The GDP Fan Charts: An Empirical Evaluation

Kevin Dowd
The GDP Fan Charts: An Empirical Evaluation

By

Kevin Dowd*

Abstract

This paper evaluates the probability density forecasts reflected in the Bank of England’s real GDP growth fan charts. Evaluation is carried out using tests that allow for data dependence and using two GDP growth estimates. Results suggest there are problems with the shorter horizon forecasts, but medium-horizon forecasts perform adequately.

JEL classification numbers: C4, N1

Key words: GDP forecasting, density forecasting, fan charts

Revised, May 22nd, 2007

An updated version of this paper is forthcoming in the National Institute Economic Review, January 2008

* Centre for Risk and Insurance Studies, Nottingham University Business School, Jubilee Campus, Nottingham NG8 1BB, UK; phone +44 115 846 6682, fax +44 115 846 6667, email: Kevin.Dowd@Nottingham.ac.uk. He thanks two anonymous referees, Ken Wallis and Martin Weale for helpful feedback, and the ESRC for support under grant RES-000-27-0014. The usual caveat applies.
1. INTRODUCTION

Since 1996, the Bank of England has been publishing ‘fan charts’ in its quarterly Inflation Report. These represent forecasts of probability density functions for a chosen macroeconomic variable – which might be inflation or real GDP growth – and show the Bank’s forecasts of the most likely outcome surrounded by forecasts of prediction intervals at various probability levels. Each interval is shaded, with the 10%-interval darkest and the shading becoming lighter as we move to broader intervals. Forecasts are given for horizons ranging from the current quarter up to 8 quarters ahead, and typically ‘fan out’ and become more dispersed as the horizon increases.

GDP fan charts have appeared in each Inflation Report since November 1997.\(^1\) Given the period that has elapsed since, it is natural to ask how good the fan-chart forecasts have turned out to be, and a number of recent papers have asked this question of the better-known inflation fan charts (see, e.g., Wallis (2003, 2004), Clements (2004), Cogley et alia (2005), Dowd (2004, 2006) and Elder et alia (2005)).\(^2\) However, little

\(^1\) The Bank publishes two fan charts for each variable: these are based on the alternative assumptions that short-term market interest rates will remain constant or follow market expectations over the forecast horizon. This paper uses data from the constant-rate version, but we get similar results with the other.

\(^2\) A number of these studies have reported that the inflation fan charts performed well over very short forecast horizons, but performed poorly over longer horizons. However, Elder et
attention has so far been paid to the GDP fan charts.³ This paper examines their performance by applying a number of forecast evaluation procedures to them. In doing so, a major complication is the need to carry out tests that take account of dependence in the data, a problem that also arises when evaluating the inflation fan charts. However, evaluation of the GDP charts is more complicated because realised GDP growth rate is never ‘observed’ in the same way in which we ‘observe’ an inflation index.⁴ This requires us to use estimates for realised GDP growth, and this raises the issue of the robustness of our results to the estimates used.

2. THE REAL GDP FAN CHARTS

³ The major exception is Elder et alia (2005). They apply Kolmogorov-Smirnov (KS) and Berkowitz-LR tests, and this latter test was carried out under the assumption that the relevant data follow an AR(1) process. However, the former is not appropriate because the data are not predicted to be independent; the latter may be reasonable, but the AR(1) assumption is arbitrary. The present paper carries out tests suitable to dependent data, identifies the best-fitting ARMA processes, and checks the robustness of test results to the fitted processes. It also addresses the issue of the ‘unobservability’ of real GDP growth by carrying out tests on alternative real GDP estimates.

⁴ Strictly speaking, one might argue that real GDP growth is ‘observed’, but only some time after the event. However, what matters here is that real GDP growth is not observed in real-time.
The Bank’s fan charts are based on an assumption that real GDP growth obeys a two-piece normal (2PN) density function. The 2PN pdf is usually defined as:

\[
\begin{align*}
  f(x; \mu, \sigma_1, \sigma_2) &= \begin{cases} 
    C \exp \left( -\frac{1}{2\sigma_1^2} (x - \mu)^2 \right), & x \leq \mu \\
    C \exp \left( -\frac{1}{2\sigma_2^2} (x - \mu)^2 \right), & x \geq \mu 
  \end{cases}
\end{align*}
\]  

(1)

where \( C = k(\sigma_1 + \sigma_2)^{-1} \), \( k = \sqrt{2/\pi} \) and \( \mu \) is the mode (see, e.g., John (1982) or Wallis (2004)). The distribution takes the lower half of a normal distribution with parameters \( \mu \) and \( \sigma_1 \), and the upper half of a normal with parameters \( \mu \) and \( \sigma_2 \). These halves are scaled to give the same mode value, and the distribution is negatively (positively) skewed if \( \sigma_1 > \sigma_2 \) (\( \sigma_1 < \sigma_2 \)).

