# Should we be surprised by the unreliability of real-time output gap estimates? Density estimates for the Euro area

James Mitchell*
National Institute of Economic and Social Research

This Revision: January 2005

**Abstract**

Recent work has found that, without the benefit of hindsight, it can prove difficult for policy-makers to pin down accurately the current position of the output gap; real-time estimates are unreliable. However, attention primarily has focused on output gap point estimates alone. But point forecasts are better seen as the central points of ranges of uncertainty; therefore some revision to real-time estimates may not be surprising. To capture uncertainty fully density forecasts should be used. This paper introduces, motivates and discusses the idea of evaluating the quality of real-time density estimates of the output gap. It also introduces density forecast combination as a practical means to overcome problems associated with uncertainty over the appropriate output gap estimator. An application to the Euro area illustrates the use of the techniques. Simulated out-of-sample experiments reveal that not only can real-time point estimates of the Euro area output gap be unreliable, but so can measures of uncertainty associated with them. The implications for policy-makers use of Taylor-type rules are discussed and illustrated. We find that Taylor-rules that exploit real-time output gap density estimates can provide reliable forecasts of the ECB's monetary policy stance only when alternative density forecasts are combined.

**Keywords**: Output gap; Real-Time; Density Forecasts; Density Forecast Combination; Taylor Rules

**JEL Classification**: E32; C53

---

*Thanks to Andrew Harvey, Kostas Mouratidis, Tommaso Proietti, Simon van Norden, Ken Wallis and Martin Weale for helpful comments. Thanks also to Yongmiao Hong for making available GAUSS code to implement his test referred to below. 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. Mitchell also acknowledges financial support from the ESRC (Award Reference: RES-000-22-0610). 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 Introduction

Policy makers require output gap estimates in real-time.[1] They do not have the luxury of being able to wait 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] Their problem can be interpreted as a forecasting one, since these real-time output gap estimates are forecasts, in the sense that they are expectations of the output gap conditional on incomplete information. Only with the arrival of additional information, such as revised historical data and data not available at the time, do the output gap estimates eventually settle down at their "final" values.[3]

Recent work has found real-time or end-of-sample output gap (point) estimates to be unreliable, in the sense that there is a large and significant revision or forecasting error; see Orphanides & van Norden's (2002) application to the US economy.[4] The revisions associated with real-time estimates are considerable; indeed for the US they were found to be as large as the output gap estimates themselves. So-called *data* revisions, explained by revisions to published GDP data, were found to be 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 it.

Clearly policy-makers misjudging the position of the business cycle in real-time, or in other words making poor forecasts, can lead to sub-optimal policy-decisions; e.g. see Nelson & Nikolov (2001), Orphanides (2001, 2003*b*) and Ehrmann & Smets (2003). The findings for the US, therefore, are worrying. Below we find that a similar picture of revisions to output gap estimates emerges when we look at the Euro area business cycle.

But, should we be surprised by this unreliability? Previous work has largely overlooked this question. With a couple of exceptions to which we turn below, focus has hitherto been on the point forecast; as in the output gap is currently 2%, for example. But it is not a question of this forecast proving to be right and another forecast proving to be wrong. Point forecasts are better seen as the central points of ranges of uncertainty. A forecast of 2% must mean that people should not be surprised if the output gap turns out to be a little larger than that. Moreover perhaps they should not be very surprised if it turns out to be much larger or indeed nothing at all. Therefore, consistent with recent

[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.

[3] Values are never truly final because data revisions and the arrival of new data are a continuous process.

[4] This unreliability, of course, has long been appreciated qualitatively; e.g. see Morgenstern's (1963) discussion (p. 268) where he views the revisions typically associated with national accounts data as "casting serious doubts on the usefulness of national income figures for business cycle analysis".

developments in the forecasting literature, it is important to provide a description of the uncertainty associated with real-time output gap estimates. Density forecasts provide a complete description of this uncertainty. Their importance can be seen by recalling that unless users of output gap estimates have symmetric, quadratic loss functions when making a decision they need to focus on not just the point estimate but the density; e.g. see Granger & Pesaran (2000). Given that the output gap is widely perceived as a leading indicator, albeit empirically perhaps a poor one, of future inflation an asymmetric inflation target implies an asymmetric loss function for the output gap.[5] In such a case the 'optimal' real-time estimate of the output gap need not equal the mean or conditional expectation: it can be 'rational' to use biased real-time estimates; e.g. see Elliott et al. (2004). Furthermore, measures of uncertainty are useful in their own right if interested in analysing and communicating, for example, risk and volatility, or the probability of a downturn.

Although previous work, such as Orphanides & van Norden (2002) and Camba-Mendez & Rodriguez-Palenzuela (2003), has provided measures of uncertainty associated with output gap estimates, *via* estimated standard errors, as indicated above focus has remained on the point estimates; the quality of real-time output gap estimates has been evaluated primarily by focusing on the degree and nature of *ex post* revisions to output gap point estimates.[6] Typically this has involved examination, for example, both of the correlation between the real-time and final point estimates and the difference between the real-time and final estimates. When presented, measures of uncertainty have been evaluated in specific ways - it has not been made explicit what defines a "good" measure of uncertainty. This can leave us unclear about the reliability of real-time output gap estimates and their uncertainty.

In this paper we provide a more general discussion of the importance of providing measures of uncertainty, *via* density forecasts, associated with real-time output gap point estimates.[7] We explicitly relate the computation and evaluation of measures of uncertainty associated with real-time output gap estimates to recently developed techniques from the forecasting literature. By comparison, previous work has left this relationship, at best, implicit and relied on *ad hoc* tests. Camba-Mendez & Rodriguez-Palenzuela (2003) test if the standard error of the real-time estimates is equal to that of the final estimates using

[5] This holds for a linear relationship between inflation at time $t+h$ (where $h$ is the forecast horizon) and the output gap at time $t$. It seems sensible, and consistent with other commentators' views, to interpret the European Central Bank's [ECB] target of keeping annual inflation at 2% or less as an asymmetric target. Only if the ECB are equally happy with inflation at, say, 0% as at 2%, implying a uniform loss function, can their target be interpreted as symmetric.

[6] Other work has noted that the standard error bands around output gap estimates are often large but not sought to evaluate the accuracy of these bands formally; e.g. see Gerlach & Smets (1999), Staiger et al. (1997) and Smets (1999). Applications, usually to the US, typically find standard errors slightly larger than 1%. This implies that an output gap point estimate must be at least $\pm 2\%$ (assuming normality) before we can view it as statistically different from zero (at around a 95% level of significance).

[7] Our work builds on pioneering contributions, such as Stone et al. (1942) and Morgenstern (1963), that look at the uncertainty of economic statistics. Morgenstern, for example, stresses the uncertainties associated with national income statistics and begins to quantify them using uncertainty bands (see Figure 10, p. 269).

what can be interpreted as a test for equal (point) forecast accuracy similar to Diebold & Mariano (1995), while Orphanides & van Norden (2002) actually use a type of Diebold-Mariano test on the second moments of the revision between real-time and final estimates to test if the variability of the revision process is consistent with what was expected *ex post* rather than *ex ante*.[8] In contrast, we consider real-time (*ex ante*) computation of measures of uncertainty associated with real-time estimates and methods of evaluation that have a clear interpretation and indeed a growing pedigree. The clarity derives from defining what constitutes a "good", or even "optimal" (in some respect), density forecast.

In this paper we also address an important practical problem faced by users of output gap estimates that is another important source of real-time output gap uncertainty, namely model uncertainty. To-date there is no consensus on the appropriate output gap estimator. Various estimators are used by policy-makers and academics; these include Hodrick-Prescott filters, band-pass filters and univariate and multivariate unobserved components models. Inevitably these paint contrasting pictures as to the position of the output gap. It is not unusual to find one estimator suggesting the economy is above trend with another indicating the reverse. Our results for the Euro area are no different; see also Orphanides & van Norden's (2002) results for the US. We consider five estimators representative of those used by applied macroeconomists and find that two suggest the output gap, according to our most recent estimates, is positive while three suggest it is negative. Similarly when looking not just at the point estimates but the whole density, we find substantial differences across the five density estimates.

This dissension across alternative (competing) estimates of the output gap will be familiar to macroeconomists. Indeed it may have contributed to anecdotal and published

[8] In the notation of Section 3, Camba-Mendez & Rodriguez-Palenzuela (2003) test if $P_{t|t} = P_{t|T}$, where $P_{t|t}$ is the variance of the real-time estimate $y^C_{t|t}$ and $P_{t|T}$ is the variance of the final estimate $y^C_{t|T}$. Note that the parameters are estimated, respectively, using current and full-sample information. Under squared error loss, Gaussianity and both the real-time and final estimates having zero mean around some 'true' estimate (we could think of the final estimate as this 'true' estimate), this amounts to a test for equal forecast performance similar to Diebold & Mariano (1995). The hypothesis is tested using an $F$-test, whose appropriateness rests on certain, unstated, rather strong conditions such as no serial correlation. Orphanides & van Norden (2002), on the other hand, examine the revision $R_{t|T} = \left(y^C_{t|T} - y^C_{t|t}\right)$ and test if $E(R_{t|T}) = 0$ and $E(R_{t|T})^2 = P_{t|t} - P_{t|T}$ *via* a Diebold-Mariano type test. The estimator for the variance of the revision, $P_{t|t} - P_{t|T}$, is not computable in real-time (*ex ante*), however, requiring full-sample information, $T$, to both derive the final or smoothed estimates and estimate the model parameters $\Theta$ on which the expression for $E(R_{t|T})^2$ given above is conditional. Nevertheless, it is interesting to compare their statistic with one of the evaluation tests considered below. Their test statistic is: $\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y^C_{t|T}-y^C_{t|t}}{\sqrt{P_{t|t}-P_{t|T}}}\right)^2 - 1\right] \rightarrow N\left(0, \frac{2\pi f(0)}{T}\right)$ where $f(0)$ is some consistent estimate of the spectral density of $\left\{\left(\frac{y^C_{t|T}-y^C_{t|t}}{\sqrt{P_{t|t}-P_{t|T}}}\right)^2\right\}_{t=1}^{T}$ at frequency zero. In contrast our test for whether uncertainty is "correctly" accounted for by the forecast [see Section 3.2] requires the scaled revision errors $\left\{\frac{y^C_{t|T}-y^C_{t|t}}{\sqrt{P_{t|t}}}\right\}_{t=1}^{T}$ to be *i.i.d.* $N(0,1)$.

evidence that the output gap is a concept of limited empirical value.[9] However we argue that any such conclusion is premature. In practice we know that policy-makers consult various estimators to inform their judgement. This provides a rationale for model-averaging. We suggest that when the output gap is to be used for a specific purpose, like modelling or forecasting interest-rates or alternatively inflation, one should consider combining information across these alternative estimators.[10] However, while it is well established (e.g. see Bates & Granger (1969) and Stock & Watson (2004)) that combining competing individual point forecasts of the same event can deliver more accurate forecasts, in the sense of a lower root mean squared error (RMSE), little attention has been paid to the combination of forecasts and their uncertainty. Accordingly, we propose a simple method of combining information across alternative density forecasts that delivers the "optimal" *pooled* or combined output gap estimator.[11]

The plan of the remainder of this paper is as follows. Section 2 illustrates the unreliability of real-time output gap point estimates in the Euro area. Parameter uncertainty is found to be a dominant source of this unreliability. We then turn to whether we should be surprised by this unreliability. To this end, Section 3 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 against the outturn. Section 4 then re-visits the simulated real-time application to the Euro area considered in Section 2. It indicates, and then evaluates, the degree of uncertainty associated with output gap estimates in real-time.

To draw out some implications of these empirical findings and introduce density forecast combination, Section 5 turns to examination of the effect of output gap uncertainty on prescriptions from monetary policy (Taylor-type) rules. It explains the circumstances in which uncertainty about real-time estimates should cause policy-makers to react less strongly to these data when they are more uncertain. The practical consequences of real-time uncertainty about the output gap for setting interest rates are then highlighted. We illustrate how uncertainty about real-time output gap estimates translates into uncertainty about the value of the policy instrument. This uncertainty, about monetary policy stance, is again best represented by a density forecast. With an eye to establishing whether these Taylor-type rules explain the ECB's (notional) behaviour, we evaluate

[9]For example in the context of forecasting inflation Orphanides & van Norden (2004) conclude that their finding that real-time output gap estimates do not provide as good forecasts of inflation out-of-sample as benchmark models "call[s] into question the practical usefulness of the output gap concept for forecasting inflation" (Abstract).

[10]For example, $N$ different output gap estimates translate (given a model relating, say, the interest rate to the output gap) into $N$ forecasts of the interest rate that can then be combined *via* least-squares regression against the outturn (the interest rate); see Granger & Ramanathan (1984). If we wish to combine the $N$ output gap estimates directly we confront the practical problem that the outturn (the 'true' value of the output gap) is not observed. One solution is to follow Smith et al. (1998) and combine the $N$ output gap estimates based on the assumption that they are all noisy estimates of some common 'true' (but unknown) value of the output gap.

[11]It may also prove convenient when communicating about the output gap to focus on one (pooled) estimator rather than many.

whether the implied density forecasts for the interest rate offer a good characterisation of actual interest-rate movements.[12] Conceptually this represents an important departure from existing studies that test how well Taylor rules fit the data with a focus on point estimates. One can potentially gain a misleading impression of these rules' explanatory/forecasting power by ignoring uncertainty.[13] Following Orphanides (2001) we distinguish between real-time and final renditions of the policy-rule. We find that individually the five individual output gap estimators considered in Section 4, for the simple rules considered, do not in real-time provide a satisfactory representation of interest rates. Furthermore, reflecting model uncertainty, there is considerable disagreement across the five estimators as to the interest rate consistent with the policy rule. Accordingly, we propose a simple method of combining information across these alternative density forecasts. The *pooled* or combined output gap estimator is defined as that linear combination of the individual density estimates that best explains actual interest rate movements. This combined density forecast is found to offer improved density forecasts for the Euro-area interest-rate. Section 6 offers some concluding comments. Details of the data and models used are confined to appendices.

