ISSN 1745-8587

Birkbeck
UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 0618

Real Time Representation of the UK Output Gap in the Presence of Trend Uncertainty

Anthony Garratt
Birkbeck, University of London

Kevin Lee
University of Leicester

Emi Mise
University of Leicester

Kalvinder Shields
University of Melbourne

August 2006

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Real Time Representation of the UK Output Gap in the Presence of Trend Uncertainty*

by

Anthony Garratt,† Kevin Lee,†† Emi Mise†† and Kalvinder Shields†††

Abstract

This paper describes an approach that accommodates in a coherent way three types of uncertainty when measuring the output gap. These are trend uncertainty (associated with the choice of model and de-trending technique), estimation uncertainty (with a given model) and data uncertainty (associated with the reliability of data). The approach employs VAR models to explain real time measures and realisations of output series jointly along with Bayesian-style 'model averaging' procedures. Probability forecasts provide a comprehensive representation of the output gap and the associated uncertainties in real time. The approach is illustrated using a real time dataset for the UK over 1961q2 – 2005q4.

Keywords: Output gap, real time data, revisions, Hodrick-Prescott trend, exponential smoothing trend, moving average trend, model uncertainty, probability forecasts.

JEL Classification: E52, E58.

*†Birkbeck College, London, UK, ††University of Leicester, UK, †††University of Melbourne, Australia. Corresponding author: Anthony Garratt, Department of Economics, Birkbeck College, Malet Street, London, WC1E 7HX, UK; email : a.garratt@bbk.ac.uk. Version dated August 2006.

1 Introduction

The measurement of the output gap, i.e. the difference between the economy's actual output level and its potential or trend level, is central to much applied macroeconometric work and to the analysis of monetary policy in particular. However, a well known difficulty with the use of the output gap is the uncertainty which surrounds its measurement. This uncertainty arises from a range of sources including: the *trend uncertainty* surrounding the choice of model and detrending technique underlying the measure (see Canova, 1998); the *estimation uncertainty* associated with any chosen model/technique and characterised by the estimated stochastic variation and variation in estimated parameters of the model; and the *data uncertainty* associated with the reliability of the data available when the gap is calculated. The latter source of uncertainty has been highlighted in a recent literature, typified by Orphanides (2001), which illustrates the importance of acknowledging that macroeconomic decisions are made in real time and on the basis of data that is frequently subject to subsequent revision.[1]

This paper describes an approach to measuring and representing the output gap that accommodates all these types of uncertainty in a coherent way, extending the work of Orphanides and van Norden (2002) [OvN] and Garratt, Lee, Mise and Shields (2006) [GLMS]. OvN highlight the unreliability in the measures of the US output gap due to data uncertainty and, using a recursive analysis of the successive vintages of data that became available over 1965q1-1997q4, they observe that revisions of the US gap have been of the same order of magnitude as the gap itself over this period. GLMS acknowledge the role of data uncertainty but argue that part of this can be offset by the use of a joint model that explains the time series of measured output as released in real time alongside the time series of revisions in measured output. Such a model enables forecasts to be

[1] In monetary policy analysis, it is now acknowledged that the use of ex post revised data can yield misleading descriptions of historical policy and can generate very different policy recommendations to those obtained on the basis of real-time data. Also, the identification and interpretation of monetary policy shocks are very sensitive to assumptions on the timing of the release of information and decisions. See, for example, Orphanides (1997), Rotemberg and Woodford (1999), Christiano *et al.* (1999), Brunner (2000), Orphanides *et al.* (2000), Amato and Swanson (2001) or Garratt, Lee, Pesaran and Shin (2006).

made of the post-revision output level (i.e. the output measure that will be released after all revisions are complete) both currently and into the future. Hence, any systematic element in the revisions are anticipated and taken into account and this directly reduces the extent of data uncertainty. Further, there is an indirect effect because the forecasts can also be used to augment the historical data in the estimation of the underlying trend. As explained in Mise et al (2005a,b) and illustrated in GLMS, this serves to reduce the end-of-sample estimation error encountered when a two-sided trend measure is calculated using a finite sample of data.

In this paper, we undertake an analysis of output gaps in the UK over the period 1961q2-2005q4. We repeat the analyses of OvN and GLMS, confirming that the problems of data uncertainty and estimation uncertainty encountered in measuring the US output gap are also found in the UK and that the procedures suggested in GLMS are appropriate for the UK too.[2] However, the analysis in this paper is extended to focus on the role of *trend uncertainty* in the measurement of the output gap. We find that the revision process is more prolonged and more complex for the UK data compared to the US data. This means that the choice of an appropriate model with which to characterise the series released in real time is more difficult and that more attention needs to be paid to the estimation of the forecasts on which the trend measures are based. We therefore propose an approach to measuring the output gap that deals with the uncertainties surrounding the choice of model and the associated detrending technique, in addition to the estimation and data uncertainties discussed previously. The approach adopts a Bayesian-style 'model averaging' procedure in order to accommodate the trend uncertainty,[3] and the focus is on combining forecast probability distribution functions to provide a comprehensive representation of the output gap and the associated uncertainties in real time.

The remainder of the paper is organised as follows. In Section 2, the proposed method for measuring the output gap is described. In this, the appropriate real time measure

[2]Related work includes that of Harvey, Trimbur and van Dijk (2006) for the US and Adams and Cobham (2005) for the UK.

[3]See Burnham and Anderson (1998) for a general discussion of the model averaging. And, for recent examples of the use of the techniques, see Pesaran and Zaffaroni (2004), Weeks and Stone (2003), and Garratt *et al.* (2003).

of the gap is discussed based on models that can jointly explain output growth and the revision process. The joint models can provide point forecasts of 'post-revision' output series and can illustrate the range of potential output outcomes that can occur using simulation methods. The section also explains how the simulation methods can be used to calculate and represent trends and gap measures taking into account data uncertainty, estimation uncertainty and trend uncertainty. Section 3 describes the application of the proposed methods to obtain output gap measures for the UK taking all the various sources of uncertainty into account and compares these with measures obtained following the procedures of OvN and GLMS. Section 4 presents some probability forecasts obtained using our modelling framework to illustrate the usefulness of our procedure. Section 5 concludes.

2 Measuring the Output Gap with Real Time Data

OvN introduce the key concepts in measuring the output gap in the presence of data uncertainty, focusing attention on the differences between 'real time' measures of the output gap based on successive vintages of output data and 'final' measures obtained from the last available vintage of data. Writing (the logarithm of) the output level at time $t-j$ by y_{t-j}, and denoting the measure of output at time $t-j$ that is released in time t by ${}_ty_{t-j}$, $j = 0, 1, 2, , ...,$ the "vintage-t" dataset is defined by $Y_t = \{{}_ty_{t-1}, {}_ty_{t-2}, {}_ty_{t-3}, ...\}$ so that it includes the time-t measure of output at time $t-1$ and before. Note that it is assumed that the first release of output data for any period takes place after a one-period delay; this corresponds to practice in the quarterly series for UK and US output, for example. OvN's real time measure of the gap is $x_t^{ro} = {}_{t+1}y_t - \widetilde{y}_t^o | Y_{t+1}$, $t = 1, ..., T-1$, where $\widetilde{y}_t^o | Y_{t+1}$ denotes the trend output level at time t calculated on the basis of the vintage-$(t+1)$ data and the 'o' superscript denotes the particular detrending method used by OvN. The corresponding 'final' output gap measure is $x_t^{fo} = {}_Ty_t - \widetilde{y}_t^o | Y_T$, where $t = 1, ..., T-1$, and where T denotes the last period for which data is available. The measure x_t^{fo} represents the most up-to-date measure of the gap available to OvN and it was the large discrepancies between x_t^{ro} and x_t^{fo} in the US data that led to the conclusion that real time measures of the gap are unreliable.

GLMS emphasise the need to use the full information set available when measuring the output gap. This information set, denoted Ω_t at time t, contains the datasets of all vintages dated at t and earlier; i.e. $\Omega_t = \{Y_t,\ Y_{t-1},\ Y_{t-2}, ...\}$. The information set grows with the addition of successive vintages of datasets by including the news on the output level in the previous period (the 'first release' of information on the output level in that period) plus news on the revisions on the output series in previous periods; i.e. $\Omega_{t+1} = \Omega_t \cup \{_{t+1}y_t,\ (_{t+1}y_{t-1} -_t y_{t-1}),\ (_{t+1}y_{t-2} -_t y_{t-2}), ...\}$. The form of the real time data is illustrated in Appendix A.

If revisions on measured output continue to occur for up to q periods, say, then the real time measure of the output gap based on full information is defined by

$$x_t^{rk} = E\left[\,_{t+q+1}y_t \mid \Omega_{t+1}\right] - \widetilde{y}_t^k \mid \Omega_{t+1} \tag{2.1}$$

where $E\left[\,_{t+q+1}y_t \mid \Omega_{t+1}\right]$ is the expectation of the post-revision measure of output at time t released in $t+q+1$, and $\widetilde{y}_t^k$ is the trend output level obtained using method k, both formed on the basis of information available at time $t+1$. Given that Ω_{t+1} includes information on revisions, the method employed to detrend the output series is also likely to focus on both the observed and expected future values of the post-revision output measure so that the definition of the trend is closely associated with the model on which forecasts of post-revision measures are based.

