W O R K I N G   P A P E R S   A N D   S T U D I E S

2003 EDITION

COPYRIGHT

Luxembourg: Office for Official Publications of the European Communities, 2003

ISBN 92-894-6855-6
ISSN 1725-4825
Cat. No. KS-AN-03-062-EN-N

Business cycle uncertainty in real-time

EUROPEAN COMMISSION

eurostat

€URO INDICATORS

Europe Direct is a service to help you find answers to your questions about the European Union

New freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2003

ISBN 92-894-6855-6
ISSN 1725-4825

4TH EUROSTAT AND DG ECFIN
COLLOQUIUM ON MODERN TOOLS FOR BUSINESS CYCLE ANALYSIS

"GROWTH AND CYCLE IN THE EURO-ZONE"

20 TO 22 OCTOBER 2003

Luxembourg, European Parliament
Hémicycle, Schuman building

Business cycle uncertainty in real-time

James Mitchell, *National Institute of Economic and Social Research, London, UK*

Business cycle uncertainty in real-time

James Mitchell*
National Institute of Economic and Social Research

21 September 2003

Abstract

In simulated out-of-sample experiments to the Eurozone this paper finds that not only are real-time point estimates of the output gap unreliable, but so are measures of uncertainty associated with them. This provides a serious challenge to users of output gap estimates.

1 Introduction

Policy makers require output gap estimates in real-time.[1] They do not have the luxury of being able to wait, say, to 2020 before deciding whether the economy is currently lying above or below its trend level. They have to decide, without the benefit of hindsight, whether a given change to output in the current period is temporary or permanent, that is whether it is a cyclical or trend movement.[2]

As shown by Orphanides & van Norden (2002) in an important study for the US economy, real time output gap estimates can be unreliable. The revisions associated with end-of-sample or real-time estimates can be considerable; indeed for the US they were found to be as large as the output gap estimates themselves. Clearly policy-makers misjudging the position of the business cycle in real-time can lead to sub-optimal policy-decisions; see Nelson & Nikolov (2001) and Ehrmann & Smets (2003). The findings of Orphanides & van Norden (2002) for the US, therefore, are worrying.

*Thanks to Martin Weale for helpful comments. This paper is part of a programme of work undertaken at the National Institute of Economic and Social Research in London with the support of EUROSTAT. Address for correspondence: James Mitchell, NIESR, 2 Dean Trench Street, Smith Square, London, SW1P 3HE, UK. Tel: +44 (0)207 654 1926. E-Mail: j.mitchell@niesr.ac.uk.

[1]Since our focus is on the "growth" business cycle rather than the "classical" cycle the "business cycle" and "output gap" are treated synonymously.

[2]In fact, in real time not only do policy makers apparently require current estimates of the output gap but future or forecasted values; see Schumacher (2002). In this paper we concentrate on obtaining in real-time current estimates of the output gap, and their uncertainty. In any case, as we argue below, one can view obtaining real-time estimates of the (current) output gap as a forecasting exercise.

This paper first considers a similar exercise to Orphanides & van Norden (2002) for the Eurozone economy.[3] Full sample or *final* estimates of the output gap are derived using data available over the (full) sample-period, 1971q1-2003q1. Then real-time estimates are computed recursively from 1981q1, in a simulated experiment designed to mimic real-time measurement of the output gap. So-called *data* revisions, explained by revisions to published GDP data, are not considered. Indeed this is expedient in an application to the Eurozone where construction of a 'real-time' data set is not readily possible. In any case, for the US Orphanides & van Norden (2002) find that these data revisions are less important than so-called *statistical* revisions. Statistical revisions are explained by the arrival of new data helping macroeconomists, with the advantage of hindsight, better understand the position of the business cycle, and also perhaps revising what model they use to identify and estimate the output gap.

Across a range of widely used univariate and multivariate estimators of the output gap we find significant differences between real-time and final output gap estimates for the Eurozone. As the future becomes the present output gap estimates are revised. It is not just in the US that real-time measurement of the output gap is difficult. This begs the question, should we be surprised by this unreliability? As forecasters know, the fact that forecasts are wrong does not mean they are misleading or useless. If within the bounds of what was expected the forecast can remain useful, similarly for output gap estimates which we view analogously to a forecast. A user of output gap estimates requires not just a real-time estimate, which we can view as analogous to a point forecast, but an indication of the degree of uncertainty associated with the output gap estimate. We should expect users of output gap estimates to care not just about one possible outcome but about the range of likely outcomes. Even if the point estimate proves to be revised considerably with the arrival of new data, estimates can still be useful to policy-makers if an accurate measure of the uncertainty associated with real-time estimates is provided. Appropriate contingency plans can be made if reliable measures of uncertainty are provided alongside point estimates. In fact, measures of uncertainty may be useful in their own right when analysing, for example, risk and volatility or the probability of a recession. Therefore, it need not be the case that the lower the *ex post* statistical revision to the output gap estimate the better the estimator, as is often implicit in studies evaluating alternative output gap estimates. There is what amounts to a trade-off between bias and variance.

Therefore in the second part of the paper we take up the challenge of providing measures of uncertainty associated with real-time estimates of the output gap. We construct both confidence intervals around the real-time estimates and density estimates.[4] Two measures of uncertainty are considered. The first, more traditional, approach relies on a state-space representation for the output gap estimator, and uses the Kalman filter recursions and an assumption of Gaussianity to estimate the covariance of the output gap.

[3] Related studies that examine real-time output gap estimates for the Eurozone are Runstler (2002) and Camba-Mendez & Rodriguez-Palenzuela (2003).

[4] Although Orphanides & van Norden (2002), (pp. 578-582), provide measures of uncertainty associated with their final (smoothed) and quasi-final (filtered) estimates they do not present them for real-time estimates.

The second quantifies the degree of uncertainty associated with the fact that future values of the (log) level of output (and other variables in multivariate measures of the output gap) are unknown but are known to affect real-time estimates. This is achieved simply by trying to capture the uncertainty about future values of the (log) level of output by forecasting the future repeatedly *via* simulation techniques. The cyclical component at time t is then derived by de-trending the observed data in the (log) level of output up to time t plus the forecasted future data. The dispersion of these estimates at time t provides an indication of the range of likely outcomes for the final output gap estimates. This approach not only provides one means of measuring the uncertainty associated with real-time estimates but can also improve the actual performance of real-time estimates. This is because the accuracy of real-time estimates is a function of how well future values of the underlying series can be forecast.

Both of these measures are then evaluated to indicate whether they offer a reliable indication of the degree of unreliability associated with output gap estimates. This is important, as otherwise all that can be said is the bands are wider for, say, output gap estimate A than estimate B. Nothing can be inferred about the appropriateness of the bands *per se*. We find that the real-time measures of uncertainty do not offer reliable indications of the degree of uncertainty associated with real-time estimates. This provides a serious challenge to users of output gap estimates. They must decide what do to given that there is mismeasurement not just of the output gap (point) estimates but their uncertainty too.

The plan of this paper is as follows. In section 2 the estimators of the output gap used in this paper are briefly described. The appendix provides further details of the precise specifications used. Section 3 then examines the reliability of real-time (point) estimates of the output gap in the Eurozone. To examine reliability twelve criteria, such as forecast accuracy as measured by RMSE, are used as diagnostics. Section 4 then considers how to measure the uncertainty associated with real-time estimates, and then evaluate them similarly to how point estimates are evaluated on the basis of their RMSE. Section 5 then re-visits the simulated real-time application to the Eurozone of Section 3. It indicates, and then evaluates, the degree of uncertainty associated with output gap estimates in real-time. Section 6 offers some concluding comments.

2 Alternative estimators of the output gap

The output gap, the difference between actual and potential output, is not observable. Various estimators have been proposed. Gerlach & Smets (1999) make the distinction between *statistical*, *structural* and *mixed* estimators. The statistical approach views the estimation of the output gap as a statistical decomposition of actual output into trend and cyclical components. This approach is univariate. The structural approach exploits economic theory to estimate the output gap, typically by using a production function to relate potential output to productivity and inputs of labour and capital; see Barrell & Sefton (1995) and *Economic models at the Bank of England* (1999). The data requirements for this approach are, however, restrictive since, for example, estimates of the capital shock are required. We consider the mixed approach.

2.1 A mixed approach to measuring the output gap: combining statistics and economics

The mixed approach combines elements of the structural and statistical approaches. Time series methods are used to study output data together with other data that economic theory suggests are closely related to the output gap. The Phillips Curve, for example, suggests that inflation data contain information about the output gap while Okun's Law suggests unemployment is important. These economic variables may contain useful information about the supply side of the economy and the stage of the business cycle. Output should not be detrended using output data alone.

Specifically we will consider the following class of multivariate estimators of the output gap: (1) Unobserved components models; (2) Multivariate Hodrick-Prescott filters and (3) Structural VAR models[5]. For details see Appendix A and Mitchell (2003).

3 Real-time analysis

The reliability of real-time output gap estimates for the Eurozone is examined by looking at the extent to which output gap estimates are revised with the arrival of new data. The *real-time* estimate of the output gap is constructed using data available only at the time an estimate is required and is the latest available output gap estimate for each point in time (in fact, given that we are not examining data revisions the real-time estimate, as defined here, is equal to the *quasi-real* estimate as defined by Orphanides & van Norden (2002)). That is, the real-time estimate at time t is calculated using observations 1 to t.[6] The *final* estimate of the output gap considers all available data (up to period T, where $t = 1, ..., t, ..., T$) and de-trends it. The difference between the real time estimate and the final estimate is the *total revision* to the output gap estimate at each point in time. This revision may have several sources and these can be decomposed further for unobserved-components models by defining the *quasi-final* estimate. This reflects the fact that UC models use the data in two ways. First, they estimate the parameters of the model, and secondly they use these estimates to obtain the smoothed estimates of the output gap that are the final estimates of the output gap. The filtered estimate of the cycle is called the quasi-final estimate. The difference between the quasi final and the real estimates reflects the use of different parameter estimates; it reflects the importance of *ex post* information in re-defining the values of the parameters. The difference between final and quasi final estimates reflects the importance of *ex post* information.

[5]Related approaches are the multivariate Beveridge-Nelson decomposition and the Cochrane approach.

[6]Given that official estimates on GDP are published at a lag strictly the output gap estimate at quarterly time period t is not available until period $(t + 1)$. Moreover, official data on GDP are subject to revisions.

3.1 Data used in examining the reliability of real-time estimates

Official Eurozone data for GDP, published by EUROSTAT, are available only from 1991. Unfortunately this does not offer a sufficiently long time-series for sensible business cycle analysis. Therefore we take the data from the ECB's Area Wide Model (AWM); see Fagan et al. (2001).[7] We use real GDP data (AWM code: YER). These data are available from 1970q1-2000q4. The data are then updated to 2003q1 using official data from EUROSTAT (via *New Cronos*).[8] All data are used in their seasonally adjusted form. This means the data have in fact been seasonally adjusted using full-sample information, that of course would not be available to policy-makers in real-time. So our results ignore this additional source of uncertainty.

For the multivariate estimators of the output gap, price, unemployment, consumption and investment data are required too. Revisions to these data are less important than for GDP data. Price data are the harmonised index of consumer prices [HICP] . Again data are updated from 2001q1 with data from New Cronos. Capacity utilisation in manufacturing, a survey based estimate, is available from New Cronos from 1985q1.

Calculations were performed using the GAUSS and `Ox` [see Doornik (1998)] programming languages. Use was made of the `SsfPack` module for `Ox` to perform many of the calculations concerning UC models; see Koopman et al. (1999). Application of a nonparametric dating rule to identify the classical business cycle found peaks at 1974q3, 1980q1, 1982q2 and 1992q1, and troughs at 1975q1, 1981q1, 1982q4 and 1993q1.[9]

3.2 Evaluating the reliability alternative output gap point estimates: the check-list

The reliability of output gap estimates is assessed against a check-list of 'economic' and 'statistical' criteria. It is has become increasingly common to seek benchmark against which alternative estimates of the output gap can be compared; see Orphanides & van Norden (2002), Camba-Mendez & Rodriguez-Palenzuela (2003) and Runstler (2002). It provides a means of objectively distinguishing between alternative estimates. One criterion not considered in this paper is the ability of output gap estimators to help forecast inflation.[10] The check-list consists of the following criteria:

[7]These data have been used in other studies of the output gap in the Eurozone such as Runstler (2002) and Camba-Mendez & Rodriguez-Palenzuela (2003).

[8]For robustness we did also consider Beyer-Doornik-Hendry [BDH] data from 1979q4-1999q3 for real GDP and inflation; see Beyer et al. (2001). Results were qualitatively similar to those reported.

[9]These results are unsurprisingly identical to those of Artis et al. (2003).

[10]The relationship between inflation and the output gap, commonly referred to as the Phillips Curve, is the basis for many models that characterise countercyclical stabilisation policy. The output gap measures the degree of excess demand in the economy; there is excess demand when the output gap is greater than zero. When the output gap is positive inflation is expected to rise. When the output gap is negative inflation is expected to fall. The monetary authorities can then set their policy instrument, typically the interest rate, in light of the latest output gap estimates. According to this theory output gap estimates, therefore, should have forecasting power over inflation. However, the degree of predictive power is an empirical issue. Indeed the relative ability of alternative output gap estimates to help forecast inflation

1. Basic summary statistics such as the mean, minimum, maximum and standard deviation for the final, quasi-final and real time estimates of the output gap. It is instructive to examine the ratio of the standard deviation of the real-time to final estimates. If the real-time data are an efficient forecast of the final data they should have a lower standard deviation than the final data; optimal forecasts are less variable than the item forecasted. Indeed, by construction for UC measures of the output gap this ratio should be less than unity in large-samples.

2. The correlation of the quasi-final and real time estimates with the final estimates.

3. Relation to known cyclical indicators. We would hope that good estimates of the output gap are well correlated with known cyclical indicators, such as the OECD's survey-based measure of capacity utilisation. Following Camba-Mendez & Rodriguez-Palenzuela (2003), we examine the correlation between our estimates of the output gap and capacity utilisation from 1985q1.

4. Properties of output gap estimates in the frequency domain. In the frequency domain we can isolate the relative importance of different frequency components. Good estimates of the output gap should exhibit cycles of duration between 6 and 32 quarters, in line with the conventional view of how long business cycles last. We compute the period (in quarters) at which the spectral density of the output gap exhibits a peak.

5. The persistence of the output gap is examined further by computing the autocorrelation coefficient (ACF) at lag one.

6. The output gap should be stationary. To test this hypothesis we consider KPSS tests in the level of the output gap; there is no reason to expect the output gap to be trending; see Kwiatkowski et al. (1992). The approximate asymptotic critical values are at the 10%, 5% and 1% levels, respectively, 0.347, 0.463 and 0.739.

7. As argued above, we can view the problem of obtaining real-time estimates of the output gap as a forecasting problem, since it is only with the arrival of new data that estimates of the output gap converge to their final values. Good real-time estimates should predict accurately the final value of the output gap. We measure the accuracy of the quasi-final and real time point estimates relative to the final estimates by their root mean squared error (RMSE). RMSE complements inference from the correlation coefficient calculated above; RMSE is sensitive to scale in contrast to correlation.