However, the Bank uses a 2PN based on an alternative parameterisation:

\[
\begin{align*}
  f(x; \mu, \gamma, \sigma) &= \begin{cases} 
    C \exp \left( -\frac{1}{2\sigma^2} (1 + \gamma)(x - \mu)^2 \right), & x \leq \mu \\
    C \exp \left( -\frac{1}{2\sigma^2} (1 - \gamma)(x - \mu)^2 \right), & x \geq \mu 
  \end{cases}
\end{align*}
\]  

(2)
where $-1 < \gamma < 1$, and $\gamma$ and $\sigma$ are related to $\sigma_1$ and $\sigma_2$ via:

\begin{align}
(1 + \gamma)\sigma_1^2 &= \sigma^2 \\
(1 - \gamma)\sigma_2^2 &= \sigma^2
\end{align}

For each horizon, the Bank publishes forecasts of the mean, median and mode ($\mu$), a skew parameter (which is not $\gamma$, but the difference between the mean and $\mu$) and uncertainty $\sigma$. This uncertainty parameter is not the same as the standard deviation, except where the skew is zero. Once the MPC specifies the values of $\mu$, $\sigma$ and the skew parameter for each horizon, the model is complete and the density forecasts can be ascertained from it.

As an example, Figure 1 reports the May 2003 GDP fan chart. This chart gives growth prediction intervals for each quarter from 03Q2 to 05Q3. Note that growth is measured as the growth rate of real GDP relative to real GDP four quarters previously; it is not the quarterly rate of growth expressed as an annualised percentage. Given that real GDP growth at the end of 03Q1 was about 2.3%, the 10% interval fans out to the range [2.35%, 2.65%] by the end of the horizon period; the 20% prediction interval fans

---

5 The parameter forecast data are downloaded from the Bank of England website at http://www.bankofengland.co.uk/inflationreport/gdpinternet.xls.

6 Thus, $\mu$ and $\sigma$ are given by the Bank, but the value of $\gamma$ needs to be derived from these two parameters and the skew parameter. Details of how this can be done are given by Wallis (2004).
out to [2.20%, 2.80%], and so forth, with the ranges gradually increasing and the 90% interval fanning out to the range [0.55%, 4.45%].

Figure 1: The May 2003 GDP Growth Fan Chart

3. A FRAMEWORK TO TEST DENSITY FORECASTS

Let \( x_t \) be the realised value of real GDP growth for quarter \( t \), and assume for the moment that \( x_t \) is observable. Each realised value is to be compared against the relevant forecasted density over each forecast horizon. Now let \( p_{t,k} \) be the value of \( x_t \) mapped to its value on the \( k \)-period-ahead forecast cdf, where \( k = 0, 1, \ldots, 8 \). This mapping is known as the Probability Integral Transform (PIT).

Under the null hypothesis that the model is adequate, \( p_{t,k} \) should be uniformly distributed over the interval \([0,1]\), so we can evaluate the forecasts by comparing the empirical distribution of \( p_{t,k} \) against the predicted uniform distribution.

However, we cannot test the null by a naïve application of standard uniformity tests (e.g., KS tests), because these presuppose that the
$p_{t,k}$ are independent, and there are two reasons why we cannot make this assumption here. First, since real GDP growth is measured as the rate of growth of real GDP over the past 4 quarters, successive growth rates thus measured will be rolling moving averages of quarterly growth rates and lack independence. But even if quarterly GDP growth was itself independent – and correlogram analysis suggests it is not – the $p_{t,k}$ would not be, except where $k=0$. For $k>0$, successive values of $p_{t,k}$ would share common random factors (i.e., the quarterly growth rates), and these create further dependence in the $p_{t,k}$. We therefore need a testing framework that allows for $p_{t,k}$ to be dependent.

Before addressing this problem, it is convenient if we follow Berkowitz (2001) and run the $p_{t,k}$ through a standard normal inverse function, viz.:

$$z_{t,k} = \Phi^{-1}(p_{t,k})$$

Under the null, these Berkowitz-transformed $z_{t,k}$ observations should be distributed as standard normal. However, they are not predicted to be iid (except for $k=0$), because lack of iid-ness in the $p_{t,k}$ implies lack of iid-ness in the $z_{t,k}$. 