If no revisions are assumed to take place, then $E\left[\,_{t+q+1}y_t \mid \Omega_{t+1}\right] = {}_{t+1}y_t$ and the measured output level can obviously be observed focusing on the most recent vintage of data only. Even if only vintage-t data is used, however, predictions of future values of the output series will be helpful in measuring the trend at the end of the sample if the time-t value of the trend is related to output in adjacent (future and past) periods. This point is forcefully made in relation to the use of the Hodrick-Prescott [HP] filter in Mise *et al.* (2005a,b). The HP filter is an exponentially-weighted moving average filter, and is two-sided symmetric in the sense that it uses both past and future observations with equal importance in order to decompose any observation in a series. The HP filter can be motivated as being the filter that minimises the change in trend growth and has the desirable property that it is optimal, in the expected squared error sense, in estimating the trend when the change in trend growth and the cycle are driven by uncorrelated white noise processes. However, the optimality properties only hold for the mid-point of the series when the series is finite. Mise *et al.* (2005a,b) note that the filter continues to provide an unbiased estimate of the trend and cycle at the endpoints of a finite series but that the estimates are inefficient. They note Burman's (1980) suggestion to augment the observed series with optimal linear forecasts and demonstrate, through their simulation exercises, that the application of the HP filter to the augmented series provides an estimate of the end-of-sample observation which is optimal. Indeed, by augmenting a series by its forecast, the standard deviation of the estimation error for the cyclical component is reduced by up to half (relative to the standard application of the HP filter) in their various simulations. The clear implication of these results is that, when the output trend is defined with respect to past and future output levels, the end-of-sample output gap measure that is of interest to decision-makers in real time should be calculated using a trend obtained by applying the filter to a forecast-augmented output series. The modelling framework to be employed in generating the forecasts is therefore most important and so this is elaborated in the section below.

2.1 Modelling Actual, Revised and Trend Output Series

GLMS suggest that, if revisions continue to occur up to q periods after the first release of data (the revision horizon), then a VAR of size $q+1$ of the following form is appropriate:

$$\begin{bmatrix} {}_ty_{t-1} - {}_{t-1}y_{t-2} \\ {}_ty_{t-2} - {}_{t-1}y_{t-2} \\ . \\ . \\ {}_ty_{t-q-1} - {}_{t-1}y_{t-q-1} \end{bmatrix} = \mathbf{a} - \mathbf{B}_1 \begin{bmatrix} {}_{t-1}y_{t-2} - {}_{t-2}y_{t-3} \\ {}_{t-1}y_{t-3} - {}_{t-2}y_{t-3} \\ . \\ . \\ {}_{t-1}y_{t-q-2} - {}_{t-2}y_{t-q-2} \end{bmatrix} - \cdots$$

$$- \mathbf{B}_p \begin{bmatrix} {}_{t-p}y_{t-p-1} - {}_{t-p-1}y_{t-p-2} \\ {}_{t-p}y_{t-p-2} - {}_{t-p-1}y_{t-p-2} \\ . \\ . \\ {}_{t-p}y_{t-p-q-1} - {}_{t-p-1}y_{t-p-q-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \xi_{1t} \\ . \\ . \\ \xi_{qt} \end{bmatrix},$$

or, equivalently,

$$\mathbf{z}_t = \mathbf{a} - \mathbf{B}_1\mathbf{z}_{t-1} - ..... - \mathbf{B}_p\mathbf{z}_{t-p} + \mathbf{e}_t \tag{2.2}$$

for $t = 1, 2, ..., T$, where $\mathbf{z}_t = ({}_ty_{t-1} - {}_{t-1}y_{t-2},\ {}_ty_{t-2} - {}_{t-1}y_{t-2},\ ...,\ {}_ty_{t-q-1} - {}_{t-1}y_{t-q-1})'$ is a vector containing the growth in the first-release data and the time-t revisions on output in the previous q periods. The model is appropriate on the reasonably uncontentious assumptions that 'actual' output is first-difference stationary (where actual output is the measured output series after all revisions are complete) and that revisions are stationary. GLMS demonstrate that the model can be rewritten in levels form, in VECM form (involving $q-1$ cointegrating vectors linking ${}_ty_{t-1}$, ${}_ty_{t-2}$, ..., ${}_ty_{t-q-1}$ pairwise with vectors of the form $(1, -1)$) and in a MA form.

The simplicity of the VAR form means that the model is straightforward to estimate and to use in forecasting. This is important here since, as we explain below, simulation methods are employed for the purposes of examining probability distribution functions of events involving the output gap and these events are, in turn, potentially complex functions of lagged and future output levels. However, given that there are $p(q+1)^2$ parameters involved in $\mathbf{B}_1, ..., \mathbf{B}_p$, the model can easily become over-parameterised if the growth dynamics are complicated and/or the revision process is protracted. This is significant from the point of view of producing forecasts since predictive accuracy can be poor in large models based on relatively small samples of data. For example, Clements and Hendry (2005) make the point that an estimated model that is *misspecified* because it omits relevant explanatory variables can have a lower mean squared forecast error (MSFE) than the estimated *true* data generating process if the parameters on the omitted variables are small (i.e. where the expected value of the t^2-test of their statistical insignificance is less than 2). Also, Harvey and Newbold (2005) show that small sample parameter estimation effects can mean that forecasts from a *true* but estimated data generating process can be enhanced by combining them with forecasts from a rival *misspecified* non-nested model. These comments suggest that, where the revision process is complicated, there is likely to be considerable model uncertainty. This means that we should take care in model selection (relating to the choice of q and p), that we might be unwise to place too much emphasis on a single forecasting approach, and that we might consider the gains from combining forecasts in producing trend and gap measures.

The difficulty involved in choosing the model with which to produce forecasts of future (post-revision) output levels generates uncertainty on the trend measure underlying the output gap (in addition to the estimation and data uncertainty observed for a given model using data that is subject to revision). This aspect of trend uncertainty is usually combined with, and compounded by, ambiguity on the form of the filter that should be applied to the forecast-augmented series. Economic theory typically provides little guidance on the nature of the trend series and, while the HP filter discussed above has been widely employed in business cycle analysis, it is just one of many potential filters that could be used.[4] To illustrate this further aspect of trend uncertainty in this paper,

[4] Woodford (2002) provides a micro-founded motivation for the trend output concept to be used in gap measures, defining the trend as the output level that would be achieved in the presence of perfect price flexibility. A clear link with a specific trend measure has not yet been made however.

we consider three detrending techniques that are linear in output, each taking the form:

$$\begin{aligned}\widetilde{y}_t^k \mid \Omega_{t+1} &= C_k(L)\ E\left[{}_{t+q+1}y_t \mid \Omega_{t+1}\right] \\ &= C_{k,0}\, E\left[{}_{t+q+1}y_t \mid \Omega_{t+1}\right] \\ &\quad + C_{k,1}\, E\left[{}_{t+q}y_{t-1} \mid \Omega_{t+1}\right] + ... + C_{k,q-1}\, E\left[{}_{t+2}y_{t-q+1} \mid \Omega_{t+1}\right] \\ &\qquad + C_{k,q}\ {}_{t+1}y_{t-q} + ... + C_{k,d}\ {}_{t-d+q+1}y_{t-d} \\ &\quad + C_{k,1}\, E\left[{}_{t+q+2}y_{t+1} \mid \Omega_{t+1}\right] + ... + C_{k,d}\, E\left[{}_{t+q+d+1}y_{t+d} \mid \Omega_{t+1}\right]\end{aligned} \tag{2.3}$$

where $C_k(L) = C_{k,d}L^d + ... + C_{k,1}L + C_{k,0} + C_{k,1}L^{-1} + ...C_{k,d}L^{-d}$ is a $(2d+1)$-order polynomial in the lag operator L, $d > q$, the weighting parameters $C_{k,i}$ $i = 0,..,d$, are defined by the choice of the k-th trend ($k = 1, 2, 3$), and $C_k(1) = 1$. The trend in (2.3) applies the filter $C_k(L)$ to the single series given by the in-sample post-revision output data for ${}_{t-d+q+1}y_{t-d}$ to ${}_{t+1}y_{t-q}$ augmented by forecasts of the post-revision output data for ${}_{t+2}y_{t-q+1}$ to ${}_{t+q+d+1}y_{t+d}$. The filter is symmetric around expected post-revision output at time t, $E\left[{}_{t+q+1}y_t \mid \Omega_{t+1}\right]$, although the trend is one-sided in the sense that it is, through the forecasts, a complicated linear function of the lagged $\mathbf{z}_t$ in Ω_{t+1}. The trend depends not only on the parameters of the specified trend filter, but also on the model with which expectations of the post-revision output series are formed (changing, therefore, as the choice of revision horizon q changes, for example). The trend measure obtained by applying the HP filter to the observed and expected post-revision output series obtained using (2.2) is denoted by $\widetilde{y}_t^{mh} \mid \Omega_{t+1}$. We shall also apply the exponential smoothing filter and a moving average model to construct output gaps from the forecast-augmented data for the purpose of comparison; these are denoted with an 'e' and 'm' superscript so that the multivariate versions of the series are $\widetilde{y}_t^{me} \mid \Omega_{t+1}$ and $\widetilde{y}_t^{mm} \mid \Omega_{t+1}$ respectively.[5]

[5]It is important to note that the model in (2.2) relates to the real time output data and is used to generate forecasts of the post-revision output series. The trend measures are obtained through the application of a filter to these (untransformed) series. It should be clear that this is quite different to the modelling and forecasting of de-trended series. The analysis of transformed data in this latter approach can provide quite misleading analysis of the time series properties of output. See Harvey and Jaeger (1993) for further details.