---

has proved a popular way to evaluate and rank alternative output gap estimates; e.g. see Camba-Mendez & Rodriguez-Palenzuela (2003). Results in Mitchell (2003) suggest that the worse performance of real-time estimates is primarily due to the difficulty of forecasting inflation, and not the unreliability of output gap estimates. This is consistent with the findings of Orphanides & van Norden (2003) for the US and Canova (2002) for the G-7 that it is the instability of the parameters of Phillips Curve equations that helps explain their poor performance relative to simple AR alternatives.

8. The proportion of times the real-time estimate predicts the sign of the final estimate correctly. We formally test for equality of the signs in both measures using the Pesaran-Timmermann test for directional change; see Pesaran & Timmermann (1992). This is a nonparametric test based on the number of correct predicted signs. The proportion of times the sign of the real-time series is the same as the final series we denote the hit-rate. The null hypothesis of the Pesaran-Timmermann test is that the forecast values (in our case the real-time estimates) have no ability to predict the sign of the outturn (in our case the final series). We report the p-value for this test.

9. Given the importance of correctly predicting turning points we will consider how well the real-time output gap estimates do at picking up business cycle turning points. Business cycle turning points are identified using a non-parametric dating rule; see Harding & Pagan (2002) and Artis et al. (2003). Application of the dating rule to the estimated output gap series yields a binary series with unity indicating a state of expansion, and zero a state of contraction.[11] Following Artis et al. (2003) when dating the deviation or growth business cycle, turning points are constrained so that peaks occur only when output is above trend, and troughs when output is below trend. We compute the correlation between the turning points of the final output gap and real-time output gap estimates.

10. The rationality of real-time estimates; see section B.3.

11. The *total* revision, i.e. the difference between the final and real-time estimates, is decomposed into *parameter* and *data* revision. Parameter revision is the difference between the filtered and real-time estimates, and data revision is the difference between the final and filtered estimates. Note that total revision is the sum of parameter and data revision. Summary statistics for the total revision, and its components, are computed such as the mean, minimum, maximum, standard deviation, first-order auto-correlation coefficient and the RMSE. We also compute the noise to signal ratio (NS), i.e. the ratio of the standard deviation of the total revision to the standard deviation of the final estimate.[12] If NS is greater than one then the revision process is deemed noisy.

12. The evolution of real-time estimates: how long do they take to converge to their final values? Orphanides & van Norden (2002) showed that the filtered and smoothed estimates differ in the middle of the sample. But we know that they will be increasingly similar, and eventually the same, as we get closer to, and eventually reach, the end of the sample. So although Orphanides & van Norden's (2002) result is useful in highlighting that we should expect considerable uncertainty (revision) in our end-of-sample (real-time) estimates, as there are important differences between real-time

[11] We would like to thank Don Harding for sending us GAUSS files to implement the Harding-Pagan rule and Tommaso Proietti for Ox files.

[12] Alternatively the RMSE could be considered.

and final estimates, it does not indicate how quickly this uncertainty should die out with the arrival of new data. Once we know this we can begin to quantify our degree of uncertainty with end-of-sample estimates that is due to the fact that future data are unknown. This issue relates to the two-sided nature of most cyclical estimators. We analyse this evolution by both simply plotting the output gap for each recursive sample and using the metric considered in Mitchell (2003) that indicates how long it takes, on average, for real-time estimates of the output gap to settle down at their "final" values.

Let us now examine in a simulated out-of-sample experiment how well the estimators reviewed in Section 2 perform on the basis of these 12 criteria.

3.3 The reliability of real-time estimates: results for the Eurozone

Full-sample, or final, estimates of the output gap are derived using data available over the full-sample period, 1971q1-2003q1. Real-time output gap estimates are computed recursively from 1981q1. This involves using data from 1971q1-1981q1, to provide an initial estimation period of 10 years to compute the real time estimate for 1981q1.[13] Then data for 1971q1-1981q2 are used to re-estimate the output gap (that involves re-estimation of the parameters of the models used to measure the output gap) and obtain real-time estimates for 1981q2. This recursive exercise, designed to simulate real-time measurement of the output gap, is carried on until data for the period 1971q1-2000q1 are used to estimate the real-time output gap for 2000q1. The last 3 years are excluded from the real-time simulation to allow for the fact that real-time estimates take time to converge to their 'final' values. This then lets us compare the real-time estimates with the final estimates that use information available over these, at least, three extra years.

To help render our results robust to the measure of the output gap chosen, as well as some representative *mixed* estimators of the output gap [see Section 2], we consider two univariate *statistical* measures. Seven alternative estimators are considered in total: (i) a Hodrick-Prescott filter (calculated by exploiting its state-space representation); (ii), a Harvey-Trimbur unobserved components cycle; (iii) a bivariate UC model; (iv) a bivariate HP filter; (v) a trivariate UC model; (vi) a trivariate SVAR model without cointegration; (vii) a trivariate SVAR model with cointegration. The first two measures are considered to provide univariate comparisons. The Hodrick-Prescott filter is perhaps the most widely used statistical approach for de-trending a time-series; we fix λ at 1600 as is common for quarterly data. Results for this filter provide an important benchmark. The Harvey-Trimbur cycle is an UC cycle, that is a generalisation of the class of Butterworth filters that have the attractive property of allowing smooth cycles to be extracted from economic time series - indeed ideal band pass filters emerge as a limiting case; see Harvey & Trimbur (2003).[14] The bivariate and trivariate estimators are discussed in Appendix A. For further

[13] As mentioned above we ignore the fact that GDP data are published with a one quarter lag.

[14] In the notation of Harvey & Trimbur (2003) we set $m = n = 2$.

discussion of the robustness of results to the chosen specification see Mitchell (2003).

Tables 1-5 summarise the real-time performance of the various UC measures of the output gap, while Tables 6-8 consider the SVAR based measures. As discussed, for the UC measures one can distinguish the filtered from the smoothed, and real-time estimates; this lets us examine the impact of parameter uncertainty on real-time estimates. The last row of the Tables, entitled "forecasted", is not discussed in this section; see Section 4.2.1 for details. Tables 9-16 then summarise some additional aspects of the revision to the real-time estimates. For the UC measures the source of revisions is decomposed into data and parameter revision. For the univariate HP filter, as the only parameter is chosen *a priori*, clearly there is no parameter uncertainty meaning that the filtered and real-time estimates are equivalent. Figures 1-10 then plot for each of the four measures of the output gap two graphs: (i) the real-time estimates alongside the final estimates and (ii) the evolution of the real-time estimates - the output gap is plotted for each recursive sample to indicate the changing view of the output gap. At this stage, it is instructive to focus on the first set of graphs to contrast the real-time and final output gap estimates.

Tables 1-8 reveal that the choice of which de-trending method we use does matter. There are differences between alternative measures of the output gap. The discrepancies show up especially in real-time. Model uncertainty, therefore, is an additional source of uncertainty associated with real-time estimates of the output gap. In fact, model uncertainty, in addition to parameter and data uncertainty, are the three sources of uncertainty affecting output gap estimates distinguished by the ECB in its October 2000 bulletin.[15]

However, across the measures the output gap estimates are highly persistent, as one should expect. This is reflected in first order autocorrelation coefficients often around 0.9 and the KPSS test sometimes indicating nonstationarity. We take this as a reflection of the power properties of unit root (and stationarity) tests, rather than an indication of genuine nonstationarity of the business cycle. Further evidence of considerable persistence is provided by the period of the business cycle, as estimated via the spectrum. Note that this measure of the length of a business cycle bears no theoretical relationship to measures based on the distance in quarters between cyclical turning points.

The correlation of the real-time estimates with the final estimates ranges from 0.2 for the univariate measures to around 0.5 for most of the multivariate measures. Only for the SVAR estimates (BQ and $p = 1$) is there a high correlation coefficient of 0.9; although in line with Camba-Mendez & Rodriguez-Palenzuela (2003), we believe this result is misleading being based on an implausible business cycle as shown by the summary statistics in Tables 6 and 14 that indicate, for example, cycle's with an amplitude in the range -0.7% to 0.5%, rather than around -2.5% to 2.5% as is typical for the other

[15]There is not a consensus on the appropriate specification of a model to identify the output gap. One simple way of optimally weighting or "pooling" alternative output gap estimates according to their reliability is to weight the alternative estimators according to the use to which the output gap estimates will be put. If one is interested in using the estimates for anticipating inflationary movements, then the weights are chosen such that the weighted output gap estimate offers the optimal forecast of inflation, in terms of RMSE. If one is interested in the fiscal balance, then the weights are chosen according to their predictive performance in this context. See Mitchell (2003) for details.

estimators.[16]

Tables 1-8 thus reveal that the real-time estimates can be quite misleading about the actual state of the economy. It is only with hindsight that a clear view about the state of the economy emerges. This picture is confirmed when one examines the graphs; see Figures 1-10. The real-time estimates had difficulty in correctly picking up the boom of the early 1990s. It is this inability, in real-time, to pick up booms and troughs that renders the job of policy makers especially difficult! Consistent evidence is provided by consulting the final column of Tables 1-8. The real-time estimates offer little or no explanatory power for turning points in the final series.

Nevertheless, it is encouraging that the move from univariate to multivariate measures of the output gap does lead to real time estimates better correlated with the final estimates. Adding 'economic information' appears to help. The univariate measures have correlation coefficients of 0.27 and 0.19, compared with coefficients of 0.46 and 0.69 for the bivariate UC models, 0.58 for the KPSW cycle and 0.46 for the trivariate UC model. The lower correlation, of 0.15, for the BQ cycle with $p = 16$ is attributable to the greater parameter uncertainty.

Parameter uncertainty, as mentioned, is an important source of the unreliability of real-time estimates. This shows up both in the higher correlation of the filtered estimates with the final estimates than the real-time estimates, and in Tables 9-16. These tables show that parameter revision systematically revises estimates of potential output. The filtered estimates often have correlation coefficients against the final estimates of around 0.9, much higher than those of the real-time estimates. This is particularly so for the multivariate measures of the output gap; e.g. while the filtered univariate HP estimate has a correlation of only 0.27 with the final estimates, the trivariate UC has a correlation coefficient of 0.91. Therefore, the problem for policy-makers is not so much not knowing the future values of the (log) level of output but not knowing what parameters to use when de-trending.

Further evidence of the unreliability of real-time output gap estimates is seen when one examines the revisions to real-time estimates. Consistent with the findings of Orphanides & van Norden (2002) for the US, revisions to the real-time output gap estimates are of the same order of magnitude as the estimated output gaps. Revisions are persistent and large. The standard deviation of the total revision is greater than that of the final estimates, i.e. $NS > 1$, except for some of the SVAR measures. A more favourable picture is painted about the reliability of the real-time estimates when one focuses on how well the real-time estimates do at picking up the sign of the final series, i.e. whether the economy was above or below its trend level. The real-time estimates have a 'hit-rate' reaching 80% for some measures. Apart from the SVAR estimates for BQ with $p = 16$, the multivariate measures of the output deliver more reliable real-time estimates of the sign

[16] SVAR estimates were found to be sensitive to the lag order chosen. For BQ for low lag orders the cycle has too small an amplitude and exhibits too little persistence. It is noteworthy that previous SVAR estimates of the output gap for the Eurozone are consistent with our results; e.g. see Camba-Mendez & Rodriguez-Palenzuela (2003). They use the SBC to choose the preferred lag length of the VAR model. In our experience a low value is selected, usually just one or two lags.

of the final estimates than the univariate measures. This comparison is strongest when comparing against the univariate HP filter, rather than the Harvey-Trimbur UC cycle. This improved performance for the multivariate measures is reflected by a higher hit-rate and more evidence, using the Pesaran-Timmermann tests, that the real-time estimates do have the ability to predict correctly the sign of the final estimates. These results are in line with Orphanides & van Norden (2002) for the US and Runstler (2002) for the Eurozone. They can be contrasted with the more positive results found by Camba-Mendez & Rodriguez-Palenzuela (2003) for the Eurozone.

The Mincer-Zarnowitz tests (column *rat*) suggest that revisions reflect news that is available at the time the real-time estimate is produced; the preliminary, or real-time, estimate can help predict the subsequent revision. This suggests that real time estimates are not rational expectations of the final estimates. This finding is also supported by the fact that the standard deviation of the real-time estimates is not always less than the standard deviation to the final estimates. The filtered estimates also are not rational; this is perhaps indicative of a small-sample problem.

We probe further at this revision by examining the evolution of the output gap estimates. Figures 2, 4, 6, 8 and 10 examine the changing views about the output gap by plotting for each recursive sample the output gap estimates. Focus on the panel entitled "smoothed" that plots the recursively computed smoothed estimates. This lets us see how quickly this revision occurs. The figures show that the revision can take some time. The dispersion of the estimates is greatest around the trough of the mid 1980s and the peak of the early 1990s; whether one thought the output gap was at a peak or trough changes considerably with the passage of time. We see, in particular for the Harvey-Trimbur cycle [see Fig. 4] that it is only around 3 years after the boom of the early 1990s that the boom is correctly identified. Indeed, see Figure 4, in mid 1992 the output gap still appears negative, rather than close to a peak of 2.5% as the final estimates indicate. The slow speed of convergence to the final estimate is seen further when one considers the metric considered in Mitchell (2003) that indicates how long it takes, on average, for real-time estimates of the output gap to settle down at their "final" values. For the Harvey-Trimbur cycle, for example, while 50% of the total revision is completed in 10 quarters, it takes 40 quarters to complete 90%. The picture is similar for the other measures of the output gap. Even 10 years after the event there is still some revision to output gap estimates, although the figures referred to earlier indicate that this may not prove qualitatively important in the sense that after about 3 years the general shape of the final output gap series becomes apparent. Certainly for all measures, the trough of the mid 1980s does not become apparent until a few years after the event and the peak of the early 1990s although detected by some of the multivariate measures fairly quickly is over-estimated by, say, the trivariate UC measure; in 1992-3 a peak in 1991 of 8%-10% is suggested, although it later proves to be only 2%; see Figure 10.

Real-time point estimates of the output gap in the Eurozone, as in the US, therefore appear unreliable, in the sense that there is a large and significant revision error. But could we have anticipated this error? In real-time was the revision error to the real-time estimate within the bounds of what we could have predicted? To address these questions we require measures of uncertainty associated with the real-time output gap estimates.

4 Uncertainty associated with output gap estimates: density estimates of the output gap

To capture fully the uncertainty associated with the real-time estimates, or forecasts, of the output gap, we construct density forecasts. Density forecasts of the realisation of a random variable at some future point in time provide an estimate of the probability distribution of the possible future values of that variable. In contrast, so-called "interval" and "event" forecasts provide specific information on forecast uncertainty that can be derived from the density forecast; interval forecasts specify the probability that the actual outcome will fall within a given interval while event forecasts focus on the probabilities of certain events, such as the probability of recession. Density forecasts of inflation in the U.K., for example, are now provided each quarter both by the Bank of England in its "fan" chart and the National Institute of Economic and Social Research (NIESR) in its quarterly forecast. Density forecasts inform the user of the forecast about the risks involved in using the forecast for decision making. Indeed, interest may lie in the dispersion or tails of the density itself; for example inflation targets often focus the attention of monetary authorities to the probability of future inflation falling outside some pre-defined target range while users of growth forecasts may be concerned about the probability of recession. Moreover, volatility forecasts, as measured by the variance, and other measures of risk and uncertainty, can be extracted from the density forecast; see Tay & Wallis (2000) for a review.

