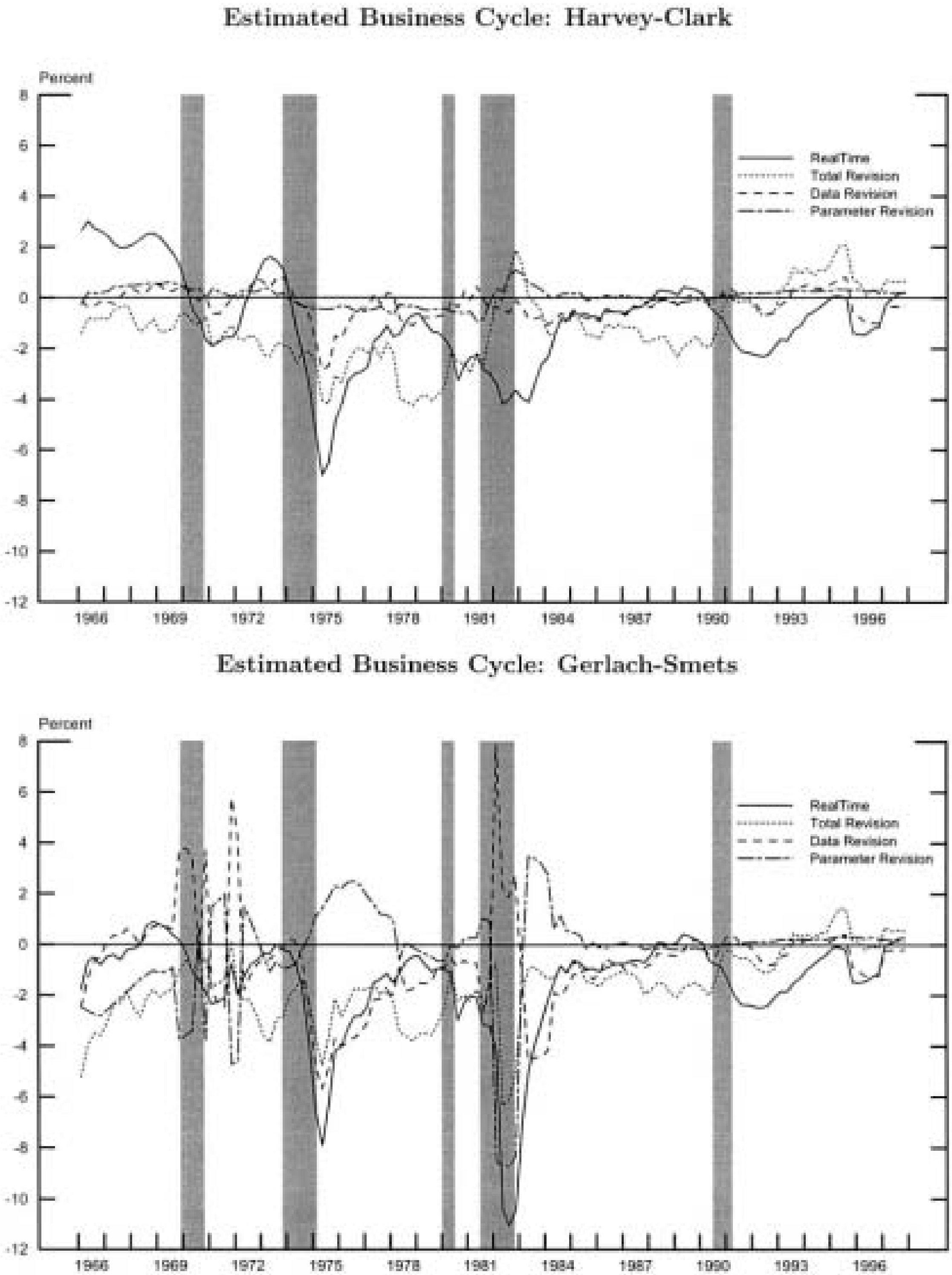
Figures 9 and 10 present the output-gap estimates and their confidence intervals from the four estimated UC models. The upper and middle panels show quasi-final and final estimates of the output gap with their corresponding 95% confidence intervals. The lower panel shows the final–quasi-final and total (final–real-time) revisions together with two sets of confidence intervals, which alternatively ignore (Kalman) and include (Ansley-Kohn) the estimated effects of parameter uncertainty.