7
We now face the problem of testing the $z_{t,k}$ for standard normality in a context where they follow an unknown dependence structure. Furthermore, because sample sizes are small, it would be unwise to rely on tests derived from large-sample theory.\footnote{These are demanding requirements. The first rules out most of the standard textbook tests, and the second rules out the more recently developed tests that can accommodate dependence, parameter risk, etc. For a survey of these, see, e.g., Corradi and Swanson (2006).}

One way forward is to postulate that $z_{t,k}$ follows an ARMA process and use Box-Jenkins analysis to obtain a parsimonious fit. For example, we might find that the dependence can be adequately represented by a first order autoregression (AR(1)):

\begin{equation}
    z_{t,k} = \mu_k + \rho_k z_{t-1,k} + \epsilon_{t,k}, \quad |\rho_k| < 1
\end{equation}

where the errors $\epsilon_{t,k}$ are iid normal, the intercept parameters $\mu_k$ are predicted to be 0 and the autoregression parameters $\rho_k$ are not expected to be 0 except where $k=0$.\footnote{AR(1) processes are assumed by Dowd (2004) and Elder et alia (2005) in their fan chart studies. However, the present paper allows for more general ARMA processes and derives the best-fits.} Alternatively, we might fit an MA\footnote{The possibility of MA features is suggested by the fact that successive values of $z_{t,k}$ share common factors and by the fact that real GDP growth is taken as a four-quarter average of quarterly growth rates. However, these considerations do not guarantee that the} or a more general ARMA process.\footnote{8}
We can now test the mean and variance predictions using the following Monte Carlo procedure. For each value of $k$:

- We obtain the PITs and put these through the Berkowitz transformation (4) to obtain a $z_{t,k}$ sample. Denote this by $\hat{z}_{t,k}$.
- We fit a parsimonious ARMA process (i.e., (5) or its higher-order AR, MA or ARMA equivalent) to the $\hat{z}_{t,k}$, taking care to ensure that the residuals appear to be independent.
- We use the fitted ARMA process to simulate a large number $m$ of possible $\hat{z}_{t,k}$ series, each of which has the dependence structure of the fitted ARMA process, and let us denote the $i^{th}$ such series as $\hat{z}_{i,k}$.

$z_{t,k}$ process will be a ‘pure’ MA because we do not know the dependence structure of the quarterly growth rates.

Whatever process we fit, the best we can do is to aim for a parsimonious approximation to it. It is therefore important to check that the residuals from our fitted process appear to be independent, and to carry out checks of the robustness of our main results to any fitted structure.

Each such series is constructed as follows: we set suitable initial values for time 0 parameters, simulate a value of $\hat{z}_{1,k}$ from the appropriate normal distribution and use the fitted ARMA process to obtain the corresponding simulated value of $\hat{z}_{1,k}$. We then simulate a value of $\hat{e}_{2,k}$ from the same distribution, and use the fitted ARMA process to obtain a simulated value of $\hat{z}_{2,k}$; and then repeat again and again until we have a complete simulated $\hat{z}_{i,k}$ path. So, for example, if the ARMA process is an AR(1), we would simulate $\hat{z}_{1,k}$ using an estimate of (5) with a zero mean, i.e., using $\hat{z}_{1,k} = \hat{\rho}_k \hat{z}_{0,k} + \hat{e}_{1,k}$, where $\hat{\rho}_k$ is our estimate of $\rho_k$. To do so, we set the initial value for $\hat{z}_{0,k}$ equal to the unconditional expected value of $z_{1,k}$, i.e., 0, simulate a value of $\hat{e}_{1,k}$ from a normal distribution with mean 0 and variance $1 - \hat{\rho}_k^2$, and then obtain $\hat{z}_{1,k} = \hat{\rho}_k \hat{z}_{0,k} + \hat{e}_{1,k}$. We then simulate a value of $\hat{e}_{2,k}$ from the same normal distribution and obtain $\hat{z}_{2,k}$ from $\hat{z}_{2,k} = \hat{\rho}_k \hat{z}_{1,k} + \hat{e}_{2,k}$, and proceed in the same way to obtain $\hat{z}_{3,k}$, $\hat{z}_{4,k}$, etc. Simulating the
• We estimate the values of the mean and variance of each \( z_{i,k} \) series.\(^{12}\) This gives us a ‘sample’ of \( m \) mean values and a ‘sample’ of \( m \)

variance values. The \( m \times 0.025^{\text{th}} \) and \( m \times 0.975^{\text{th}} \) highest means then
give us the 95% confidence interval for the mean test statistic, and the \( m \times 0.025^{\text{th}} \) and \( m \times 0.975^{\text{th}} \) highest variances give us the 95%

confidence interval for the variance test statistic, under the null hypothesis that \( z_{i,k} \) is standard normal but has the dependence

structure of the fitted ARMA process.

