Why are survey forecasts superior to model forecasts?

Michael P. Clements

No 954

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Why are survey forecasts superior to model forecasts?

Michael P. Clements

Department of Economics

University of Warwick

September 30, 2010

Abstract

We investigate two characteristics of survey forecasts that are shown to contribute to their superiority over purely model-based forecasts. These are that the consensus forecasts incorporate the effects of perceived changes in the long-run outlook, as well as embodying departures from the path toward the long-run expectation. Both characteristics on average tend to enhance forecast accuracy. At the level of the individual forecasts, there is scant evidence that the second characteristic enhances forecast accuracy, and the average accuracy of the individual forecasts can be improved by applying a mechanical correction.

Keywords: consensus forecast, model-based forecasts, long-run expectations. JEL classification: C53, E37.

1 Introduction

Survey expectations are often found to be superior to model-based forecasts, where 'the' survey forecast is often taken to be the median of the individual respondents' forecasts. For example, Ang, Bekaert and Wei (2007) show that surveys outperform other methods for forecasting annual inflation one-year ahead.[1] Ang *et al.* (2007, p.1207) attribute this as being likely due to a combination of 'the pooling of large amounts of information; the efficient aggregation of that information; and the ability to quickly adapt to major changes in the economic environment such as the great moderation.' However, the quotation from Ang *et al.* (2007) is rather too general and would appear to be true almost by definition. In this paper we wish to discover what are the specific characteristics of survey forecasts that account for their relative superiority. Hence we begin with the median or consensus forecast, and then consider the extent to which the characteristics of these forecasts, which enhance accuracy, are also a characteristic of the individual forecasts. Or is it that the aggregation *per se* is instrumental in delivering the greater accuracy?

Two related ideas underpin the empirical analysis. Model-based forecasts are backward-looking in that they project forward past patterns in the data. One aspect of this concerns the mean of the model. It is well known that model-based forecasts will approach the unconditional mean of the model as the forecast horizon increases, assuming that the model is of the equilibrium-correction class, see e.g., Clements and Hendry (2006), who also establish the generality of this class of model. Clements and Hendry (2006) argue that this property is one of the main causes of forecast failure when there is a structural break, because a model's forecasts subsequently 'correct' towards a mean that is no longer appropriate. This is part of the failure of the model to 'quickly adapt' to the changed circumstances. Survey forecasts might outperform model-based forecasts if, by incorporating a forward-looking element, they were able to foresee changes in the economic environment. This is the idea behind the use of long-run inflation expectations in the forecasting models of Clark and McCracken (2008), for example, where the survey information captures perceived changes in the

[1] The other methods are time-series ARIMA models, Phillips curve models with real-activity variables, and term-structure models.

long-run mean. Without the survey information, their models' forecasts would 'correct' towards the long-run mean that characterised the past data. Hence we are interested in determining whether survey forecasts do incorporate perceived changes in the outlook which contribute to more accurate (shorter-term) forecasts.

Secondly, we document the finding that survey forecasts do not always move monotonically to the long-run expectation as the forecast horizon increases. For example, the forecaster may expect faster growth in the near future before the economy slows to its (perceived) long-run rate of expansion. Of interest is whether departures from the path to the long-run position enhance forecast accuracy, or simply constitute noise: do they contain useful information or they are uninformative about the variable being forecast? To address this issue we exploit the pattern or term structure of forecasts from 1 to h-steps ahead made from a given point in time. We name shorter-run survey forecasts that are off the path of convergence to the long-run forecast 'non-convergent' (henceforth NC), and define below precisely how these are calculated. Whether or not such forecasts enhance accuracy is an empirical question. Forecasters might find it difficult to do better than following the trend. We evaluate whether NC-forecasts enhance accuracy by constructing simple counterfactual forecasts which do not have the NC characteristic, and compare the forecast accuracy of the actual forecasts with that of the counterfactuals. This avoids the pitfall of directly comparing the accuracy of the sets of NC forecasts and non-NC forecasts, which is that more (less) predictable observations might be systematically associated with NC-forecasts.[2]

To further motivate the issues we are interested in, consider figures 1 and 2. These portray the median forecasts of the year-ahead quarter-on-quarter growth rate of real GDP, and the corresponding model-based forecasts (defined in section 2), and the median forecasts of the year-ahead annualised quarter-on-quarter rate of CPI inflation, with the model forecasts. These forecasts, as well as the individual forecast data used in this paper, are taken from the Survey of Professional

[2] For example, a forecast of next quarter might be a NC forecast if the forces expected to result in a blip in inflation next quarter are in train and known (a pre-announced rise in indirect taxation, say). And that quarter's inflation rate may be markedly less uncertain (i.e., easier to predict) than on average. Alternatively, we could imagine cases where NC forecasts are typically of less predictable observations: there is expected to be a temporary dip, but the magnitude of the dip is very uncertain.