2.2 Representing the Output Gap under Uncertainty

Having estimated a chosen model of the form (2.2), it is relatively straightforward to obtain point forecasts of the terms in (2.3) and to use these in (2.1) to obtain a real time output gap measure. But the simple statistic obtained in this way obviously does not convey the estimation and data uncertainties associated with the output gap measure, and these are potentially significant here given that forecasts of the revised and unrevised series are used in various different ways in the construction of the measure. Moreover, these uncertainties are likely to be compounded when there is a protracted revision process because of the trend uncertainties discussed above. Certainly it is not immediately obvious how a simple measure of the gap can reflect the choice of detrending technique or the choice of model used to obtain forecasts of post-revision output.

In fact, however, all of these issues can be accommodated in a relatively straightforward way if we choose to represent the output gap through estimates of its probability density function (pdf's) rather than through simple point estimates. Indeed, providing a richer probabilistic description of the output gap not only helps in conveying the uncertainties associated with the gap more clearly but also allows us to provide statements on the likely occurrence of specified events that involve the gap and that may be of interest to particular decision-makers. For example, in monetary policy decisions, the point estimates of the gap will be sufficient only in the special circumstances of the "LQ problem" where the monetary authority's objective function is quadratic in the gap and any constraints are linear. In practice, it might be more realistic to assume that the authority's objective function is concerned with 'booms' and 'recessions' (i.e. whether the output gap is positive or negative over some period, irrespective of size), say, or with whether conditions are improving or deteriorating (i.e. with the gap rising or falling).[6] In this case, interest focuses on joint events involving the gap in successive periods, and the events are unlikely to be easily inferred from point estimates of the gap. Rather, direct statements of the likelihood of the probability of these events will be helpful and probabilistic representations of the output gap will be required.

[6]See Svensson (2001, 2002), Cukierman and Gerlach (2003) and Walsh (2003) for further discussion of the form of monetary authorities' objective functions.

Event probability forecasts and pdf's of this sort are straightforward to calculate using simulation methods so long as the underlying data generating process is relatively simple. This is the case with the VAR model in (2.2). Garratt *et al.* (2003) provide a detailed description of the methods, but the ideas are simple to explain. For example, consider the case where we abstract from parameter uncertainty for the time being and focus on the model of (2.2), denoted M_q. Then one can use the estimated version of the model, based on the observed data $\mathbf{Z}_{t+1} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_{t+1})'$, to generate S replications of the future vintages of data, denoted $\widehat{\mathbf{Z}}^{(s)}_{t+2,t+H} = (\widehat{\mathbf{z}}^{(s)}_{t+2}, \widehat{\mathbf{z}}^{(s)}_{t+3}, ..., \widehat{\mathbf{z}}^{(s)}_{t+H})'$ for $s = 1, ..., S$, on the assumption that the model continues to hold over the forecast horizon $t+2, ..., t+H$. These S simulated future vectors of variables provide the pdf of $\mathbf{Z}_{t+2,t+H}$ conditional on the observations available at the end of period $t+1$ and on model M_q, which we will denote $\Pr\left(\mathbf{Z}_{t+2,t+H} \mid \mathbf{Z}_{t+1},\, M_q\right)$. In particular, the simulations include values of the forecast post-revision output level at time t, ${}_{t+q+1}\widehat{y}^{(s)}_t$, and of the subsequent d observations on which the trend measure $\widetilde{y}^{k(s)}_t|\Omega^{(s)}_{t+1}$ can be based. The simulated distribution of $x^{rk(s)}_t = {}_{t+q+1}\widehat{y}^{(s)}_t - \widetilde{y}^{k(s)}_t|\Omega^{(s)}_{t+1}$, $s = 1, .., S$, obtained in this way directly provides the estimated pdf of the real time output gap measure. Equally, counting the number of times an event occurs in these simulations provides a forecast of the probability that the event will take place; for example, the fraction of the simulations in which $\{\widehat{x}^{rk(s)}_{t-1} < 0\}$ *and* $\{\widehat{x}^{rk(s)}_t < 0\}$ provides a real time estimate of the forecast probability that there is a recession in time-t (defined in this case as two consecutive periods where the gap is negative). Extending the simulation exercise to accommodate parameter uncertainty for the given model simply involves an additional iteration of the simulation procedure in which replications of the historical data and of the model parameters are also produced (see Garratt *et al.* (2003) for more details). And the process can be conducted recursively moving through the sample of data, up to the most recently-available data, so that final vintage estimates of the pdf and probability forecasts involving the gap can be simulated too. In this way, a complete characterisation of the output gap measure can be obtained, accommodating the various elements of data and estimation uncertainty.

Perhaps even more importantly here though, it is also straightforward to accommodate the difficulties involving model choice and the choice of detrending method, as highlighted by (2.3), in the calculation of estimated pdfs and event probability forecasts involving the gap. This is achieved adapting the 'Bayesian model averaging' approach of Draper (1995) and Hoeting *et al.* (1999) and it is important because, as we have noted, there are potential problems in placing too much emphasis on a single forecasting approach so that pooling of forecasts might be advisable. Specifically, assuming that there are Q different models, denoted M_q, $q = 1, ..., Q$, then the pdf of $\mathbf{Z}_{t+2,t+H}$ conditional on $\mathbf{Z}_{t+1}$ and accommodating model uncertainty is provided by the "Bayesian model averaging" formula,

$$\Pr\left(\mathbf{Z}_{t+2,t+H} \mid \mathbf{Z}_{t+1}\right) = \sum_{q=1}^{Q} \Pr\left(M_q \mid \mathbf{Z}_{t+1}\right) \Pr(\mathbf{Z}_{t+2,t+H} \mid \mathbf{Z}_{t+1}, M_q). \tag{2.4}$$

The $\Pr\left(\mathbf{Z}_{t+2,t+H} \mid \mathbf{Z}_{t+1},\, M_q\right)$ are given directly by the simulation exercises described above for each model. Further, Draper (1995) suggests the use of the familiar Schwarz Bayesian information criterion to obtain model weights $w_{q,t+1}$ given by

$$\Pr\left(M_q \mid \mathbf{Z}_{t+1}\right) = \frac{\exp(SBC^{*}_{q,t+1})}{\sum_{j=1}^{Q} \exp(SBC^{*}_{j,t+1})} \tag{2.5}$$

where $SBC^{*}_{q,t+1} = SBC_{q,t+1} - \max_j(SBC_{j,t+1})$, $SBC_{q,t+1} = LL_{q,t+1} - \left(\frac{k_q}{2}\right)\ln(t)$ is the Schwarz Bayesian information criterion, and $LL_{q,t+1}$ is the maximized value of the log-likelihood function for model M_q based on data available at time $t+1$. Alternatively, following Burnham and Anderson (1998), one could use Akaike weights, using AIC in place of SBC in (2.5). These assumptions allow $\Pr\left(\mathbf{Z}_{t+2,t+H} \mid \mathbf{Z}_{t+1}\right)$ to be estimated straightforwardly using (2.4) based on ML estimation of the candidate models.

The pdf of $\mathbf{Z}_{t+2,t+H}$ given in (2.4) and (2.5) can be used to estimate pdfs of functions of the variables in $\mathbf{Z}_{t+2,t+H}$ and events involving these, including functions defining gaps. In practice, this means that we simply count the number of times the event occurs in the simulations generated using all the candidate models, weighting the number of simulations according to (2.5).[7] A similar procedure will also allow us to pool pdfs to accommodate any uncertainty over the choice of detrending technique; i.e. the choice of detrending

[7]In such an exercise, it is important that the events considered are the same in all simulations. It is therefore worth noting that, although the trends given by (2.3) are defined according to the revision horizon q, any events of interest involving the gap will be the same across all models (so that the pooling is appropriate) because the trends are defined with regard to the post-revision output series. This is clear if we note that $E\left[{}_{t+q+1}y_t \mid \Omega_{t+1},\, M_q\right] = E\left[{}_{t+Q+1}y_t \mid \Omega_{t+1},\, M_q\right]$ for all $q < Q$ (the longest of the

technique k and associated parameters $C_{k,d}$ in (2.3). Here, however, there are no obvious candidates with which to decide on $\Pr(k \mid \mathbf{Z}_{t+1})$ so that simulations involving the different detrending techniques are likely to be simply given equal weights.

3 Output Gaps in the UK

In this section, the methods described above are applied to the real time dataset on quarterly UK output data constructed by the Bank of England. The dataset includes 180 vintages of data, with the first vintage dated 1961q2 and the final vintage dated 2006q1.[8] All vintages of data run from 1955q1 up to one period prior to the release date; i.e. $Y_t = \{_t y_{1955q1}, ...,_t y_{t-1}\}$, $t = 1961q2, ...2006q1$.