• We estimate the values of the mean and variance of our ‘real’ sample

\( \hat{z}_{i,k} \); for each of these two statistics, the forecasts pass the test if the sample value lies within the estimated confidence interval, and

otherwise fail.\(^ {13}\)

---

\(^{12}\) We focus on the mean and variance predictions because the ARMA framework assumes that the \( \varepsilon_{i,k} \) are normal, and this undermines any rationale for using it to test for departures from normality, i.e., skewness and excess kurtosis.

\(^{13}\) The tests suggested in the text have the attractions that they take account of the dependence structure of the data and suffer from no discernable small sample problems. However, two alternatives should be noted. (1) We could decompose the \( \hat{z}_{i,k} \) sample into bins and carry out a textbook chi-squared test of whether observed frequencies within each bin are sufficiently close to their predicted values. I preferred not to use this test because it does not take account of the \( \hat{z}_{i,k} \) dependence structure (and the simulation-based tests used in this paper do take account of it) and because of doubts about its small sample properties when \( \hat{z}_{i,k} \) is dependent: in particular, if \( \hat{z}_{i,k} \) is dependent and the sample is small, then observations are likely to be more clustered around the initial starting value than they ‘should’ be under the independence assumption on which the chi-squared test is predicated. Applying a chi-squared test therefore fails to allow for this clustering. (2) Another possible approach is an ‘iid resample’ method recently proposed by Dowd (2007): this test makes use of a bootstrap algorithm which chooses resamples that are iid by construction. Using
4. DATA ISSUES

Testing the GDP fan charts requires quarterly observations, or estimates, of real GDP growth between the quarter in question and the comparable quarter a year before. Ideally, we would like a series that is accurate and timely, in the sense that it was also available to the MPC at or close to the time when the forecasts become due. The need for an accurate series is obvious, but the timeliness criterion is also important to avoid anachronism.\textsuperscript{14} However, any available series involves a trade-off between these two desirable characteristics.

From the accuracy point of view, the most natural series is the latest 4-quarter growth (IHYR) series produced by National Statistics. However, this series is subject to revisions, and these can be substantial.

---

this bootstrap would allow one to apply standard iid-based tests to the resamples drawn from the original sample. However, the ‘cost’ of this approach is loss of power, and tests based on this approach turn out to have very little power in samples as small as the ones available to us here.

\textsuperscript{14} For example, it might be that the MPC has been forecasting well using the data it had at the time, and some researcher comes to different conclusions years later using a revised GDP series. The researcher might be using a better series, but the relevance of the exercise would be doubtful: this would be akin to criticizing the builders of the pyramids because they didn’t use hydraulic cranes that were invented long afterwards.
ones made years later. Hence, there is no a priori guide over which vintage of IHYR series to use: the most accurate can appear years later and have no timeliness. On the other hand, we can also use timely series, but these may be inaccurate, because they cannot take account of revisions made afterwards. Given that no series meets both criteria, this study uses two alternatives:

- The first is the latest available IHYR series, as of August 3, 2006. This has the advantage of being the most accurate series available at the time of writing, but its disadvantage is its lack of timeliness.

- The second is the MPC’s ‘best estimate’ each quarter, using the information then available to it, of that quarter’s year-on-year real GDP growth rate. This series has the advantage of being very timely, but its disadvantage is that it can take no account of later revisions to the GDP growth rate.

---

15 The first preliminary estimates are made in the quarter concerned. These are followed by the first release, seven weeks after the end of a quarter, the National Accounts quarterly series 12 weeks after the end of a quarter, and then the various Blue Book and Post Blue Book estimates. To add to which, past data are sometimes periodically revised later (e.g., to incorporate the change to the European System of Accounts (ESA95) in 1998, to incorporate changes in chain-weighting, and so on). These changes have led to notable increases in estimates of real GDP growth over the period from late 1998 to late 2001. For more on these issues, see, e.g., Akritidis (2003) and Elder et alia (2005).


17 This series is obtained from the fan chart parameters as the current-period forecast of the real-growth mode for that quarter, the mode being that of the constant-rate fan chart.