Comparing confidence intervals for the Harvey-Clark and Gerlach-Smets models in figure 9, we see that the Kalman bands are somewhat narrower for the Gerlach-Smets model, but the Ansley-Kohn bands are considerably wider, on average.[9] This suggests that, in the absence of parameter uncertainty, incorporating information from the Phillips curve based on the final data helps narrow the uncertainty of the estimated output gaps. However, this narrowing is reversed when parameter uncertainty is taken into account.

Perhaps more importantly, both sets of confidence bands include zero in virtually every quarter from 1966 to 1997.

[9] For the 1966:1–1997:4 period shown in the figure, the average Kalman standard errors for the quasi-final estimates from the Harvey-Clark and Gerlach-Smets models are 2.32% and 1.93%, respectively. By contrast, the corresponding average Ansley-Kohn standard errors are 2.46% and 8.78%.

FIGURE 8.



This is true for both the final and quasi-final gaps and for both models. Thus, these gap estimates are virtually never significantly different from zero in this sample. The situation must be worse for real-time estimates, since these face additional effects of parameter uncertainty and data revision not allowed for in these bands.

Examining the revisions in the lower panel suggests that neither the final plus quasi-final nor the total revisions appear unusually large relative to their confidence intervals. This impression is confirmed by the results in table 6. The first two columns give the *RMS* revisions (final–real-time, from table 2) and the mean of the Ansley-

TABLE 5.—REVISION STATISTICS AT NBER PEAKS: FINAL VERSUS REAL-TIME ESTIMATES

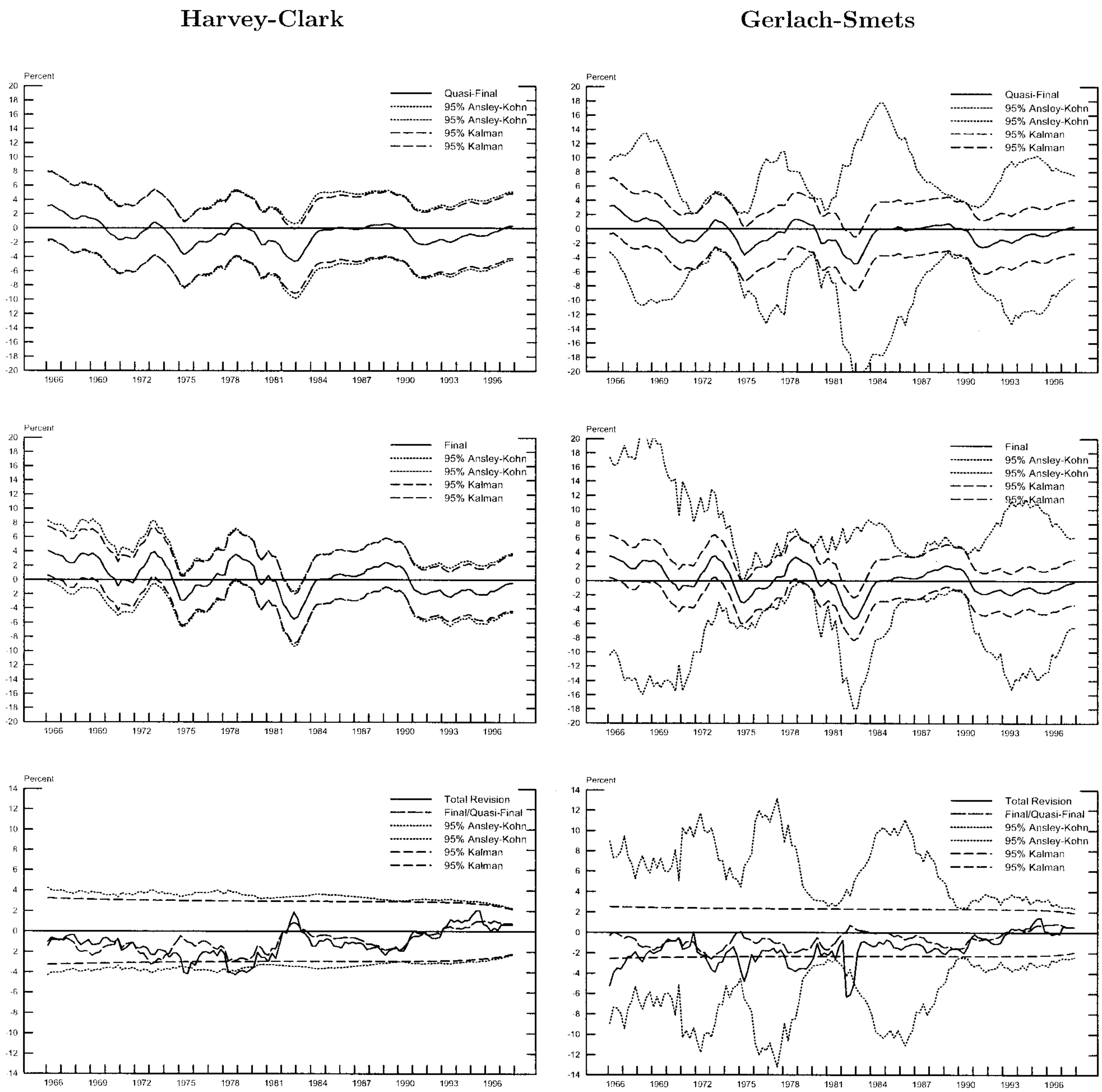
Method	*MEAN*	*SD*	*RMS*	*MIN*	*MAX*
Hodrick-Prescott	2.38	0.76	2.49	0.64	3.44
Breaking trend	0.67	0.55	0.86	−0.27	1.35
Quadratic trend	2.86	2.07	3.48	−0.95	5.20
Linear trend	5.40	1.38	5.56	3.57	7.50
Watson	2.83	1.25	3.08	1.19	4.86
Kuttner	4.37	1.29	4.55	2.11	6.58
Harvey-Clark	1.82	0.97	2.04	0.42	3.80
Gerlach-Smets	1.82	0.84	1.99	0.47	3.06