Forecasters (SPF, see Data Appendix for details). The year-ahead forecasts are regarded as a measure of long-run expectations. Figure 1 for output growth shows little change in the model forecast, whereas the survey forecast ranges from 1.1 to 0.4% quarterly growth in the 1980s, captures the slowdown in the late 1990s, and the pickup in the early part of this century prior to the recent recession. In the case of inflation (Figure 2), the long-run survey expectation declines from around 8% per annum at the beginning of the period to around 2% at the end, with some reversals (such as in the late eighties). By way of contrast, the mean of the model forecasts is little changed over the period (although there is more variability than in the forecasts of output growth). [3] These figures serve to illustrate the changes that occur over time in the long-run outlook as given by the median year-ahead forecasts. We will consider whether these changes are associated with more accurate forecasts of output and inflation, as well as considering a range of other variables routinely reported in the SPF.

The second focus of our investigation is highlighted by figures 3 and 4. These again display the long-run survey expectations (depicted as the triangles) for output growth and CPI inflation, but in addition we have plotted the lagged value of the variable at each forecast origin (the circles) as well as the forecast of the current quarter (the squares). Figures 1 and 2 portray the evolution of long-run expectations over time. Figures 3 and 4 show the 'term structure of forecasts' over time. Each figure consists of four panels, where the top left corresponds to surveys run in the first quarter of each calendar year, the top right to second quarter surveys, etc. In fact, there are no substantive differences between the quarters of the year the survey falls in for our current purposes. The division of forecasts by survey quarter is simply to aid readability. It is generally the case that the current quarter forecast lies between the lagged value (where we are when we start forecasting) and the long-run expectation, so that forecasts (here just the current forecast, although we consider all the intermediate forecasts in what follows) converge to the long-run expectation. In terms of the figures, the square lies between triangle and the circle. However, there are exceptions. Consider

[3] Of course 'the' model forecasts could be made more adapative by considering models with time-varying parameters and a rolling forecasting scheme - as explained in section 2, the model forecast here is a fixed-parameter AR using a recursive forecasting scheme - but the essential feature that model forecasts are 'backward-looking' will remain.

inflation, and the 2008:Q4 median forecasts (figure 4, bottom right panel, last observation). At the time the survey forecasts are filed, last quarter's inflation rate is estimated at close to 7% (circle), and the long-run expectation is for inflation of around 2% (triangle). However, the forecast of inflation in the current quarter (square) is of a rate below -3%. This is an extreme example of a NC-forecast; others are discernible both for inflation and for output growth (consider e.g., the same observation), and occur in 'normal times' as well as periods of financial turmoil. Of interest is whether NC-forecasts generally improve forecast accuracy. We consider a range of variables in addition to real output growth and CPI inflation, and a range of forecast horizons in addition to the current quarter horizon depicted in these figures. We have used as a motivating example the median forecasts, as we are primarily interested in explaining the outcomes of forecast comparisons that use the median or consensus forecast. We will also be interested in the individual-respondent forecasts, as these are amenable to a behavioural interpretation.[4]

The plan of the remainder of the paper is as follows. Section 2 consider whether the median survey forecasts respond to perceived changes in the long-run outlook in a way that enhances the accuracy of the shorter-horizon forecasts. Section 3 defines NC-forecasts and the impact of such behaviour on the accuracy of the consensus forecast. Section 4 analyses whether the accuracy-enhancing characteristics of the consensus forecast are also found at the level of the individual forecasts. Section 5 offers some concluding remarks. Details on the SPF data and real-time data sources are given in Appendix 1, and Appendix 2 provides some details on the pooled regression estimators of section 4.