18 However, when using this series as a proxy for GDP growth in some quarter, we can only use it to evaluate forecasts made in earlier periods: we cannot take the MPC contemporary forecast as a proxy for realised GDP growth and then use this proxy to
So the first series is good on accuracy, but poor on timeliness, whilst the second series is good on timeliness, but poor on accuracy, i.e., the two series are polar opposites on the accuracy and timeliness criteria. They also make for a good comparison for another reason: a priori, we might expect the first series to produce results that are biased against the Bank’s model (because the Bank cannot anticipate later revisions to the GDP series), and we might expect the second series to be biased in favour of the Bank’s model (because this series is itself a set of forecasts from the model being tested).19

Figure 2 provides plots of these series over our sample period. The series exhibit roughly similar shapes, but the latest National Statistics (NS) series shows greater economic growth over the earlier part of our period; this reflects subsequent changes which have revised growth upwards.

Table 1 provides some summary statistics. The NS series has a higher mean (as we would expect). Both series have similar variances, differ

---

19 The Bank has done work on the construction of real-time databases for real GDP, and a collection of different ‘vintages’ of data is available on its website (at www.213.225.136.206/statistics/gdpdatabase/index.htm). Such a dataset would in principle allow a more thorough comparison of different real GDP growth series, but is of limited use for our purposes because it ends in 01Q4. For more details, see Castle and Ellis (2002).
somewhat in their extremes, and are positively (but not highly positively) correlated.

\textbf{Insert Table 1}

The other data used are the fan chart parameter forecasts. These consist of 35 sets of parameter forecasts, one for each published chart. Each set consists of 9 values for each of the mode, skew and uncertainty parameters, for \( k=0,1,\ldots,8 \). As explained in section 2, we use these to obtain the density forecasts, and this gives us 9 sets of density forecasts for each quarter, for horizons ranging from \( k=0 \) to \( k=8 \).

5. RESULTS

\textit{Preliminary results}\textsuperscript{20}

We first report the results of some preliminary analysis. Figure 3 shows plots of the predicted and empirical \( p_{t,k} \) for each horizon. Under the null, we would expect the plots to be ‘close’ to the 45\(^\circ\) line. The notes also show

\textsuperscript{20} Calculations were carried out using specially written MATLAB functions, which are available on request.
the least-squares slopes of the empirical \( p_{t,k} \) series, which should be close to 1 under the null. However, most plots are not ‘close’ to these expectations, but give the impression that plots are better for medium- and long-horizon forecasts. For their part, the estimated slopes suggest that performance is best for medium-horizon forecasts.

**Insert Figure 3 here**

Summary statistics for the \( \hat{z}_{k,t} \) sample moments are shown in Table 2. The sample values tend not to be ‘close’ to the predicted values, and the low-horizon forecasts tend to perform more poorly than the others. The variances also clearly suggest that the better performing forecasts are the medium-horizon ones. However, the two GDP series conflict in other ways (e.g., the NS means suggest that performance steadily improves as the horizon increases, whereas the MPC means do not).

**Insert Table 2 here**

*Test results*

Table 3 shows the ‘best fitting’ ARMA processes, their estimated parameters and their standard errors, and P-values of a portmanteau statistic of the ARMA residuals. The best-fitting process is an AR(1) process if we
use the NS series, and is usually (but not always) an MA(1) if we use the
MPC series. (The two exceptions are an iid process for \( k=1 \) and an AR(1)
for \( k=8 \).) The P-values for the portmanteau statistic tend to suggest that the
residuals have no significant dependence structure, and confirm the
goodness of the fits. The Table also shows the P-values of the mean and
variance tests using the Monte-Carlo procedure outlined earlier.

**Insert Table 3 here**

In interpreting these results, we might note that under the null
hypothesis, the \( \hat{z}_{0,j} \) (available only for the NS GDP proxy) is predicted to be
iid \( \text{N}(0,1) \), and this prediction is rejected: the AR(1) parameter is significant
(which rejects the iid prediction) and the P-values of the mean and variance
tests clearly reject the ‘standard’ aspect of the standard normality
prediction. As for the predictions as they apply to other \( \hat{z}_{k,j} \) series, the
forecasts easily pass the mean tests, but often have difficulty with the
variance ones. More precisely, the NS series fail the variance prediction
over very low horizons and the MPC series always (strongly) fails the
variance prediction over all horizons. Comparing results across the two
GDP estimates: if we use the GDP proxy, the model performs adequately

---

21 Standard textbook tests of these predictions are also applicable and give similar results.
except for very short horizons; and if we use the MPC proxy, the model performs poorly.

These results are open to error in the event of a misspecified dependence structure, so it is important to check their robustness. Accordingly, Table 4 reports the P-values for these two sets of tests under each of three possible alternative dependence structures applied to all $\hat{z}_{k,t}$: no dependence (i.e., iid), AR(1) and MA(1). We are not suggesting that these provide good fits in each case: we merely postulate them here to assess the robustness of our earlier ‘best-fit’ results to changes in the assumed dependence structure.