A simple analysis of the time series properties of the series, using augmented Dickey-Fuller tests, shows that the assumptions necessary to apply our modelling approach (namely that actual output growth and output revisions are stationary) appear to hold. However, a brief review of the process by which the output series are collated and revised suggests that obtaining an adequate model of the form given in (2.2) will not be straightforward.[9] Skipper (2005) notes that the "preliminary" release of GDP figures in the UK is published 25 days after the end of the relevant quarter. This is claimed to be the fastest in the world but it is acknowledged that the information used to construct the estimate is approximately 44% data while the rest is based on forecasts using Holt-Winter/ARIMA methods. As more survey data becomes available, these forecasts are progressively replaced and initial estimates are revised. Actual data used to estimate GDP increases to 67% by the time of the "first release" of the UK output, income and expenditure series, published around 55 days after the quarter in question. This figure reaches 80% by the time of the publication of the UK quarterly national accounts, about

alternative revision horizons that are considered) since, under model M_q, it is assumed that there are no further revisions between $t + q + 1$ and $t + Q + 1$.

[8]The real time database available at http//www.bankofengland.co.uk/statistics/gdpdatabase provides vintages to 2002q1. The Bank also provided corresponding vintages up to 2006q1 on request.

[9]This is in contrast to GLMS's analysis of US data where a simple model with a revision horizon of two quarters and lags in the VAR of order 2 was sufficient to adequately model the real time output data at all recursions.

85 days after the quarter. By this time, most of the forecasting in GDP(O) is in the government categories, which are constructed using information that is still only 36% actual data. Other information also continues to arrive after these revisions, however, particularly that relating to imports and exports. This is certainly a more complicated and uncertain data collection/revision process than in the US, therefore, and Garratt and Vahey (2006) find evidence of mean revisions continuing to be significant after eight quarters in the UK. This suggests that our focus on trend uncertainty and the potential role of pooling of forecasts will be appropriate for the UK.

3.1 Forecast Augmentation

The first exercise undertaken on the UK data aims to find whether the unreliability of the real time measure of the US output gap found by OvN carries over to the UK and, if it does, whether the forecast augmentations suggested in GLMS help to offset the problem. To this end, we first follow OvN and consider the successive vintages of data in turn, restricting attention to trends based on the HP filter (using a smoothing parameter of 1600), to derive the 'real-time measure' $\widetilde{y}_t^o|Y_{t+1}$, $t = 1961q2, ...2005q4$ as the end-of-sample observation of the trend in each of the recursions. To identify the potential role of forecast augmentation (and to highlight the end-of-sample estimation problem associated with the use of two-sided filters), we also derive the corresponding HP trends based on data augmented by forecasts generated by a univariate AR(8) specification applied to the time-t vintage of data (i.e. ignoring the potential information captured by the revisions over time). This trend, and associated gap measure, is denoted by a 'u' superscript.

Figure 1 shows two of the output gaps considered by OvN, namely $x_t^{ro} = ({}_{t+1}y_t - \widetilde{y}_t^o|Y_{t+1})$ and $x_t^{fo} = ({}_Ty_t - \widetilde{y}_t^o|Y_T)$, for $t = 1961q1, ..., 2005q4$, and $T = 2006q1$. The figure illustrates the considerable differences between the real time and final vintage measures of the gap arising out of data revisions and the end-of-sample effects on the underlying trends. Table 1 shows that the correlation between the real time and final measures of OvN is just 0.419, and the two measures agree on whether output growth is above or below trend in only 61.7% of the sample period. Taking the final measure x_t^{fo} as the best indicator of the true output gap available, the poor performance of the x_t^{ro} measure in reflecting the true output gap would lead us to the same conclusion for the UK as OvN's conclusion for the US: namely, that real time measures of the gap are unreliable.

Table 1 also describes the effect of employing the forecast-augmentation method of calculating the trends on the gap measures, where the augmentation is based on the simple univariate model estimated using the vintage-t output data. There is a substantial impact on the variability of the output gap series, reducing the standard deviation of the real time output gap measure by around 20%. This illustrates that the forecast-augmentation is having a considerable impact on the trend measure, reducing the estimation error variance associated with the application of the HP filter at the end-of-sample. The effect is to raise the correlation between the final measure x_t^{fuh} and the real time measure, x_t^{ruh} to 0.589. The agreement on the occurrence of booms and recessions rises to 74%. The improvement in reliability using the forecast-augmentation method is pronounced, indicating that the forecast augmentation suggested in GLMS is useful here too.

3.2 Accommodating Trend Uncertainty

We turn next to the main novelty of the UK data analysis which is concerned with the treatment of trend uncertainty. The forecast-augmentation method obviously relies on choosing an appropriate model of output growth and the revision process. Further, the impact of this choice might differ depending on the choice of detrending technique. Neither of these decisions is straightforward in the case of the UK.

In deciding on the appropriate multivariate model, our *a priori* view was that the revision process for the UK is protracted and complex so that a relatively sophisticated model of the data might be required (i.e. large p and large q in (2.2)). On the other hand, one purpose of the model is to provide forecasts with which to augment the in-sample post revision data and we were also aware of the dangers of relying too heavily on a single forecasting model, particularly in the presence of small samples where a large VAR model could rapidly become over-parameterised. To acknowledge our uncertainty on the appropriate model, we therefore consider a set of eight alternative models for UK output growth and revisions, defined according to the specified revision horizon (i.e. the period after which no further revision takes place). Hence, the models each take the form given in model (2.2) with $q = 1, ..., 8$, and are denoted M_1 to M_8 respectively; see Appendix for detail. The maximum lag length considered in each model is $p = 4$ and, to deal with potential over-parameterisation, we also obtain a set of restricted models following a specification search on each of M_8 to M_1 considered in turn at each recursion. In this search, we impose a zero coefficient restriction on the variable with the lowest t-ratio in each model until all the remaining variables have t-ratios of 1.25 or more.[10]

Once again, we investigate the models in real time through recursive estimation. Given the large number of explanatory variables in each equation ($37 = 1 + 9 \times 4$ in each of the unrestricted equations of model M_8 for example), our first recursion is based on the full information set available in 1985q1. This means that, including the final recursion dated in 2006q1, we consider 84 recursions in total. As evidence to support our choice of a maximum revision horizon of 8 and lag length of 4, we calculated two sets of tests. *First*, we computed $\chi^2_{LM}(4)$ tests of the joint hypothesis that the parameters on the lags of the eighth revision are equal to zero in each of the nine equations in model M_8 at each recursion. The rejection rate at the 5% level of significance, calculated over the 84 recursions, was close to zero for all but the seventh and eighth equations, but rejection rates here were 93% and 87% respectively. This confirms that the revision process is indeed very protracted in the UK, with the revisions made up to two years after the first release of data still containing systematic and predictable content. *Second*, we computed $\chi^2_{LM}(9)$ tests of the joint hypothesis that the parameters on the fourth lags of the explanatory variables were zero in each equation at each recursion. Again working at the 5% level of significance, equations 1, 2 and 7 were found to have rejection rates close to zero considered over the recursions. However, the remaining equations had very high rejection rates, indicating that a lag length of 4 is required in most equations in the models.

Before proceeding with our analysis, we also checked that the sophistication of the multivariate model is helpful and that gap measures based on these will outperform those based on the univariate model. This is confirmed by in-sample root mean squared errors

[10] We estimate the models using OLS for each equation in turn rather than SUR estimation. This is because SUR would deliver no gains in terms of efficiency over OLS in the unrestricted VAR and little in the restricted version whilst complicating the estimation and search procedures considerably.

(RMSE), calculated using the difference between the real time and final vintage output gap measures; i.e. $RMSE^u = \sqrt{\sum_{t=1}^{T-q-1}(x_t^{ruh} - x_t^{fuh})^2}$ for the univariate model and the equivalent statistic, obtained by replacing x_t^{ruh} with x_t^{rmh}, for the multivariate model. Specifically, we find that $RMSE^u = 0.014971$ and that this is larger than the corresponding figure obtained in all of the multivariate models that we consider; the smallest improvement in RMSE over the univariate counterpart is the 50% improvement achieved by model M_7 and the largest gain is model M_4 with an 89% improvement.

To gain an overview of the relative levels of support for the various alternative multivariate models, Table 2 reports the SBC and AIC weights for the eight models based on (2.5) calculated as an average over the 84 recursions. According to these averages, models M_4-M_6 have most (and broadly equal levels of) support based on SBC, while models M_6-M_8 have most support according to AIC. The fact that five of the eight models considered gain some support by one or other criteria, according to these average statistics, confirms that an adequate treatment of model uncertainty is important here. Moreover, these averages obscure the considerable variability in support for the alternative models that occurs over time (and which is important in real time decision making). Given the similarity between the models and the formula in (2.5), we find that relatively small differences in likelihoods across models can translate into quite large changes in weights across the range of models we consider. But even if we mitigate this effect by considering a ten-period moving average of the model weights (e.g. $\overline{w}_{it}^{sbc} = \frac{1}{10}(w_{it}^{sbc} + ... + w_{it-9}^{sbc})$, $i = 1, .., 8$), we still find that the support for the models shifts considerably over time. Figure 2a illustrates this effect, showing that, according to SBC, model M_5 is most strongly supported in the early recursions but is overtaken by model M_6 through the middle of the sample period, and that model M_4 has most support by the end of the period. Figure 2b displays similar shifts in support, from model M_8 to M_6 and then to M_7, according to the AIC weights $\overline{w}_{it}^{aic}$. The substantial time-variation in the weights provides a further argument for the use of the model averages in preference to any individual model since the time-varying weights can be used to accommodate this aspect of model uncertainty also.