# 2 The unreliability of real-time output gap point estimates illustrated. An application to the Euro area

Euro area data are taken from the ECB's Area Wide Model (AWM) database; see Appendix A for more details.[14] It is expedient in an application to the Euro area to focus on statistical revisions alone as construction of a 'real-time' data set is not readily possible. We do, however, make an attempt at construction of an "approximate" real-time data set for GDP; see Appendix A.1.[15] This is meant merely to be suggestive and in the main body of the paper attention is restricted to the AWM data, i.e. to statistical revisions alone, by considering what amounts to a 'final' data vintage.

To illustrate the unreliability of real-time output gap estimates the following experi-

[12] If the rule does replicate actual interest-rate movements then by using the rule as a benchmark the ECB ensure a degree of continuity in monetary policy behaviour pre and post EMU. For a discussion of some normative aspects of these rules (e.g. how 'good' are they?) with a focus on the ECB see Peersman & Smets (1999).

[13] In a related literature, authors have argued that the failure to find empirical support out-of-sample for nonlinear business cycle forecasts may be explained by the traditional focus on point forecasts and RMSE. They argue that nonlinear models may do better at forecasting higher moments, that are captured by density forecasts; see Clements et al. (2003).

[14] Related studies that examine real-time output gap estimates for the Euro area are Runstler (2002) and Camba-Mendez & Rodriguez-Palenzuela (2003). They ignore data revisions, and in fact use AWM data as in the majority of this paper. These AWM data involve aggregating the available national data using the so-called index method that, for example, defines the log of Euro area GDP as the weighted sum of the log of country-level GDP. The weights are based on relative GDP shares; see Fagan et al. (2001) for further details.

[15] Although data on other variables are used to help estimate the output gap we should expect data revisions to them to be minor compared with those typically experienced by GDP.

ment is undertaken. 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.[16] 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 mimic 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.

When the latest vintage of data is used throughout the experiment, what we call the real-time estimate is strictly the *quasi-real* estimate of Orphanides & van Norden (2002). As indicated, we do also provide some tentative indication of the impact of data revisions by, from 1993q3, also de-trending using the data vintage available at the time rather than the final one. This is the first time data unreliability, albeit imperfectly proxied, has been considered in the context of Euro area output gap estimation.

To give an indication of what the final output gap estimates (using the AWM data) look like, Figure 1 presents them for two univariate and three multivariate unobserved components (UC) estimators.[17] These estimators are representative of those used both by policy-makers and in related studies.[18] For a review of these estimators and the specifications used see Appendix B. Although the five different estimates share the same shape, picking up the trough in the mid 1980s and the peak of the early 1990s, Figure 1 reminds us that inference is sensitive to measurement.[19] The bivariate HP estimator produces cycles with a far greater, indeed implausible, amplitude than the other estimators; this perhaps illustrates the dangers of imposing, rather than estimating, model parameters.

[16] We ignore the fact that GDP data are published with, at least, a one quarter lag.

[17] It is considered important to consider forecasts from more than one model given that model uncertainty 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.

[18] We did also experiment with structural VAR based estimators of the output gap; see Mitchell (2003). Results are not presented since they were not dissimilar to those presented in Table 1 for the bivariate and trivariate UC estimators. In any case, the methodological differences between VAR and UC estimators are, to a degree, illusionary since the VAR model can be re-written in state-space form; e.g. see DeSerres & Guay (1995). We should note, however, that when using the well known Blanchard-Quah identification method for VAR models, results were found to be extremely sensitive to the chosen lag order, $p$. Only with very high $p$ did we obtain *a priori* plausible looking cycles; in this case the performance of the VAR cycle in real-time was similar to that of the multivariate UC estimators in Table 1. But when $p = 1$ the correlation coefficient in Table 1 increased to 0.9. Although similar results, suggesting that reliable output gap estimates can be obtained in real-time for the Euro area, have been obtained by Camba-Mendez & Rodriguez-Palenzuela (2003), see their Figure 3, we believe this result is misleading. It is based on an implausible looking business cycle with, for example, an amplitude in the range -0.5% to 0.5%, rather than around -2.5% to 2.5% as is plausible and indeed typical for the other estimators.

[19] Other studies of the Eurozone business cycle, such as Runstler (2002) and Artis et al. (2004), find similarly shaped cycles to Figure 1.

We can also see the end-of-sample problem; the different estimators in Figure 1 do not agree about the current position of the cycle.

The real-time unreliability of output gap point estimates then is summarised in Table 1 for the five representative output gap estimators. Reflecting the fact that the five estimators in Table 1 can be interpreted within an UC framework [see Appendix B] and that UC models use the data in two ways, as well as considering the real-time estimate, Table 1 also examines the filtered estimates (or the *quasi-final* estimates in the parlance of Orphanides & van Norden (2002)).[20] Let us denote the real-time estimates, based on recursively updated estimates of the parameters of a given UC model $\Theta$, by $y^C_{t|t}$; let $y^C_{t|T}$ denote the final estimates of the output gap at time $t$, that use full-sample information $T$ to estimate $\Theta$, and then let $y^C_{t|t}(\widehat{\Theta}_T)$ denote the filtered estimates of the output gap at time $t$ using these full-sample based parameter estimates. For the univariate HP filter, as the only parameter is chosen *a priori*, clearly there is no parameter uncertainty; the filtered and real-time estimates are then equivalent; $y^C_{t|t} = y^C_{t|t}(\widehat{\Theta}_T)$.[21] The reliability of the real-time and filtered point estimates is summarised by their correlation against the final estimates and the noise-to-signal ratio (NSR). The NSR is the ratio of the RMSE of the revision (the difference between the real-time or filtered estimate and the final estimate) to the standard deviation of the final estimate of the output gap.[22] To provide some indication of how sensitive results are to the chosen sample period, Table 1 contrasts the performance of the real-time and filtered estimates computed over the period 1981q1-2000q1, with those computed over the period 1993q3-2000q1. This latter period was chosen as it is after both the peak in economic activity and the exchange-rate crises in the early 1990s. The final column of Table 1 provides a tentative indication of the impact of data revisions on real-time output gap estimation and presents the correlation of the real-time estimates, using real-time data vintages, against the final estimates using the final data vintage.[23]

Table 1 indicates that as the future becomes the present output gap estimates are

[20] UC models use the data in two ways in the sense that first they estimate the parameters of the model, denoted by the vector $\Theta$, and secondly they use these estimates to obtain the filtered and smoothed estimates of the output gap, namely the quasi-final and final estimates of the output gap, respectively. The filtered and smoothed estimates are the expected value of the output gap conditional on information available at time $t$ ($t = 1, 2, ..., T$) and $T$, where $T \geq t$, respectively.

[21] In general, as the parameters of the UC model are estimated by maximum likelihood (ML) and since ML estimation, under standard regularity conditions, delivers consistent estimators asymptotically ($\widehat{\Theta}_t \xrightarrow{p} \Theta$, $\forall t$) we should not expect a difference between real-time and filtered estimates. This holds as long as we view the UC model under consideration as the 'true' data-generating-process. In practice over time, in the presence of deterministic and stochastic nonstationarities, we should expect macroeconomists to revise the UC model they use to estimate the output gap.

[22] It is of theoretical interest that for correctly specified UC models this ratio in the case of the filtered estimates (i.e. if we ignore the fact that we recursively re-estimate the parameters of the UC model) depends on the type of UC model considered; see Proietti (2004).

[23] It should be noted that the filtered estimates using the real-time data set (not reported in Table 1) are not identical, as they should be, to those using the AWM data. They are, however, qualitatively similar. Due to its method of construction, detailed in Appendix A.1, the final column of the real-time data matrix for GDP is not identical to the AWM database.

revised. Real-time point estimates of the output gap in the Euro area, as in the US, are unreliable, in the sense that there is a large and important revision error. This is reflected by the correlation coefficients, over the period 1981q1-2000q1, being in the range 0.27-0.76 and the noise-to-signal ratio exceeding unity for three of the five output gap estimators, and for the remaining estimators being greater than 0.8. Reflecting the importance of *ex post* information in re-defining the parameter values the filtered estimates are more reliable than the real-time estimates: correlation is higher and NSR lower. Parameter uncertainty appears to be a dominant source of the unreliability of real-time estimates. Interestingly, as in the US, data revisions do not appear to be a major source of revisions. The real-time estimates using real-time data were found to have similar properties to the real-time estimates that use final (AWM) data. This is reflected here in the final column of Table 1 that indicates that use of the real-time data, rather than the final data considered in the sixth column, leads to estimates correlated similarly with the final estimates.

There is some variation across estimators and the sample period. Comparing the estimators across the period 1981q1-2000q1, correlation ranges from 0.27 for the univariate HP measure to around 0.6-0.7 for the multivariate measures. In this sense 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. However, the multivariate estimators have higher noise-to-signal ratios than the univariate UC estimator.

This unreliability of real-time point estimates also shows up over the period 1993q3-2000q1. Although the univariate estimators are more reliable over this later period than 1981q1-2000q1, some of the multivariate estimators appear to be less reliable. This may reflect the fact that they do well at picking up the big movements, such as the cyclical peak in the early 1990s, but are worse at picking up the less volatile movements that appear to characterise the *modern* business cycle.

Table 1: The unreliability of real-time output gap point estimates. The correlation (*cor*) of alternative real-time and filtered estimates of the Eurozone output gap against the final estimates and the noise-to-signal ratio ($NSR$)

| | AWM data | | | | | | | | real-time data |
|---|---|---|---|---|---|---|---|---|---|
| | 1981q1-2000q1 | | | | 1993q3-2000q1 | | | | 1993q3-2000q1 |
| | real-time | | filtered | | real-time | | filtered | | real-time |
| | cor | NSR | cor | NSR | cor | NSR | cor | NSR | cor |
| Uni HP | 0.273 | 1.43 | 0.273 | 1.43 | 0.875 | 1.17 | 0.875 | 1.17 | 0.896 |
| Uni UC | 0.600 | 0.82 | 0.838 | 0.64 | 0.904 | 0.57 | 0.898 | 0.71 | 0.863 |
| Bi UC | 0.599 | 0.87 | 0.941 | 0.56 | 0.864 | 0.70 | 0.923 | 0.58 | 0.894 |
| Bi HP | 0.688 | 3.31 | 0.727 | 1.35 | 0.585 | 3.81 | 0.599 | 1.87 | 0.579 |
| Tri UC | 0.760 | 1.79 | 0.921 | 0.57 | 0.352 | 2.15 | 0.795 | 1.07 | 0.309 |

The findings of Table 1 beg the question, should we be surprised by this unreliability? To address this question, we therefore first take up the challenge of providing, in real-time, measures of uncertainty associated with real-time (point) estimates of the output gap, *via* density forecasts. Secondly we evaluate them to ascertain whether the unreliability of output gap point estimates is surprising.[24] Could we have anticipated this revision error? In real-time was the revision error to the real-time estimate within the bounds of what we could have predicted? It is important to evaluate whether these measures of uncertainty offer a reliable indication of the degree of unreliability associated with output gap estimates 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*.

# 3 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.[25] Density forecasts of the output gap provide an estimate of the probability distribution of its possible future values. 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.

Questions then arise over how the density forecasts should be constructed. We consider a simple approach that relies on a state-space representation for the output gap estimator. In Section 3.1.1 we draw out some characteristics of this approach by noting that it can be seen to minimise forecast errors, in the sense that it is based on optimal forecasts.

## 3.1 Measuring the uncertainty associated with the output gap

Conditional on Gaussianity (of the disturbances driving the components of the state vector) and knowledge of the covariance matrix of the estimated state vector confidence intervals around the output gap can be presented, and density estimates derived.[26] Before continuing our discussion we note that other distributions could be considered, although we confine attention in the rest of this paper to the Gaussian case since this both has the advantage of simplicity/familiarity and is sufficient to illustrate the role of density

[24] 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, or evaluate, them for real-time estimates.

[25] For a review of density forecasting see Tay & Wallis (2000).

[26] 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. This approach has also been followed, for example, by Orphanides & van Norden (2002).

forecasting.[27] For example, since 1996q1 the Bank of England has published the so-called "fan" chart of inflation in the UK, that allows for skewness. The fan chart is based analytically on the two-piece normal distribution; see Wallis (2004). This distribution has the convenient property that it can capture any asymmetries between upside and downside risk but still be computed using standard normal tables. Alternatively, rather than making some distributional assumption one could let the data *decide*; one could analyse the historical forecast errors. Past forecast errors are commonly used as a practical way of forecasting future errors; e.g. see Wallis (1989), pp. 55-56. Given the backward looking and mechanistic nature to this method of determining the distribution and its moments, it is important what historical sample period is chosen to analyse the errors. Historical errors are expected to provide a poor guide to future behaviour in the presence, for example, of regime changes.[28]

Continuing our discussion, let $P_{t|t}$ denote the Kalman filter based variance of the output gap at time $t$, $y^C_{t|t}$, using information available up to time $t$; the output gap is one of the elements of the state vector in the UC model. Then conditional on Gaussianity, the density is $N(y^C_{t|t}, P_{t|t})$. Denote the revision between the real-time and final estimates, the focus of this paper, by $R_{t|T} = y^C_{t|T} - y^C_{t|t}$. It is helpful to recall that with known and constant parameters (i.e. for the *filtered* rather than *real-time* estimates of Table 1), this variance $P_{t|t'}$ decreases as $t'$ increases; $P_{t|t} = P_{t|T} + Var(R^*_{t|T})$. The variance of the filtered (or one-sided) estimate is greater than that of the smoothed (or two-sided) estimate, $P_{t|T}$, by a positive scalar equal to the variance of the revisions between the filtered and smoothed estimates $R^*_{t|T}$ where $R^*_{t|T} = y^C_{t|T} - y^C_{t|t}(\widehat{\Theta}_T)$. This is the familiar result that the variance of the filtered (forecasted) estimate is equal to the variance of the outturn, $P_{t|T}$, plus the square of the bias, $Var(R^*_{t|T})$. This assumes the revision process has mean zero so that $Var(R^*_{t|T}) = E(R^*_{t|T})^2$.