![Insert Table 4 here](image)

The results in Table 4 suggest that the poor performance of the NS $\hat{z}_{0,t}$ is fairly robust: these forecasts fail 5 of the 6 tests. These results also confirm that there are problems with low-horizon NS forecasts. However, for the most part, they also confirm that the medium- and long-horizon NS forecasts perform adequately.22 Turning to the MPC forecasts, these clearly and robustly perform well when evaluated with the mean test.

---

22 The exceptions relate to the variance test results for the MA(1), which are poor for horizons of 4 quarters ahead or more. However, one would be inclined not to put too much weight on these given the other results and given that these forecasts perform well under our ‘best-fitting’ AR(1) dependence structures.
And, for their part, the variance tests applied to the MPC forecasts suggest a robust failing of at least the $k=1$ forecasts and some evidence against other short-horizon forecasts, but paint a mixed picture for longer-horizon forecasts.

6. CONCLUSIONS

Any overall assessment of these results is a matter of judgement. However, I would summarise the results as suggesting that there is strong evidence against at least some of the low-horizon forecasts, there is not-so-strong evidence against some of the long-horizon forecasts and there is relatively weak evidence against the medium-term forecasts. These suggest that a supporter of the Bank’s forecasts would find it difficult to mount a convincing defence of the performance of the short-horizon forecasts, but a critic would find it difficult to make a compelling case against the performance of the medium-horizon forecasts – and the performance of the long-horizon forecasts is less clear-cut one way or the other.\(^{23}\)

---

\(^{23}\) If the Bank wish to facilitate external scrutiny of its forecasts, it might consider establishing its own ‘official’ GDP growth series for the purposes of validating these forecasts. Such a series would also be useful for the Bank itself when it evaluates its own forecasts. Using such a series would make evaluation more transparent and verifiable, and
remove the ‘smoke’ involved with trying to evaluate forecasts of an unobserved variable. A possible methodology that could be used to do this is suggested by Ashley et alia (2005).
REFERENCES


Figure 1: The May 2003 GDP Fan Chart

Notes: The Figure shows previous realised values and predicted intervals for GDP growth, measured as the percentage increase in GDP over the previous 12 months, for the constant rate model. The realised GDP series differs slightly from the one in the May 2003 Inflation Report because of revisions to the GDP growth series. The Figure is reproduced with permission from the Bank of England.
Figure 2: Alternative Measures of Real GDP Growth

Notes: The National Statistics headline series is the latest available data for series IHYR, and the contemporary MPC estimates are the MPC’s mode forecasts of contemporary real GDP growth. Both series refer to growth of real GDP over the previous four quarters, and the sample period is 97Q4 to 06Q2.
Notes: Based on the real GDP series in Figure 2 and Bank of England real GDP fan chart forecasts for 97Q4 to 06Q2. The continuous lines refer to PITs for the NS GDP series and the ‘dashed-dot’ lines refer to those of the MPC series as defined earlier. The least-squares estimates of the slopes for the NS series are 1.1, 1.1, 1.1, 1.1, 0.97, 0.92, 0.84, 0.80 and 0.75, going from the 0-quarter-ahead to 8-quarter-ahead forecasts; for the MPC series, they are 0.67, 0.83, 0.84, 0.86, 0.82, 0.78, 0.74 and 0.67 going from 1-quarter to 8-quarter-ahead forecasts.
TABLES

Table 1: Summary Statistics for Different Real GDP Growth Series

<table>
<thead>
<tr>
<th>Summary statistic</th>
<th>Latest NS series</th>
<th>MPC series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.786</td>
<td>2.349</td>
</tr>
<tr>
<td>Variance</td>
<td>0.512</td>
<td>0.537</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.600</td>
<td>0.790</td>
</tr>
<tr>
<td>Maximum</td>
<td>4.300</td>
<td>4.020</td>
</tr>
<tr>
<td>N</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td></td>
<td>0.438</td>
</tr>
<tr>
<td>Rank correlation</td>
<td></td>
<td>0.441</td>
</tr>
</tbody>
</table>

Notes: The headline series is the latest available data for National Statistics’ series IHYR; and the contemporary MPC estimates are the MPC’s ‘best estimates’ of contemporary real GDP growth, as given in the Bank’s GDP growth fan chart parameters. Both series refer to growth of real GDP over the previous four quarters, and there are 35 observations of each in our sample period of 98Q4 to 06Q2.
Table 2: Summary Statistics for Standard Normal Inverses of $\hat{p}_{k,p}$ Series