Questions then arise over how the density forecasts should be constructed. We consider two approaches. The first relies on a state-space representation for the output gap estimator while the second quantifies the degree of uncertainty through forecasting. Let us consider each of these measures in turn before in Section 4.3 considering how one can evaluate them.

4.1 Measuring uncertainty using the Kalman filter

Traditionally when using unobserved-components based measures of the output gap two measures of uncertainty associated with the cyclical component of output are distinguished: filter and parameter uncertainty. Conditional on Gaussianity (of the disturbances driving the components of the state vector) confidence intervals around the output gap then can be presented given knowledge of the covariance matrix of the estimated state vector.[17] This is the approach followed, for example, by Orphanides & van Norden (2002). For a review see Appendix B. As Orphanides & van Norden (2002) acknowledge these measures of uncertainty do not capture the effects of data revision nor the greater

[17] The Kalman filter recursions automatically return estimates of the covariance matrix of the state vector; see Harvey (1989). The diagonal elements of these matrices then can be used to construct the confidence intervals and density estimates.

parameter uncertainty associated with estimation in real-time using shorter samples.[18] Nevertheless, it is important to evaluate how well the implied interval and density forecasts perform relative to the outturn (i.e. the final estimate). Using the notation of Appendix B, let $P_{t|t}$ denote the covariance matrix of the output gap at time t, $(y_{t|t} - y^*_{t|t})$, calculated at time t (the output gap is one of the elements of the state vector). Then conditional on Gaussianity confidence intervals can be derived, as can density estimates. The Gaussian density is $N((y_{t|t} - y^*_{t|t}), P_{t|t})$. In Section 4.3 we examine how to evaluate these interval and density forecasts once the final output gap estimate, $(y_{t|T} - y^*_{t|T})$ where $T \rightarrow \infty$, has become available.[19]

4.2 Measuring uncertainty by repeatedly forecasting the future

We quantify the degree of uncertainty associated with the fact that future values of the (log) level of output (and other variables in multivariate measures of the output gap) are unknown but are known to affect real-time estimates. Quantification is achieved simply by trying to capture the uncertainty about future values of the (log) level of output by forecasting the future repeatedly, say R times, *via* a bootstrap procedure that allows for parameter uncertainty in real-time estimates. The cyclical component at time t is then derived by de-trending the observed data in the (log) level of output up to time t plus the forecasted data for periods $(t+1), (t+2), ..., (t+h)$. Given the R set of forecasts from $(t+1)$ to $(t+h)$, R estimates of the output gap at time t can be computed. The dispersion of these estimates at time t provides an indication of the range of likely outcomes for the final output gap estimates, $(y_{t|T} - y^*_{t|T})$, that can be computed once actual data from period $(t+1)$ to T are published. The R estimates of the output gap can be used to construct confidence intervals (around say the median estimate) and density estimates.

Questions then arise over how the density forecasts should be constructed *via* forecasting. What forecasting model should be used, should this model change over time, how should we account for uncertainty, how far ahead should we forecast etc.? We do not take a comprehensive look at these important questions here; we focus on particular models and ways of deriving the density forecasts that are merely illustrative. Uncertainty about the parameter estimates, stemming from the knowledge that with the arrival of new data the parameters will change, is quantified by in each case generating repeated forecasts from the model using a bootstrap approach. Alternatives are to look at the past revision error variance.

Specifically, for univariate measures of the output gap we use the bootstrap proce-

[18] An alternative approach is adopted by Fabiani & Mestre (2001). Their measure of uncertainty is based on the use of stochastic simulation. Although they provide confidence intervals they are not evaluated in any way making it difficult to know what they measure and how they should be used.

[19] We do not present measures of uncertainty associated with SVAR estimates. Note that since SVAR models can be interpreted within a state-space set-up it should prove possible to define similar measures of uncertainty for the SVAR models, as the UC models considered in this paper. An alternative means of quantifying the uncertainty associated with SVAR output gap estimates would be to undertake a stochastic simulaton. Experience suggests that this would deliver very wide confidence bands.

dure of Pascual et al. (2001) to study the impact of parameter uncertainty on prediction densities; see Mitchell (2003) for more details. Parameter uncertainty is expected to contribute significantly to any unreliability that may be associated with real-time estimates. Structural changes in the economy, for example, will cause parameter uncertainty. As structural changes often cannot be detected until some time afterwards, it is only with the arrival of additional data that parameter uncertainty is expected to decrease. Forecasts are produced from ARMA models. Although the parametric methods of measuring the output gap could themselves be used directly for forecasting, the motivation for considering ARMA models is that they can be rationalised as the reduced form of unobserved components models. For example, an ARIMA(0,2,2) model in the log-level of GDP corresponds to the reduced form of a local linear trend UC model; see Harvey (1989), p. 180. Forecasting via the reduced form, rather than structural model, is more agnostic; indeed, forecasts from atheoretical models have been found often to outperform those from structural models; e.g. see Clements & Hendry (1999).

We focus on ARMA models estimated in the second difference of the log-level of output, say y_t. We experiment both forecasting from ARMA models where the number of lags is chosen recursively using some information criterion and using fixed lags chosen *a priori*. Results below are presented for an ARMA(0,2) model in the second differences of output. For muultivariate models of the output gap we require forecasts not just of the log-level of output (GDP) but the additional economic variables. The same procedure as documented for univariate models is followed except that a VAR(p) model is considered. A VAR(p) model, with large p, can be seen as an approximation to an underlying VARMA process. In fact, we chose p using the Bayesian information criterion, that is known to select a parsimonious model. However, parsimonious models are potentially more robust to unforeseen structural breaks; e.g. see Clements & Hendry (1999). Future work will consider alternative forecasting models.

Even having decided upon a forecasting model, or perhaps a collection of models, how far ahead should one forecast? Results, undoubtedly, will be sensitive. If one was concerned in only forecasting, or perhaps we should say "now"casting, current GDP, since it is available at a one quarter lag, so-called FLASH estimators could be used that exploit, for example, industrial production data since they are published within the quarter. We, however, consider forecasting further ahead. We limit discussion to forecasts four quarters ahead. An alternative is to choose the forecast horizon with reference to how far ahead the chosen filter looks so that a two-sided filter can be applied even at the end of the sample of 'hard' data; see Harvey & Koopman (2000) and Koopman & Harvey (2003).

4.2.1 Can forecasting also deliver improved real-time (point) estimates?

Forecasting not only provides one means of measuring the uncertainty associated with real-time estimates but may help us improve the accuracy of our real-time point estimates. This is because the accuracy of real-time estimates is a function of how well future values of the underlying series can be forecast; see Appendix B.2 for details. The better one can forecast (in real-time) the future values of the (log) level of output (and other variables in multivariate measures of the output gap) the less the revision. Indeed, if one could forecast perfectly there would, of course, be no revision at all.

The potential advantages of forecasting can be appreciated in another, related, way. Forecasting future values prior to de-trending facilitates use of a less one-sided de-trending filter. De-trending filters, whether parametric or nonparametric, can be seen to involve application of a moving-average filter to the raw data; e.g. see Harvey & Koopman (2000) and Koopman & Harvey (2003) who present algorithms for deriving the weights for UC models. Most de-trending filters, such as the HP filter, imply a smooth two-sided moving average in the middle of the sample with no time-series observation receiving a large weight relative to its close neighbours. However, at the end of the sample the moving average becomes one-sided. It is well known that application of a one-sided filter will lead to more volatile estimates at the end of the sample as the weight on the last observation will be much higher than any of the weights associated with application of the filter in the middle of the sample; e.g. see van Norden (1997). Forecasting prior to de-trending can reduce the impact of the last observation. Naturally, if the future is forecast incorrectly this may not deliver more reliable output gap estimates. However, as shown in a stylised univariate framework by van Norden (1997), (section 1.5), even under the assumption that past output growth does not help predict present and future output growth, when the output gap does not Granger-cause output, forecasted values receive no weight. In such a case the one-sided filter will be reliable. Extending van Norden's (1997) argument, if output does not Granger-cause the output gap then real-time, or one-sided, output gaps will be unreliable as in real-time our only way of trying to guess what will happen (in the future) to the output gap is to look at past output growth. If this offers no information about the output gap then we should expect real-time estimates to be unreliable.

Re-visiting the real-time experiment of Section 3.3 the final row of Tables 1-7 indicates whether forecasting prior to de-trending can help deliver more accurate estimates in real-time. We focus on the behaviour of the median estimate across 99 bootstrap replications. We also report the *forecast* revision; see Tables 9-16. The forecast revision is the difference between the final and the median of the forecasted real-time estimates. The tables indicate that forecasting ahead in real-time, prior to de-trending, can help. The correlation against the final estimates is often higher, and RMSE lower, for the forecasted output gap estimates than the real-time estimates. Clearly these results are merely suggestive since they are sensitive to choices about the forecasting model, forecast horizon etc.[20] However, they suggest that there is some value to trying to forecast ahead prior to de-trending.

[20]For the HP filter we did experiment with forecasting 25 periods ahead based on the view that this is roughly the length of the filter implied by HP. The correlation of the real-time and final estimates fell to 0.230.

4.3 Evaluation of interval and density forecasts of the output gap

4.3.1 Evaluation of interval forecasts

Interval forecasts are evaluated following Christoffersen (1998).[21] A 'good' interval forecast should have correct conditional coverage, so that in volatile periods the interval is wider than in less volatile periods. Define I_t as an indicator variable that takes the value 1 if the outcome falls within the interval forecast at time t, and the value 0 otherwise. Consider an interval forecast for coverage probability p. Then Christoffersen (1998) defines a set of *ex ante* forecasts as being "efficient" with respect to the information set (say, Ω_{t-1}) if $E(I_t \mid \Omega_{t-1}) = p$. If $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, ...\}$ then this implies $\{I_t\}$ is *i.i.d.* Bernoulli with parameter p.

Then a test for unconditional coverage, ignoring independence, is defined as the test for whether actual coverage equals nominal: i.e. $E(I_t) = p$. A LR test is then defined:

$$LR_{uc} = -2 \log \frac{(1-p)^{n_0} p^{n_1}}{(1-\widehat{\pi})^{n_0} \pi^{n_1}} \sim \chi^2(1), \tag{1}$$

where $n_1 = \sum_{j=1}^{n} I_j$, $n_0 = n - n_1$ and $\widehat{\pi} = n_1/(n_0 + n_1)$ is the proportion of successes in the sample and $\widehat{\pi} \xrightarrow{p} \pi$.

(1) tests whether the coverage is correct but does not have power against the alternative that the 0's and 1's come clustered together in a time-dependent fashion. To remedy this omission, (1) is supplemented with a test for independence.

Define the first order Markov chain:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \tag{2}$$

where $\pi_{ij} = P(I_t = j \mid I_{t-1} = i)$. Under independence (2) equals

$$\Pi_1 = \begin{bmatrix} 1-\pi_1 & \pi_1 \\ 1-\pi_1 & \pi_1 \end{bmatrix}. \tag{3}$$

A LR test for independence (note that in fact only second moments are being considered) is then given by:

$$LR_{ind} = -2 \log \frac{(1-\widehat{\pi}_1)^{(n_{00}+n_{10})} \widehat{\pi}_1{}^{(n_{01}+n_{11})}}{(1-\widehat{\pi}_{01})^{n_{00}} \widehat{\pi}_{01}^{n_{01}} (1-\widehat{\pi}_{11})^{n_{10}} \widehat{\pi}_{11}{}^{n_{11}}} \sim \chi^2(1), \tag{4}$$

where n_{ij} is the number of times event i is followed by event j.

A joint test for correct conditional coverage can then be defined as $LR_{cc} = LR_{uc} + LR_{ind}$.

[21] An alternative regression based approach for evaluating interval forecasts is proposed by Clements & Taylor (2003). This approach allows for more general dependence structures than the first order Markov chain process considered below.

4.3.2 Evaluation of density forecasts

Density forecasts are evaluated *ex post* using the probability integral transform; see Diebold. et al. (1998).[22] They popularised the idea of evaluating a sample of density forecasts based on the idea that a density forecast can be considered "optimal" if the model for the density is correctly specified. One can then evaluate forecasts without the need to specify a loss function. This is attractive as it is often hard to define an appropriate general (economic) loss function. These tests evaluate the whole densities. Sometimes interest may lie in a particular region of the density, such as the probability of a recession. In these cases it is sensible to evaluate just the event of concern rather than the whole density; for an application see Clements (2002*b*).

A sequence of estimated density forecasts, $\{p_t(y_t)\}_{t=1}^{T}$, for the realisations of the process $\{y_t\}_{t=1}^{T}$, coincides with the true densities $\{f_t(y_t)\}_{t=1}^{T}$ when the sequence of z_t is independently and identically distributed (*i.i.d.*) with a uniform distribution, U(0,1), where,

$$z_t = \int_{-\infty}^{y_t} p_t(u)du, \ , t = 1, ..., T \tag{5}$$

Therefore, to test whether the density forecasts are optimal and do capture all aspects of the distribution of y_t one must test whether the z_t are both *i.i.d.* and U(0,1). Departures from *i.i.d.* U(0,1) reveal useful information about model failures. Deviations from uniform *i.i.d.* indicate that the forecasts have failed to capture some aspect of the underlying data generating process. For example, serial correlation in the z-sequence (in the first moment, squares, third powers etc.) would indicate poorly modelled dynamics, whereas non-uniformity indicates improper distributional assumptions or poorly modelled dynamics, or both.

By taking the inverse normal CDF transformation of the $\{z_t\}$ to give, say, $\{z_t^*\}$ the test for uniformity can be considered equivalent to one for normality on $\{z_t^*\}$; see Berkowitz (2001). This is useful as normality tests are widely seen to be more powerful than uniformity tests. However, testing is complicated by the fact that the impact of dependence on the tests for uniformity/normality is unknown, as is the impact of non-uniformity/normality on tests for dependence. For example, consider the Gaussian density $N(c_{t|t}, P_{t|t})$. The probability integral transforms are $z_t = \Phi((\underset{T\to\infty}{c_{t|T}} - c_{t|t}), \sqrt{P_{t|t}})$, i.e. $z_t = \Phi((\underset{T\to\infty}{c_{t|T}} - c_{t|t})/\sqrt{P_{t|t}})$. The Berkovitz series $\{z_t^*\}$ are then the scaled revision errors, $(\underset{T\to\infty}{c_{t|T}} - c_{t|t})/\sqrt{P_{t|t}}$ that are $N(0,1)$ under the null.