Below in Section 3.2 we examine how to evaluate these interval and density forecasts once the final output gap estimate, $y^C_{t|T}$ where $T \to \infty$, has become available.[29]

[27]The non-Gaussian case may require modification of the traditional state-space architecture. Durbin & Koopman (2001) provide an account of how non-Gaussian (and nonlinear) state-space models can be handled using simulation techniques.

[28]Blake (1996) considered how stochastic simulation, with a coherent policy structure, can be used as an alternative to historical errors to measure the uncertainty associated with, in his application, the inflation rate. It is explained that this is expected to deliver a better measure of uncertainty if a new policy regime (say a new target for inflation or EMU) has been adopted.

[29]These measures of uncertainty do not capture parameter uncertainty. Recall that for UC estimators two measures of uncertainty associated with the cyclical component of output can be distinguished: filter and parameter uncertainty. If $\Theta$ 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 $\Theta$ is estimated, say by $\widehat{\Theta}_t$; i.e. $P_{t|t}(\widehat{\Theta}_t) > P_{t|t}(\Theta)$ and $P_{t|T}(\widehat{\Theta}_t) > P_{t|T}(\Theta)$. Similarly to Orphanides & van Norden (2002) we did experiment with using analytical approximations (see Quenneville & Singh (2000)) to capture the parameter uncertainty. We should expect this to add to uncertainty most for those measures of the output gap that estimate most parameters, namely the multivariate measures. However, results were qualitatively similar to those presented here. In fact, as we will shall see in Figure 2, in general, filter uncertainty alone appears to often over-estimate the degree of

### 3.1.1 Understanding revisions to real-time output gap point estimates

In the absence of data revisions, revisions to real-time output gap estimates are explained by forecasting errors. This is seen as follows. Consider the final estimate of the output gap to be a known weighted linear function, that is say centered and symmetric, of the underlying GDP data, say $y_t$:

$$y^C_{t|T} = B(L,F)y_t, \tag{1}$$

where $B(L,F) = b_0 + \sum_{j=1}^{\infty} b_j(L^j + F^j)$ and $Ly_t = y_{t-1}$ and $Fy_t = y_{t+1}$.

Assuming the output gap is a weighted linear function of $y_t$ as in (1) involves no loss in generality. De-trending filters, whether interpreted as parametric or nonparametric, can be seen to involve application of a moving-average filter to the raw data.[30] 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 since at the end of the sample the filter is unable to distinguish temporary from permanent shocks; e.g. see St-Amant & van Norden (1997).

The final estimate of the output gap is then given by:

$$y^C_{t|T} = B(L,F)y_t = b_0 + \sum_{j=1}^{\infty} b_j y_{t-j} + \sum_{j=1}^{\infty} b_j y_{t+j}. \tag{2}$$

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

$$y^C_{t|t} = b_0 + \sum_{j=1}^{\infty} b_j y_{t-j} + \sum_{j=1}^{\infty} b_j E(y_{t+j}|\Omega_t). \tag{3}$$

Subtracting (3) from (2) the revision is obtained as a function of the forecasting error, $(y_{t+j} - E(y_{t+j}|\Omega_t))$,

$$y^C_{t|T} - y^C_{t|t} = R_{t|T} = \sum_{j=1}^{\infty} b_j(y_{t+j} - E(y_{t+j}|\Omega_t)). \tag{4}$$

---

uncertainty and lead to excessively wide confidence bands.

[30] Traditionally business cycle analysts have often distinguished between parametric and nonparametric de-trending methods. However, in reality this distinction is somewhat blurred. Many nonparametic filters, such as the Hodrick-Prescott and band-pass filters, can be rationalised as parametric; see Harvey & Jaeger (1993) and Harvey & Trimbur (2003). Furthermore, the parametric methods, like the nonparametric ones, are also 'simply' taking weighted averages of the data. For example, Harvey & Koopman (2000) and Koopman & Harvey (2003) present algorithms for deriving the moving-average weights for parametric (UC) models.

It follows that if these forecast errors are reduced the revision will decrease. Forecasting future values prior to de-trending facilitates use of a less one-sided de-trending filter.[31]

Crucially the density $N(y^C_{t|t}, P_{t|t})$, considered above, can be seen implicitly to forecast future values optimally. This follows from the fact that for a correctly specified model, application of the one-sided Kalman filter is equivalent to application of the two-sided filter (smoother) to $y_t$ extended infinitely into the future with optimal forecasts. These optimal, minimum mean square error [MSE], forecasts are derived *via* the Kalman filter, exploiting the state-space representation for a given output gap estimator. There is therefore no need for forecast extensions.

Only when the UC model is misspecified can forecast extensions, using an alternative forecasting model, help to attenuate revisions [lessen estimation MSE]. The importance of real-time testing for model misspecification is therefore apparent; see Harvey & Koopman (1992) for appropriate diagnostic tests for UC models. Of course, it is well known that good in-sample fit need not translate into good out-of-sample performance. For completeness and as a reference we do note here how forecast extensions might be used to derive an alternative density to $N(y^C_{t|t}, P_{t|t})$. See Mitchell (2003) for further discussion and an empirical application.

Draw a large number $R$ of realisations from the predictive density of future observations $f_{t+l} = y_{t+l}|y_1, ..., y_t$ $(l = 1, ...h)$, based on an assumed forecasting model, say a simple autoregressive (AR) or some other robust forecasting model.[32] Then apply the Kalman smoother to the series $(y_1, ..., y_t)$ extended by the $R$ forecasts to get $R$ point estimates $\widehat{y}^C_{t|t+h} = E\left[y_t \mid y_1, ..., y_t, f_{t+1}, ..., f_{t+h}\right]$. The variance of these explicitly quantifies the degree of uncertainty associated with the fact that future values of the (log) level of output are unknown but known to affect real-time estimates. However, the variance of this Monte-Carlo density must have less dispersion than that of $N(y^C_{t|t}, P_{t|t})$.[33] This follows from the 'law of conditional variances' that states that $P_{t|t}$ can be decomposed into two components, only the second of which is the variance $Var[\widehat{y}^C_{t|t+h}]$:[34]

$$P_{t|t} = Var(\widehat{y}^C_t \mid y_1, ..., y_t) = E[Var\left[y_t \mid y_1, ..., y_t, f_{t+1}, ..., f_{t+h}\right]] + Var[\widehat{y}^C_{t|t+h}]. \quad (5)$$

The first component $E[Var\left[y_t \mid y_1, ..., y_t, f_{t+1}, ..., f_{t+h}\right]]$ is the average across $R$ of the variance $P_{t|t+h}$, delivered by the Kalman smoother.

Forecast extensions are not commonly used by analysts when using UC models to estimate the output gap. However, they have been employed when using nonparametric filters in real-time, whether for output gap estimation or seasonal extraction; e.g. see Wallis (1982), Stock & Watson (1999) and the discussion in van Norden (2004). Motivation

[31] In fact, as the forecasts are often a linear combination of the observed data, the filter applied is still one-sided in the original data.

[32] It is well known that simple AR models can guard against unforeseen events such as structural breaks; see Clements & Hendry (1999). See Garratt et al. (2003) for a discussion in terms of VAR models of how simulation techniques can be used to calculate density forecasts.

[33] Thanks to a referee for explaining this point.

[34] This 'law' states that for the random variables $y_t$ and $i$: $Var(y_t) = E[Var(y_t|i)] + Var[E(y_t|i)]$.

rests on the fact that the "optimal" real-time or one-sided nonparametric filter, in the sense of delivering the minimum mean square revision error, involves application of the two-sided filter to the actual data extended by optimal linear forecasts; for a discussion in the context of seasonal adjustment see Wallis (1982).[35]

## 3.2 Evaluation of density forecasts of the output gap

While there exist well established techniques for the *ex post* evaluation of point forecasts, often based around the RMSE of the forecast relative to the subsequent outturn, only recently has the *ex post* evaluation of density forecasts attracted much attention. Currently, following Diebold et al. (1998), the most widespread approach is to evaluate density forecasts statistically using the probability integral transform, itself a well-established result.[36] Diebold et al. 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.

We follow this approach and propose to evaluate real-time density forecasts of the output gap *ex post* with respect to the final estimates for the output gap $y^C_{t|T}$, where $T \to \infty$.[37] By treating these final estimates of the output gap as the 'outturn', evaluation of real-time output gap density forecasts is analogous to evaluation of a sequence of rolling one-step ahead density forecasts, the case commonly considered in this recent density forecast evaluation literature.[38]

[35]See also van Norden (2004) who, when discussing nonparametric filters, neatly shows that extending the series $\{y_t\}$ with optimal forecasts prior to de-trending can be seen as equivalent to deriving the *optimal* one-sided filter, in terms of minimising $E\left[\sum_{j=0}^{T-1}\widehat{b}_j y_{T-j} - \sum_{j=-\infty}^{\infty} b_j y_{T-j}\right]^2$ with respect to $\widehat{b}_j$, where $b_j$ are the optimal "ideal" weights given as $b_j = \frac{\sin ju - \sin jl}{\pi j}$ for $|j| \geq 1$ and $b_j = \frac{u-l}{\pi}$ for $j = 0$, where $u$ and $l$ are the bands outside of which the spectral gain is zero, and inside of which the gain is unity.

[36]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; e.g. see Granger & Pesaran (2000) and Clements (2004). Other work has evaluated density forecasts using scoring rules; e.g. see Giacomini (2002). Alternatively, density forecasts can be evaluated by reducing them to an interval forecast. For example, Mitchell (2003) focused on the central 50% or inter-quartile-range forecast and the 95% confidence intervals implied by the density forecast. Christoffersen (1998) and Wallis (2003) have proposed likelihood-ratio and, asymptotically equivalent, Pearson chi-squared tests for the evaluation of such interval forecasts.

[37]Of course since the final estimates are themselves estimated, and not known with certainty, we could consider evaluation of the real-time estimates not just against the final (point) estimate, as is traditional, but against the density of the final estimates. Mitchell (2003) did carry out such an exercise. However, it is perhaps little cause for celebration for policy-makers when analysts can only provide them with reliable interval and density forecasts of the output gap when the target (the final output gap they are trying to forecast) is itself acknowledged to be uncertain.

[38]In fact our real-time output gap estimates are contemporaneous rather than one-step ahead. Irre-

Real-time output gap density estimates are optimal, i.e. able to capture all aspects of the distribution of $y_{t|T}^C$, only when the sequence of probability integral transforms (pit's), $z_t$, defined below, is independently and identically distributed (*i.i.d.*) with a uniform distribution, U(0,1):

$$z_t = \int_{-\infty}^{y_{t|T}^C} p_t(u)du, \text{ or specifically in our case} \tag{6}$$

$$z_t = \Phi((y_{t|T}^C - y_{t|t}^C)/\sqrt{P_{t|t}}) , \tag{7}$$

($t = 1, 2, ...$), where $\{p_t(y_{t|t}^C)\}$ are the sequence of estimated real-time density forecasts and $\Phi$ is the standard normal cumulative density function (CDF); in our case $p_t(y_{t|t}^C) = N(y_{t|t}^C, P_{t|t})$.

By taking the inverse normal CDF transformation of $\{z_t\}$ to give, say, $\{z_t^*\}$ where $z_t^* = (y_{t|T}^C - y_{t|t}^C)/\sqrt{P_{t|t}}$ are the standardised revision errors, a 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.

In the empirical application below we abstract from any perceived uncertainties regarding the appropriateness of specific tests; we take an eclectic approach to testing *i.i.d.* uniformity/normality. We consider a range of statistical tests that have been used in empirical studies.[39] They have been used to detect misspecification in the mean, variance,

spective of this, for a rolling sequence of optimal forecasts we should not expect serial dependence in the forecast errors $\left\{y_{t|T}^C - y_{t|t}^C\right\}$ across $t$, yet alone in the standardised forecast errors $\{z_t^*\}$, defined below. In contrast we should expect dependence when evaluating a sequence of rolling optimal $h$-step ahead point forecasts or optimal fixed-event point forecasts; e.g. see Clements & Hendry (1998), pp. 56-62. Recall that $y_{t|t}^C$ are the *real-time* point estimates of the output gap considered in Table 1 (involving recursive re-estimation of the parameters of the UC model), not the filtered or *quasi-final* estimates. These use information available at time $T$, where $T > t$, to estimate the parameters $\Theta$ of the UC model. We continue to denote these filtered estimates $y_{t|t}^C(\widehat{\Theta}_T)$. Theoretically for a correctly specified UC model we should expect dependence in the revisions between the filtered and smoothed estimates $\left\{y_{t|T}^C - y_{t|t}^C(\widehat{\Theta}_T)\right\}$. With fixed (constant and known) parameters, application of the Kalman smoother, used to derive the final estimates $y_{t|T}^C$, will by construction induce serial dependence in these revisions across $t$. This is seen, for example, by considering a simple UC model, say the random walk plus noise model. Fixed-point smoothing and fixed-interval smoothing will result in the revisions being serially correlated across $t$; e.g. see Harvey (1989) equation 3.6.15 (p.153) and equation 3.6.17 (p.154). The nature of this dependence is dictated by the nature of the process governing the determination of $y_t$; for further discussion see Proietti (2004). As Proietti explains the smoother the trend the greater the revision. When no parameters are estimated, as with the HP filter, and the real-time and filtered estimates are equivalent in finite samples (as well as asymptotically which they are for all UC models) it should be no surprise (and in this sense constitute no violation of optimality) that revisions are serially correlated; it is a statistical artefact of the method chosen to identify the final estimates of the output gap.