3.2.1 Output Gap Measures Based on the Multivariate Models

For each recursion, we are able to compute forecasts of the post-revision output series and to estimate output gap measures as in (2.1) and (2.3) using any one of the models $M_1 - M_8$ or their averages. Table 3 provides summary statistics on the output gap measures derived on the basis of the eight individual models plus measures based on three alternative 'pooled' forecasts; namely, the SBC and AIC model averages plus, for the purposes of further comparison, an equal-weights model average. The first column reports on gap measures which continue to use the HP detrending technique and are directly comparable to those in Table 1. The figures show that the output gap measures based on any of the multivariate model are rather less volatile than those discussed in Table 1, with standard deviations in the range [0.092, 0.011] compared to 0.016 and 0.013 for x_t^{ro} and x_t^{ruh} respectively. Importantly, the results also show that the advantages of the forecast-augmentation remain using the multivariate model. For models M_1-M_8, the correlations between the real time measure of the gap and the corresponding final measure are in the range [0.571, 0.717] and the agreement on booms and recessions between the real time and final vintage gap measures is between 70% and 75% across all the models. It is perhaps worth noting that the correlation between the real time and final vintage measures are highest among the models with shorter revision horizons (i.e. $q = 1, 2, 3, 4$). This latter observation is reflected also by the correlations for the AIC, SBC and equal weight average models, which are equal to 0.571, 0.647 and 0.648 respectively. Hence, the lowest correlation is found for the AIC average model which we know to be dominated by M_6-M_8, and this suggests that SBC might be a more appropriate selection criterion in this context.

Table 3 also provides statistics relating to the output gap measures obtained using two alternative methods for measuring the trend in place of the HP filter. The measure denoted x_t^{rme} refers to the gap obtained in real time and based on an exponential smoothing (ES) filter. The filter is again applied to the post-revision series augmented with forecasts based on our multivariate model and the 'smoothing' parameter was set equal to 20. This means that 67% of the weight is on observations one year either side of the observation of interest and 87% on two years either side which, in total, broadly corresponds to the

HP weights. The third measure, x_t^{rmm}, applies a seventeen-quarter moving average to the same series. Hence, the three filters are comparable in the sense that they all focus attention on a four-year period centred around the date of interest but differ according to the parameters in the $C_k(L)$ of (2.3).[11]

The results in Table 3 show that the (relatively) reassuring results obtained for the gaps based on the HP filter are also found with the other two smoothers. Using the ES smoother, the correlation between the real time and final vintage gap measures are comparable to those obtained with HP, with the highest correlation observed for the SBC average and the associated models with shorter revision horizons. The results obtained using the seventeen-period moving average smoother provide correlations between the real time and final measure ranging which are reasonably high for the shorter revision-horizon models, but are rather smaller for the longer revision-horizon models.

Figures 3a and 3b show that a reasonably consensual picture of the state of the macroeconomy would have been obtained in real time using any of the alternative gap measures based on the multivariate models. Figure 3a illustrates the variability in gap measures arising out of the model uncertainty, focusing on the models that were preferred according to SBC at various points in the sample; i.e. $M_4 - M_6$. The figure shows a high degree of agreement over most of the sample, with the three series lying within 0.2-0.3% of each other at most times. The exceptions are for some observations at the beginning of the sample, when the measure based on model M_4 lies around 0.5% below the other two, and in 2003/4 when the measure based on M_6 falls to unreasonably low levels compared to the other measures. As it happens, however, these observations are given zero weight in the SBC-based average, indicating that the unusual forecasts obtained from these models at these times are the result of a poorly fitting model, and highlighting the advantages of allowing for time-varying weights in the model averages in this real time exercise. Figure 3b shows that there is more variability in gap measures introduced through the choice

[11] The gains in using a multivariate model over a univariate model identified previously by comparing $RMSE$s based on HP-filtered series are also found with these two alternative detrending techniques. Hence, for ES and MA, we find $RMSE^u = 0.01255$ and 0.01394 respectively. The figure is reduced by between 60% (M_7) and 105% (M_4) for the gaps based on the ES filter, and comparable improvements are found in the MA measures.

among the three detrending techniques we consider. Even here, however, the gap measures continue to be within 0.5% of each other for the most part with the most significant discrepancies arising during the recessionary period 1991-1994 where the three measures differ in their perception of when output begins to recover and return to its trend level.

4 Probability Forecasts of the Output Gap

Figure 3b also includes a plot of the output gap obtained using the final vintage of data applying the HP filter to the series, augmented at the end of the sample with forecasts based on the SBC average multivariate model. Apart from the end of the sample, this series corresponds to the final vintage measure of Table 1 and represents the best available indicator of the output gap based on HP. Comparison with the other series in the Figure shows that, although the correspondence between the real time and final measures is better than in Figure 1, there remain some substantial discrepancies between the series. While forecast augmentation and model averaging help improve the real time measures, therefore, it remains very important to convey accurately the uncertainty associated with the measures when they are reported for decision-making purposes.

In Section 2, we noted that information on the size and the precision of measures of the gap can be conveyed directly through the use of pdf's of the gap measured at different forecast horizons and, using these, through the use of forecasts of the probability of specified events involving the output gap occurring. Figure 4 provides an illustration of the pdf's that are obtained using the methods described in Section 2 showing densities relating to measures of the output gap formed in 2003q2 (based on the SBC-average multivariate model and applying the HP filter). This was a particularly significant period for data revisions in the UK because there was a sequence of revisions to GDP figures that attracted considerable public comment and generated criticism of the Office of National Statistics. The media reaction to the revisions was seen as a potential threat to confidence in official statistics and in the organisations and organisational arrangements responsible for them and the Statistics Commission reviewed the background to the revisions as a result, publishing its conclusions in the 'Mitchell Report' (Statistics Commission, 2006). It is true that the 2003q2 revision was, at just over 0.3 percentage points (quarterly growth rate), the largest since the 1980s and will have had a major impact on decision-maker's perception of the state of the economy. But Figure 4 places the revision in perspective showing, for example, that the inter-quartile range for the annualised gap measure in 2003q2 was [-1.3%, 0.6%], entirely compatible with a measurement error of 1.2% therefore. Indeed, the probability that the gap was at least 1.2% greater than the point estimate at that time was around 20%. A revision of the size observed in 2003q2 is unusual, therefore, but not extraordinary. As the Mitchell Report emphasises, the relatively expert users who rely on these statistics generally understand that revisions are to be expected. But the undue reliance on point estimates of output series and associated gap measures can be misleading and could be easily avoided with the publication of pdf's of the form in Figure 4.

The simulations underlying the pdf of the contemporaneous (or 'now-cast') gap measure for 2002q3 plotted in Figure 4 accommodate both stochastic uncertainty and model uncertainty (since we use the SBC-weighted model).[12] No data for output in 2002q3 had been released (the first release data in that period corresponds to output in 2003q1) and the gap measure depends directly on a forecast of the post-revision observation of y_{2002q3} to be released nine periods later. Moreover, the forecasts of post-revision series further into the future also impact on this measure through their influence on the estimated trend. With forecast horizons of two-four years involved, it is not surprising that the gap measure is calculated with considerable uncertainty. Of course, this uncertainty increases if we move further into the future to obtain pdfs of forecast future gap measures. To illustrate this, Figure 4 also shows pdfs associated with gap measures one year beyond the end of the sample and two years beyond. These are flatter than the 2003q2 density to reflect the rising uncertainty although the changes are not dramatic, indicating that the largest part of the uncertainty involved in these measures is common to all three.

The sequence of pdf's plotted in Figure 4 can also be used to read the likelihood

[12]We also computed the pdfs accounting for parameter uncertainty but the numbers were very similar. The results are based on 10,000 replications where the innovations were obtained from a multivariate normal distribution chosen to match the observed correlation of the estimated residuals for each of the models over the full sample period (a parametric bootstrap).