To test for normality of the $\{z_t^*\}$ series we employ the Doornik-Hansen test; see Doornik & Hansen (1994). To test for independence of the $\{z_t^*\}$ series we use the Ljung-Box test

[22] This methodology seeks to obtain the most "accurate" density forecast, in a statistical sense. It can be contrasted with economic approaches to evaluating forecasts that evaluate forecasts in terms of their implied economic value, which derives from postulating a specific (economic) loss function; see Granger & Pesaran (2000) and Clements (2002*a*).

for auto-correlation; see Harvey (1989), p. 259. Since dependence may occur in higher moments we consider $(z_t^* - \overline{z}^*)^j$ for $j = 1, 2, 3$.[23]

5 The uncertainty of real-time estimates in the Eurozone

To illustrate what degree of uncertainty traditional measures indicate about the final output gap estimates, Figure 11 plots the smoothed Harvey-Trimbur output gap estimates, plus 95% confidence intervals allowing for filter and parameter uncertainty. Figure 11 illustrates that allowing for both parameter and filter uncertainty is important. When allowing for both sources of uncertainty there is considerable uncertainty about the output gap; we can rarely say the output gap estimates, at a 95% significance level, are statistically different from zero.

Consider now the two proposed measures of uncertainty, computed in real-time, associated with the real-time point estimates, based upon the density $N((y_{t|t} - y^*_{t|t}), P_{t|t})$ and the simulated density derived from forecasting four quarters ahead 99 times for each recursive sample. Note in the former case we allow for only filter uncertainty; parameter uncertainty could be added in too. We should expect this to matter most for those measures of the output gap which estimate most parameters, namely the multivariate measures. Figures 12-16 provide an indication of the degree of estimated real-time uncertainty for the UC based measures of the output gap by computing 95% confidence intervals based upon the two alternative methods of computing the density (forecast) of the real-time output gap estimates. The confidence bands are wider for the Gaussian bands derived from the Kalman filter recursions than those derived via simulation. Indeed, the Gaussian bands are rarely significantly different from zero; policy-makers on this basis could never be sure about the position of the business cycle. The simulated confidence intervals are narrower, and sometimes do suggest that the output gap is statistically significant.

Given these differences between the two alternative estimates of real-time uncertainty it is important to ask, which measure of uncertainty, if any, is best? Is the finding that the Gaussian bands nearly always cover zero in fact correctly quantifying the degree of uncertainty associated with real-time estimates. We analyse this formally using the statistical tests considered in Section 4.3. First, however, it is useful to simply contrast the confidence bands for each measure with the actual outturn (printed in bold face). Apart from the Harvey-Trimbur measure, the Gaussian bands always 'cover' the final estimate while the simulated bands do not. For the Gaussian bands this comes at the expense of often very wide intervals. The simulated bands tend to fail to cover the boom

[23] Alternatively use could be made of recently proposed joint tests for uniformity and independence; see Hong (2002). Care should be exercised in interpreting results from density evaluation tests. Clements et al. (2003) have found, using Monte-Carlo experiments, that traditional density evaluation tests are not useful in sample sizes typical to macroeconomics; specifically, they found that the tests failed to detect non-linearities present under the data-generation-process. Graphical analysis, using P and Q plots, can also be useful in explaining why density forecasts have failed.

of the early 1990s; in real-time they under-estimate, in a statistically significant manner, the degree of the upturn to economic activity. This is an important failing. But, at other times, they perform quite well. Contrast the two measures of real-time uncertainty in Figure 15. While both bands cover the final estimate, the simulated bands do so keeping the bands much narrower. The uncertainty associated with the bivariate HP filter is, in fact, also relatively large. Also looking at the Harvey-Trimbur cycle, Figure 13, the simulated bands actually do a better job at picking up the boom of the early 1990s than the Gaussian bands by indicating a greater degree of uncertainty, particularly on the upside.

Tables 17-26 then formally evaluate the measures of uncertainty computed for each output gap estimator by testing the interval and density estimates. Density estimates of the output gap are evaluated both over the full period, 1981q1-2000q1, and over a restricted period 1991q1-2000q1. Results are discouraging, in the sense that the interval and density estimates, in general, are rejected in both the simulated and Gaussian cases. There appears to be no common cause for these failures. The density estimates are sometimes rejected on the basis of the independence test and sometimes due to the normality test. Equally, sometimes the interval estimates fail the unconditional test, and sometimes the independence test. It is of interest to note that the wide confidence bands implied by the bivariate HP filter in the Gaussian case, Figure 15, do appear to offer a worse characterisation of the uncertainty than the simulated measure; see the density evaluation results in Table 23.

These results suggest that the real-time measures of uncertainty considered in this paper do not prove to offer reliable indications of the degree of uncertainty associated with real-time estimates. However, this is not to say alternative measures of uncertainty may not do better. This is an area for research. We could further analyse the reasons for these failures through graphical analysis of the probability integral transform.

6 Concluding comments

This paper stresses the distinction between point estimates of the output gap, and measures of uncertainty associated with them. In an application to the Eurozone results indicate that not only are real-time estimates of the output gap unreliable, but so are their measures of uncertainty. This provides a serious challenge to users of output gap estimates. They must decide what do to given that there is mismeasurement not just of the output gap (point) estimates but their uncertainty too. However, in related research we find that although real-time output gap estimates often have little forecasting power over inflation relative to simple autoregressive alternatives this does not appear to be due to the unreliability of output gap estimates but rather the difficulties of forecasting inflation *per se*; see Mitchell (2003).

7 Bibliography

References

Apel, M. & Jansson, P. (1999), 'A theory-consistent approach for estimating potential output and the NAIRU', *Economics Letters* **64**, 271–275.

Artis, M., Marcellino, M. & Proietti, T. (2003), 'Dating the Euro area business cycle', *CEPR Discussion Paper No. 3696* .

Astley, M. & Yates, T. (1999), Inflation and real disequilibria. Bank of England Working Paper.

Barrell, R. & Sefton, J. (1995), 'Output gaps, some evidence from the UK, France and Germany', *National Institute Economic Review* **151**, 65–73.

Berkowitz, J. (2001), 'Testing density forecasts, with applications to risk management', *Journal of Business and Economic Statistics* **19**, 465–474.

Beyer, A., Doornik, J. & Hendry, D. (2001), 'Constructing historical Euro-zone data', *Economic Journal* **111**, 308–327.

Blanchard, O. & Quah, D. (1989), 'The dynamics of aggregate demand and supply disturbances', *American Economic Review* **79**, 655–673.

Boone, L. (2000), 'Comparing semi-structural methods to estimate unobserved variables: the hpmv and kalman filters approaches'. OECD Economics Department Working Papers No. 240.

Bullard, J. & Keating, J. (1995), 'The long-run relationship between inflation and output in postwar economies', *Journal of Monetary Economics* **36**, 477–496.

Camba-Mendez, G. & Rodriguez-Palenzuela, D. (2003), 'Assessment criteria for output gap estimates', *Economic Modelling* **20**, 529–562.

Canova, F. (2002), G-7 inflation forecasts. Mimeo, Universitat Pompeu Fabra.

Chagny, O. & Lemoine, M. (2002), The impact of the macroeconomic hypothesis on the estimation of the output gap using a multivariate Hodrick-Prescott filter: the case of the euro-area. Eurostat Working Paper.

Christoffersen, P. (1998), 'Evaluating interval forecasts', *International Economic Review* (39), 841–862.

Clements, M. (2002*a*), 'Evaluating the Bank of England density forecasts of inflation', *Warwick University Discussion Paper* .

Clements, M. (2002*b*), 'An evaluation of the survey of professional forecasters probability distributions of expected inflation and output growth', *Warwick University Discussion Paper* .

Clements, M., Frances, P., Smith, J. & van Dijk, D. (2003), 'On SETAR non-linearity and forecasting', *Journal of Forecasting* **22**, 359–375.

Clements, M. P. & Hendry, D. F. (1999), *Forecasting Non-stationary Economic Time Series*, The MIT Press, London, England.

Clements, M. & Taylor, N. (2003), 'Evaluating interval forecasts of high-frequency financial data', *Journal of Applied Econometrics* **18**, 445–456.

DeSerres, A. & Guay, A. (1995), Selection of the truncation lag in structural VARs (or VECMs) with long run restrictions. Bank of Canada Working Paper 95-9.

Diebold., F. X., Gunther, A. & Tay, K. (1998), 'Evaluating density forecasts with application to financial risk management', *International Economic Review* **39**, 863–883.

Doornik, J. (1998), *Object-oriented Matrix Programming using Ox*, Timberlake Consultants.

Doornik, J. & Hansen, H. (1994), A practical test for univariate and multivariate normality. Discussion Paper, Nuffield College, Oxford.

Dupasquier, C., Guay, A. & St-Amant, P. (1999), 'A survey of alternative methodologies for estimating potential output and the output gap', *Journal of Macroeconomics* **21**, 577–595.

Economic models at the Bank of England (1999).

Ehrmann, M. & Smets, F. (2003), 'Uncertain potential output: implications for monetary policy', *Journal of Economic Dynamics and Control* **27**, 1611–1638.

Fabiani, S. & Mestre, R. (2001), A system approach for measuring the Euro area NAIRU. European Central Bank Working Paper No. 65.

Fagan, G., Henry, J. & Mestre, R. (2001), 'An area-wide model (AWM) for the eu11', *European Central Bank* .

Faust, J. & Leeper, E. (1997), 'When do long run identifying restrictions give reliable results?', *Journal of Business and Economic Statistics* **15**, 345–353.

Faust, J., Rogers, J. & Wright, J. (2000), News and noise in G-7 GDP announcements. Board of Governors of the Federal Reserve System.

Gerlach, S. & Smets, F. (1999), 'Output gaps and monetary policy in the EMU area', *European Economic Review* **43**, 801–812.

Gordon, R. J. (1997), 'The time-varying NAIRU and its implications for economic policy', *Journal of Economic Perspectives* **11**, 11–32.

Granger, C. & Pesaran, M. (2000), 'Economic and statistical measures of forecast accuracy', *Journal of Forecasting* **19**, 537–560.

Hamilton, J. D. (1986), 'A standard error for the estimated state vector of a state-space model', *Journal of Econometrics* **33**, 387–397.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Harding, D. & Pagan, A. (2002), 'Dissecting the cycle: A methodological investigation', *Journal of Monetary Economics* **49**, 365–381.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A. C. & Jaeger, A. (1993), 'Detrending, stylized facts and the business cycle', *Journal of Applied Econometrics* **8**, 231–247.

Harvey, A. & Koopman, S. (2000), 'Signal extraction and the formulation of unobserved components models', *Econometrics Journal* **3**, 84–107.

Harvey, A. & Trimbur, T. (2003), 'General model based filters for extracting trends and cycles in economic time series', *Review of Economics and Statistics* **85**, 244–255.

Hong, Y. (2002), 'Evaluation of out-of-sample probability density forecasts', *Cornell University Discussion Paper* .

Kaiser, R. & Maravall, A. (2001), *Measuring business cycles in economic time series*, Springer.

King, R. G., Plosser, C., Stock, J. & Watson, M. (1991), 'Stochastic trends and economic fluctuations', *American Economic Review* **81**, 819–840.

Koopman, S. & Harvey, A. (2003), 'Computing observation weights for signal extraction and filtering', *Journal of Economic Dynamics and Control* **27**, 1317–1333.

Koopman, S. J., Shephard, N. & Doornik, J. A. (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal* **2**(1), 107–60.

Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?', *Journal of Econometrics* **54**, 159–178.

Laxton, D. & Tatlow, R. (1992), 'A simple multivariate filter for the measurement of potential output', *Bank of Canada* **59**.

Mitchell, J. (2003), The use of output gap estimates in real-time. mimeo, NIESR.

Morley, J. (1999), A note on constraining AR(2) parameters in estimation. mimeo, Washington University.

Nelson, E. & Nikolov, K. (2001), UK inflation in the 1970s and 1980s: The role of output gap mismeasurement. Bank of England Working Paper no. 148.

Orphanides, A. & van Norden, S. (2002), 'The unreliability of output-gap estimates in real time', *The Review of Economics and Statistics* **84**, 569–583.

Orphanides, A. & van Norden, S. (2003), 'The reliability of inflation forecasts based on output gap estimates in real time'. CIRANO Working Paper 2003s-01.

Pascual, L., Romo, J. & Ruiz, E. (2001), 'Effects of parameter estimation on prediction densities: a bootstrap approach', *International Journal of Forecasting* **17**, 83–103.

Pesaran, M. H. & Timmermann, A. (1992), 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics* **10**(4), 461–465.

Quah, D. & Vahey, S. (1995), 'Measuring core inflation', *Economic Journal* **105**, 1130–1144.

Runstler, G. (2002), The information content of real-time output gap estimates: an application to the Euro area. European Central Bank Working Paper No. 182.

Schumacher, C. (2002), 'Forecasting trend output in the Euro area', *Journal of Forecasting* **21**(8), 543–558.

Staiger, D., Stock, J. & Watson, M. (1997), 'The NAIRU, unemployment and monetary policy', *Journal of Economic Perspectives* **11**, 33–49.

Tay, A. & Wallis, K. (2000), 'Density forecasting: a survey', *Journal of Forecasting* (19), 235–254.

van Norden, S. (1997), 'Measurement of the output gap: a discussion of recent research at the Bank of Canada'. Bank of Canada Working Paper.

A Appendix. Alternative output gap estimators

A.1 Multivariate Unobserved Components filters

We consider two representative multivariate unosbserved components (UC) or state space models. The first is along the lines of Gerlach & Smets (1999) and Runstler (2002) and uses information on output and inflation. The second also considers unemployment; see Apel & Jansson (1999) and Fabiani & Mestre (2001).

To avoid having to fix the signal to noise ratio [see Gordon (1997) and *Economic models at the Bank of England* (1999)] assumptions can be made about the nature of the cyclical process for unemployment or output; see also Gerlach & Smets (1999), Fabiani & Mestre (2001) and Runstler (2002). This typically takes the form of assuming a stationary cyclical process. Without such an assumption, the trend component typically accounts for all the variation in the level of the variable and soaks up all residual variation. Alternatively, as with the multivariate HP filter, *a priori* restrictions are placed on the variances with the aim of obtaining plausible looking cycles. This approach has been adopted, for example, by Chagny & Lemoine (2002).

A.1.1 A bivariate UC model of output and inflation

To give the output gap a more economic interpretation than in univariate unobserved components models, it is related to data on inflation. This is sensible as the causal relationship between the output gap and inflation is central to the Phillips Curve and the conduct of monetary policy.

We consider general models of the form:

$$\begin{aligned} y_t &= y_t^* + y_t^C && (6)\\ \Theta(L) y_t^C &= \varepsilon_t^y && (7)\\ \Gamma(L)\Delta\pi_t &= \lambda y_{t-1}^C + \varepsilon_t^\pi && (8)\\ y_t^* &= y_{t-1}^* + \beta_{t-1} + \varepsilon_t^{y^*} && (9)\\ \beta_t &= \beta_{t-1} + \varepsilon_t^\beta && (10) \end{aligned}$$

where y_t is the log of actual output, y_t^* is its trend level, y_t^C is the output gap, π_t is quarterly inflation in percentage points measured at an annual rate and ε are the disturbances. All disturbances are assumed *i.i.d.* Gaussian.