[39] Alternatively, graphical means of exploratory data analysis are often used to examine the quality of density forecasts; see Diebold et al. (1998) and Diebold et al. (1999).

skewness and/or kurtosis of the forecasts. They differ not just in terms of their motivation, which we do not discuss here, but with respect to what they test. Some test for misspecification in all of the first four moments, while others focus on specific moments. Below we summarise relevant aspects of the tests:

1. Kolmogorov-Smirnov (KS) test for uniformity of $\{z_t\}$. The KS test relies on random sampling. As noted, for example, by Diebold et al. (1999) the effect of dependence on the distribution of the KS test statistic is unknown.

2. Anderson and Darling (AD) test for uniformity of $\{z_t\}$. Using Monte-Carlo Noceti et al. (2003) found the AD test to have more power to detect misspecification in the first four moments than the KS test (and related distributional tests).

3. The Bai and Ng (BN) 'robust' test of normality on $\{z_t^*\}$; see Bai & Ng (2003). Bai and Ng extend traditional (Jarque-Bera) normality tests, designed for *i.i.d.* data, to weakly dependent data. Essentially, any serial dependence is taken into account by consistently estimating the long-run variance (the spectral density at frequency zero) using a $HAC$ estimator that is robust to serial dependence (and heteroscedasticity). This test looks at the coefficients of skewness and kurtosis and therefore has no power to detect misspecification in the first two moments.

4. A Ljung-Box (LB) test of independence of $\{z_t\}$. To test for independence of the $\{z_t\}$ series we use the Ljung-Box test for auto-correlation; see Harvey (1989), p. 259. Since dependence may occur in higher moments we consider $(z_t - \overline{z})^j$ for $j = 1, 2, 3$; results are denoted LB1, LB2 and LB3, respectively.

5. The Hong joint test for uniformity and independence applied to $\{z_t\}$; see Hong (2002). This test is theoretically attractive as it offers a joint test. This means one can control the size of the test, something that cannot easily be done using separate tests for uniformity/normality and independence. Hong's joint test for *i.i.d.* U(0,1) is based on generalised spectral analysis. We estimate the spectrum nonparametrically using the Bartlett kernel and, following Hong, use a data-driven approach to determine the bandwidth. This requires us to choose a preliminary bandwidth. Hong proposes two test statistics, $M_0$ and $M_1$. Hong recommends $M1$ in small samples; accordingly we consider this. Hong also proposes a test for independence robust to non-uniformity.

# 4 The uncertainty of real-time output gap estimates in the Euro area: density forecasts

Figure 2 provides a visual indication of the degree of estimated real-time uncertainty for the five UC based estimators of the output gap by plotting 95% confidence intervals around the point estimates. There are striking differences in the degree of uncertainty (predicted in real-time) associated with the real-time estimates across the alternative estimators of the output gap. As is to be expected, uncertainty is lowest for the Hodrick-Prescott estimator. It is largest for the bivariate Hodrick-Prescott estimator where the degree of uncertainty is, incredibly, large.[40] This reflects the imposition of *a priori* parametric restrictions leading to a poorly defined model, in a statistical sense, where there is considerable uncertainty about the values of the remaining free parameters.

The Gaussian bands are very rarely significantly different from zero; policy-makers on this basis could never be sure about the position of the business cycle. With estimates of the standard error often in excess of 1%, point estimates for the output gap must be at least $\pm 2\%$ to be statistically significant, at traditional significance levels. But is the finding that these bands nearly always cover zero, in fact, correctly quantifying the degree of uncertainty associated with real-time estimates? This can be analysed formally using the statistical tests outlined in Section 3.2. First, however, it is useful simply to contrast the confidence bands for each measure with the actual outturn (printed in bold face) by looking again at Figure 2. There are again important differences across the alternative output gap estimators. The outturn frequently falls outside the 95% confidence bands for the Hodrick-Prescott estimator.

Table 2 complements Figure 2 by providing the results of the formal evaluation tests for the predicted measures of uncertainty. Density evaluation tests are performed both over the periods 1981q1-2000q1 and 1993q3-2000q1. Perhaps not unsurprisingly given the evidence in Figure 2, the consensus across these tests is that the computed measures of uncertainty do not provide a statistically satisfactory indication of the degree of uncertainty associated with real-time output gap estimates when evaluated against the point estimate for the outturn.

The most apparent reason for failure of the density forecasts from Table 2 is serial dependence in the first moment of $\{z_t\}$, although there is some evidence that this dependence is less acute over the 1993q3-2000q1 period.[41] This shows up in both the Ljung-Box and robust tests for *i.i.d.* (Hong iid). This serial dependence is consistent with our finding of considerable persistence in the (scaled) revisions. It is accordant with the view that even if the model used to estimate the output gap is well-specified in-sample, out-of-sample

[40] The narrow bands for the univariate HP estimator, and the implausible estimates for the bivariate HP estimator, may also be explained by misspecification of the UC model underlying the HP filter. This misspecification, as discussed above, will deliver sub-optimal inference. Empirically we did find support for this view; we subjected the estimated models to a range of misspecification tests as suggested by Harvey & Koopman (1992). The HP models, univariate and bivariate, appeared to suffer from excessive residual correlation suggesting that the models fail to pick up the dynamics adequately. However, we continue to consider both models as they provide useful benchmarks.

[41] We also considered the interval evaluation tests referred to above. Like the density evaluation tests they comprise both distributional and dependence aspects. They require a 'good' interval forecast to both have correct (unconditional) coverage, i.e. a 'hit rate' equal to the nominal rate, and for these hits and misses to exhibit no temporal pattern (independence). These tests did not reject the 95% and central 50% interval forecasts for the trivariate UC model. There is also some support for the bivariate UC interval forecasts; see Mitchell (2003). The density evaluation results suggest, however, that not all possible interval forecasts, and there are an infinite number of them, will offer a reliable indication of the degree of uncertainty.

Table 2: Evaluation of real-time output gap density estimates

| | 1981q1-2000q1 | | | | | 1993q3-2000q1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Uni HP | Uni UC | Bi UC | Bi HP | Tri UC | Uni HP | Uni UC | Bi UC | Bi HP | Tri UC |
| KS | 0.2848 | 0.1861 | 0.1629 | 0.2781 | 0.3589 | 0.5147 | 0.4371 | 0.3360 | 0.3813 | 0.4597 |
| AD | 25.167 | 6.340 | 3.308 | 9.924 | 24.830 | 12.930 | 7.168 | 5.900 | 6.421 | 14.480 |
| BN | 0.921 | 0.483 | 0.395 | 0.985 | 0.483 | 0.332 | 0.161 | 0.842 | 0.558 | 0.324 |
| LB1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.014 | 0.000 |
| LB2 | 0.000 | 0.000 | 0.000 | 0.017 | 0.000 | 0.254 | 0.024 | 0.231 | 0.940 | 0.150 |
| LB3 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.557 | 0.031 | 0.032 | 0.786 | 0.044 |
| Hong M1 | 0.156 | 0.053 | 0.068 | 0.013 | 0.659 | 0.358 | 0.012 | 0.003 | 0.006 | 0.350 |
| Hong iid | 86.660 | 73.650 | 72.880 | 14.130 | 29.880 | 9.114 | 9.432 | 8.086 | 1.641 | 10.510 |

Notes: KS is the Kolmogorov-Smirnov statistic with associated 95% critical value 0.25; AD is the Anderson-Darling statistic which has an associated 95% critical value of 2.502; BN is the p-value of the Bai-Ng robust test for normality; LB1, LB2 and LB3 are the p-values for the Ljung-Box tests for serial correlation in the first, second and third power; Hong M1 is the Hong joint test statistic for uniformity and independence with associated 95% critical value of 0.052; Hong iid is the Hong test statistic for independence that is robust to non-uniformity with associated 95% critical value of 1.96

(unforeseen) structural shocks lead one to revise the model (in particular the parameter estimates) used to identify and estimate the cycle resulting in serially correlated revisions to real-time estimates.[42] Although it is difficult to disentangle the joint test for *i.i.d.* uniformity/normality, there does appear to be some support for the assumed distribution. Although the KS and AD tests are largely unsupportive, the distributional test of Bai and Ng, that in contrast to these tests is designed to be robust to the dependence, provides support with $p$-values greater than 0.15 in all cases.

Therefore it appears that not just, as indicated in Table 1, are real-time point estimates of the output gap unreliable but so are measures of uncertainty associated with them. Alternative measures of uncertainty, of course, may do better than the one considered here. Future work should consider other ways of measuring the uncertainty associated with real-time estimates with the aim of finding the 'correct' density estimate. Re-calibrating or bias correcting the density forecasts is one option.

# 5 Implications of output gap uncertainty

This section draws out some implications of our empirical finding that not just are real-time output gap point estimates unreliable but so are their associated measures of uncertainty. We focus on the effect of this uncertainty on prescriptions from policy rules.

[42] As noted above if we believe the UC model underlying the HP filter is correctly specified (which as indicated we do not) we should expect the revisions associated with these optimal real-time HP estimates to be serially correlated since the real-time estimates equal the filtered estimates.

Alternatively, given the much debated importance of the output gap in forecasting inflation, we could consider the effect on forecasts of inflation. Analogously to the discussion below we should expect real-time uncertainty about the output gap to translate into (even greater) uncertainty about future values of inflation, in other words a wider *fan-chart*.

## 5.1 Taylor-type rules

Taylor-type rules model the interest rate as a function of both the deviation of inflation from its target level and the output gap. They have been found both to match the historical behaviour of interest rates in many countries and to perform well (in the sense of minimising some type of loss function over output, inflation and interest rate variations) relative to optimal feedback rules in simulation exercises; e.g. see Clarida et al. (1998) and Peersman & Smets (1999), respectively.[43] Recently Taylor rules have also been discussed with reference to the ECB; e.g. see Peersman & Smets (1999) and Gerlach & Schnabel (2000). They could be a useful informal benchmark to analyse and predict how the ECB sets interest rates. We address this issue in this section focusing on how the considerable uncertainty that appears to exist about real-time output gap estimates translates into uncertainty about monetary policy. This uncertainty is again best represented by a density forecast.

The principle of certainty equivalence might lead us to expect that uncertainty about real-time output gap estimates should have no effect on the optimal response to them in monetary policy rules.[44] However, we know from empirical applications to the US that the difference between real-time and final estimates is not irrelevant when analysing and using monetary policy rules; see Orphanides (2001, 2003*b*). At a theoretical level, uncertainty will also impact on the optimal coefficients in Taylor-rules. Uncertainty is expected to attenuate the coefficient on the real-time output gap estimate.[45] As long as we view the problem in real-time as one of extracting from the real-time data a signal (our best guess of the 'final' value of the output gap) this holds even when we consider optimal, as well as restricted or simple, policy rules; see Swanson (2004).

Signal extraction implies that the real-time estimate is additively composed of the final estimate, the signal $y^C_{t|T}$, plus a noise component representing measurement error; i.e. $y^C_{t|t} = y^C_{t|T} + \eta_t$, where $\eta_t$ is distributed independently of $y^C_{t|T}$. While in a linear-quadratic model, optimal policy is certainty equivalent with respect to the best guess $E(y^C_{t|T})$, that is optimal policy reacts to $E(y^C_{t|T})$ equivalently to $y^C_{t|T}$ treating $E(y^C_{t|T})$ as

[43]In establishing that Taylor-type rules often provide a good characterisation of interest rate behaviour focus has been on the point estimates from the Taylor rule. As we stress in this paper it is better to view these estimates as central points of ranges of uncertainty.

[44]This principle states that in the linear-quadratic case (quadratic loss function and linear constraints) the optimal decision rule is independent of the problem's noise statistics. In other words, policy-makers can ignore uncertainties about the shocks that hit the economy and set policy as if they were certain. For a textbook discussion and proof see Ljungqvist & Sargent (2000), pp. 57-59. For explicit discussion with respect to optimal monetary policy rules see, for example, Swanson (2004).

[45]See, for example, Aoki (2003), Ehrmann & Smets (2003), Orphanides (2003*a*), Rudebusch (2001) and Smets (1999).

if certain, it is not certainty equivalent with respect to the actual observed real-time output gap estimates $y^C_{t|t}$. This is because $y^C_{t|t} \neq E(y^C_{t|T})$; *via* the Wiener-Kolmogorov (or equivalently Kalman) filter $E(y^C_{t|T}) = ay^C_{t|t}$ where $a = \frac{ACF(y^C_{t|T})}{ACF(y^C_{t|t})}$, where ACF denotes the autocovariance generating function.[46] Only when we consider the "news", rather than "noise", model: $y^C_{t|T} = y^C_{t|t} + \eta_t$ where $\eta_t$ is mean zero and distributed independently of $y^C_{t|t}$, does $E(y^C_{t|T}) = y^C_{t|t}$ and certainty equivalence hold.[47]

In the noise model it is optimal for policy-makers to attenuate their response to variables about which uncertainty has increased; see Swanson (2004). This attenuation has been found to bring the optimal coefficients into line with historical (empirical) estimates. In the absence of uncertainty the optimal coefficients are often larger; see Smets (1999) and Rudebusch (2001).