2 Median SPF forecasts and the changed outlook

Our sample consists of the quarterly SPF surveys from 1981:Q3 to 2008:4. Each respondent provides forecasts of the current quarter, and for each of the next four quarters (so the longest is a forecast of the survey quarter in the following year). Let t denote the survey quarter, so that t is one of

[4] In the sense that the forecasts, and in particular the decision of whether to report an NC forecast, can be viewed as a conscious decision by the individual forecaster.

1981:Q3 to 2008:Q4. The forecast of the current (survey) quarter t is essentially a 1-quarter ahead forecast based on $t-1$, which we denote by $y_{t|t-1}$. Then the forecasts are given by $y_{t-1+h|t-1}$, for $h = 1, 2, \ldots, 5$, for each survey quarter t. So for the 1981:Q3 survey, $h = 1$ refers to a forecast of 1981:Q3, and $h = 5$ to a forecast of 1982:Q3. We consider the forecasts made by the regular respondents[5], and calculate the consensus forecasts from this subset. The model-based forecasts are from an AR model estimated on the vintage of data available at the time of the corresponding survey forecast (namely, the t-data vintage containing data through $t-1$ for forecasts matching the survey t forecasts). The data are taken from the real-time datasets (RTDSMs): see the Data Appendix. Hence the model-based forecasts corresponding to the 1981:Q3 survey forecasts were from a model estimated on the 1981:Q3 vintage of data, containing observations from 1947:Q2 through 1981:Q2. The model order was selected by BIC. Forecasts from subsequent origins were generated using a recursive scheme (an expanding window of data), whereby the model order was selected and the parameters were estimated anew at each forecast origin, the last being 2008:Q4 (using data from 1947:Q2 through 2008:Q3). We generated 1 to 5-step ahead forecasts: so for the first forecast origin, these were a 1-step forecast of 1981:Q3 up to a 5-step ahead forecast of 1982:Q3.

Table 1 compares the accuracy of the median survey forecasts and model forecasts on MSFE. The first three columns report the MSFEs for each forecast horizon, and the ratio of the median survey to model MSFE. We find the survey forecasts are markedly more accurate at the shorter horizons for all the eight variables we consider: real GDP, five GDP component series, the GDP deflator and the CPI (full descriptions are given in Appendix A). For example, for $h = 1$, the median survey forecast MSFEs range from as little as two fifths to three-quarters of the model MSFEs. The MSFEs are computed using estimates of the actual values published in the second quarter following the data being forecasted.[6] Notice that the survey forecasts will be informed by

[5] Regular respondents are those who responded to 12 or more surveys.

[6] Following Romer and Romer (2000) a number of authors have made this assumption: it helps to ensure that the actual values are measured according to the accounting practices prevalent at the time the forecast was made, rather than reflecting the impact of subsequent benchmark revisions, while also ensuring the actuals are based on more complete information than the advance or preliminary estimates (see, e.g., Landefeld, Seskin and Fraumeni (2008)). For the calculation of forecast accuracy, the last forecast origin is 2007:Q2, as for this origin the longest forecast is of

knowledge of developments up to the middle of the quarter t,[7] whilst the model forecasts use data only through quarter $t-1$. It is unsurprising that the timeliness and breadth of the information that survey forecasts can draw on results in superior forecasts. We are interested in whether the way in which the superior information benefits the short-term forecasts can be ascribed either to its effect on the long-run outlook (this section) or to induced departures from the equilibrium path (the following section). That is, how is this information used? However, a simple check of whether the superiority of the median survey forecasts holds, when the timing convention instead favours the model forecasts, was performed by replacing the SPF forecasts with 'next quarter' forecasts from the previous survey quarter. The effective forecast horizon of the survey forecasts now exceeds one quarter. The relative superiority of the survey forecast is generally diminished but remains.

We consider whether the greater adaptability of the survey forecasts to changes in the long-run outlook accounts for the relative superiority of the survey (shorter-term) forecasts as follows. We regress the difference in the accuracy of the short-term survey and model forecasts on the change in the long-run outlook as measured by the longer-horizon survey forecasts. That is, we estimate by OLS the following regression, and report HAC (heteroscedasticity and autocorrelation consistent) standard errors to account for possible heteroscedasticity and the overlapping nature of the forecasts:

$$\text{MAE}\left(y^{Med}_{t|t-1} + y^{Med}_{t+1|t-1}\right) - \text{MAE}\left(y^{Mod}_{t|t-1} + y^{Mod}_{t+1|t-1}\right) = \beta_1 + \beta_2 \left|\sum_{i=2}^{4}\left(y^{Med}_{t+i|t-1} - y^{Med}_{t-1+i|t-2}\right)\right| + \zeta_t \quad (1)$$

where t runs over the surveys from 1981:Q3 onwards. Our measure of forecast accuracy is mean absolute error (MAE) to guard against a few large forecast errors unduly influencing the findings (as might occur with squared error loss). The dependent variable consists of the difference in accuracy between the survey and model forecasts. Each component is the MAE of the sum of the