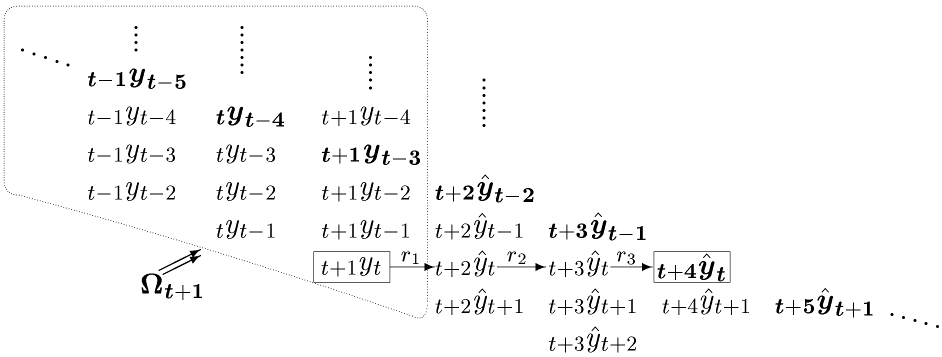
of particular events of interest and how these change over time. So, for example, the probability that the output gap is negative is forecast to be approximately the same in 2003q2 and 2004q2, at 60%, falling to around 55% in 2004q2. Such statements reflect the fact that the point forecasts are all in the region -0.4% to -0.2%. But if a decision-maker is concerned with whether the gap is positive or negative, the statements present the forecast information in a way that is directly useful and convey far more precisely the strength of conviction with which the forecaster predicts the event will or will not occur. Moreover, probability forecasts can also be used to explain the likelihood of complicated joint events occurring which is difficult (if not impossible) based on point forecasts only. So, for example, a decision-maker might be interested in making an investment decision but only if the economy avoids a recession. This concept can be measured in many ways but one possibility, say, is that the five-period moving average of the gap, centred on time t, is positive. A sequence of point estimates and forecasts of the gap over the five periods would provide little insight on this because the averaging of the forecast at each point in time cannot convey information on the *sequence* of forecasts over the five periods that might be observed and which is important to the decision-maker. However, the simulations underlying the pdf's of Figure 4 can be used directly to evaluate the likelihood of this event. Figure 5 illustrates precisely this idea, plotting the probability of 'avoiding a recession' as calculated in real time. The first of the plots here is based on the SBC-average multivariate model and, while it clearly reflects the pattern of the point estimates given in Figure 3, it again provides slightly different information and in a way that is more directly useful to decision-makers. The second plot is based on the average of the results obtained using the three alternative detrending techniques (with equal weights). This corresponds relatively closely to the SBC-average series, but it is worth emphasising that this latter plot accommodates all aspects of trend uncertainty as well as the stochastic, estimation and model uncertainty captured in the SBC-average. And finally, we plot here the same probability forecast based on the univariate model considered earlier. This series clearly differs from the other two and, while the extent to which these differences are important depends on the decision-making context, this provides a further illustration of the advantages of working with the multivariate model.

5 Conclusions

The analysis of this paper starts from the point that output gap measures are an essential element of many decisions but that they are measured with considerable uncertainty. This is because of the imprecision of the output data available at the time decisions have to be made and because of the difficulties in establishing a precise measure of trend output. We have shown that these uncertainties can be mitigated by modelling the output process alongside the revision process, making use of forecasts of current and future post-revision output levels, to obtain more precisely estimated measures of the gap for use in real time decision-making. The use of a model averaging approach means that 'trend uncertainties' surrounding the choice of model and detrending technique can be readily quantified, alongside the more usual stochastic and estimation uncertainties. And the representation of the gap measures using forecast pdf's and event probabilities ensures that these uncertainties are conveyed in a straightforward way. This is important if the output gap measures are to be used by decision-makers in any context other than special circumstances of the "LQ problem".

Appendix A: The Form of the Real Time Data

The diagram below illustrates the form of the real time data set employed. We illustrate the use of the data set available at time $t+1$ assuming that there are three revisions (i.e. $q = 3$).



In the diagram, the bordered data represents the full information set $\mathbf{\Omega_{t+1}}$. The first release of information on output at t, $_{t+1}y_t$, is expected to be revised three times so that the forecast of the post-revision series is $_{t+4}\hat{y}_t$, with the "^" denoting the forecast here. The emboldened series represents the actual and forecast values of the post-revision output series.

Appendix B: Restrictions Defining Models M_1-M_8

The general form of forecasting model is given by:

$$\mathbf{z}_t = \mathbf{a} - \mathbf{B}_1\mathbf{z}_{t-1} - ..... - \mathbf{B}_p\mathbf{z}_{t-p} + \mathbf{u}_t$$

where $\mathbf{z}_t$ is $((q+1)\times 1)$, $\mathbf{a}$ is $((q+1)\times 1)$, $\mathbf{B}_1, ...\mathbf{B}_p$ and $\mathbf{u}_t$ are $((q+1)\times 1)$ matrices. In our application, $q = 1, ..., 8$ and $p = 4$. For the unrestricted version of model M_8, the $\mathbf{a}$ vector and $\mathbf{B}_p$ matrices are given by:

$$\mathbf{B}_p = \begin{bmatrix} b^p_{11} & b^p_{12} & . & . & b^p_{19} \\ b^p_{21} & b^p_{22} & . & . & b^p_{29} \\ . & & . & & . \\ . & & & . & . \\ b^p_{91} & . & . & . & b^p_{99} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ a_9 \end{bmatrix}$$

The preferred model M_8 is the outcome of a specification search in which only variables whose coefficients have (absolute) t-values in excess of 1.25 are retained. The restrictions for each of our models, in reverse order, M_7 to M_1 are:

<u>Model M_7</u>: Row $(b^p_{91}, b^p_{92}, ..., b^p_{99}) = 0$ and column $(b^p_{19}, b^p_{29}, ..., b^p_{99})' = 0$, $p = 1, ..4$. Also $a_9 = 0$. This amounts to 69 restrictions compared to unrestricted M_8. The preferred model M_7 is based on a specification search over this unrestricted M_7.

<u>Model M_6</u>: M_7 restrictions plus row $(b^p_{81}, b^p_{82}, ..., b^p_{88}) = 0$ and column $(b^p_{18}, b^p_{28}, ..., b^p_{88})' = 0$, $p = 1, ..4$. Also $a_8 = 0$. Total restrictions $= 130$ compared to the unrestricted M_8.

Models M_5, M_4, M_3, M_2 and M_1 are similarly defined, with total restrictions numbering 183, 228, 256, 294 and 315 respectively compared to the unrestricted M_8.

It is worth noting that, in calculating the likelihood of the preferred model M_7, for use in the weighting formula (2.5) for example, the model is supplemented with the ninth equation obtained in the preferred model M_8 so that the systems are directly comparable (although the effect of the ninth equation is neutral as far as the selection criterion is concerned). Similarly for calculating the likelihoods of models M_6,.., M_1.

References

Adam, C. and D. Cobham (2005), "Real-Time Output Gaps in the Estimation of Taylor Rules: A Red Herring", *mimeo.*

Amato, J. D., and N. R. Swanson (2001) "The Real-Time Predictive Content of Money for Output", *Journal of Monetary Economics*, 48, 3-24.

Brunner, A.D. (2000), "On the Derivation of Monetary Policy Shocks: Should we Throw the VAR Out With the Bath Water", *Journal of Money Credit and Banking*, 32, 254-279.

Burman, J.P. (1980), "Seasonal Adjustment by Signal Extraction", *Journal of Royal Statistical Society* A, 143, 321-337.

Burnham, K. P. and Anderson, D. R. (1998), *Model Selection and Inference: A Practical Information-Theoretic Approach*, New York: Springer-Verlag.

Canova, F. (1998), "Detrending and Business Cycle Facts", *Journal of Monetary Economics*, 41, 3, 475-512.

Christiano, L.J., M. Eichenbaum and C.L. Evans (1999),"Monetary Policy Shocks: What Have We Learned and to What End?", Chapter 2 in J.B Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, Volume 1A. North-Holland, Elsevier: Amsterdam.

Clements, M. and D. Hendry (2005), "Information in Economic Forecasting", *Oxford Bulletin of Economics and Statistics*, 67 (Supplement), 713-753.

Cukierman, A. and S. Gerlach (2003), "The Inflation Bias Revisited: Theory and Some International Evidence", *Manchester School*, 71, 3, 541-565.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty," *Journal of Royal Statistical Society Series B*, 58, 45-97.

Garratt, A., K.C. Lee, E, Mise and K Shields (2006), "Real Time Representations of the Output Gap", *mimeo*, University of Leicester.

Garratt, A., K. Lee, M.H. Pesaran, and Y. Shin, Y. (2003), "Forecast Uncertainties in Macroeconometric Modelling: An Application to the UK Economy", *Journal of American Statistical Association*, 98, 464, 829-838.

Garratt, A., K.C. Lee, M.H. Pesaran and Y. Shin, (2006), *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*, forthcoming Oxford University Press, Oxford.

Garratt, A. and S.P. Vahey (2006), "UK Real-Time Macro Data Characteristics", *Economic Journal*, 116, 509, F119-F135.

Harvey, A.C. and A. Jaeger (1993), "Detrending, Stylised facts, and the Business Cycle", *Journal of Applied Econometrics*, 8, 3, 231-247.

Harvey, A.H., Trimbur, T. and H. van Dijk (2006), "Trends and Cycles in Economic Time Series: A Bayesian Approach", forthcoming, *Journal of Econometrics.*

Harvey, D.I. and P. Newbold (2005), "Forecast Encompassing and Parameter Estimation", *Oxford Bulletin of Economics and Statistics*, 67 (Supplement), 815-835.

Hodrick, R.J. and E.C. Prescott (1997), "Postwar U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29, 116-130.

Howery, E.P. (1978) "The Use of Preliminary Data in Econometric Forecasting", *Review of Economic Statistics*, 6, 193-200.

Mise, E., T-H. Kim and P. Newbold (2005a), "On the Sub-Optimality of the Hodrick-Prescott Filter, *Journal of Macroeconomics*, 27(1), 53-67.