The empirical support in this paper for the view that, across widely used output gap estimators, real-time output gap point estimates are unreliable, specifically support for the noise model, implies that in real-time policy-makers do require reliable estimates of the uncertainty associated with these estimates. These estimates should dictate the strength of their optimal reaction to movements in real-time output gap data at a given point in time. Informally we know that central bankers do attach health warnings to data; they know data are imperfect and subject to revision. Only when policy-makers believe the output gap estimates they consult in real-time to be reliable (i.e. consistent with the news model) should they ignore the uncertainty associated with their estimates; the empirical evidence presented in this paper suggests that perhaps it would be unwise for the ECB

[46] When $y^C_t$ is a stationary Gaussian random variable then the ACF is defined as $ACF(y^C_t) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j(L^j + F^j)$, where $Cov(y^C_t, y^C_{t-k}) = \gamma_k$ for $k = 0, \pm 1, \pm 2, ....$

[47] It is clearly important to establish whether real-time output gap estimates are consistent with the noise or news paradigm. Therefore, empirically we tested for news versus noise *via* a series of Mincer-Zarnowitz tests; e.g. see Clements & Hendry (1998) pp. 56-59. These essentially test if revisions are unbiased (weakly rational) as well as efficient or (strongly) rational. The null hypothesis of news amounts to testing, *via* a Wald or F-test that is robust to serial correlation and heteroscedastictiy, the joint hypothesis that $b_0 = 0$ and $b_1 = 0$ in the regression: $R_{t|T} = b_0 + b_1 y^C_{t|t} + \epsilon_t$. Under the noise model, $\eta_t$ is correlated with $y^C_{t|t}$ and will therefore help predict the subsequent revision $R_{t|T}$, implying a rejection of the null. Across the five output gap estimators we found no support for the news model ($p$-values equal to 0.000). 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. For completeness we note that for a correctly specified UC model the filtered estimates $y^C_{t|t}(\widehat{\Theta}_T)$ are rational expectations of the final estimates $y^C_{t|T}$. This follows from the fact that $y^C_{t|t}(\widehat{\Theta}_T)$ is the minimum MSE estimator of the true (unknown) $y^C_t$. Given normality, these conditional expectations are equivalent to orthogonal projections. Therefore, $y^C_{t|t}(\widehat{\Theta}_T)$ is the orthogonal projection of $y^C_t$ on $\Omega_t$. Similarly $y^C_{t|T}$ is the orthogonal projection of $y^C_t$ on $\Omega_T$. *Via* the law of iterated expectations $E(y^C_{t|T} \mid \Omega_t) = E(E(y^C_t \mid \Omega_T) \mid \Omega_t) = E(y^C_t \mid \Omega_t) = y^C_{t|t}(\widehat{\Theta}_T) \Rightarrow E(y^C_{t|T} - y^C_{t|t}(\widehat{\Theta}_T) \mid \Omega_t) = 0$. So, the filtered estimates can be also seen as the orthogonal projection of the final estimates on $\Omega_t$. This implies both that the revisions (between the filtered and final estimates) are unbiased, and that this revision is orthogonal to the filtered estimate (rationality).

to believe this.

Figure 2 also suggests that these uncertainty estimates are heteroscedastic; $Var(\eta_t) = \sigma^2_{\eta t} \neq \sigma^2_{\eta}$. For the trivariate UC estimator, in particular, we observe an increase in uncertainty as the economy approached the turning point of the early 1990s, and a decrease thereafter. However, the pertinent result to extract from the empirical analysis in Section 4 is that typical output estimators when applied to Euro area data do not provide policy-makers, in real-time, with reliable estimates of this uncertainty $\sigma^2_{\eta t}$. More work along the lines of Rudebusch (2002), p. 424, is therefore required to establish using simulated models the consequences of using unreliable estimates of the uncertainty estimate in real-time. Perhaps the unreliability in real-time of uncertainty, as well as point, output gap estimates will strengthen support for policy rules that ignore the output gap, such as nominal income rules; for further discussion see Rudebusch (2002). Alternatively given the apparent scale of the uncertainty associated with output gap estimates in real-time perhaps one should consider a nonlinear policy rule where policy-makers respond more aggressively as the uncertainty reduces; e.g. see Meyer et al. (2001). We find that typically it takes about three years for output gap estimates to begin to settle down at their 'final' values.

It is therefore important for policy-makers to look at the uncertainty of real-time output gap estimates. The remainder of this section draws out three practical implications of this uncertainty. First, we compare real-time policy recommendations with those based on final output gap data. Notwithstanding the limitations of having to base our analysis on AWM data, it is important to gain some indication of how well Taylor-type rules, using both final and real-time output gap data, explain the ECB's notional behaviour. Existing work, comparing real-time and final data, has focused on the US; see Orphanides (2001, 2003*b*). Secondly, we illustrate how uncertainty about real-time output gap estimates translates into uncertainty about the value of the policy instrument. In accordance with our focus in this paper, we then evaluate whether the implied density forecasts for the interest rate offer a good characterisation of actual movements in the interest rate. Conceptually this represents an important departure from existing studies that test how well the Taylor rule fits the data with a focus on point estimates. Thirdly, we propose a simple means of potentially improving real-time density forecasts of the monetary policy stance by combining forecasts across the five output gap estimators considered in Section 4.

Two simple but representative Taylor-type rules are considered in these three examples. The rules determine the reaction of the policy instrument, the short-term nominal interest rate $r_t$, to the four-quarter inflation rate (in percent), $\pi_t$, and the output gap $y_t^C$. These rules, both assumed linear as well as time-invariant in the coefficients $\alpha_\pi$ and $\alpha_y$, are (i) a traditional (static or long-run) Taylor-type rule and (ii) a partial-adjustment rule that accommodates interest rate smoothing by the ECB:[48]

[48]For convenience we continue to ignore publication lags and therefore consider only contemporaneous rather than lagged values of $\pi_t$ and $y_t^C$ on the right-hand-side of the policy rules.

$$\text{Policy Rule (i)} \quad : \quad r_t = r_t^* + \pi_t + \alpha_\pi(\pi_t - \pi_t^*) + \alpha_y y_t^C; \tag{8}$$
$$\text{Policy Rule (ii)} \quad : \quad r_t = r_{t-1} + \rho[r_t^* + \pi_t + \alpha_\pi(\pi_t - \pi_t^*) + \alpha_y y_t^C - r_{t-1}], \tag{9}$$

where $r_t^*$ is the equilibrium real interest rate assumed constant, $\pi_t^*$ is the inflation target, that consistent with ECB's stated objectives is assumed to be 2%, and $\rho$ $[0 \leq \rho \leq 1]$ is the smoothing parameter. The term in square brackets in (9) is therefore the target interest rate minus the interest rate in the previous quarter. Consistent with the findings for the Euro area of Gerlach & Schnabel (2000), we assume $r_t^* = 3.55\%$. We set $\rho = 0.25$; this implies that the interest rate is only moved by 25% of the target change. Again this is broadly consistent with previous empirical findings that find a large coefficient on the lagged interest rate in estimated policy rules. The coefficients $\alpha_\pi$ and $\alpha_y$ are set equal to their traditional values: $\alpha_\pi = 0.5$ and $\alpha_y = 0.5$; Gerlach & Schnabel (2000) in fact found similar weights for the Euro area when they estimated these coefficients.[49]

We should note that these policy-rules are often considered in forward-looking forms, to reflect the fact that monetary policy operates with a lag. Typically $\pi_t$ is replaced with a forecasted future value $E(\pi_{t+h} \mid \Omega_t)$ where $h$ is the forecast horizon; see Clarida et al. (1998). Accordingly, we did experiment with rules where $\pi_{t+h}$ (with $h = 4$ and 8) replaces $\pi_t$ (i.e. we assume a perfect forecast); results were qualitatively similar. In any case there is a forward looking element to the rules (8) and (9), in the sense that, as discussed, $y_t^C$ is widely interpreted as a leading indicator of inflation. Also we can interpret (8) and (9) as the reduced form, expressed in terms of observable variables, of a forward-looking rule; see Clarida et al. (1998).

## 5.2 Explaining interest rates in the Euro area using Taylor-type rules

The policy rules (8) and (9) are used to forecast the interest rate based on both real-time $y_{t|t}$ and final $y_{t|T}$ estimates of the output gap $y_t^C$; denote these forecasts $r_{t|t}$ and $r_{t|T}$, respectively. Note that $r_{t|t}$ is effectively out-of-sample.

Table 3 examines the accuracy of these alternative forecasts. Accuracy is summarised by the RMSE of the forecasts against the actual interest-rate. As is common practice, a benchmark AR forecast is considered also. Specifically we consider a first-order AR: $r_t = r_{t-1} + \tau_t$, where $\tau_t$ is a mean zero disturbance. We also indicate the RMSE between the interest rates implied by the real-time and final output gap estimates; $(r_{t|T} - r_{t|t})$ can be taken as an indicator of the degree to which the unreliability of output gap point estimates in real-time translates into policy-makers misjudging the state of the economy and setting the 'wrong' interest rate.

[49] We might consider estimating these weights ourselves using final and real-time data. Acknowledging the uncertainty associated with the real-time estimates this would deliver a ("thick") distribution of estimated coefficients.

Table 3 shows that the dynamic policy rule (9), rather than the static rule, proves a good characterisation of the evolution of actual interest rates.[50] This holds especially over the latter sample period, when the Taylor rule often beats the AR model. Consistent with the volatility of the bivariate HP estimator seen in Figure 1, Table 3 reveals high RMSE for this estimator: as we see below this estimator delivers patently absurd predictions. The evidence is mixed, however, on whether the real-time or final data provide a better explanation of actual interest rates; cf. columns $(r_t - r_{t|t})$ and $(r_t - r_{t|T})$. In contrast in an application to US data, Orphanides (2001) using a similar rule to (8) found that final data provide a better description of actual interest rate than the real-time output gap data.

Table 3: The fit of Taylor rules based on real-time and final data: RMSE

| | | 1981q1-2000q1 | | | 1993q3-2000q1 | | |
|---|---|---|---|---|---|---|---|
| | output gap estimator | $r_t - r_{t\|t}$ | $r_t - r_{t\|T}$ | $r_{t\|T} - r_{t\|t}$ | $r_t - r_{t\|t}$ | $r_t - r_{t\|T}$ | $r_{t\|T} - r_{t\|t}$ |
| Policy Rule (i) | uni UC | 1.482 | 1.715 | 0.693 | 0.755 | 0.900 | 0.219 |
| | uni HP | 1.79 | 1.793 | 0.596 | 0.994144 | 0.813 | 0.335 |
| | Bi UC | 1.701 | 1.86 | 0.688 | 0.738938 | 0.905 | 0.277 |
| | Bi HP | 9.766 | 2.771 | 8.089 | 6.137826 | 1.837 | 5.017 |
| | Tri UC | 1.554 | 1.754 | 1.120 | 1.066165 | 0.881 | 0.784 |
| Policy Rule (ii) | uni UC | 0.585 | 0.583 | 0.173 | 0.356 | 0.373 | 0.055 |
| | uni HP | 0.630 | 0.622 | 0.148 | 0.404 | 0.358 | 0.084 |
| | Bi UC | 0.611 | 0.600 | 0.172 | 0.351 | 0.340 | 0.069 |
| | Bi HP | 2.290 | 0.672 | 2.022 | 1.428 | 0.427 | 1.254 |
| | Tri UC | 0.618 | 0.609 | 0.280 | 0.500 | 0.365 | 0.196 |

Notes: RMSE of AR model is 0.566 over 1981q1-2000q1 and 0.421 over 1993q3-2000q1

Table 3 also indicates that the differences between the real-time and final renditions of the rule $(r_{t|T} - r_{t|t})$ are not as great as those of $r_{t|T}$ and $r_{t|t}$ individually against actual interest rates. Nevertheless, over the full-sample period with the static policy rule the RMSE is often 60 basis points, indicating that there remains considerable scope for setting the 'wrong' interest rate in real-time. As stated, for example, by Orphanides (2003*b*): "it has been tempting to associate good macroeconomic performance with setting policy based on the Taylor rule and even associate deviations of the [interest rate] from such rules as policy 'mistakes'" (p.637). Giving in to this temptation, it is nevertheless perhaps going too far to attribute deviations of real from final point renditions of the policy rule to mistakes. In reality we know that policy-makers, cognisant of probable revisions, will (at least informally) recursively in real-time form a density forecast for the interest-rate implied by the Taylor-rule. We therefore need to acknowledge the uncertainty associated with their real-time predictions $r_{t|t}$. In the following section we illustrate and then evaluate

[50] Gerlach & Schnabel (2000) also conclude that the Taylor rule captures the behaviour of interest rates in the Euro area in the 1990-97 period, with the exception of the period 1992-3 characterised by exchange-rate volatility.

this uncertainty. Evaluation is performed against the outturn $r_t$; alternatively we might consider evaluation against the final estimates $r_{t|T}$. By focusing on $r_t$ we test the ability of real-time estimates to forecast actual interest rate. This helps us establish whether these policy rules, using real-time output gap estimates, can be used in real-time as useful indicators of monetary policy stance.