2008Q2 ($h = 5$). As we use second estimate actuals from the 2008:Q4 RTDSM, the latest period for which we have the second estimate of the actual is 2008:Q2.

[7] The survey responses for quarter t are filed around the middle of the middle month of quarter t. Information on the quarter being forecast would be expected to improve the forecasts of that quarter. See, for example, Montgomery, Zarnowitz, Tsay and Tiao (1998) and Clements and Galvão (2008), who consider this issue in the context of model-based forecasts.

current and next-quarter forecast errors. This was adopted as a more robust measure of short-horizon forecast accuracy than considering either the current or next-quarter forecasts separately. Formally, $\text{MAE}\left(y^{For}_{t|t-1} + y^{For}_{t+1|t-1}\right) = \left|y^{t+2}_{t} + y^{t+3}_{t+1} - y^{For}_{t|t-1} - y^{For}_{t+1|t-1}\right|$ (where For denotes either a survey (Med) or model (Mod) forecast, and y^{t+2}_{t} denotes an estimate of the value of y in quarter t (the subscript) taken from the data vintage available two quarters later in quarter $t+2$ (the superscript)). The slope is the absolute value of the difference between the longer-horizon forecast issued from survey quarter t relative to the previous survey quarter $t-1$.[8] The longer-horizon forecasts sum over the 2, 3 and 4 quarter ahead forecasts to provide a more robust measure of the longer-term outlook. As well as running the regression described above, we also calculate the Spearman rank correlation test between the dependent and slope variables of (1). This tests the null hypothesis of no correlation between the two series under weaker assumptions than are required for the regression. Given our sample size, the probability that a standard normal random variable exceeds the test statistic value provides the p-value, and both the test statistic and p-value are recorded in table 2 along with the regression results.

For the first five variables in table 2 there is evidence that large changes in the long-run outlook are associated with more accurate short-horizon survey forecasts, as indicated by the negative value β_2. The regressions indicate that β_2 is significantly different from zero in two of these cases, and in four of the five cases the Spearman test rejects the null of no correlation (at the 10% level). For the two price variables, PGDP and CPI, there is no evidence to reject the null that $\beta_2 = 0$ or of no correlation on the the rank test, but for both these variables the significantly negative constant term signals the greater average accuracy of the survey forecasts, consistent with the findings reported in table 1. Hence broadly speaking the results split by variable type: for the two price measures, there is no association between changes in the long-term outlook and short-horizon forecast performance; for all the GDP components (other than two sub-divisions of investment, RNRESIN and RSLGOV) there is a positive correlation between changes in the long-run outlook and the accuracy of the short-horizon consensus forecasts.

[8] Recall that our timing convention is that $y_{t+i|t-1}$ denotes a forecast from survey quarter t (not survey quarter $t-1$).

3 Median SPF forecasts and 'non-convergent' forecasts

In this section we consider whether departures of shorter-horizon forecasts from the path of convergence to the long-run position 'add value', in that they enhance forecast accuracy. We begin by defining what is meant by a 'non-convergent' (NC) forecast. We then examine the extent to which the median survey forecasts portray this property, before estimating its impact on forecast accuracy.

3.1 Defining NC forecasts

Recall that we have survey forecasts given by $y_{t-1+h|t-1}$, for $h = 1, 2, \ldots, 5$, for each survey quarter t. We define forecasts $h = 1, \ldots, 4$ as being NC if they move further from the long-run expectation (given by $h = 5$) than the starting point (the latest value at the time of forecasting, y_{t-1}), or if they 'overshoot' the long-run position. Figure 5 illustrates. As drawn, the long-run expectation $y_{t+4|t-1}$ exceeds y_{t-1}. But the $h = 1$ forecast $y_{t|t-1}$ is for a value lower than y_{t-1}. We call this type of NC forecast a forecast that 'bucks the trend' (btt). The other intermediate forecast shown in the figure is for $h = 4$. This forecast, $y_{t+3|t-1}$, exceeds the long-run expectation, and as y_{t-1} was below the long-run expectation, we say that the $h = 4$ forecast 'overshoots' (os). Hence we subdivide the NC forecasts into two categories, to allow that these two types of deviations from the trend may have different characteristics. Convergent forecasts are those which remain within the tunnel defined by the horizontal lines through y_{t-1} and $y_{t+4|t-1}$.