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 Tables 1 and 2.

Kohn standard errors for the revisions. The third column reports the test statistic for the null hypothesis that these revisions are consistent with these standard errors. The statistic is approximately normally distributed, so that rejection against the one-sided alternative that the revisions are larger than expected requires large positive values. For the Harvey-Clark and Gerlach-Smets models, it shows no significant evidence that the revisions are larger than one should expect.

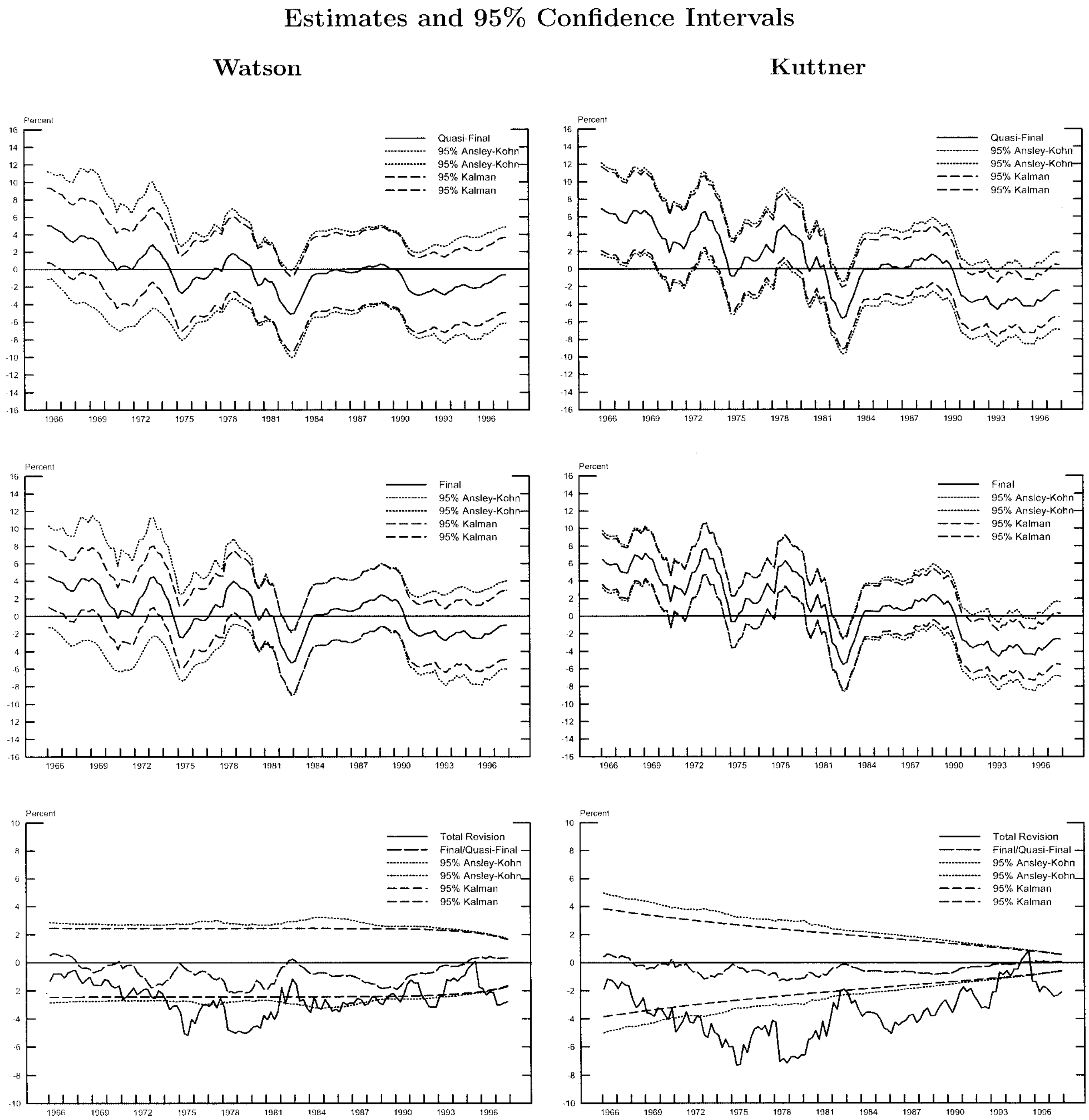
FIGURE 9.