The Phillips Curve equation, (8), provides a link between inflation and aggregate demand (measured here by the output gap). Since inflation is commonly assumed to depend only on nominal factors in the long-run we can see (8) to be imposing a long run homogeneity restriction. This can be seen by appreciating that underlying (8) is the following equation, $\Gamma^*(L)\pi_t = \lambda(L) y_{t-1}^C + \varepsilon_t^\pi$ expressed in the level of inflation. Long run homogeneity requires $\Gamma^*(1) = 0$, implying $\Gamma^*(L) = \Gamma(L)(1-L)$, meaning the Phillips Curve relationship is expressed in the first differences of inflation, $\Delta\pi_t$.

We allow for the lag polynomial $\Theta(L)$ to be second order. The roots are constrained to be stationary, but importantly we do allow for complex roots; see Morley (1999) for an account of how the roots can be constrained through a simple re-parameterisation. Experimentation suggested that restricting attention to a first order polynomial led to less sensible looking cycles. $\Gamma(L)$ is assumed to be first order. We consider a smooth trend specification where $\sigma^2_{\varepsilon_t^{y^*}} = 0$.[24]

A.1.2 A trivariate UC model of output, inflation and unemployment

It is also common to consider as an additional economic variable, unemployment; see Apel & Jansson (1999) and Fabiani & Mestre (2001). This is based on the view that the output and unemployment gaps are closely related. Inflation is likely to contain information about the size of both gaps. Restrictions can then be imposed in an unobserved components model that identify not just the Phillips Curve, linking measures of excess demand to inflation, but Okun's Law, that relates the unemployment and output gaps.

Define u_t and u_t^* as unemployment and trend unemployment, respectively. We consider the following model, based on Apel & Jansson (1999) and Fabiani & Mestre (2001):

$$\begin{aligned} \Delta\pi_t &= \Gamma(L)\Delta\pi_{t-1} + \Psi(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^\pi && (11) \\ y_t - y_t^* &= \gamma(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^{qgap} && (12) \end{aligned}$$

where

$$\begin{aligned} y_t^* &= y_{t-1}^* + m_{t-1}^1 + \varepsilon_t^{y^*} && (13) \\ u_t^* &= u_{t-1}^* + m_{t-1}^2 + \varepsilon_t^{u^*} && (14) \end{aligned}$$

and

$$\begin{aligned} m_t^1 &= m_{t-1}^1 + \varepsilon_t^{m^1} && (15) \\ m_t^2 &= m_{t-1}^2 + \varepsilon_t^{m^2} && (16) \end{aligned}$$

The unemployment gap is modelled as an autoregressive process:

$$(u_t - u_t^*) = \delta(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^{ugap} \quad (17)$$

[24] Other options are to allow the disturbances to follow moving average processes; e.g. if we let ε_t^π follow a MA process then a far richer dynamic process is possible than with a purely auto-regressive specification; such an approach is followed by Runstler (2002). We also address the issue of whether inflation should depend on lagged, but not current, or current as well as lagged, output (or unemployment) gap. Gerlach & Smets (1999) and Staiger et al. (1997), for example, consider lagged values only. However, current values are also considered, see for example Gordon (1997). This means that inflation and the output gap can be affected by shocks simultaneously, which appears reasonable; see, for example, Astley & Yates (1999) for further discussion of why it is important to allow for such endogeneity. Experimentation revealed that in practice although this assumption affects the timing of the cycles by one period, the shape is largely unaffected.

Equation (11) is a version of Gordon's triangle Phillips Curve model; it relates inflation to movements in the unemployment gap. Expectations are implicit in the inflation dynamics. Equation (12) is an Okun's Law relationship, relating cyclical unemployment and output movements.

Equations (13) and (14) are assumed to follow a local linear trend model. This representation was used with success by Fabiani & Mestre (2001) in an application to the Eurozone. It implies that the trend of output and unemployment (the NAIRU) are I(2) processes. We consider a first order polynomial for $\gamma(L)$. $\delta(L)$ is second order and constrained to be stationary but allowed to be complex (this constraint is empirically important). $\Psi(L)$ is first order. $\Gamma(L)$ is second order. Again we consider a smooth trend specification where $\sigma^2_{\varepsilon_t^{y*}} = 0$.

It should be noted that the above system does not allow for full endogeneity between real disequilibria and inflation, that would involve neither real disequilibria causing inflation nor vice-versa. A restrictive path for the transmission of demand shocks is implied as demand shocks lead to inflation via the unemployment gap, and then and from the unemployment gap to the output gap; see Astley & Yates (1999).

A.1.3 Estimation of multivariate UC models

The parameters of the univariate and multivariate unobserved components models are estimated by maximum likelihood exploiting their state-space form. Computations are performed using the beta version of `SsfPack 3` for `Ox`; see Koopman et al. (1999) for a discussion of the earlier version of `SsfPack`. Importantly, in contrast to Fabiani & Mestre (2001), for example, all observable variables are put in the state vector to ensure parameter uncertainty is fully accounted for; see Harvey (1989), pp. 366-368. In the context of Fabiani & Mestre (2001), see their Appendix, this means that inflation is also placed in the state vector rather than left in the measurement equation. See Mitchell (2003) for details of how this is achieved.

A.2 Multivariate Hodrick Prescott filters

Laxton & Tatlow (1992) proposed an extension to the Hodrick-Prescott (HP) filter which incorporates economic information. Additional, so-called economic, constraints are imposed on the minimisation from which the HP filter is defined. The residuals from a structural equation, such as the Phillips Curve or Okun's Law, are added to the minimisation problem that the univariate HP filter seeks to solve. Just as the univariate or traditional HP filter can be interpreted within an unobserved components framework [see Harvey & Jaeger (1993)], so can the multivariate HP filter; see Boone (2000). This facilitates estimation by maximum likelihood and inference since confidence bands around the estimates can be derived from the Kalman filter recursions.

A.2.1 A bivariate HP model of output and inflation

If one does not assume a particular parametric process for the cyclical component, as is commonly done with the UC models, to obtain plausible looking cycles typically one will need to constrain the variance of the disturbances driving the elements of the state vector. This is the approach taken by the HP filter, albeit implicitly when the filter is interpreted as a nonparametric filter.[25] To illustrate, consider the following UC model where potential output is assumed to follow, say, a smooth trend representation [see Harvey & Jaeger (1993)] and the economic constraint is based on the Phillips Curve:

$$
\begin{aligned}
y_t &= y_t^* + y_t^C && (18)\\
\Gamma(L)\Delta\pi_t &= \lambda y_{t-1}^C + \varepsilon_t^\pi && (19)\\
y_t^* &= y_{t-1}^* + \beta_{t-1} && (20)\\
\beta_t &= \beta_{t-1} + \varepsilon_t^{y^*} && (21)
\end{aligned}
$$

where all disturbances are $i.i.d.$ Gaussian. Alternatives representations for potential output such as the local linear trend and the random walk representation can be considered.

Writing (18) in its state-space form, the transition equation for the state-vector (for expositional ease only making some specific assumptions about the form of the lag polynomials in (18)) is given by:

$$
\begin{bmatrix} \Delta\pi_t \\ y_t^* \\ y_{t-1}^* \\ y_t^C \end{bmatrix} = \begin{bmatrix} * & 0 & 0 & * \\ 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\pi_{t-1} \\ y_{t-1}^* \\ y_{t-2}^* \\ y_{t-1}^C \end{bmatrix} + \begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^{y^*} \\ 0 \\ \varepsilon_t^{y^C} \end{bmatrix} \qquad (22)
$$

Consistent with how for the univariate HP filter the signal to noise ratio determines the smoothness of the trend series $(\sigma_{y^*}^2/\sigma_{y^C}^2)$ [see Harvey & Jaeger (1993) and Kaiser & Maravall (2001)], the relative variance of the disturbances in (22) controls the smoothness of the cycle and the fit of the economic relationship (the second equation in (18)). As $\sigma_{y^*}^2$ tends to infinity the more explanatory power is given to the unobserved variable and the less the importance of the cycle.

Let λ_1 and λ_2 denote these relative variances, that control the problem, where

$$
\begin{aligned}
\lambda_1 &= \sigma_{y^C}^2/\sigma_{y^*}^2 && (23)\\
\lambda_2 &= \sigma_{y^C}^2/\sigma_{\pi}^2. && (24)
\end{aligned}
$$

[25] The variances of the disturbances, of course, can be estimated by maximum likelihood. It is the decision not to estimate them, but assume *a priori* values, that we take to be the defining characteristic of the multivariate HP approach. This contrasts the multivariate UC approach where these variances are estimated.

Set, without loss of generality, $\sigma^2_{y^C} = 1$, then

$$\lambda_1 = 1/\sigma^2_{y^*} \tag{25}$$
$$\lambda_2 = 1/\sigma^2_{\pi} \tag{26}$$

Then λ_1 controls the smoothness of the trend component of output. As $\lambda_1 \to 0$ the trend becomes very volatile and can soak up all cyclical variation. As $\lambda_1 \to \infty$ the trend trends to a deterministic (smooth) trend. Traditionally, as with the univariate HP filter, $\sigma^2_{y^*} = 1/1600 \Leftrightarrow \lambda_1 = 1600$.

λ_2 controls the fit of the Phillips Curve relationship. As $\lambda_2 \to 0$ ($\sigma^2_\pi \to \infty$) the worse the fit of the the economic relationship, implying less information provided by the economic relationship. Traditionally $\sigma^2_\pi \simeq 1/25 \Leftrightarrow \lambda_2 = 25$ or $\sigma^2_\pi \simeq 1 \Leftrightarrow \lambda_2 = 1$; see Boone (2000) and Chagny & Lemoine (2002). Note that the better the fit of the economic relationship the better the explanatory power lagged values of the output gap have over inflation.

A.3 Structural VAR (SVAR) models

Following the seminal paper by Blanchard & Quah (1989), VAR models with economic restrictions imposed on the long-run have been used to estimate output gaps; see Bullard & Keating (1995) and Camba-Mendez & Rodriguez-Palenzuela (2003), and for a related exercise Quah & Vahey (1995).[26] A perceived advantage of SVAR models, relative to UC models, is that SVAR models can be viewed as one-sided filters; in this sense they overcome the end-point problem associated with UC model that can be seen to involve application of a two-sided filter.

In the SVAR approach restrictions are imposed on the matrix of long-run multipliers so that demand (or transitory) shocks can be distinguished from supply (or permanent shocks). Transitory shocks are identified by imposing the restriction that they have no long-run effect on the level of the variable of concern. The output gap is the cumulative sum of the transitory shocks to output. Potential output is then the cumulative sum of the permanent shocks.

In contrast to the other multivariate methods of estimating the output gap, the SVAR approach gives the components of output an economic interpretation. Potential output is not assumed to follow a random walk as in UC models. Furthermore, both real disequilibria and inflation are modelled as endogenous variables.

We distinguish two broad types of SVAR model: those with and without cointegration. Specifically, we consider models of the type proposed by Blanchard & Quah (1989) and King et al. (1991) that deal with the case of no cointegration and cointegration, respectively. In cointegrating VAR models there is an additional identification problem taking two related forms: (i) identification of the cointegrating vectors and (ii) identification of

[26]There are related multivariate approaches to estimating output gap based on the multivariate Beveridge-Nelson decomposition and the Cochrane approach; e.g. see Dupasquier et al. (1999).

the common stochastic trends. Both the cointegrating vectors and the common stochastic trends offer characterisations of the "long run" in a VAR model. A popular approach to identification in the presence of cointegration is King et al. (1991). Cointegration provides information on the number of permanent and transitory shocks. Atheoretical alternatives to the identification strategy of King et al. (1991) have been considered also; for an application to the Eurozone see Schumacher (2002).

A.3.1 Blanchard-Quah SVAR models

We consider SVAR models in line with and Astley & Yates (1999) and Camba-Mendez & Rodriguez-Palenzuela (2003). We will consider trivariate VAR models of output, inflation and unemployment; see Camba-Mendez & Rodriguez-Palenzuela (2003). Restrictions then can be imposed on the matrix of long-run multipliers so that inflation is determined in the long-run by only one structural shock (the inflation shock), whereas output is determined by two structural shocks (the inflation shock plus a supply shock), and unemployment is determined by three structural shocks (the inflation shock, the supply shock and a demand shock). The output gap is then the cumulative sum of the demand shocks to output.

It is important that the estimated (reduced-form) VAR model, that provides the basis for the identification of the output gap, is stationary. This requires the variables in the VAR to be appropriately transformed prior to estimation. Following Camba-Mendez & Rodriguez-Palenzuela (2003) in their application to the Eurozone we consider the first differences of inflation, GDP and unemployment. We did, however, experiment with representations in the level of inflation and unemployment.

A.3.2 King-Plosser-Stock-Watson SVAR models

Rather than estimating a stationary (perhaps first-differenced) VAR model a cointegrating VAR model is considered. We follow King et al. (1991) and consider a three variable system of GDP, consumption and investment. There are two cointegrating vectors: the ratios of consumption and investment to GDP are stationary. Therefore, there is one permanent shock and two transitory shocks. The output gap is the cumulative sum of the two transitory shocks to output.

A.3.3 The importance of lag length

The long-run restrictions implicit to identification require estimation of the matrix of long run responses (the sum of the lag polynomial in the moving-average representation corresponding to the VAR model). This is based on estimation of the sum of the coefficients in the VAR model. Therefore the reliability of SVAR estimates of the output gap rests on reliable estimation of the AR coefficients; see Faust & Leeper (1997) for details. A key issue then is selection of lag order in the VAR model. Using too small a lag order can lead to significant biases in estimation of the permanent and transitory components; see DeSerres & Guay (1995). DeSerres & Guay (1995) found that use of information criteria leads to too low an estimated lag order. This is consistent with our findings in the sense that the BIC tended to select a lag order of one, but it was only with much larger lag orders that sensible looking output gap estimates were generated.

A.3.4 Uncertainty associated with SVAR estimates

SVAR models require estimation of a large number of parameters. This is expected to give rise to considerable uncertainty about the estimates, as in a changing world we expect considerable parameter uncertainty. One interesting approach to reduce this uncertainty is to impose rank restrictions on the estimated VAR model that reduce the number of estimated parameters; see Camba-Mendez & Rodriguez-Palenzuela (2003). One means of quantifying the uncertainty associated with SVAR estimates of the output gap is to undertake a series of stochastic simulation. An alternative is to appreciate that underlying the SVAR model is a state-space representation; e.g. DeSerres & Guay (1995) exploit a state-space representation that is close in spirit to the BQ approach to investigate the importance of lag order on output gap estimation via Monte-Carlo simulation.