Mise, E., T-H. Kim and P. Newbold (2005b), "Correction of the Distortionary end-effect of the Hodrick-Prescott Filter: Application", *mimeo*, downloadable from http://www.le.ac.uk/economics/staff/em92.html.

Orphanides, A. (2001) “Monetary Policy Rules Based on Real-Time Data”, *American Economic Review*, 91, 964-985.

Orphanides, A. and S. van Norden (2002) “The Unreliability of Output-Gap Estimates in Real Time”, *Review of Economics and Statistics*, 84(4), 569-583.

Orphanides, A., R.D. Porter, D. Refschneider, R. Tetlow and F. Finan (2000), *Journal of Economics and Business*, 52, 117-141.

Pesaran, M.H. and P. Zaffaroni (2004), “Model Averaging and Value at Risk Based Evaluation of Large Multi Asset Volatility Models for Risk Management”, *mimeo*, downloadable at http://www.econ.cam.ac.uk/faculty/pesaran.

Rotemberg, J. and M. Woodford (1999), “Interest Rate Rules in an Estimated Sticky Price Model”, pages , 57-119 in Taylor J.B. (ed.) *Monetary Policy Rules*, University of Chicago Press: Chicago.

Skipper, H. (2005), “Early Estimates of GDP: information content and forecasting methods”, *Economic Trends,* 617, April.

Statistical Commission (2004). *Revisions to Economic Statistics: The Mitchell Report*, Report no. 17, Vols. 1, 2 and 3, April.

Svensson, L.E.O.,(2001), “Inflation Targeting: Should It Be Modelled as an Instrument Rule or a Target Rule?”, *European Economic Review*, 46, 771-780.

Svensson, L.E.O. (2002), “What is Wrong with Taylor Rules ? Using Judgement in Monetary Policy through Targeting Rules”, *Journal of Economic Literature*, 41, 426-477.

Walsh, C.E. (2003), “Speed Limit Policies: The Output Gap and Optimal Monetary Policy”, *American Economic Review*, 93, 265-278.

Weeks, M. and M Stone (2001), “Financial Crisis, Balance Sheets and Model Uncertainty”, *IMF Working Paper* 01/162.

Table 1: Univariate Hodrick-Prescott Output Gap Measures: 1961q1 – 2005q4

	x_t^{ro}	x_t^{fo}	x_t^{ruh}	x_t^{fuh}
Mean	-0.003	0.000	-0.004	0.000
SD	0.016	0.014	0.013	0.014
Min	-0.040	-0.030	-0.033	-0.030
Max	0.054	0.051	0.042	0.051
x_t^{ro}	1	0.419	0.837	0.420
x_t^{fo}	*0.617*	1	0.698	0.999
x_t^{ruh}	*0.811*	*0.739*	1	0.588
x_t^{fuh}	*0.616*	*0.988*	*0.739*	1

Notes: Output gaps are denoted by x_t The 'r', 'q' and 'f' terms refer to real-time, quasi-real time and final measures respectively, as described in the text; the 'o' and 'uh' superscripts refer, respectively, to trend measures based on methods described in OvN and MKN, using an eighth-order univariate autoregression for forecasts, again described in the text. Summary statistics in the upper panel refer to the mean, standard deviation, minimum and maximum values respectively. Figures in the lower panel refer to correlation coefficients and, in italics, proportion of the sample for which there is agreement that the output gap is positive or negative.

Table 2: Average SBC and AIC Model Weights

Model	AIC Weights	SBC Weights
Model M_1	0.00	0.00
Model M_2	0.00	0.00
Model M_3	0.00	0.00
Model M_4	0.00	0.456
Model M_5	0.005	0.233
Model M_6	0.242	0.273
Model M_7	0.319	0.038
Model M_8	0.434	0.00

Notes: The weights reported here are the averages of the 84 weights computed for each of the recursions starting in 1961q2 and ending in period t, where t=1985q1,..,2005q4 .

Table 3: Correlations and % Agreement of Ups and Downs Between Real Time and Final Multivariate Output Gaps: 1985q1 – 2003q4

Model	HP		ES		MA(17)	
	Corr. [%UD]	S.D.	Corr. [%UD]	S.D.	Corr. [%UD]	S.D.
1	0.664 [75.0%]	0.0096	0.595 [72.4%]	0.0070	0.574 [71.1%]	0.0082
2	0.665 [72.4%]	0.0092	0.594 [71.1%]	0.0069	0.567 [71.1]	0.0081
3	0.665 [72.4%]	0.0099	0.583 [72.4%]	0.0071	0.557 [68.4%]	0.0081
4	0.717 [76.3%]	0.0010	0.667 [75.0%]	0.0072	0.620 [75.0%]	0.0082
5	0.612 [69.7%]	0.0098	0.520 [68.4%]	0.0070	0.481 [64.5%]	0.0080
6	0.570 [75.0%]	0.0106	0.492 [73.7%]	0.0075	0.451 [67.1%]	0.0086
7	0.571 [72.4%]	0.0110	0.482 [73.7%]	0.0077	0.446 [72.4%]	0.0088
8	0.577 [73.40%]	0.0111	0.515 [72.4%]	0.0078	0.473 [68.4%]	0.0088
AIC	0.571 [75.0%]	0.0111	0.504 [73.7%]	0.0079	0.462 [69.7%]	0.0089
SBC	0.647 [75.0%]	0.0099	0.572 [75.0%]	0.0070	0.533 [73.7%]	0.0080
EQ	0.648 [73.7%]	0.0099	0.579 [72.4%]	0.0070	0.543 [71.1%]	0.0080

Notes: Output gaps are denoted by x_t^{rmk} and x_t^{fmk}, k=h, m, and e, relating to the HP, ES and MA detrending techniques respectively. The 'r' and 'f' superscripts refer to real-time and final measures described in the text. Corr denotes correlations between x_t^{rmk} and x_t^{fmk}, %UD denotes percentage of agreement in positive and negative gaps and S.D. refers to the standard deviation of x_t^{rmk}. See also notes to Table 1.

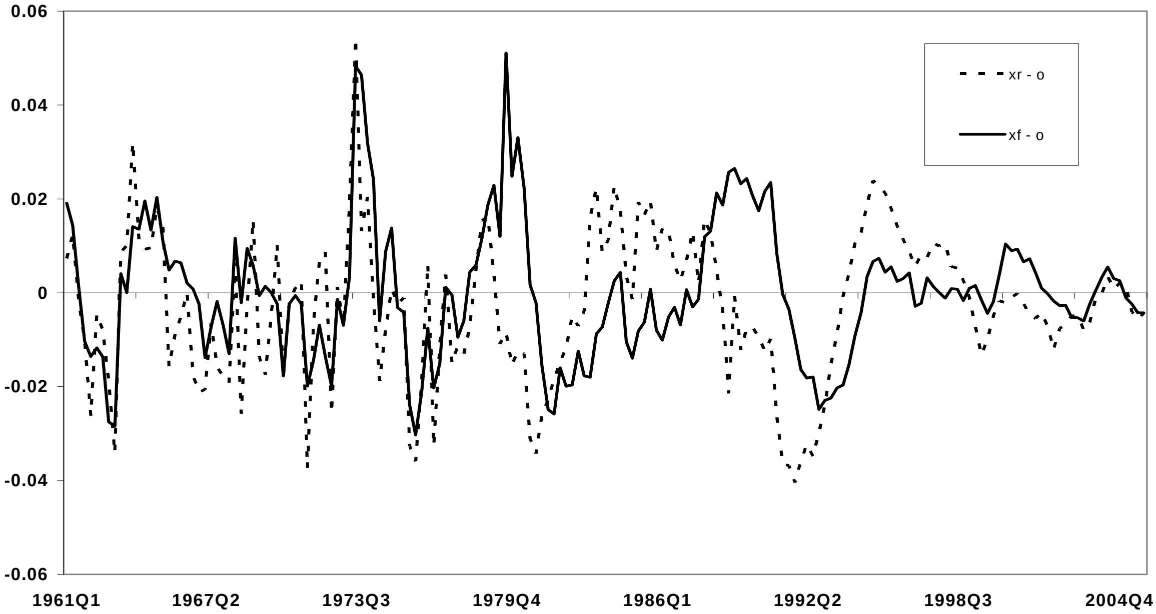


Figure 1: Real Time and Final Output Gap Measures OvN.