A similar analysis follows when instead $y_{t-1} > y_{t+4|t-1}$, giving the formal statement of the conditions for NC as follows. For $h = 1, 2, 3$ and 4, we say that the h-step ahead forecast, $y_{t-1+h|t-1}$, is NC in the sense of 'bucking the trend' (btt) if:

$$y_{t-1+h|t-1} : y_{t-1+h|t-1} < y_{t-1} < y_{t+4|t-1} \text{ or: } y_{t-1+h|t-1} > y_{t-1} > y_{t+4|t-1}$$

where y_{t-1} is the value of the quarterly growth rate in the period prior to the survey quarter.[9]

[9] This growth rate is taken from the survey-quarter RTDSM data vintage (see Data Appendix), and is the vintage of data that would have been available to the respondent to survey t. The survey respondents' supply their 'estimate'

For $h = 1, 2, 3$ and 4, we say that the h-step ahead forecast, $y_{t-1+h|t-1}$, is NC in the sense of 'overshooting' (os) the long-run expectation if:

$$y_{t-1+h|t-1} : y_{t-1} < y_{t+4|t-1} < y_{t-1+h|t-1} \text{ or: } y_{t-1} > y_{t+4|t-1} > y_{t-1+h|t-1}. \tag{2}$$

Rather than comparing $y_{t-1+h|t-1}$ to the first value (y_{t-1}) and the longest horizon forecast ($y_{t+4|t-1}$), one might compare $y_{t-1+h|t-1}$ to adjacent forecasts, and signal a forecast that has the btt property if (again for $h = 1, 2, 3$ and 4):

$$y_{t-1+h|t-1} : y_{t-1+h|t-1} < y_{t-1+h-1|t-1} < y_{t-1+h+1|t-1} \text{ or: } y_{t-1+h+1|t-1} > y_{t-1+h-1|t-1} > y_{t-1+h|t-1} \tag{3}$$

where for $h = 1$, $y_{t-1+h-1|t-1}$ is y_{t-1}. And similarly for os forecasts.

However, this is problematic: suppose $y_{t+1|t-1} > y_{t+2|t-1} > y_{t+4|t-1}$; but $y_{t+1|t-1} > y_{t+3|t-1} > y_{t+2|t-1}$. Then both $y_{t+2|t-1}$ and $y_{t+3|t-1}$ are btt on the 'local' definition, whereas neither is on the 'global' definition. Suppose the forecaster is largely indifferent to whether $y_{t+2|t-1} \gtreqless y_{t+3|t-1}$: the local definition is less robust to small changes in $y_{t+2|t-1}$ relative to $y_{t+3|t-1}$.

Finally we might declare all $\{y_{t-1+h|t-1}\}$, $h = 1, \ldots, 5$ to be NC if the forecasts are neither monotonically increasing or decreasing from y_{t-1}, but this is especially susceptible to the problem of falsely signalling NC.

We assume that y refers to the quarter-on-quarter growth rate. For variables which have unit roots or near-unit roots, it seems sensible to define conditions (2) and (3) in terms of the growth rates, as the growth rates will have well-defined long-run expectations. Hence we define NC in terms of growth rates for the level of real GDP and its components, as well as for the GDP deflator and CPI inflation. The CPI forecasts are of the annualised rate of inflation, and we consider

of the level of the variable for the quarter prior to the survey quarter, and this is used in the construction of the $h = 1$ survey forecast, but the quarterly growth rate for the quarter prior to the survey quarter (i.e., y_{t-1}) is from the RTDSMs (except for the CPI). For the surveys after 1990:Q3, when the Philadelphia Fed assumed control of the administration of the survey, the respondents were always provided with last period's value. Consequently, the respondents' 'estimates' invariably match the values recorded in the RTDSMs. Prior to 1990, there is more variability in the estimates of this value, and in the empirical work we check to see whether this affects the findings.

these forecasts directly.[10] The forecasts of GDP and its components are evaluated in terms of forecasts of the quarter-on-quarter changes, and the CPI forecasts are evaluated directly in terms of (annualised) quarter-on-quarter changes.