B Appendix. Uncertainty and UC models

B.1 Confidence bands

Let $a_{t|T} = E(\alpha_t|\Omega_T)$ and $a_{t|t} = E(\alpha_t|\Omega_t)$ denote the optimal (minimum mean squared error [MSE]) estimators of the state vector of the UC model, α_t, based on observations (or more generally information set Ω) up to and including T and t, respectively, conditional on the parameter vector Θ that is estimated by maximum likelihood. Then let $P_{t|t} = V(a_{t|t} - \alpha_t)$ and $P_{t|T} = V(a_{t|T} - \alpha_t)$ denote their covariance matrices. Then, assuming covariances between $a_{t|t}$ and $a_{t|T}$ are zero, we know

$$P_{t|t} = V(a_{t|t} - \alpha_t) = V(a_{t|t} - a_{t|T} + a_{t|T} - \alpha_t) = P_{t|T} + V(a_{t|T} - a_{t|t}) \quad (27)$$

$$P_{t|t} = P_{t|T} + V(a_{t|T} - a_{t|t}) \quad (28)$$

so the MSE of the filtered (or one-sided) estimates is greater than that of the smoothed (or two-sided) estimates by a positive semidefinite matrix, $V(a_{t|T} - a_{t|t})$. This MSE decreases as t increases.

If Θ were known $P_{t|t}$ and $P_{t|T}$ would indicate the uncertainty in the Kalman-filter recursions. We call this uncertainty, filter uncertainty. However, there is an additional source of uncertainty if Θ is estimated, say by $\widehat{\Theta}$; i.e. $P_{t|t}(\widehat{\Theta}) > P_{t|t}(\Theta)$ and $P_{t|T}(\widehat{\Theta}) > P_{t|T}(\Theta)$. We account for both of these types of uncertainty following Hamilton (1986).[27]

[27] An alternative approach is to use analytical approximations to capture the parameter uncertainty like Orphanides & van Norden (2002).

B.2 Revisions

Define the revision $R_{t|T}$ as $R_{t|T} = a_{t|T} - a_{t|t}$. Then we may re-write (27) as:

$$P_{t|t} = P_{t|T} + V(R_{t|T}). \tag{29}$$

The MSE of the revisions is a lower bound for the MSE of $P_{t|t}$. This, as stressed by Runstler (2002), is useful as it provides an indication of the MSE for estimators that cannot be cast in the state-space form. The variance of the revision allowing for parameter uncertainty as well as filter uncertainty is:

$$V(R_{t|T})(\widehat{\Theta}) = P_{t|t}(\widehat{\Theta}) - P_{t|T}(\widehat{\Theta}). \tag{30}$$

We can compute the variance of the revision as a function of the distance between t and T:

$$\begin{aligned} V(R_{t|t+1}) &= P_{t|t} - P_{t|t+1}, & (31)\\ V(R_{t|t+2}) &= P_{t|t} - P_{t|t+2}, & (32)\\ V(R_{t|t+3}) &= P_{t|t} - P_{t|t+3}. & (33) \end{aligned}$$

B.2.1 Relating revisions to forecast errors

The revisions can be related to the forecast error. Since [see Hamilton (1994), p.396]:

$$\begin{aligned} a_{t|T} &= a_{t|t} + P_{t|t}T'P^{-1}_{t+1|t}\left[a_{t+1|T} - a_{t+1|t}\right] & (34)\\ R_{t|T} &= P_{t|t}T'P^{-1}_{t+1|t}\left[a_{t+1|T} - a_{t+1|t}\right] & (35) \end{aligned}$$

where T is the matrix in the transition equation of the state vector, the revision is a function of the forecast error $a_{t+1|t}$. The closer $a_{t+1|t}$ is to the final estimate $a_{t+1|T}$ the smaller the revision and the less the difference between the filtered and smoothed estimates. Therefore, the better the one-step ahead forecasts of the state vector the less the revision.

We can see this in another way.[28] Consider the final estimate of the output gap as a known weighted linear function, that is say centered and symmetric, of the underlying data, say x_t:

$$a_{t|T} = C(L, F)x_t \tag{36}$$

where $C(L, F) = c_0 + \sum_{j=1}^{\infty} c_j(L^j + F^j)$ and $Lx_t = x_{t-1}$ and $Fx_t = x_{t+1}$. Then the final estimate of the output gap is given by:

$$a_{t|T} = C(L, F)x_t = c_0 + \sum_{j=1}^{\infty} c_j x_{t-j} + \sum_{j=1}^{\infty} c_j x_{t+j}. \tag{37}$$

[28]See also Kaiser & Maravall (2001), p. 118.

But in real-time the future values of x_t, namely $x_{t+1}, x_{t+2}...$ are unknown. To apply the two-sided filter future values need to be forecasted. Denote the forecasts of $x_{t+1}, x_{t+2}...$ made at time t by $E(x_{t+j}|\Omega_t)$. Then we may define the real-time estimate as:

$$a_{t|t} = c_0 + \sum_{j=1}^{\infty} c_j x_{t-j} + \sum_{j=1}^{\infty} c_j E(x_{t+j}|\Omega_t). \tag{38}$$

Subtracting (38) from (37) the revision is obtained as a function of the forecasting error, $(x_{t+j} - E(x_{t+j}|\Omega_t))$,

$$a_{t|T} - a_{t|t} = R_{t|T} = \sum_{j=1}^{\infty} c_j(x_{t+j} - E(x_{t+j}|\Omega_t)). \tag{39}$$

It follows that if these forecast errors are reduced the revision will decrease.

B.3 Testing the "rationality" of real-time estimates

Expectations are *rational* if they are the same as the prediction of theory. More formally, expectations of y_t, y_t^*, are rational if they are "optimal" predictions in the sense that they are the best approximation of y_t based upon the information set Ω_{t-1}[29]:

$$y_t^* = \varepsilon(y_t|\Omega_{t-1}). \tag{40}$$

The optimal (linear) prediction of y_t is the orthogonal projection of y_t on Ω_{t-1}. Therefore the REH can alternatively be characterised by the orthogonality condition

$$E(y_t - y_t^*)y_t^* = 0. \tag{41}$$

If deterministic terms (in particular a constant) are allowed in Ω_{t-1} rational expectations are also unbiased

$$E(y_t - y_t^*|\Omega_{t-1}) = 0. \tag{42}$$

We know that $a_{t|t}$ is the minimum MSE estimator of α_t, and given that we have assumed normality, these conditional expectations are equivalent to orthogonal projections. Therefore, $a_{t|t}$ is the orthogonal projection of α_t on Ω_t. Similarly $a_{t|T}$ is the orthogonal projection of α_t on Ω_T. Therefore, both $a_{t|t}$ and $a_{t|T}$ are orthogonal projections of α_t, with respect to the appropriate information set.

Via the law of iterated expectations

$$\begin{aligned} E(a_{t|T} \quad | \quad \Omega_t) &= E(E(\alpha_t \mid \Omega_T) \mid \Omega_t) = E(\alpha_t \mid \Omega_t) = a_{t|t} && (43) \\ \Rightarrow \quad & E(a_{t|T} - a_{t|t} \mid \Omega_t) = 0. && (44) \end{aligned}$$

[29] By using $\varepsilon(y|\Psi)$ to denote the optimal prediction we distinguish it from the conditional expectation, $E(y|\Psi)$.

So, $a_{t|t}$ can be also seen as the orthogonal projection of $a_{t|T}$ on Ω_t. This means the filtered estimates are rational expectations of the smoothed estimates $a_{t|T}$. This implies both $E(R_{t|T} \mid \Omega_t) = 0$, i.e. the revisions are unbiased, and $E(R_{t|T}a_{t|t}) = 0$; i.e. the revision is orthogonal to the real-time estimate $a_{t|t}$.

We can test this rationality implication by estimating the following regression, see also Runstler (2002),

$$a_{t|T} = a_0 + a_1 a_{t|t} + u_t, \tag{45}$$

and testing the joint null hypothesis that $a_0 = 0$ and $a_1 = 1$ *via* a Wald or F-test that is robust to serial correlation and heteroscedastictiy. Equivalently, we can see the test to be of the form:

$$R_{t|T} = b_0 + b_1 a_{t|t} + u_t, \tag{46}$$

where the joint null hypothesis that $b_0 = 0$ and $b_1 = 0$ now should be tested.

Tests of these sort, so-called Mincer-Zarnowitz tests, have been used in the related context of testing the rationality of the revision process to official releases of GDP data; see Faust et al. (2000). If $b_0 = 0$ and $b_1 = 0$ then revisions reflect news not available at the time the preliminary (or in our context real-time) estimate is produced; the preliminary, or real-time, estimate cannot predict the revised or final estimate.

Table 1: Output gap summary statistics for HP filter

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.001	-0.019	0.018	0.008	1.000	0.789	1.000	30.800	0.813	0.265	0.000	1.000	0.000	1.000
filtered	-0.000	-0.033	0.015	0.011	0.273	0.523	0.000	∞	0.894	0.393	0.012	0.429	0.763	0.243
realtime	-0.000	-0.033	0.015	0.011	0.273	0.523	0.000	∞	0.894	0.393	0.012	0.429	0.763	0.243
forecasted	0.001	-0.011	0.010	0.005	0.283	0.572	0.001	∞	0.930	0.404	0.009	0.442	0.608	0.141

Notes: mean, min, max and sd refer to the mean, minimum, maximum and standard deviation of the estimated output gap. r final: correlation of output gap against final estimate; r CAP: correlation against capacity utilisation; rat: p-value of test for rationality; period: length in quarters of cycle derived from spectrum; ACF: autocorrelation coefficient at lag one; KPSS: test for stationarity, test statistic reported; RMSE: root mean squared error; PT: Pesaran-Timmermann test; r tp: correlation of turning points against those of final estimates

Table 2: Output gap summary statistics: Harvey Trimbur

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.003	-0.026	0.029	0.016	1.000	0.692	1.000	∞	0.960	0.609	0.000	1.000	0.000	1.000
filtered	-0.009	-0.028	0.013	0.011	0.890	0.783	0.000	∞	0.966	1.101	0.010	0.883	0.000	0.649
realtime	-0.003	-0.018	0.010	0.008	0.179	0.551	0.003	∞	0.928	0.769	0.017	0.623	0.040	-0.100
forecasted	0.001	-0.012	0.011	0.006	0.190	0.366	0.075	∞	0.825	0.769	0.017	0.675	0.000	0.423

Table 3: Output gap summary statistics: Bivariate UC model

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.008	-0.037	0.023	0.016	1.000	0.682	1.000	∞	0.929	0.768	0.000	1.000	0.000	1.000
filtered	-0.010	-0.024	0.005	0.008	0.941	0.696	0.000	∞	0.937	1.188	0.009	0.870	0.000	0.596
realtime	-0.018	-0.068	0.009	0.016	0.457	0.249	0.000	∞	0.388	0.913	0.019	0.714	0.076	0.111
forecasted	-0.014	-0.054	0.005	0.010	0.462	0.110	0.065	∞	0.541	0.872	0.016	0.714	0.060	0.318

Table 4: Output gap summary statistics: Bivariate HP filter

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.036	-0.127	0.036	0.049	1.000	0.726	1.000	∞	0.975	1.571	0.000	1.000	0.000	1.000
filtered	-0.032	-0.221	0.167	0.092	0.727	0.543	0.000	∞	0.699	0.508	0.066	0.831	0.000	0.317
realtime	-0.056	-0.434	0.391	0.190	0.688	0.485	0.000	∞	0.654	0.417	0.162	0.779	0.000	0.131
forecasted	-0.039	-0.235	0.119	0.076	0.515	0.476	0.000	38.500	0.528	0.259	0.066	0.805	0.000	0.089

Table 5: Output gap summary statistics: Trivariate UC model

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.004	-0.021	0.022	0.012	1.000	0.545	1.000	∞	0.934	0.430	0.000	1.000	0.000	1.000
filtered	0.001	-0.015	0.026	0.010	0.906	0.336	0.039	∞	0.944	0.574	0.007	0.857	0.000	0.162
realtime	0.004	-0.025	0.070	0.019	0.455	-0.031	0.000	∞	0.623	0.718	0.019	0.597	0.005	0.066
forecasted	-0.008	-0.050	0.016	0.016	0.537	0.552	0.000	∞	0.814	1.744	0.014	0.688	0.004	-0.303

Table 6: Output gap summary statistics: BQ model with p=1

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.001	-0.007	0.005	0.002	1.000	-0.343	1.000	∞	0.783	1.571	0.000	1.000	0.000	1.000
realtime	-0.002	-0.008	0.004	0.003	0.901	-0.584	0.000	∞	0.786	0.722	0.002	0.896	0.000	0.795
forecasted	-0.002	-0.009	0.005	0.003	0.910	-0.581	0.000	∞	0.785	0.767	0.002	0.896	0.000	0.732

Table 7: Output gap summary statistics: BQ model with p=16

output gap	mean	min	max	sd	r final	r CAP	rat	period	ACF	KPSS	RMSE	hit rate	PT	r tp
smoothed	-0.001	-0.013	0.017	0.007	1.000	-0.274	1.000	18.500	0.819	0.472	0.000	1.000	0.000	1.000
realtime	0.002	-0.012	0.012	0.006	0.153	-0.529	0.000	14.800	0.569	0.253	0.009	0.459	0.590	0.238
forecasted	0.005	-0.005	0.014	0.005	0.232	-0.330	0.000	12.333	0.331	0.299	0.009	0.541	0.017	0.244

Notes: Results for p=16 are over the restricted period starting in 1991q1 rather than 1981q1

Table 8: Output gap summary statistics: KPSW model

output gap	mean	min	max	sd	r final	r CAP	rat.	period	ACF	KPSS	RMSE	hit rate	PT	r
smoothed	-0.000	-0.025	0.019	0.011	1.000	0.098	1.000	99999.000	0.968	0.859	0.000	1.000	0.000	1.0
realtime	0.004	-0.034	0.030	0.012	0.575	-0.026	0.002	30.800	0.585	0.294	0.011	0.792	0.000	0.2

Table 9: Revision summary statistics for HP filter

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	-0.001	-0.018	0.025	0.012	0.967	0.012	1.430
data revision	-0.001	-0.018	0.025	0.012	0.967	0.012	0.000
parameter revision	0.000	0.000	0.000	0.000	0.000	0.000	0.000
forecast revision	-0.003	-0.022	0.020	0.009	0.865	0.009	0.000

Table 10: Revision summary statistics: Harvey Trimbur

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	0.000	-0.034	0.032	0.017	0.940	0.017	1.038
data revision	0.006	-0.015	0.026	0.008	0.881	0.010	0.000
parameter revision	-0.006	-0.024	0.010	0.010	0.974	0.012	0.000
forecast revision	-0.004	-0.032	0.027	0.016	0.934	0.017	0.000

Table 11: Revision summary statistics for Bivariate UC model

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	0.010	-0.034	0.044	0.016	0.223	0.019	1.028
data revision	0.002	-0.021	0.022	0.009	0.799	0.009	0.000
parameter revision	0.008	-0.018	0.048	0.012	0.137	0.015	0.000
forecast revision	0.006	-0.033	0.032	0.014	0.612	0.016	0.000