Figure 2a: SBC Weights (10 period moving average)

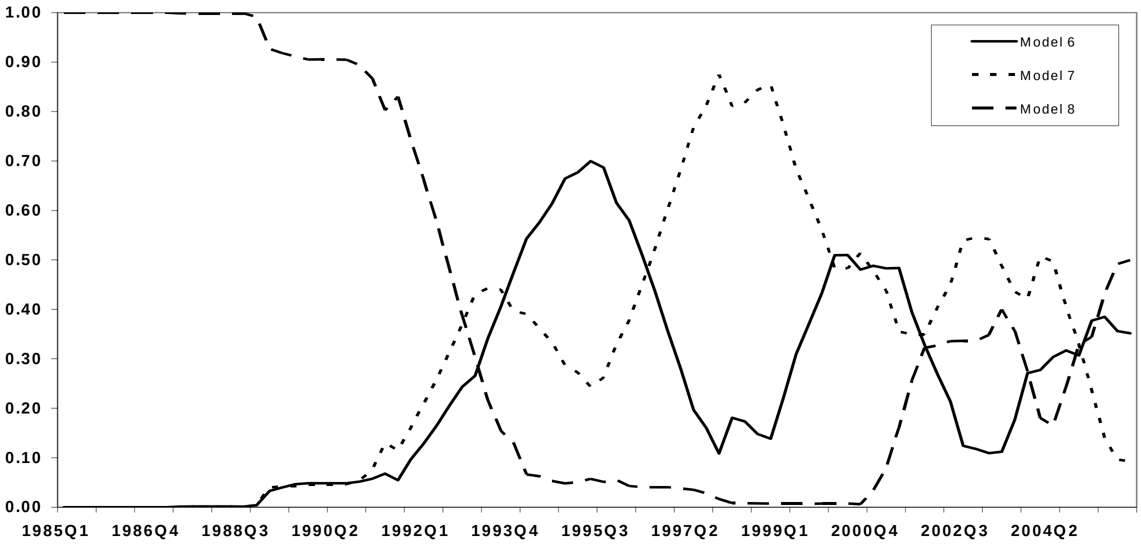


Figure 2b: AIC Weights (10 period moving average)

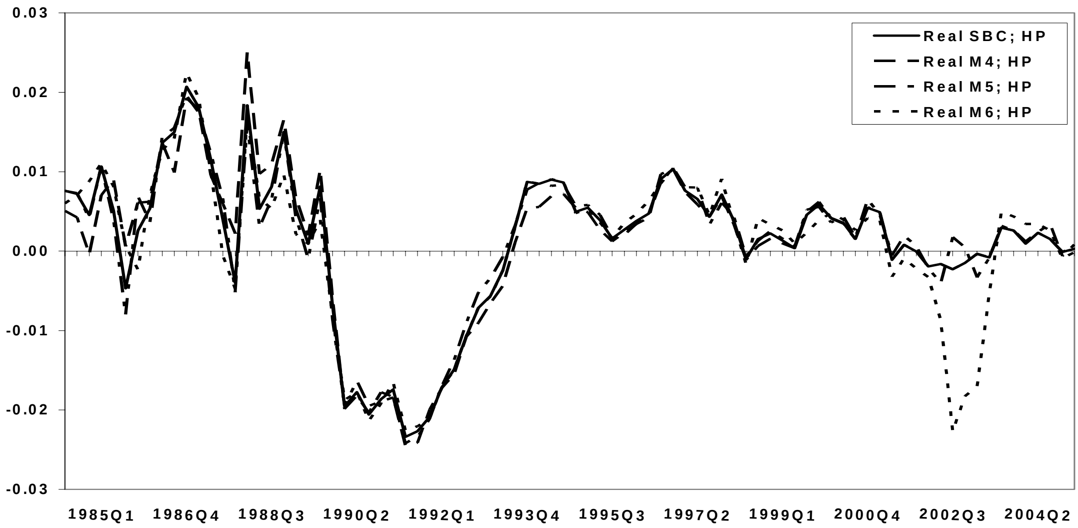


Figure 3a: Real Time HP Gap Measures Based on Alternative Multivariate Models.

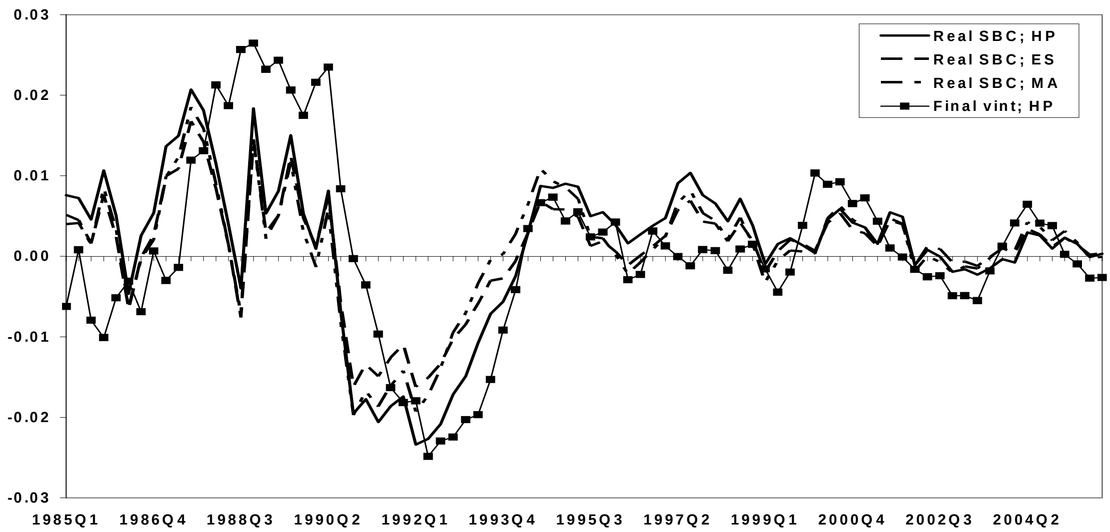


Figure 3b: Final Vintage HP Gap Measures and Real Time Gap Measures Based on SBC Average Using Alternative Filters.

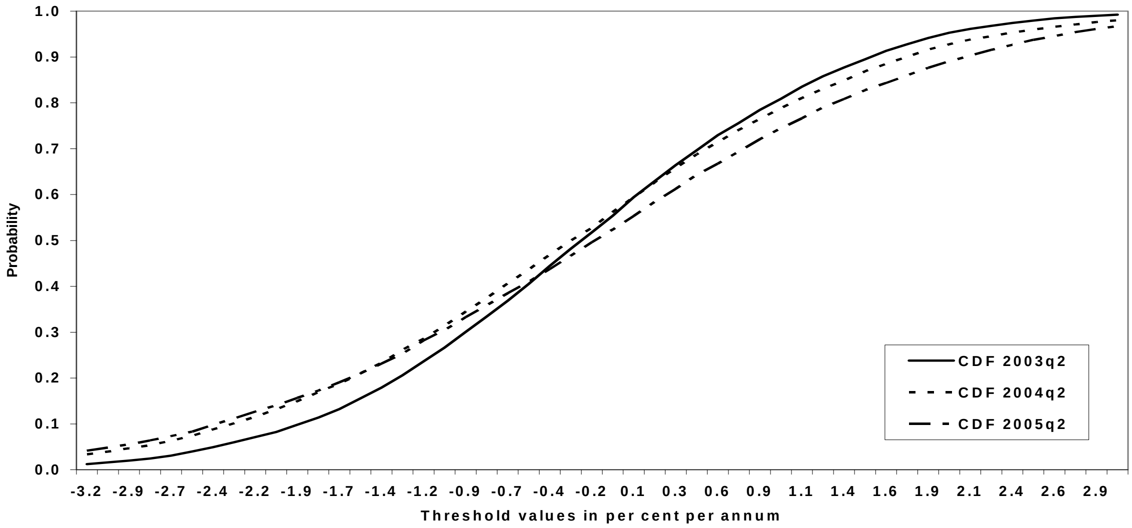


Figure 4: Cumulative Density Functions for Output Gap Measures Calculated in 2003q2.

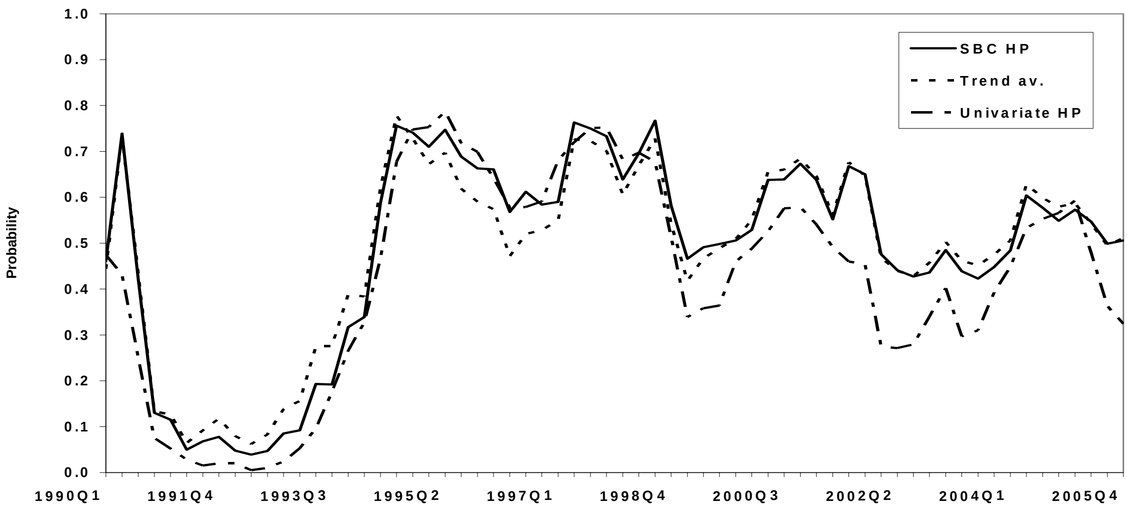


Figure 5: Real Time Probability Forecasts of 'Avoiding a Recession'.