Estimates and 95% Confidence Intervals



The upper and middle panels show the quasi-final and final estimates of the output gap, and the bottom panel shows the total and final–quasi-final revisions for the indicated UC models. Two sets of 95% confidence intervals are also shown. Kalman is based on the Kalman-filter variances assuming no parameter uncertainty. Ansley-Kohn is based on an approximation that also incorporates parameter uncertainty.

FIGURE 10.



See notes to Figure 9.

TABLE 6.—STANDARD ERRORS AND TEST FOR TOTAL REVISIONS

Method	Revision *RMS*	Mean *SE*: Ansley-Kohn	Revision Size Test Statistic
Watson	2.79	1.88	2.84*
Kuttner	3.98	1.32	3.23*
Harvey-Clark	1.82	1.75	0.07
Gerlach-Smets	2.18	3.20	−1.25

The root mean square (*RMS*) of the total revisions is from Table 2. The mean standard error (*SE*) and test statistic are computed for 1966:1–1997:4 as detailed in the Appendix. The test statistic is for the hypothesis that the size of the revisions is consistent with the estimated standard errors against the alternative that they are bigger, on average.

* Indicates rejection at the 0.1% significance level.

Figure 10 presents the corresponding estimates and confidence bands for the Watson and Kuttner models.[10] The gaps for the Watson model are (with few exceptions) not significantly different from zero. The Kuttner model gives much more evidence of significant output gaps, including, perhaps surprisingly, most of the first half of the 1990s.

The final–quasi-final revision falls within the Kalman bands in virtually every quarter for both models. However,

[10] The average Kalman (Ansley-Kohn) standard errors for the quasi-final estimates from the Watson and Kuttner models are 1.81% (2.42%) and 1.83% (2.20%), respectively.

the total revisions are frequently outside the Ansley-Kohn bands. This is reflected by the test statistics in table 6, which strongly reject the null in favor of the alternative that total revisions are more volatile than these standard errors predict.[11] This suggests that the calculated confidence intervals understate the degree of uncertainty associated with real-time estimates of the output gap for these two models.

V. Conclusions

We have examined the reliability of several detrending methods for estimating the output gap in real time. In doing so, we have focused on the extent to which output-gap estimates are updated 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 and give lower bounds on the estimation error associated with any given method.

Our results suggest that the reliability of output-gap estimates in real time tends to be quite low. The revisions are of the same order of magnitude as the estimated output gap itself for all the methods examined. The size of the measurement error is compounded by a high degree of persistence of the revisions. Although these results are based on a mechanical application of simple models, they mirror results based on the revision of output-gap series produced by the Federal Reserve staff during the 1980s and early 1990s.