Table 12: Revision summary statistics: Bivariate HP filter

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	0.020	-0.385	0.342	0.160	0.520	0.162	3.306
data revision	-0.003	-0.198	0.139	0.066	0.430	0.066	0.000
parameter revision	0.024	-0.223	0.217	0.100	0.610	0.103	0.000
forecast revision	0.003	-0.147	0.155	0.066	0.374	0.066	0.000

Table 13: Revision summary statistics: Trivariate UC model

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	-0.008	-0.058	0.019	0.017	0.467	0.019	1.451
data revision	-0.005	-0.018	0.006	0.005	0.715	0.007	0.000
parameter revision	-0.003	-0.044	0.025	0.015	0.436	0.015	0.000
forecast revision	0.004	-0.017	0.045	0.014	0.727	0.014	0.000

Table 14: Revision summary statistics: BQ model with p=1

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	0.002	-0.000	0.005	0.001	0.858	0.002	0.571
forecast revision	0.002	-0.001	0.005	0.001	0.855	0.002	0.000

Table 15: Revision summary statistics: BQ model with p=16

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	-0.003	-0.016	0.018	0.008	0.676	0.009	1.196
forecast revision	-0.005	-0.018	0.012	0.007	0.652	0.009	0.000

Table 16: Revision summary statistics: KPSW model

output gap revision	mean	min	max	sd	ACF	RMSE	NS
total revision	-0.004	-0.036	0.048	0.011	0.436	0.011	0.943
data revision	-0.004	-0.036	0.048	0.011	0.436	0.011	0.000
parameter revision	-0.004	-0.036	0.048	0.011	0.436	0.011	0.000
forecast revision	-0.004	-0.036	0.048	0.011	0.436	0.011	0.000

Table 17: Density forecast evaluation for HP filter

	LB(1)	LB(2)	LB(3)	DH
boot: 81q1-00q1	1.01E-12	3.30E-07	8.72E-08	0.99605
boot: 91q1-00q1	1.24E-05	0.427327	0.58163	0.908341
norm: 81q1-00q1	2.22E-51	6.85E-35	1.05E-38	4.50E-06
norm: 91q1-00q1	1.21E-24	3.98E-16	6.04E-18	4.19E-07

Notes: p-values are reported. LB(j): Ljung-Box for j-th order independence; DH: Doornik-Hansen test for normality; boot: simulated measure of uncertainty; norm: measure of uncertainty derived from Kalman filter recursions

Table 18: Interval forecast evaluation for HP filter

	LR_{uc}	LR_{ind}	LR_{cc}
boot	1.99E-44	2.28E-08	5.79E-50
norm	4.95E-03	1.00E+00	0.019263

Table 19: Density forecast evaluation: Harvey Trimbur

	LB(1)	LB(2)	LB(3)	DH
boot: 81q1-00q1	5.22E-20	6.82E-07	1.08E-10	0.76397
boot: 91q1-00q1	2.58E-07	0.04881	0.000275	0.095061
norm: 81q1-00q1	4.31E-45	3.14E-23	1.33E-23	0.677475
norm: 91q1-00q1	5.58E-11	1.80E-05	0.000219	3.26E-11

Table 20: Interval forecast evaluation: Harvey Trimbur

	LR_{uc}	LR_{ind}	LR_{cc}
boot	5.23E-29	2.13E-09	1.21E-35
norm	9.66E-28	1.24E-16	1.74E-41

Table 21: Density forecast evaluation: Bivariate UC model

	LB(1)	LB(2)	LB(3)	DH
boot: 81q1-00q1	9.80E-15	5.40E-02	0.003324	4.44E-05
boot: 91q1-00q1	2.42E-07	0.015164	0.002204	0.000355
nor: 81q1-00q1	3.78E-02	0.999392	0.983388	2.47E-10
nor: 91q1-00q1	2.39E-06	4.04E-06	0.11093	0.641939

Table 22: Interval forecast evaluation: Bivariate UC model

	LR_{uc}	LR_{ind}	LR_{cc}
boot	2.62E-15	1.47E-05	2.21E-18
norm	7.76E-02	8.70E-01	0.20803

Table 23: Density forecast evaluation: Bivariate HP filter

	LB(1)	LB(2)	LB(3)	DH
boot: 81q1-00q1	0.016819	0.212568	0.249791	0.030296
boot: 91q1-00q1	0.908287	0.386204	0.875054	0.453636
nor: 81q1-00q1	6.86E-09	0.061551	0.034555	0.354069
nor: 91q1-00q1	0.025885	0.005814	0.188037	0.266507

Table 24: Interval forecast evaluation: Bivariate HP filter

	LR_{uc}	LR_{ind}	LR_{cc}
boot	3.15E-05	0.083247	3.86E-05
nor	0.004946	1	0.019263

Table 25: Density forecast evaluation: Trivariate UC model

	LB(1)	LB(2)	LB(3)	DH
boot: 81q1-00q1	1.92E-23	0.017166	6.67E-06	0.00183
boot: 91q1-00q1	9.92E-08	0.190281	0.006059	0.096363
nor: 81q1-00q1	3.05E-10	1.59E-10	0.00117	0.244611
nor: 91q1-00q1	0.274434	0.002689	0.313163	0.094095

Table 26: Interval forecast evaluation: Trivariate UC model

	LR_{uc}	LR_{ind}	LR_{cc}
boot	9.57E-22	3.81E-12	3.93E-31
nor	5.65E-01	0.401173	0.595523

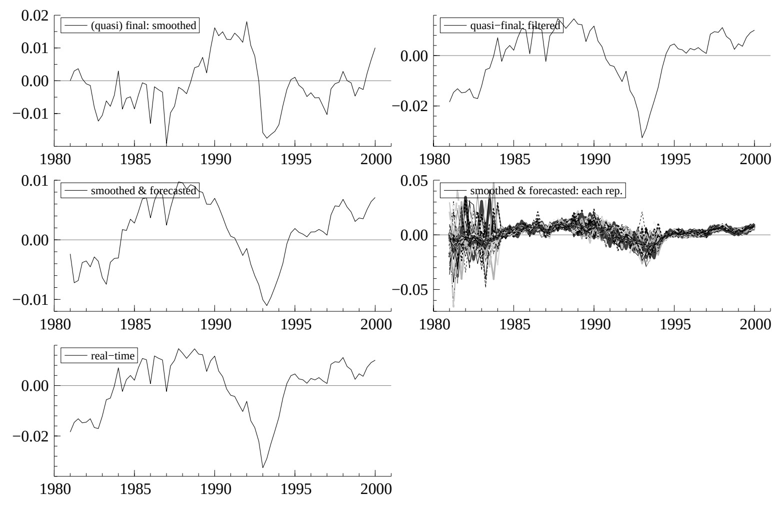

Figure 1: The real-time behaviour of the output gap using the HP filter

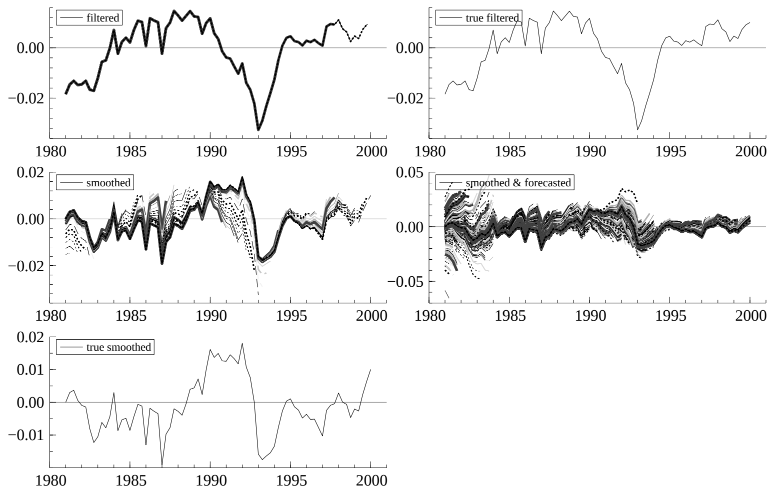

Figure 2: The evolution of real-time estimates using the HP filter

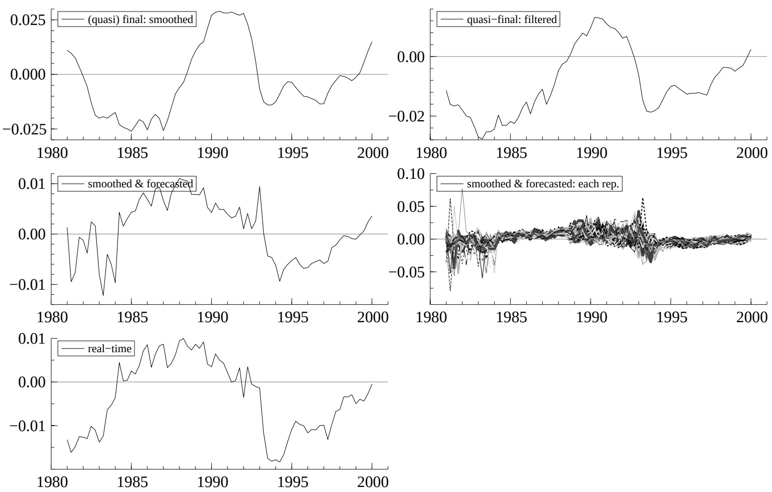

Figure 3: The real-time behaviour of the output gap using the Harvey-Trimbur filter

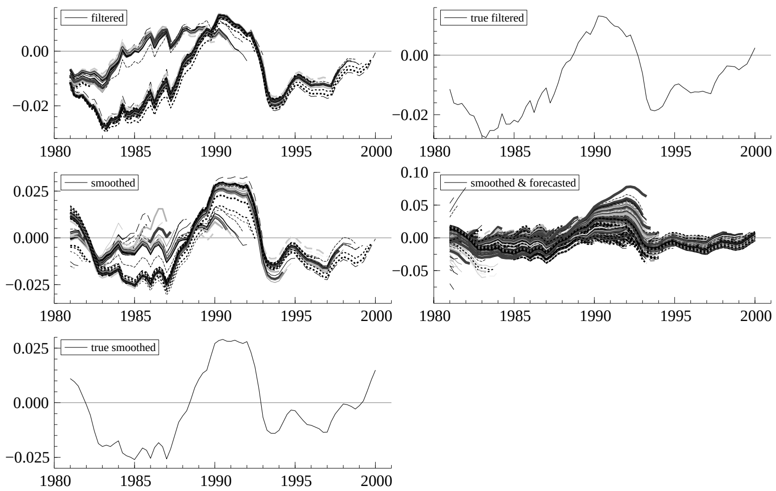

Figure 4: The evolution of real-time estimates using the Harvey-Trimbur filter

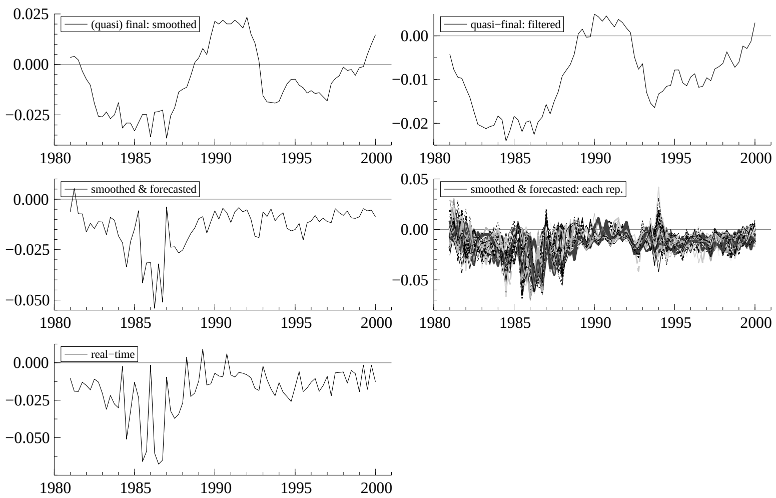

Figure 5: The real-time behaviour of the output gap using the bivariate UC filter

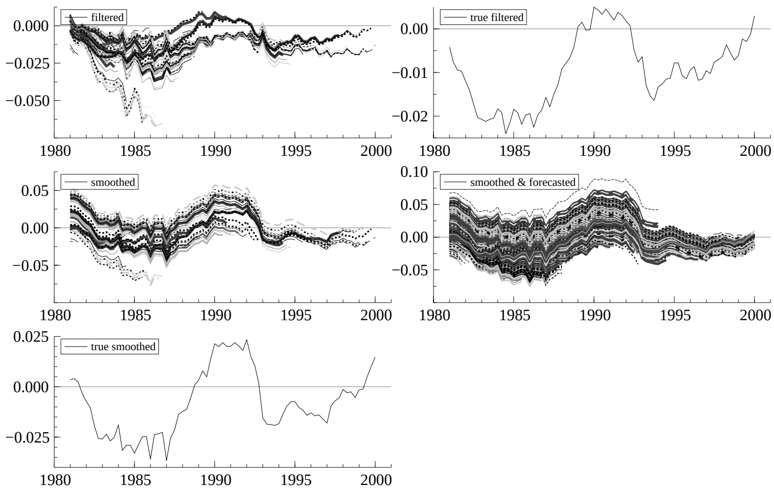

Figure 6: The evolution of real-time estimates using the bivariate UC filter

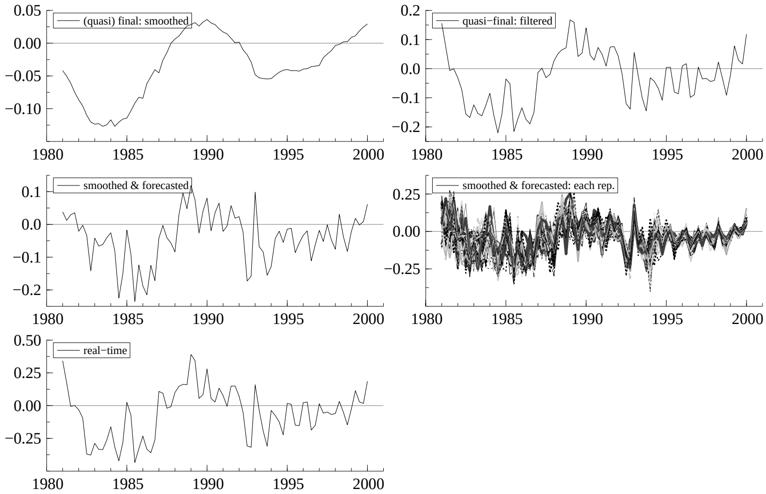

Figure 7: The real-time behaviour of the output gap using the bivariate Hodrick-Prescott filter