3.2 Median forecasts and the NC property

Our interest is in whether the consensus forecast (which we take to be the median) exhibits NC-behaviour, and if so, whether the NC-behaviour improves or worsens the accuracy of the consensus forecast.

Table 3 reports the proportions of surveys (from 1981:Q3 to 2008:Q4) for which the forecasts possessed the btt and os properties, separately for each forecast horizon h. We begin with btt forecasts. Such forecasts are more common at the short horizons. For example, 25% of the 1-step median survey forecasts of real output (RGDP) are btt, declining to around 3% (i.e., 3 of the 110 forecasts) for $h = 4$. Fewer residential investment (RRESINV) and government expenditure forecasts (both Fed, and State & Local: RFEDGOV, RSLGOV) are btt. As expected, far fewer model forecasts are btt: in the case of output growth, only 2% are btt $h = 1$ (equating to 2 of the 110), and none at longer horizons.

Turning now to the price variables, we report two sets of results for CPI inflation. The first set uses the RTDSM estimate of last period's inflation rate to calculate NC-forecasts, as for the real variables. The second uses the median forecast of the inflation rates in the previous period reported by the survey respondents. Using the RTDSM estimate results in around a quarter of all $h = 1$ forecasts being classified as btt, with a halving of this number when the median 'forecast' of y_{t-1} is used. The mean difference between the RTDSM value and the median forecast of y_{t-1} is 0.05, compared to an average annualised quarter-on-quarter inflation rate of 3.4% over the whole period (1981–2008). Further investigation revealed the discrepancy between the RTDSM and median forecast values of y_{t-1} to be mainly due to the period prior to 1990, before the Philadelphia Fed

[10] For the CPI, we use the RTDSM value of the quarterly growth rate for the quarter prior to the survey quarter (i.e., y_{t-1}) as for all the other variables, but as a sensitivity check we also use the SPF 'forecast' values of y_{t-1} in an additional exercise.

assumed control and always provided respondents with the latest estimate of the last quarter's value. The model forecasts of the two price variables differ from those of the real variables in that there are many more btt-forecasts. This we attribute to the greater persistence of these variables, and the possibility that inflation is close to being $I(1)$ (so that the forecasts converge more slowly in h).[11]

In terms of os, we find quite different patterns: generally a quarter to a half of the median survey forecasts are os, and the proportion does not depend upon the horizon. There is again generally less evidence of os for the model forecasts (none for output growth), but there are os-forecasts for some variables such as consumption and the government variables across horizons.

In summary, there is evidence that a significant proportion of the shorter-horizon median survey forecasts 'buck the trend'. This is especially true for output growth and inflation, the key 'headline' macro indicators. There is generally more overshooting of the median survey forecasts across all horizons. There is less evidence of the model forecasts having these properties, with some exceptions.

To measure the effect on forecast accuracy of NC behaviour, it is tempting to calculate the average forecast accuracy of the NC and non-NC forecasts separately, and to compare the two. However, this approach is flawed unless NC forecasts are issued independently of the current and prospective state of the economy: if NC forecasts were made at times of greater macroeconomic fluctuations, for example, we would underestimate the beneficial effect of NC behaviour.

We get around this by comparing the NC forecasts to simple counterfactual forecasts which do not possess the NC-property. We construct the artificial forecasts by replacing the btt and os forecasts by forecasts which are as close as possible to the originals subject to them not being NC. This is most easily understood in terms of figure 5. Letting $\widetilde{y}_{t-1+h|t-1}$ denote the artificial forecast, we set $\widetilde{y}_{t|t-1} = y_{t-1}$, and $\widetilde{y}_{t+3|t-1} = y_{t+3|t-1}$. Again in terms of the figure, the accuracy of the counterfactual forecast, $\widetilde{y}_{t|t-1}$, will improve when y_t (the actual value in period t) is closer to y_{t-1} than $y_{t|t-1}$, i.e., provided $y_t > \frac{1}{2}\left(y_{t-1} + y_{t|t-1}\right)$. Although *ad hoc*, this way of constructing

[11]Note that inflation is sometimes modelled as an $I(1)$ process (see, e.g., Stock and Watson (2008)) indicating that the deflator is an $I(2)$ variable.