For UC models, we find that multivariate methods that incorporate information from inflation to estimate the output gap are not more reliable than their univariate counterparts. Though the information from multivariate methods may be useful in principle, their added complexity introduces additional sources of parameter uncertainty and instability which may offset the potential improvement in real time.

Although important, the revision of published data does not appear to be the primary source of revisions for the methods we examined. Rather, the bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the output trend. Thus, even if the reliability of the underlying real-time data were to improve, real-time estimates of the output gap would remain unreliable.

Our findings suggest that output-gap mismeasurement may pose a serious policy problem, one that can be especially acute for economic stabilization policy. Policy experiments in macroeconometric models suggest that a strong systematic policy response to the output gap could greatly stabilize economic fluctuations—provided a reliable measure of the output gap is available for policymakers to use.[12] However, policy actions based on incorrect measures of the output gap can inadvertently cause instability. Policy design based on the erroneous presumption of reliability can lead to flawed policy recommendations.[13] In light of the unreliability of real-time estimates of the output gap, great caution is required in their use.

[11] The same was not true for test statistics (not reported) using final and quasi-final revisions and Kalman standard errors.

[12] See, for example, Taylor (1999) for a recent survey of policy evaluations of this nature.

[13] An informative illustration of this pitfall in the context of linear-quadratic-gaussian (LQG) models of optimal control is provided in Orphanides (1998).

REFERENCES

Ansley, C. F., and R. Kohn, "Prediction Mean Square Error for State Space Models with Estimated Parameters," *Biometrika* 73 (1986), 467–474.
Clark, Peter K., "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics* 102:4 (1987), 797–814.
Council of Economic Advisers, *Economic Report of the President* (Washington, DC: U.S. Printing Office, 1977).
Croushore, Dean, and Tom Stark, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* November (2001), 111–130.
Diebold, Francis X., and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–265.
Gerlach, Stefan, and Frank Smets, "Output Gaps and Inflation: Unobservable-Components Estimates for the G-7 Countries," Bank for International Settlements mimeograph, Basel (1997).
Hamilton, James, "A Standard Error for the Estimated State Vector of a State-Space Model," *Journal of Econometrics,* 33 (1986), 387–397.
——— , *Time Series Analysis* (Princeton University Press, 1994).
Harvey, Andrew C., "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics* 3 (1985), 216–227.
——— , *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, 1989).
Harvey, A. C. and A. Jaeger, "Detrending, Stylized Facts, and the Business Cycle," *Journal of Applied Econometrics* 8 (1993), 231–247.
Henderson, R. "A New Method of Graduation," *Actuarial Society of America Transactions,* 25 (1924), 29–40.
Hodrick, R., and E. Prescott, "Post-war Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking* 29 (1997), 1–16.
King, Robert G., and Sergio Rebelo, "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control* 17:1–2 (1993), 207–231.
Kuttner, Kenneth N., "Monetary Policy with Uncertain Estimates of Potential Output," *Economic Perspectives* (Federal Reserve Bank of Chicago) 16 (1992), 2–15.
——— , "Estimating Potential Output as a Latent Variable," *Journal of Business and Economic Statistics* 12:3 (1994), 361–368.
McCallum, Bennett, and Edward Nelson, "Performance of Operational Policy Rules in an Estimated Semi-classical Structural Model," in John Taylor (Ed.), *Monetary Policy Rules* (Chicago: University of Chicago Press, 1999).
Orphanides, Athanasios, "Monetary Policy Evaluation with Noisy Information," Federal Reserve Board Finance and Economics Discussion Series no. 1998-50 (1998).
——— , "The Quest for Prosperity without Inflation," European Central Bank working paper no. 15 (2000).
——— , "Monetary Policy Rules Based on Real-Time Data," *American Economic Review* 92:2 (2001), 115–120.
Orphanides, Athanasios, and Simon van Norden, "The Reliability of Output Gap Estimates in Real Time," Federal Reserve Board Finance and Economics Discussion Series no. 1999-38 (1999).
Persons, Warren M., "Indices of Business Conditions," *The Review of Economic Statistics* 1:1 (1919), 5–107.
Pfeffermann, Danny, and Richard Tiller, "Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters," manuscript (2000).
Quenneville, B., and Singh, A. C., "Bayesian Prediction Mean Square Error for State Space Models with Estimated Parameters," *Journal of Time Series Analysis* 21 (2000), 219–236.
Smets, Frank, "Output Gap Uncertainty: Does It Matter for the Taylor Rule?" BIS Working Paper no. 60 (1998).