### 5.3 Illustrating and evaluating the inherent uncertainty of monetary policy

In this section we indicate the degree to which uncertainty about real-time output gap estimates translates into uncertainty about monetary policy. This is based on examination of the density forecast for the interest rate $r_{t|t}$. This density is normal with variance derived from that of the real-time output gap estimates $P_{t|t}$. To isolate the effect of real-time output gap uncertainty, we ignore other sources of uncertainty. In the specific context of the policy-rules, (8) and (9), this means we ignore uncertainty associated with $r_t^*$ and $\pi_t$, plus parameter uncertainty. Naturally we should expect uncertainty about these variables, in particular about inflation or the inflation forecast in forward-looking variants of the policy-rules, to translate into even greater uncertainty about interest rates. Indeed we should expect the output gap and inflation forecast to be correlated. The density forecasts for $r_{t|t}$ we consider, and the implied confidence intervals we present, are therefore very much a lower bound on the degree of uncertainty that in reality would be associated with the ECB's assessment of monetary policy stance. Nevertheless, they provide some indication of the range of values for the interest rate that the ECB might have considered in real-time.

Before evaluating the density forecasts for the interest rate implied by our output gap estimates, Figure 3 provides a visual impression of the range of outcomes for the interest rate implied by the uncertainty estimates for the output gap. It plots the 95% confidence intervals for the interest rate based on the static rule (8), for each of the output gap estimators. Even confining attention to a given output gap estimator, Figure 3 indicates considerable uncertainty about the interest rate that should be set. For example, acknowledging the uncertainty (predicted in real-time) associated with the univariate UC estimator of the output gap results in a point forecast for the interest rate in 2000q1 of 5.9% with a 95% confidence interval ±2%. Across estimators (models) we observe even greater uncertainty, even when ignoring the bivariate HP results. The predictions implied by Bivariate HP are clearly implausible. Not just are they excessively uncertain, but the nominal interest rate falls below zero! Accordingly, we do not consider this output gap estimator further. Although not illustrated in these graphs smoothing, namely use of the dynamic rule (9), reduces uncertainty. This is expected since $Var(r_t) = \rho^2 \alpha_{1y}^2 P_{t|t}$. Smoothing or *gradualism* is discussed widely as a means of reducing uncertainties associated with monetary policy; e.g. see Martin (1999).

We now turn to an evaluation of the accuracy of the real-time density forecasts $r_{t|t}$ with respect to actual interest rates $r_t$. Focus is on the period 1993q3-2000q1, after both the cyclical peak of the early 1990s and the exchange-rate crises of 1992. Table 4 presents the results. Following Clements (2004) we again consider a benchmark AR (density) forecast; it is assumed Gaussian with mean equal to the actual interest rate one quarter previously (so that it is known in real-time) and variance equal to that estimated from the available sample for actual interest rates.

Table 4 shows that for the five output gap estimators in both static and dynamic cases there is evidence to suggest that the density forecasts are not satisfactory. But the reasons for failure differ across (8) and (9). The static rule tends to fail the independence tests but look better distributionally while the dynamic rule tends to fail the distributional tests but pass (at least at 99%) the independence tests. So individually none of the alternative output gap estimators deliver accurate density forecasts for the interest-rate.

Given the results of these evaluation tests and the finding that, when accounting for their uncertainty, the alternative output gap estimators imply considerable uncertainty about the Taylor-rule interest rate it is tempting to conclude that in this sense the output gap (in real-time) is an uninformative concept. It does not appear to offer a satisfactory representation of interest rate behaviour.

### 5.4 Combining alternative output gap estimates: density forecast combination

This conclusion is perhaps premature. Not only might alternative output gap estimators and/or alternative policy-rules to those considered here deliver a better characterisation of observed interest rate movements, but it is clearly implausible to suppose that the ECB only consults output gap estimators, and their implied interest rate forecasts, individually. It seems more reasonable to suppose that the ECB has various models to-hand when deliberating over its interest rate decision. Certainly we know that the Bank of England adopt what they call a "suite of models approach", relying on several modelling approaches rather than one alone to inform their judgement; see Bank of England (1999). This provides a rationale for model averaging.

Accordingly, in this section we consider means of combining alternative output gap estimates and their uncertainty. While it is well established that combining competing individual point forecasts of the same event can deliver more accurate forecasts, in the sense of a lower RMSE, little attention has been paid to the combination of density forecasts. This brief discussion therefore brings together two important but hitherto largely unrelated areas of the forecasting literature in economics, density forecasting and forecast combination.[51]

Let $g_{it}$ denote the density forecast, assumed continuous, of the interest rate $r_t$ at time $t$ $(t = 1, ..., T)$ made by model $i$ $(i = 1, ..., N)$. We consider aggregating these $N$ density forecasts directly. The combined density, also known as the linear opinion pool, is defined

[51] Related work has considered the combination of event, interval and quantile forecasts; see Clements (2002) and Granger et al. (1989). These inevitably involve a loss of information compared with consideration of the 'whole' density; e.g. only as the number of quantiles examined reaches infinity is no information about the density lost.

as the finite mixture distribution:

$$p_t(y_t) = \sum_{i=1}^{N} w_i g_{it}(y_t), \tag{10}$$

where $w_i$ are a set of non-negative weights that sum to unity. We ignore any constraints forcing, or truncating, the density to ensure positive nominal interest rates.

Inspection of (10) reveals that taking a weighted linear combination of the $N$ densities can generate a combined density with characteristics quite distinct from those of the individual densities. For example, if all the $N$ densities are normal, but with different means and variances, then the combined density will be mixture normal. Mixture normal distributions can have heavier tails than normal distributions, and can therefore potentially accommodate skewness and kurtosis. If the true (population) density is non-normal we can begin to appreciate why combining individual density forecasts, that are normal, may mitigate misspecification of the individual densities. Equally, if the true distribution is normal combining using (10) will, in general, get the distribution wrong; for further discussion see Hall & Mitchell (2004*a*).

The key practical issue then is to determine $w_i$. How we measure the accuracy of forecasts is central to how we choose to combine them "optimally". Point forecasts are traditionally evaluated on the basis of their RMSE relative to the outturn. Then point forecasts can be optimally combined to achieve the most "accurate" combined forecast, in the sense of minimum RMSE; this amounts to choosing the optimal weights *via* OLS estimation of the outturn on the competing point forecasts. Following Hall & Mitchell (2004*b*) we propose a simple data-driven approach for optimally combining density forecasts that extends this logic and is motivated by the desire to obtain the most "accurate" density forecast, in a statistical sense. This involves numerically searching for that set of weights that for a given density evaluation test, say the AD test, deliver the minimum value of the test statistic.[52]

To illustrate the proposed methods, we consider combining the alternative density forecasts implied by the dynamic policy rule. This means we can conveniently abstract from issues of dependence, as the individual densities appear to be satisfactory on this basis, and focus on examining whether combination can improve the distributional fit. By construction use of optimal weights ensures the combined density performs at least as well (with respect to the statistical test considered) as the best of the individual density forecasts. Therefore, it is when the optimal weights are not unity and zero that density combination helps; in a sense, one individual density forecast does not encompass the others.

Focusing on the AD test for illustrative purposes only, we found the optimal weights to be 0.7 on the bivariate UC estimator and 0.3 on the benchmark density. These weights were chosen by considering combination of the five densities jointly and consideration of

[52] Alternative ways to determine $w_i$ such as Bayesian model averaging have been suggested also; e.g. see Garratt et al. (2003). Use of equal weights, $w_i = 1/N$, has also been found to be empirically and theoretically beneficial; e.g. see Hendry & Clements (2004).

all bivariate combinations of the individual densities. This combined density delivered an AD test statistic of 2.748, far closer to the 95% critical value of 2.502 than any of the individual densities; moreover the combined density continued to 'pass' the dependence tests. This linear combination of the individual densities delivers the best explanation of actual interest rate behaviour in the Euro area, and indeed one that is far closer to being statistically satisfactory. In this sense, we might view this combination as the most likely combination of the individual densities that the ECB used when deliberating over its interest rate decision.

We note that use of equal weights did not help, yielding an AD test statistic of 8.589. These preliminary results are therefore encouraging; they suggest that pooling information across density forecasts, using an appropriate method, can deliver empirical gains. This is consistent with previous findings about point forecasts.

Table 4: Evaluation of real-time density forecasts for the Euro area interest rate

| | Policy rule (i) | | | | Policy rule (ii) | | | | AR |
|---|---|---|---|---|---|---|---|---|---|
| | uniUC | uniHP | biUC | triUC | uniUC | uniHP | biUC | triUC | |
| KS | 0.236 | 0.554 | 0.227 | 0.397 | 0.369 | 0.742 | 0.298 | 0.609 | 0.410 |
| AD | 1.974 | 47.290 | 2.111 | 15.210 | 8.873 | 177.341 | 5.309 | 59.530 | 7.939 |
| Bai/Ng | 0.125 | 0.774 | 0.132 | 0.768 | 0.595 | 0.478 | 0.596 | 0.660 | 0.606 |
| Hong M1 | 0.014 | 0.304 | 0.004 | 0.171 | 0.152 | 0.536 | 0.099 | 0.326 | 0.005 |
| Hong iid | 11.410 | 8.281 | 9.386 | 9.817 | 3.217 | 2.258 | 3.349 | 3.202 | 3.659 |
| LB1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.014 | 0.121 | 0.017 | 0.016 | 0.004 |
| LB2 | 0.000 | 0.009 | 0.002 | 0.002 | 0.130 | 0.197 | 0.147 | 0.016 | 0.990 |
| LB3 | 0.000 | 0.016 | 0.001 | 0.001 | 0.056 | 0.225 | 0.029 | 0.011 | 0.896 |

# 6 Concluding comments

This paper stresses the distinction between point estimates of the output gap, and measures of uncertainty associated with them. Interpreting real-time output gap estimates as forecasts we introduce the idea of evaluating the quality of output gap density forecasts. This is based on treating the 'final' value of the output gap as the 'true' value against which we can evaluate the statistical quality of the density forecasts. This extends previous work that has evaluated the quality of output gap point estimates, as in this paper using techniques familiar to us from the forecasting literature, but either ignored or at best not evaluated the uncertainty associated with the forecast. Moreover, we explain the importance of providing ranges of uncertainty associated with real-time output gap estimates.

In a simulated out-of-sample application to the Euro area our results indicate that not only are real-time point estimates of the output gap unreliable, consistent with previous research for the US, but so in general are their measures of uncertainty. Notwithstanding the poor performance of these real-time density forecasts we suggest, in contrast to current practice, that analysts when estimating the output gap in real-time should also routinely indicate the degree of confidence in their estimates *via* interval and preferably density forecasts. They could be published in tandem with density forecasts for inflation that are being increasingly published by central banks; e.g. see Clements (2004). Even, if the predicted degree of uncertainty proves to be incorrect *ex post*, at least the policy-maker has been warned *ex ante* of the dangers associated with the real-time output gap point estimates. Our results for some representative univariate and multivariate output gap estimators suggest that typically the predicted degree of uncertainty associated with the point estimates will be so large that it is impossible to reject the hypothesis that the output gap is statistically insignificant from zero.

We then explain and illustrate how uncertainty about real-time output gap estimates will impact on policy-makers use of policy-rules (e.g. Taylor rules) that rely on real-time output gap estimates; in the presence of the degree of uncertainty discovered in this paper we should expect a much reduced response to the output gap. But given that it is clearly difficult in real-time to measure accurately the degree of uncertainty associated with the real-time output gap estimates it is hard to know how much policy makers should down-weight their response. Future work is required to investigate this further. Despite this, we find that real-time output gap density estimates when combined can be used to obtain reliable forecasts of the ECB's monetary policy stance. This suggests that density forecast combination, introduced in this paper, is also an interesting and potentially useful area for further research.

Our findings also point to obvious problems in defining fiscal rules with reference to the economic cycle. Performance against a rule which requires the current budget to be balanced "over the cycle" cannot be assessed for several years. It is likely that the true position can be established only after at least two or three years. Similarly reforming the Stability and Growth Pact to take account of the cycle runs into obvious problems since the cycle cannot be measured accurately on a timely basis.

# 7 Bibliography

## References


Altissimo, A., Bassanetti, A., Cristadoro, R., Forni, M., Lippi, M., Hallin, M., Reichlin, L. & Veronese, G. (2001), EuroCOIN: A real time coincident indicator of the Euro area business cycle. CEPR Discussion Paper Series No. 3108.

Aoki, K. (2003), 'On the optimal monetary policy response to noisy indicators', *Journal of Monetary Economics* **50**, 501–523.

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. (2004), 'Dating business cycles: a methodological contribution with an application to the Euro area', *Oxford Bulletin of Economics and Statistics* **66**, 537–574.

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

Bai, J. & Ng, S. (2003), 'Tests for skewness, kurtosis and normality for time series data', *Journal of Business and Economic Statistics* . Forthcoming.

Bank of England (1999), *Economic Models at the Bank of England*.

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

Bates, J. M. & Granger, C. W. J. (1969), 'The combination of forecasts', *Operational Research Quarterly* **20**, 451–468.

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.

Blake, A. (1996), 'Forecast error bounds by stochastic simulation', *National Institute Economic Review* **156**, 72–79.

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

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

Canova, F. (1998), 'Detrending and business cycle facts', *Journal of Monetary Economics* **41**(3), 475–512.

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.

Clarida, R., Gali, J. & Gertler, M. (1998), 'Monetary policy rules in practice: some international evidence', *European Economic Review* **42**, 1033–1067.

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

Clements, M. P. (2004), 'Evaluating the Bank of England density forecasts of inflation', *Economic Journal* **114**, 844–866.

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

Clements, M. P. & Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge University Press: Cambridge.

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

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.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.

Diebold, F. X., Tay, A. S. & Wallis, K. F. (1999), Evaluating density forecasts of inflation: the Survey of Professional Forecasters, *in* R. Engle & H. White, eds, 'Cointegration, causality and forecasting: a festschrift in honour of Clive W. J. Granger', Oxford University Press.

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

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.