(a) Using latest National Statistics real GDP growth series

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k=0$</th>
<th>$k=1$</th>
<th>$k=2$</th>
<th>$k=3$</th>
<th>$k=4$</th>
<th>$k=5$</th>
<th>$k=6$</th>
<th>$k=7$</th>
<th>$k=8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.7495</td>
<td>0.6228</td>
<td>0.4603</td>
<td>0.3163</td>
<td>0.2532</td>
<td>0.1790</td>
<td>0.1455</td>
<td>0.1274</td>
<td>0.0708</td>
</tr>
<tr>
<td>Variance</td>
<td>2.2028</td>
<td>1.8012</td>
<td>1.5563</td>
<td>1.3609</td>
<td>1.0878</td>
<td>0.8280</td>
<td>0.6087</td>
<td>0.4873</td>
<td>0.3777</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.6831</td>
<td>-0.3196</td>
<td>-0.1950</td>
<td>-0.1768</td>
<td>0.0631</td>
<td>0.2597</td>
<td>0.1392</td>
<td>0.0373</td>
<td>0.0356</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.3962</td>
<td>1.8945</td>
<td>2.0552</td>
<td>2.4895</td>
<td>2.9123</td>
<td>2.7034</td>
<td>2.7925</td>
<td>2.3423</td>
<td>1.6434</td>
</tr>
<tr>
<td>$N$</td>
<td>35</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>31</td>
<td>30</td>
<td>29</td>
<td>28</td>
<td>27</td>
</tr>
</tbody>
</table>

(b) Using MPC contemporaneous mode real GDP growth forecasts

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k=0$</th>
<th>$k=1$</th>
<th>$k=2$</th>
<th>$k=3$</th>
<th>$k=4$</th>
<th>$k=5$</th>
<th>$k=6$</th>
<th>$k=7$</th>
<th>$k=8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>NA</td>
<td>0.0170</td>
<td>-0.0423</td>
<td>-0.1059</td>
<td>-0.0940</td>
<td>-0.1368</td>
<td>-0.1268</td>
<td>-0.1009</td>
<td>-0.1084</td>
</tr>
<tr>
<td>variance</td>
<td>NA</td>
<td>0.3636</td>
<td>0.5618</td>
<td>0.6489</td>
<td>0.6020</td>
<td>0.5135</td>
<td>0.4516</td>
<td>0.3892</td>
<td>0.3040</td>
</tr>
<tr>
<td>skewness</td>
<td>NA</td>
<td>-0.6005</td>
<td>0.2439</td>
<td>0.4675</td>
<td>0.3846</td>
<td>0.4136</td>
<td>0.3212</td>
<td>-0.1874</td>
<td>0.0545</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>NA</td>
<td>3.7493</td>
<td>2.7303</td>
<td>3.3678</td>
<td>2.4558</td>
<td>2.3103</td>
<td>2.3094</td>
<td>2.0328</td>
<td>2.2555</td>
</tr>
<tr>
<td>$n$</td>
<td>35</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>31</td>
<td>30</td>
<td>29</td>
<td>28</td>
<td>27</td>
</tr>
</tbody>
</table>

Notes: As per notes to Table 1. Results refer to sample parameters of PIT series put through standard normal inverse transformations, where PITs are obtained using Bank of England real GDP growth fan chart forecasts over $k$ quarters ahead, with sample sizes $n$. Under the null we would expect the mean and skewnesses to be 0, the variances to be 1, and the kurtoses to be 3.
Table 3: Results for Mean and Variance Tests Based on ‘Best Fitting’
ARMA Representations of the $\hat{z}_{k,l}$ Dependence Structure

<table>
<thead>
<tr>
<th>Horizon</th>
<th>‘Best fit’</th>
<th>Parameters</th>
<th>Q-stat</th>
<th>Test stat PVs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>const. AR(1) MA(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>AR(1)</td>
<td>0.905 (0.428) 0.516 (0.139) NA</td>
<td>0.359</td>
<td>0.0050** 0.0006**</td>
</tr>
<tr>
<td>1</td>
<td>AR(1)</td>
<td>0.693 (0.656) 0.768 (0.114) NA</td>
<td>0.116</td>
<td>0.0780 0.0165*</td>
</tr>
<tr>
<td>2</td>
<td>AR(1)</td>
<td>0.447 (0.664) 0.789 (0.112) NA</td>
<td>0.321</td>
<td>0.1540 0.0368*</td>
</tr>
<tr>
<td>3</td>
<td>AR(1)</td>
<td>0.079 (0.613) 0.794 (0.106) NA</td>
<td>0.370</td>
<td>0.2516 0.0650</td>
</tr>
<tr>
<td>4</td>
<td>AR(1)</td>
<td>-0.075 (0.511) 0.795 (0.095) NA</td>
<td>0.080</td>
<td>0.2979 0.1497</td>
</tr>
<tr>
<td>5</td>
<td>AR(1)</td>
<td>-0.035 (0.505) 0.804 (0.106) NA</td>
<td>0.062</td>
<td>0.3585 0.3035</td>
</tr>
<tr>
<td>6</td>
<td>AR(1)</td>
<td>0.005 (0.420) 0.781 (0.116) NA</td>
<td>0.101</td>
<td>0.3806 0.4310</td>
</tr>
<tr>
<td>7</td>
<td>AR(1)</td>
<td>0.076 (0.356) 0.748 (0.129) NA</td>
<td>0.136</td>
<td>0.3887 0.2166</td>
</tr>
<tr>
<td>8</td>
<td>AR(1)</td>
<td>0.012 (0.316) 0.747 (0.131) NA</td>
<td>0.604</td>
<td>0.4411 0.0933</td>
</tr>
</tbody>
</table>