St-Amant, Pierre, and Simon van Norden, "Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada," Bank of Canada technical report no. 79 (1998).

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

———, "The NAIRU, Unemployment and Monetary Policy," *Journal of Economic Perspectives* 11:1 (1997b), 33–49.

Taylor, John B., "The Robustness and Efficiency of Monetary Policy Rules as Guidelines for Interest Rate Setting by the European Central Bank," *Journal of Monetary Economics* 43:3 (1999), 655–679.

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

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

APPENDIX: CONFIDENCE INTERVALS FOR UC MODELS

A. Revisions

Let $S_{t|\tau}$ denote the estimate of the unobserved state vector S_t conditional on the parameter vector θ of the UC model as well as on all data available through time τ. Since we do not observe θ, we replace it by its maximum likelihood estimator $\hat{\theta}$. For convenience, we will refer to $S_{t|t}(\hat{\theta})$ as the *filtered* estimate $S_{t|t}$, and to $S_{t|T}(\hat{\theta})$ as the *smoothed* estimate $S_{t|T}$. For a given $\hat{\theta}$, the revision in this estimate may be defined as $R_{t|T} = S_{t|T} - S_{t|t}$.[14]

If we have $P_{t|t} = \text{Var}\,(S_{t|t} - S_t)$ and $P_{t|T} = \text{Var}\,(S_{t|T} - S_t)$, then it follows that

$$P_{t|t} = \text{Var}\,(S_{t|t} - S_t) = \text{Var}\,((S_{t|t} - S_{t|T}) + (S_{t|T} - S_t)),$$
$$P_{t|t} = \text{Var}\,(R_{t|T}) + P_{t|T} + \text{Cov}\,(S_{t|t} - S_{t|T}, S_{t|T} - S_t) \quad \text{(A.1)}$$

Provided this last term is zero, the variance of the revision must be

$$\text{Var}\,(R_{t|T}) = P_{t|t} - P_{t|T}. \quad \text{(A.2)}$$

$S_{t|t} - S_{t|T}$ will be orthogonal to $S_{t|T} - S_t$, because $S_{t|T}$ incorporates all information available up to time T.[15] In what follows, we use equation (A.2) as the basis of our calculations for the confidence intervals surrounding revisions $R_{t|T}$.

B. Standard Errors

If θ were known, $P_{t|t}(\theta)$ and $P_{t|T}(\theta)$ might be calculated using the usual Kalman-filter equations. In reality, however, we have only $\hat{\theta}$, and its estimation uncertainty therefore adds to the uncertainty in our estimated output gaps. We therefore require estimates of $P_{t|t}(\hat{\theta}) > P_{t|t}(\theta)$ and $P_{t|T}(\hat{\theta}) > P_{t|T}(\theta)$.

Hamilton (1986) suggests a Bayesian simulation approach to the problem. It draws n i.i.d. parameter vectors θ_i from a multivariate normal distribution $N(\hat{\theta}, \hat{\Sigma}_\theta)$,[16] then uses the simulated values of $(1/n) \sum_{i=1}^{n}(S_{t|t}(\hat{\theta}) - S_{t|t}(\theta_i))^2$ and $(1/n)\sum_{i=1}^{n}(S_{t|T}(\hat{\theta}) - S_{t|T}(\theta_i))^2$ as estimates of $P_{t|t}(\hat{\theta}) - P_{t|t}(\theta)$ and $P_{t|T}(\hat{\theta}) - P_{t|T}(\theta)$.[17] Alternatively, Ansley and Kohn (1986) suggest using the first-order approximation.[18]

$$P_{t|t}(\hat{\theta}) - P_{t|t}(\theta) = \left(\frac{d}{d\theta} S_{t|t}(\theta)|_{\theta=\hat{\theta}}\right)\hat{\Sigma}_\theta\left(\frac{d}{d\theta} S_{t|t}(\theta)|_{\theta=\hat{\theta}}\right)',$$
$$P_{t|T}(\hat{\theta}) - P_{t|T}(\theta) = \left(\frac{d}{d\theta} S_{t|T}(\theta)|_{\theta=\hat{\theta}}\right)\hat{\Sigma}_\theta\left(\frac{d}{d\theta} S_{t|T}(\theta)|_{\theta=\hat{\theta}}\right)'. \quad \text{(A.3)}$$

Quenneville and Singh (2000) stress that both of these methods are simply approximations, but their simulations suggest that the Ansley-Kohn approach performs better in small samples.[19] We experimented with both methods but found the Hamilton approach problematic in our case. For some parameter draws with some models, the output gap became explosive, generating extremely large standard errors or numeric overflows. For these reasons, the results we report are based exclusively on the Ansley-Kohn method.