Durbin, J. & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

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

Elliott, G., Komunjer, I. & Timmermann, A. (2004), 'Estimation and testing of forecast rationality under flexible loss', *Review of Economic Studies* . Forthcoming.

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* .

Garratt, A., Lee, K., Pesaran, M. H. & Shin, Y. (2003), 'Forecast uncertainties in macroeconometric modelling: an application to the UK economy', *Journal of the American Statistical Association* **98**, 829–838.

Gerlach, S. & Schnabel, G. (2000), 'The Taylor rule and interest rates in the EMU area', *Economics Letters* **67**, 165–171.

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

Giacomini, R. (2002), 'Comparing density forecasts via weighted likelihood ratio tests: asymptotic and bootstrap methods', *UCSD Discussion Paper 2002-12* .

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

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

Granger, C. W. J. & Ramanathan, R. (1984), 'Improved methods of combining forecasts', *Journal of Forecasting* **3**, 197–204.

Granger, C. W. J., White, H. & Kamstra, M. (1989), 'Interval forecasting: an analysis based upon ARCH-quantile estimators', *Journal of Econometrics* **40**, 87–96.

Hall, S. G. & Mitchell, J. (2004*a*), Density forecast combination. National Institute of Economic and Social Research Discussion Paper No. 249.

Hall, S. G. & Mitchell, J. (2004*b*), "Optimal" combination of density forecasts. National Institute of Economic and Social Research Discussion Paper No. 248.

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. C. & Koopman, S. J. (1992), 'Diagnostic checking of unobserved-components time series models', *Journal of Business and Economic Statistics* **10**, 377–389.

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

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

Hendry, D. F. & Clements, M. P. (2004), 'Pooling of forecasts', *Econometrics Journal* **7**, 1–31.

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.

Koopman, S. J. & Harvey, A. C. (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.

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

Ljungqvist, L. & Sargent, T. J. (2000), *Recursive Macroeconomic Theory*, The MIT Press.

Martin, B. (1999), Caution and gradualism in monetary policy. Bank of England Working Paper No. 105.

Meyer, L. H., Swanson, E. T. & Wieland, V. W. (2001), 'NAIRU uncertainty and nonlinear policy rules', *American Economic Review* **91**, 226–231.

Mitchell, J. (2003), Should we be surprised by the unreliability of real-time output gap estimates? Density estimates for the Eurozone. National Institute of Economic and Social Research Discussion Paper No. 225.

Morgenstern, O. (1963), *On the Accuracy of Economic Observations*, Princeton University Press, New Jersey.

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.

Noceti, P., Smith, J. & Hodges, S. (2003), 'An evaluation of tests of distributional forecasts', *Journal of Forecasting* **22**, 447–455.

Orphanides, A. (2001), 'Monetary policy based on real-time data', *American Economic Review* **91**, 964–985.

Orphanides, A. (2003*a*), 'Monetary policy evaluation with noisy information', *Journal of Monetary Economics* **50**, 605–631.

Orphanides, A. (2003*b*), 'The quest for prosperity without inflation', *Journal of Monetary Economics* **50**, 633–663.

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. (2004), 'The reliability of inflation forecasts based on output gap estimates in real time', *Journal of Money, Credit and Banking* . Forthcoming.

Peersman, G. & Smets, F. (1999), 'The Taylor rule, a useful monetary policy benchmark for the Euro area?', *International Finance* **1**, 85–116.

Proietti, T. (2004), On the model-based interpretation of filters and the reliability of trend-cycle estimates. Working Paper University of Udine.

Quenneville, B. & Singh, A. (2000), 'Bayesian prediction mean squared error for state space models with estimated parameters', *Journal of Time Series Analysis* **21**, 219–236.

Rudebusch, G. D. (2001), 'Is the Fed too timid? Monetary policy in an uncertain world', *Review of Economics and Statistics* **83**, 203–217.

Rudebusch, G. D. (2002), 'Assessing nominal income rules for monetary policy with model and data uncertainty', *Economic Journal* **112**, 402–432.

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.

Smets, F. (1999), Output gap uncertainty: does it matter for the Taylor rule?, *in* B. Hunt & A. Orr, eds, ‘Monetary Policy under Uncertainty’, Wellington, New Zealand: Reserve Bank of New Zealand.

Smith, R. J., Weale, M. R. & Satchell, S. E. (1998), ‘Measurement error with accounting constraints: Point and interval estimation for latent data with an application to U.K. Gross Domestic Product’, *Review of Economic Studies* **65**, 109–134.

St-Amant, P. & van Norden, S. (1997), ‘Measurement of the output gap: a discussion of recent research at the Bank of Canada’. Bank of Canada Technical Report No. 79.

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

Stock, J. & Watson, M. (1999), Business cycle fluctuations in U.S. macroeconomic time series, *in* J. B. Taylor & M. Woodford, eds, ‘The Handbook of Macroeconomics’, Elsevier, Amsterdam, pp. 3–64.

Stock, J. & Watson, M. (2004), ‘Combination forecasts of output growth in a seven-country data set’, *Journal of Forecasting* **23**, 405–430.

Stone, J., Champernowne, D. & Meade, J. (1942), ‘The precision of national income estimates’, *Review of Economic Studies* **9**, 111–125.

Swanson, E. T. (2004), ‘On signal extraction and non-certainty-equivalence in optimal monetary policy rules’, *Macroeconomic Dynamics* **8**, 27–50.

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

van Norden, S. (2004), Optimal band-pass filtering and the reliability of current analysis. Eurostat Working Paper.

Wallis, K. F. (1982), ‘Seasonal adjustment and revision of current data: linear filters for the X-11 method’, *Journal of the Royal Statistical Society A* **145**, 74–85.

Wallis, K. F. (1989), ‘Macroeconomic forecasting: a survey’, *Economic Journal* **99**, 28–61.

Wallis, K. F. (2003), ‘Chi-squared tests of interval and density forecasts, and the Bank of England’s fan charts’, *International Journal of Forecasting* (19), 165–175.

Wallis, K. F. (2004), ‘An assessment of Bank of England and National Institute inflation forecast uncertainties’, *National Institute Economic Review* **189**, 64–71.

# A Data

Official Euro area 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). 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*).[53] 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 and unemployment 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]. These data are taken from the AWM and updated from 2001q1 using *New Cronos*.

## A.1 Construction of an "approximate" real-time data set for Euro area GDP

A real-time data set for Euro area GDP is constructed from consecutive quarterly issues of EUROSTAT's *Quarterly National Accounts*, starting from the 1993q1 publication. In each consecutive publication figures on quarterly real GDP growth are published, in general, for the previous 7 quarters although sometimes the previous 9 quarters. This means the data are only revised for these periods. We back-date all the vintages with values to 1990q3 (the earliest observation from the 1993q1 publication) using information that would have been available at the time; i.e. the previous vintages are used to back-date more recent vintages beyond their first published value. From 1998q1 GDP data are published with a one quarter lag; i.e. the 1998q1 publication's most recent estimate is for 1997q4. Prior to 1998q1 GDP data were published with a two quarter lag, with the odd exception. Viewing each vintage as a column of a matrix with the first row of each column corresponding to a common starting point, we obtain an upper triangular matrix, that is account for this change, by inserting a "fake" real-time data vintage between 1997q4 and 1998q1 - this consists of the 1997q4 vintage (which has estimates up to 1997q2) plus the 1997q3 value from the 1998q1 vintage (for this vintage the 1997q4 value is now the most recent estimate). Since there were not always four publications per year (in 1995, for example, there were only two in 1995q1 and 1995q4) the missing vintages are assumed equal to the more recent vintage but observations that would not have been available

[53]After the completion of this paper an updated version of the AWM database (to 2002q4) was released by Jerome Henry. The real GDP series from this source appears very similar to the series considered in this paper (updated using official data). For example, the correlation of the Harvey-Trimbur cycles extracted from the two series is 0.996 over the period 1971q1-2002q4. An alternative data source for aggregate Eurozone real GDP data from 1979q4-1999q3 is Beyer et al. (2001). Given the later starting point this series is not considered.

at the time (given the typical publication lag) are deleted. For example, the 1995q3 vintage (which is not published) is assumed to be the 1995q4 vintage (which is published) minus the value for 1995q2 and the 1995q2 unpublished vintage is the 1995q4 vintage minus the values for 1995q1 and 1995q2. There are also definitional changes regarding the composition of the Euro area. If there is a choice we opt for the definition closest to the Euro12.

To facilitate business cycle analysis the real-time data set needs to go back beyond 1990q3. In this sense our real-time data set is "approximate" in that we backcast using the values from the AWM data-set. That is, the AWM data from 1970q1-1990q2 are updated using the consecutive vintages (the level of real GDP is backed out from the quarterly growth rate). This delivers 43 vintages, with data back to 1970q1, that can be used to examine the output gap in 'real-time' from 1992q3 to 2003q1. Although using the AWM data to backcast is *ad hoc* and does not reflect information truly available at the time, at least the most recent values for a given vintage, the ones most likely to impact on real-time output gap estimates, are genuine.

# B Appendix. Alternative output gap estimators

The output gap, the difference between actual and potential output, has an importance in the popular debate which can tend to run ahead of the problems in measuring it; the output gap is not observable. The choice of what measure, or estimator, of the output gap to use is more than a dry academic issue. As Canova's (1998) analysis, for example, showed inference can be sensitive to measurement. Various estimators have been proposed. Gerlach & Smets (1999) make the distinction between *statistical, structural* and *mixed* estimators.[54] 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 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 statistical and mixed approaches.

[54] An alternative approach, "euroCOIN" published by the CEPR, constructs a so-called "indicator" of the Eurozone business cycle by extracting the unobserved common component of Eurozone GDP growth from a panel, with a large number of variables, using dynamic principal components; see Altissimo et al. (2001). This cross-sectional information makes it possible to remove high frequency components with less lead and lag observations than traditionally used in time-series filters. But since this indicator is extracted from the first difference of GDP, a filter that is known to induce phase shift, rather than directly from the (log) level of GDP this indicator can less obviously be interpreted as a measure of the business cycle than the estimators considered in this paper. It is perhaps better interpreted as the trend, or likely future path, of GDP growth. Given this interpretation the euroCOIN indicator can be evaluated in terms of its forecast performance. In any case, although euroCOIN does provide a timely estimate of GDP growth, the output gap estimators considered in this paper have been more widely used by policy-makers for business cycle analysis; hence our emphasis on them in this paper.

Two statistical estimators are considered, largely for comparative purposes. The Hodrick-Prescott filter is perhaps the most widely used statistical approach for de-trending a time-series; we fix $\lambda$ 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).[55] Calculations in this paper were performed using the GAUSS and `Ox` [see Doornik (1998)] programming languages. Use was made of the beta version of `SsfPack 3` for `Ox`; see Koopman et al. (1999).

## B.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 and (2) Multivariate Hodrick-Prescott filters.[56]

## B.2 Multivariate Unobserved Components filters

We consider two representative multivariate unobserved 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 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 used, for example, by Chagny & Lemoine (2002).

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

[56] Related multivariate approaches to estimating the output gap are based on structural VAR models, the multivariate Beveridge-Nelson decomposition and the Cochrane approach; e.g. see Dupasquier et al. (1999).

### B.2.1 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 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:

$$\begin{aligned} y_t &= y_t^* + y_t^C && (11) \\ \Theta(L) y_t^C &= \varepsilon_t^y && (12) \\ \Gamma(L)\Delta\pi_t &= \lambda y_{t-1}^C + \varepsilon_t^\pi && (13) \\ y_t^* &= y_{t-1}^* + \beta_{t-1} + \varepsilon_t^{y^*} && (14) \\ \beta_t &= \beta_{t-1} + \varepsilon_t^\beta && (15) \end{aligned}$$

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

The Phillips Curve equation, (13), 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 (13) to be imposing a long run homogeneity restriction. This can be seen by appreciating that underlying (13) 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 $\sigma^2_{\varepsilon_t^{y^*}} = 0$.[57]

[57] Other options are to allow the disturbances to follow moving average processes; e.g. if we let $\varepsilon_t^\pi$ 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.

### B.2.2 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):

$$\Delta\pi_t = \Gamma(L)\Delta\pi_{t-1} + \Psi(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^{\pi} \tag{16}$$

$$y_t - y_t^* = \gamma(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^{qgap} \tag{17}$$

where

$$y_t^* = y_{t-1}^* + m_{t-1}^1 + \varepsilon_t^{y^*} \tag{18}$$

$$u_t^* = u_{t-1}^* + m_{t-1}^2 + \varepsilon_t^{u^*} \tag{19}$$

and

$$m_t^1 = m_{t-1}^1 + \varepsilon_t^{m^1} \tag{20}$$

$$m_t^2 = m_{t-1}^2 + \varepsilon_t^{m^2} \tag{21}$$

The unemployment gap is modelled as an autoregressive process:

$$(u_t - u_t^*) = \delta(L)(u_{t-1} - u_{t-1}^*) + \varepsilon_t^{ugap} \tag{22}$$

Equation (16) 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 (17) is an Okun's Law relationship, relating cyclical unemployment and output movements.

Equations (18) and (19) are assumed to follow a local linear trend model. This representation was used with success by Fabiani & Mestre (2001) in an application to the Euro area. 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 from the unemployment gap to the output gap; see Astley & Yates (1999).

### B.2.3 Estimation of UC models

The parameters of the univariate and multivariate unobserved components models are estimated by maximum likelihood exploiting their state-space form. 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.

## B.3 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.

### B.3.1 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.[58] 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:

$$
\begin{aligned}
y_t &= y_t^* + y_t^C && (23)\\
\Gamma(L)\Delta\pi_t &= \lambda y_{t-1}^C + \varepsilon_t^{\pi} && (24)\\
y_t^* &= y_{t-1}^* + \beta_{t-1} && (25)\\
\beta_t &= \beta_{t-1} + \varepsilon_t^{y^*} && (26)
\end{aligned}
$$

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.