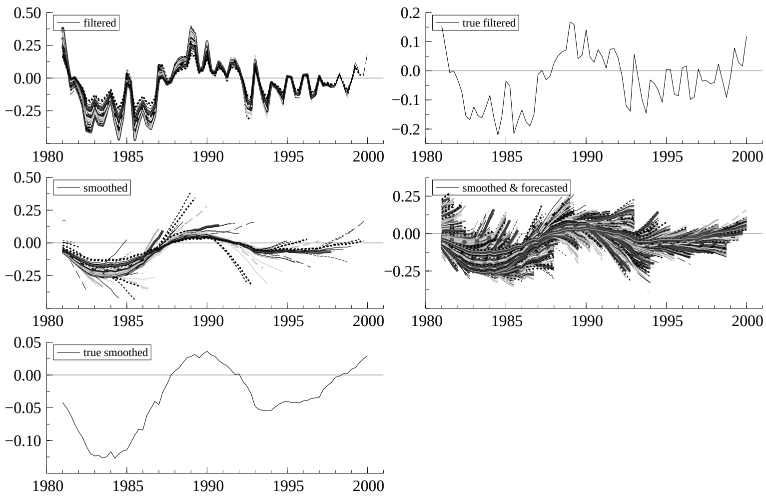

Figure 8: The evolution of real-time estimates using the bivariate Hodrick-Prescott filter

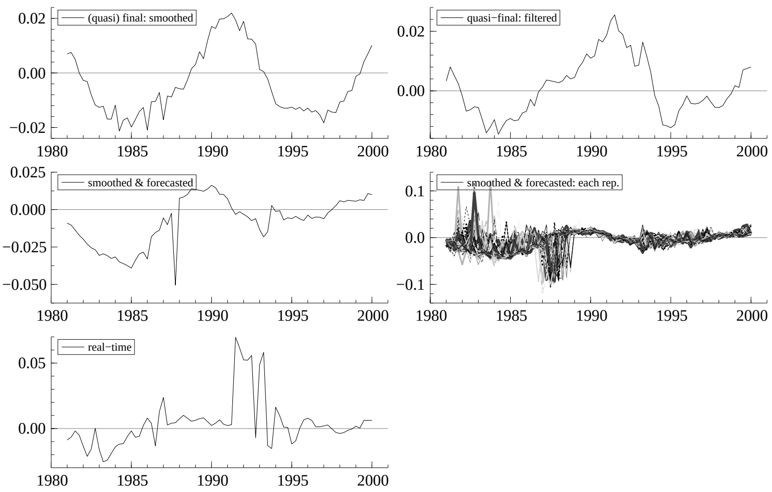

Figure 9: The real-time behaviour of the output gap using the trivariate UC model

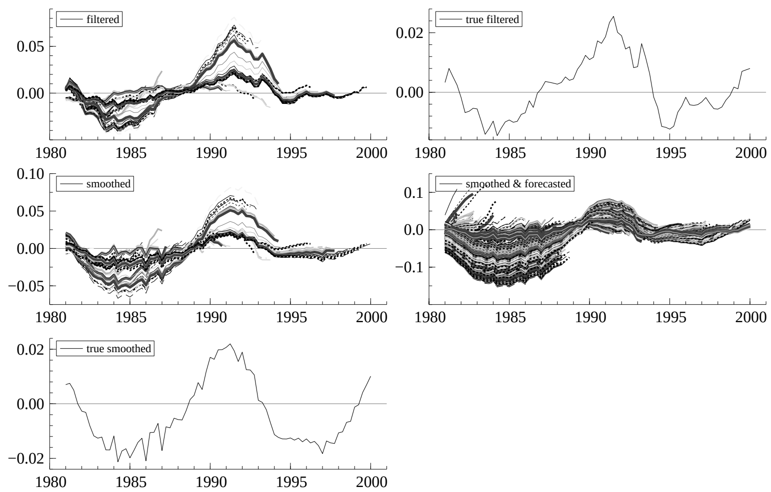

Figure 10: The evolution of real-time estimates using the trivariate UC model

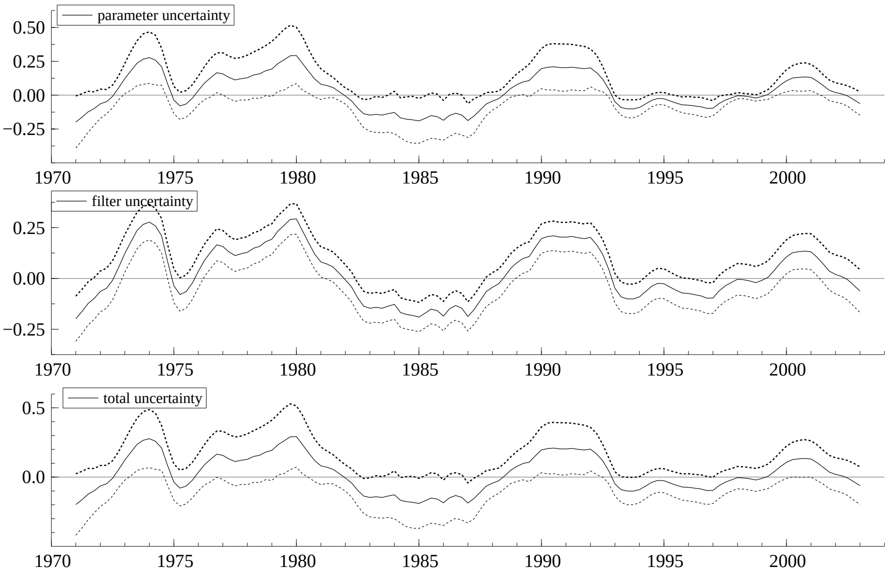

Figure 11: Filter, parameter and total uncertainty associated with Harvey-Trimbur smoothed estimates, 1971q1-2003q1: 95% confidence intervals

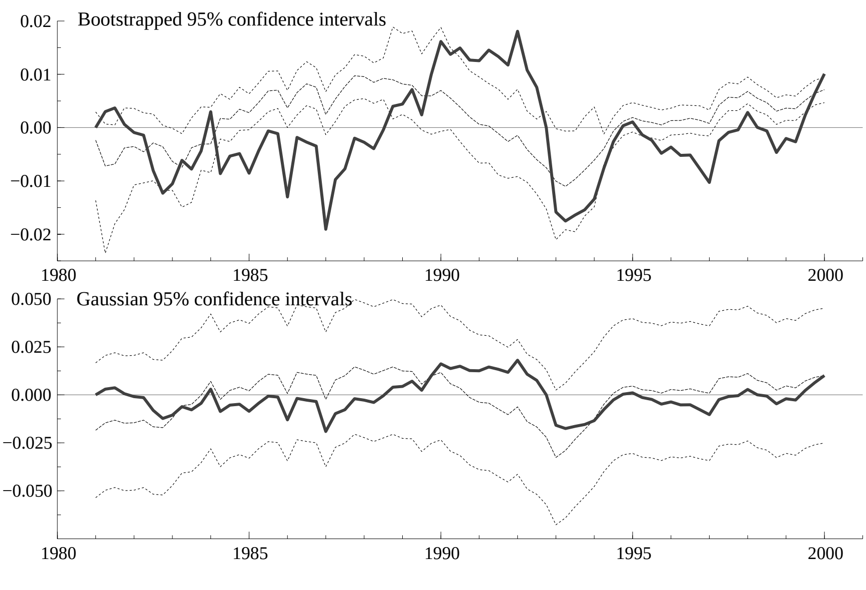

Figure 12: Uncertainty associated with real-time estimates using the HP filter. Note: final estimate in bold

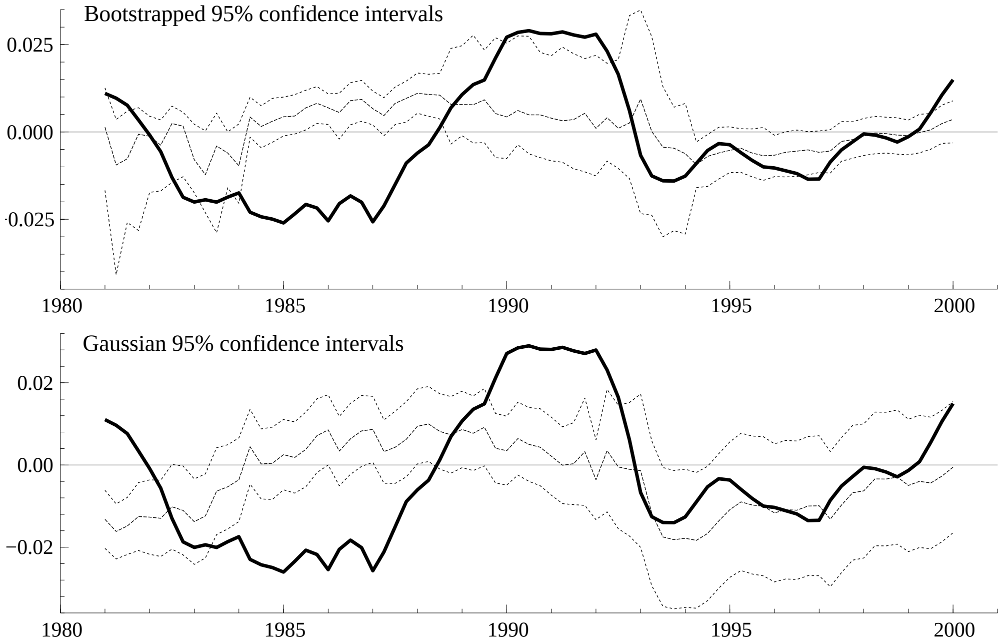

Figure 13: Uncertainty associated with real-time estimates using the Harvey-Trimbur filter. Note: final estimate in bold

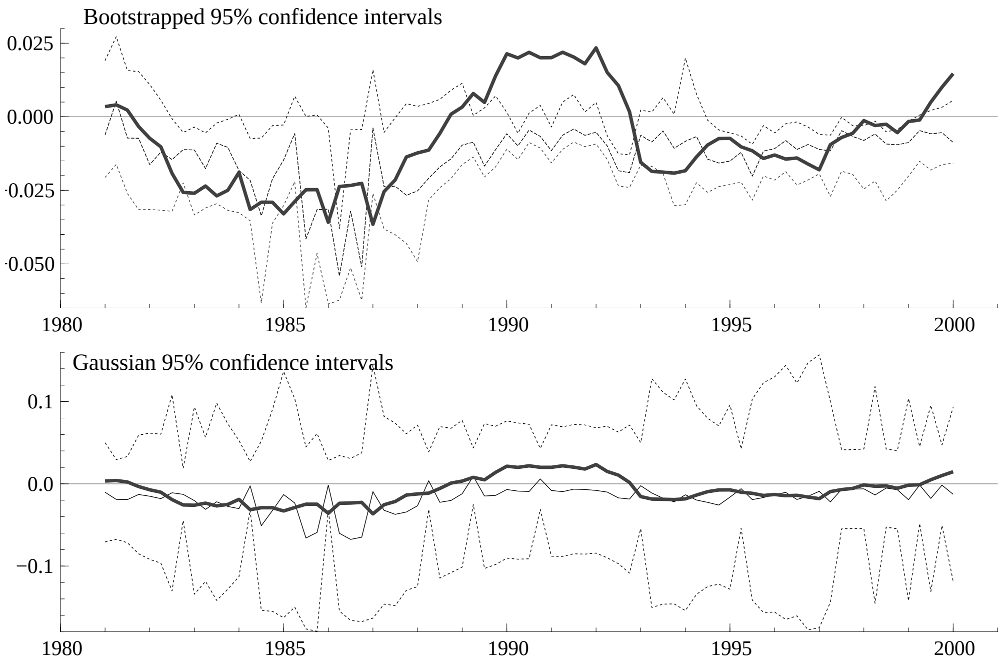

Figure 14: Uncertainty associated with real-time estimates using the bivariate UC filter. Note: final estimate in bold

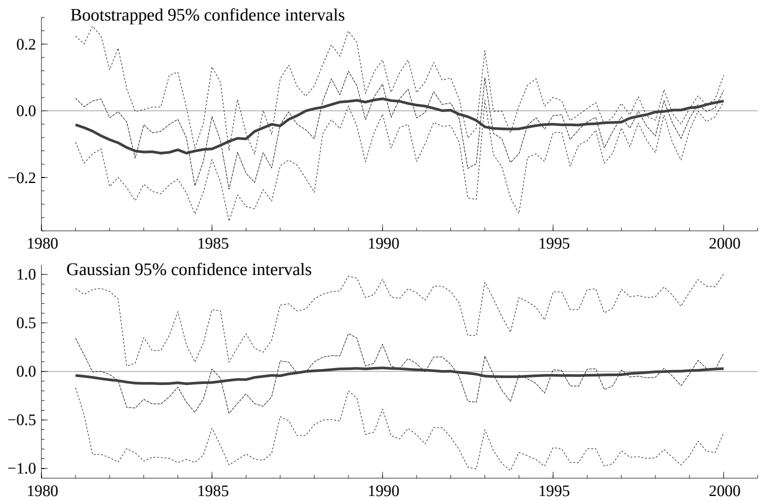

Figure 15: Uncertainty associated with real-time estimates using the bivariate Hodrick-Prescott filter. Note: final estimate in bold

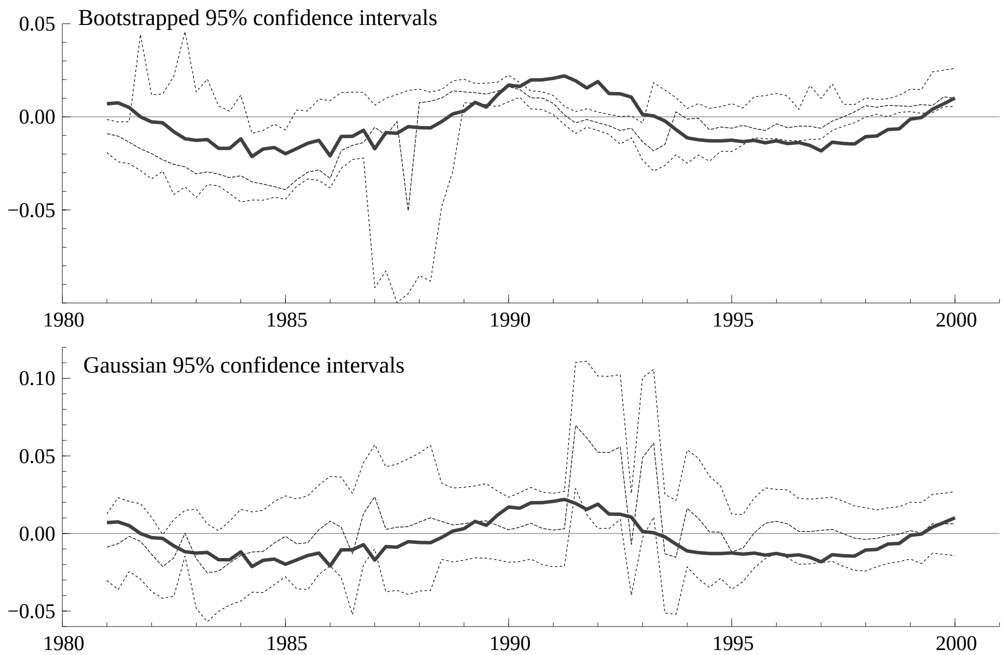

Figure 16: Uncertainty associated with real-time estimates using the trivariate UC mode. Note: final estimate in bold