As noted in equation (A.2), the implied variability of $R_{t|T}$ is simply the difference of the variances of the filtered and smoothed estimates. When θ is known, the Kalman filter guarantees that this difference is always positive semidefinite. The same cannot be said of the approximations above for $P_{t|t}(\hat{\theta})$ and $P_{t|T}(\hat{\theta})$; neither Hamilton's nor Ansley and Kohn's method guarantees $P_{t|t}(\hat{\theta}) > P_{t|T}(\hat{\theta})$.[20] However, a logical extension which guarantees this result is to simply use

$$\text{Var}\,(R_{t|T}) = (P_{t|t}(\theta) - P_{t|T}(\theta)) + \frac{d}{d\theta}(S_{t|T}(\theta) - S_{t|t}(\theta))|_{\theta=\hat{\theta}}\hat{\Sigma}_\theta \times \frac{d}{d\theta}(S_{t|T}(\theta) - S_{t|t}(\theta))'|_{\theta=\hat{\theta}}. \quad \text{(A.4)}$$

Accordingly, (A.4) is used to generate the implied confidence intervals for the revisions under the null hypothesis that $R_{t|T} \sim N(0, \text{Var}\,(R_{t|T}))$.

C. Test Statistics

Equation (A.4) allows us to construct a confidence interval for the revision at any specific point in time. We also wish to test whether the variability of $R_{t|T}$ over the entire observed sample is consistent with what we would expect from our UC model and its parameter estimates. One way to test this would be to standardize the revisions by their estimated standard errors $\sigma_{t|T} = \sqrt{\text{Var}\,(R_{t|T})}$ and test the variance of the resulting process. A problem here is that $R_{t|T}$ is almost certain to be serially correlated. We correct for this using a heteroskedasticity- and autocorrelation-consistent (HAC) estimator along the same lines as Diebold and Mariano (1995). Specifically, we construct the test statistic

$$D = \left(\frac{1}{T}\sum_{t=1}^{T}(R_{t|T}/\sigma_{t|T})^2 - 1\right)\omega^{-1}, \quad \text{(A.5)}$$

which has an asymptotic standard normal distribution under the null hypothesis that $E(R_{t|T}) = 0$ and $E(R_{t|T}{}^2) = \sigma_{t|T}{}^2$. Here, ω is a HAC estimate of the standard deviation of $(R_{t|T}/\sigma_{t|T})^2$. Approximate p-values for D were simply calculated using the standard normal cdf. For the results reported in Table 6, we computed the statistic using eight lags and a Bartlett kernel. We found similar results using a Parzen kernel and using lag truncation parameters from 5 to 10.

[14] If $\hat{\theta}$ is fixed at its full-sample (final) estimate, this corresponds to the revision from the quasi-final to the final estimate of the state vector.

[15] More generally, this would continue to hold if we replaced $S_{t|t}(\hat{\theta})$ with $S_{t|t}(\theta)$ and $S_{t|T}(\hat{\theta})$ with $S_{t|T}(\theta)$ for any arbitrary θ.

[16] $\hat{\Sigma}_\theta$ is simply the estimated variance-covariance matrix of $\hat{\theta}$ about its true value θ.

[17] See Hamilton (1994, section 13.7, pp. 397–399) for a more detailed exposition.

[18] See Harvey (1989, p. 149). These derivatives are typically not available in closed form, but may be easily computed numerically.

[19] Quenneville and Singh (2000) and Pfeffermann and Tiller (2000) both suggest more sophisticated approximations. However, the former's simulations show that their method works only about as well as Ansley and Kohn's, while the latter method is too computationally intensive to be practical in our context.

[20] In practice, this proved to be a problem for both methods, although the problem was more common for the Hamilton method.