[58] 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.

Writing (23) 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 (23)) is given by:

$$\begin{bmatrix} \Delta\pi_t \\ y_t^* \\ y_{t-1}^* \\ y_t^C \end{bmatrix} = \begin{bmatrix} * & 0 & 0 & * \\ 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\pi_{t-1} \\ y_{t-1}^* \\ y_{t-2}^* \\ y_{t-1}^C \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{\pi} \\ \varepsilon_t^{y^*} \\ 0 \\ \varepsilon_t^{y^C} \end{bmatrix} \tag{27}$$

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 (27) controls the smoothness of the cycle and the fit of the economic relationship (the second equation in (23)). 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 $\lambda_1$ and $\lambda_2$ denote these relative variances, that control the problem, where

$$\lambda_1 = \sigma_{y^C}^2/\sigma_{y^*}^2 \tag{28}$$

$$\lambda_2 = \sigma_{y^C}^2/\sigma_{\pi}^2. \tag{29}$$

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

$$\lambda_1 = 1/\sigma_{y^*}^2 \tag{30}$$

$$\lambda_2 = 1/\sigma_{\pi}^2 \tag{31}$$

Then $\lambda_1$ 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 tends to a deterministic (smooth) trend. Traditionally, as with the univariate HP filter, $\sigma_{y^*}^2 = 1/1600 \Leftrightarrow \lambda_1 = 1600$.

$\lambda_2$ controls the fit of the Phillips Curve relationship. As $\lambda_2 \to 0$ ($\sigma_{\pi}^2 \to \infty$) the worse the fit of the the economic relationship, implying less information provided by the economic relationship. Traditionally $\sigma_{\pi}^2 \simeq 1/25 \Leftrightarrow \lambda_2 = 25$ or $\sigma_{\pi}^2 \simeq 1 \Leftrightarrow \lambda_2 = 1$; 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.

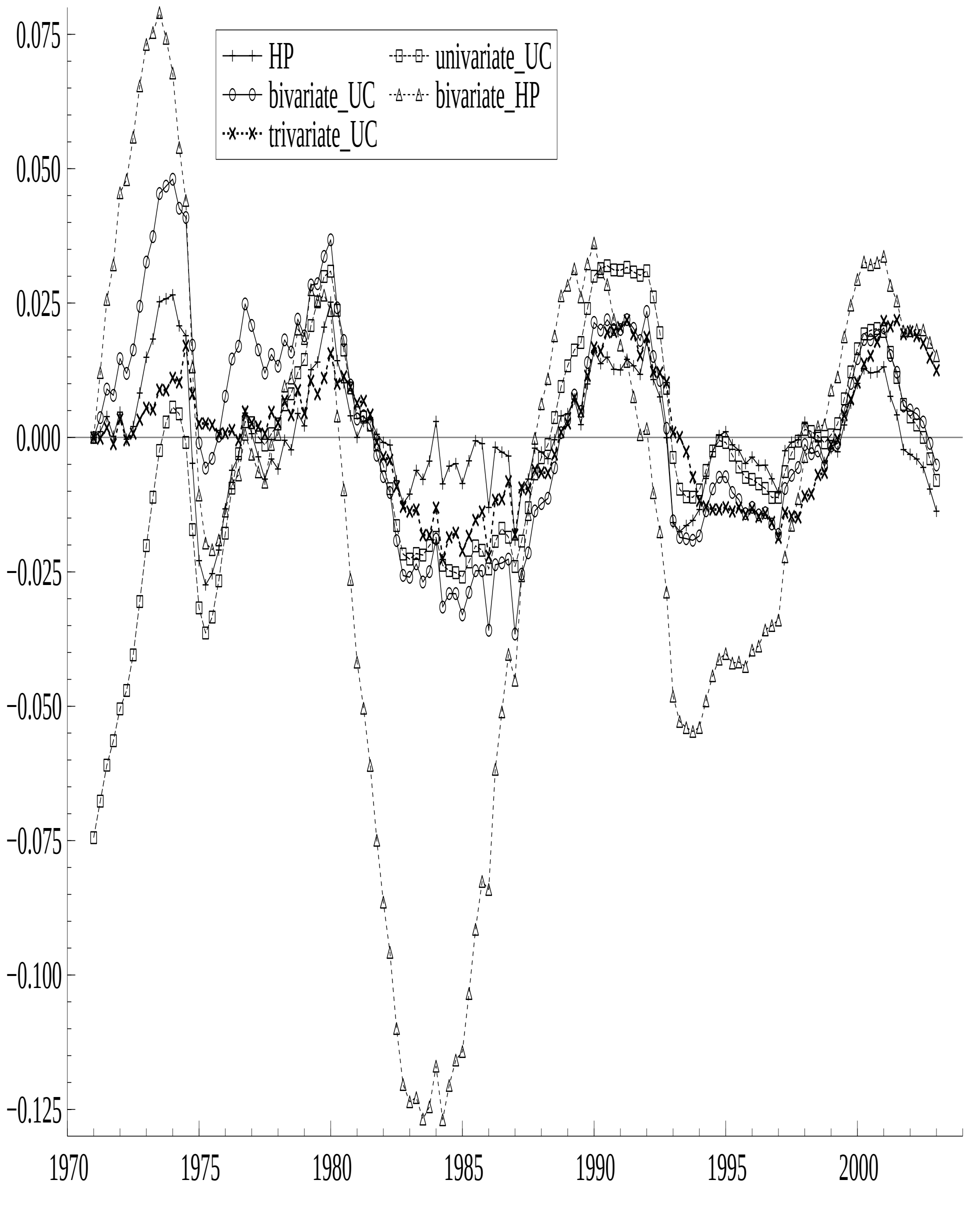


Figure 1: Final estimates of the Eurozone business cycle: a comparison of different estimators

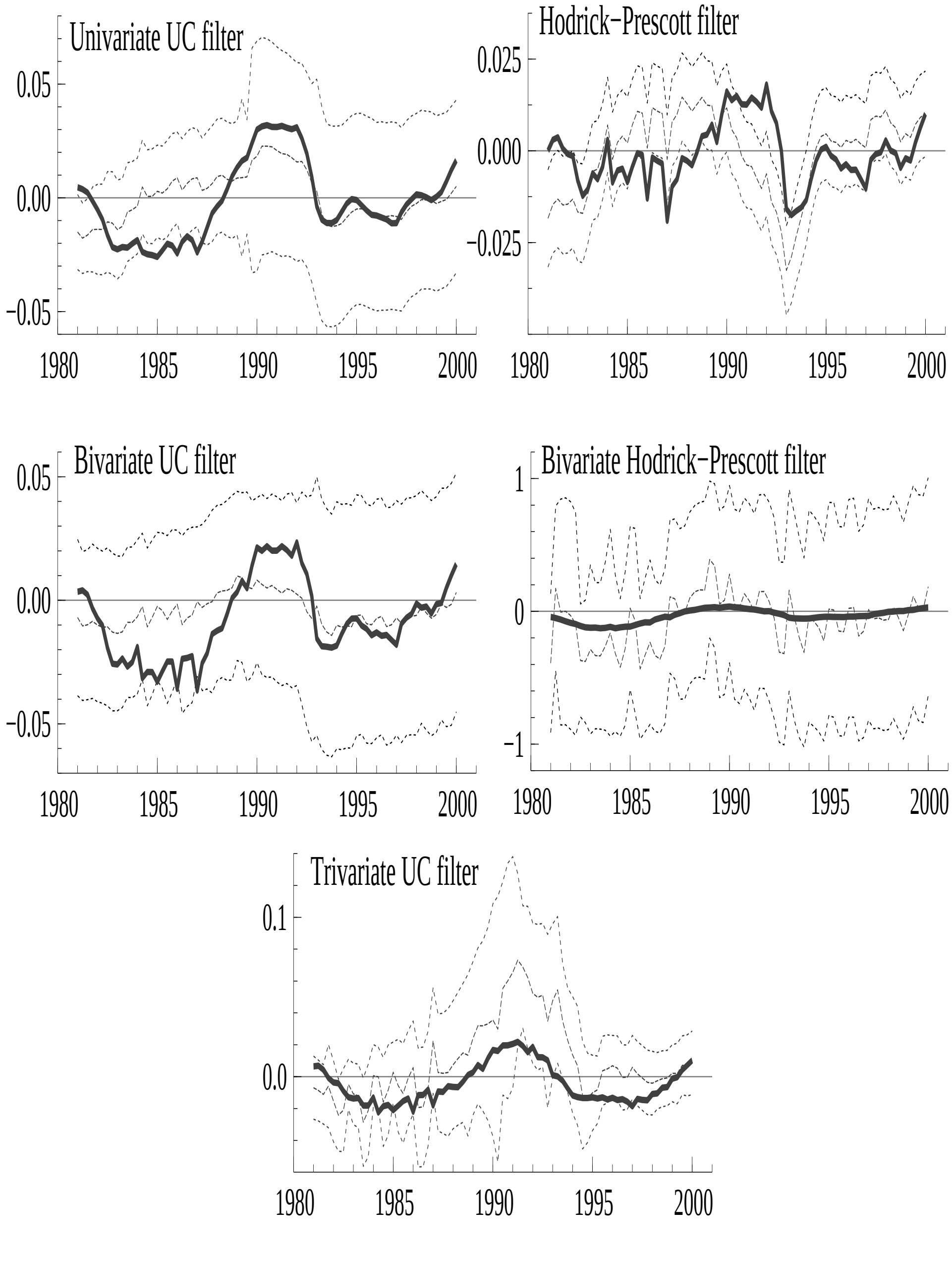


Figure 2: Uncertainty associated with real-time estimates using five different measures of the output gap: 95 per cent confidence intervals. Note: final estimate is in bold face

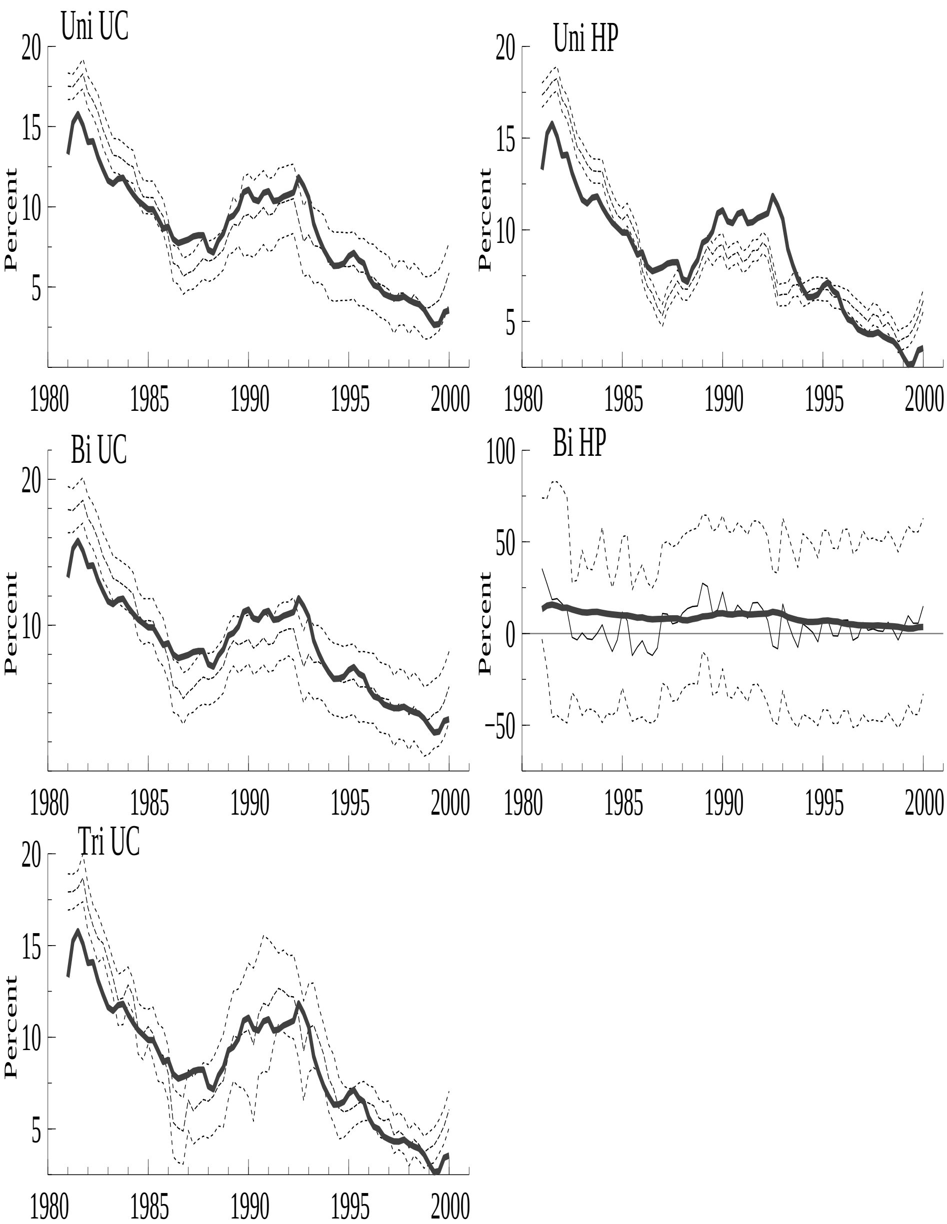


Figure 3: Real-time Taylor rule predictions for the Euro area interest-rate and the associated 95% confidence intervals. Note: actual interest rates are in bold face