(a) Using latest National Statistics real GDP growth series

(b) Using MPC contemporaneous mode real GDP growth forecasts
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>MA(1)</td>
<td>-0.091</td>
<td>NA</td>
<td>0.604</td>
<td>0.365</td>
<td>0.3517</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.156)</td>
<td></td>
<td>(0.148)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>AR(1)</td>
<td>-0.083</td>
<td>0.618</td>
<td>NA</td>
<td>0.507</td>
<td>0.3871</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.235)</td>
<td></td>
<td>(0.165)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: As per earlier Tables. For columns 3-5, the Table gives estimated parameter values obtained using EViews 5 followed by their estimated standard errors in brackets. Column 5 gives the P-value for the portmanteau (or Q-stat) for the residuals taken up to 3 lags. ‘PV’ refers to ‘P-value’. P-values are calculated using the relevant fitted process using 20000 simulation trials. * indicates significance at 5% level, ** indicates significance at 1% level.
Table 4: P-values for Mean and Variance Tests Based on Alternative Assumed ARMA Representations of the $\hat{z}_{t,\tau}$ Dependence Structure

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Assuming iid</th>
<th>Assuming AR(1)</th>
<th>Assuming MA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean test</td>
<td>Variance test</td>
<td>Mean test</td>
</tr>
<tr>
<td>(a) Using latest National Statistics real GDP growth series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0**</td>
<td>0.0001**</td>
<td>0.0050**</td>
</tr>
<tr>
<td>1</td>
<td>0.0001**</td>
<td>0.0032**</td>
<td>0.0780</td>
</tr>
<tr>
<td>2</td>
<td>0.0034**</td>
<td>0.0263*</td>
<td>0.1540</td>
</tr>
<tr>
<td>3</td>
<td>0.0394*</td>
<td>0.0864</td>
<td>0.2516</td>
</tr>
<tr>
<td>4</td>
<td>0.0795</td>
<td>0.3398</td>
<td>0.2979</td>
</tr>
<tr>
<td>5</td>
<td>0.1655</td>
<td>0.2671</td>
<td>0.3585</td>
</tr>
<tr>
<td>6</td>
<td>0.2175</td>
<td>0.0521</td>
<td>0.3806</td>
</tr>
<tr>
<td>7</td>
<td>0.2524</td>
<td>0.0121*</td>
<td>0.3887</td>
</tr>
<tr>
<td>8</td>
<td>0.3557</td>
<td>0.0019**</td>
<td>0.4411</td>
</tr>
<tr>
<td>(b) Using MPC contemporaneous mode real GDP growth forecasts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.4623</td>
<td>0.0002**</td>
<td>0.4651</td>
</tr>
<tr>
<td>2</td>
<td>0.4004</td>
<td>0.0214*</td>
<td>0.4349</td>
</tr>
<tr>
<td>3</td>
<td>0.2745</td>
<td>0.0681</td>
<td>0.3610</td>
</tr>
<tr>
<td>4</td>
<td>0.3016</td>
<td>0.0395</td>
<td>0.3827</td>
</tr>
<tr>
<td>5</td>
<td>0.2301</td>
<td>0.0153</td>
<td>0.3477</td>
</tr>
<tr>
<td>6</td>
<td>0.2491</td>
<td>0.0061</td>
<td>0.3667</td>
</tr>
<tr>
<td>7</td>
<td>0.3006</td>
<td>0.0019</td>
<td>0.3851</td>
</tr>
<tr>
<td>8</td>
<td>0.2881</td>
<td>0.0001</td>
<td>0.3871</td>
</tr>
</tbody>
</table>

Notes: See notes to Table 3.