counterfactuals has the merit of assessing whether the NC characteristic enhances accuracy relative to the closest forecast that does not possess this property.[12]

The columns of table 1 headed '$\widetilde{y}_{btt}$' and '$\widetilde{y}_{os}$' record the results of replacing the NC forecasts by the artificial forecasts. They report the MSFE of the $\widetilde{y}$ forecasts to the MSFE of the reported forecasts (noting that the artificial forecasts are identical to the reported forecasts for non-NC forecasts). The adjusted forecasts are generally worse at $h = 1$, and for some variables markedly so, for both btt and os. For CPI inflation, for example, the ratio of the MSFE of the artificially-adjusted os-forecasts to that of the published forecasts is around 1.5 at $h = 1$, and 1.10 at $h = 2$, indicating that overshooting markedly improves accuracy.

As a check that the accuracy-enhancing NC-characteristic is specific to the survey forecasts, we also report the ratio of the MSFE of the artificial model forecasts (corrected as for the median survey forecasts) to the model forecasts. As expected, the ratio is always close to one, so that the effect of NC-behaviour on the model forecasts is neutral.

Our findings suggest that both the forward-looking nature of the survey forecasts, and especially the 'non-convergent' characteristic, contribute to their superiority over the model forecasts. Changes in the long-run outlook improve the accuracy of the GDP and most component forecasts, whereas the NC-characteristic generally enhances the accuracy of all the short-horizon consensus forecasts (relative to the model forecasts), including those of the two price variables.

4 Individual-level analysis of NC-characteristic

To what extent are the accuracy-enhancing characteristics of the median survey forecasts present at the level of the individuals' forecasts? We consider the NC-characteristic as this is found to improve the forecasts of nearly all the variables at the aggregate level (including CPI inflation). Table 4 reports the proportions of all the individual forecasts over all surveys that are either btt or os, separately for each h. Across all variables, roughly one fifth of all $h = 1$ forecasts are btt,

[12] One could view $\widetilde{y}$ as a simple linear combination of y_{t-1} and $y_{t+4|t-1}$, i.e., $\omega y_{t-1} + (1-\omega)\, y_{t+4|t-1}$, where we set $\omega = 0$ (for NC os forecasts) or $\omega = 1$ (for NC btt forecasts). Viewed in this way, other values of ω satisfying $0 < \omega < 1$ are possible, but our choice minimises $\left|y_{t-1+h|t-1} - \widetilde{y}_{t-1+h|t-1}\right|$.

declining to just less than half this fraction at $h = 4$. Roughly one third of all forecasts are os, and this fraction is largely the same across forecast horizons. These are not very different from the findings for the consensus forecasts.

The results regarding the impact on forecast accuracy are recorded in table 5. The table records the results for btt and os-forecasts separately. For each forecast horizon $h = 1, \ldots, 4$ we report: the average accuracy of all the individual forecasts, where we average the squared forecast errors over all respondents and surveys; the number of NC and non-NC forecasts (either btt or os); and the results of replacing the NC forecasts by the artificial forecasts. The columns headed 'btt[os]-ratio MSFE' report the MSFE of the $\widetilde{y}$ forecasts to the MSFE of the reported forecasts. By and large, the adjusted forecasts are generally more accurate than the originals, with the exception of the CPI forecasts, indicating that individuals' NC-behaviour worsens forecast performance. For CPI inflation, on the other hand, the 'smoothed' counterfactual forecasts are roughly 10% to 15% less accurate. These results are clearly at odds with those for the consensus forecasts for all variables other than the CPI. To investigate further, we calculate the average absolute forecast error over all respondents and surveys (instead of the squared error) to check whether the average measure is being unduly influenced by a few large errors (possibly resulting from idiosyncratic reporting errors, etc): see the columns headed 'btt[os]-ratio MAE'. There is now less evidence that NC-behaviour clearly harms forecast accuracy, but by and large little evidence for the positive effect found for the median forecasts (except for the CPI).

Note that the results of the two ways of calculating NC behaviour for CPI inflation match closely. NC behaviour enhances accuracy of the CPI forecasts on average whether the NC forecasts are determined on the basis of the RTDSM estimates of period $t - 1$ inflation (row labelled CPI_1) or the individuals' reported values (row labelled CPI_2).

We also experimented with only using surveys from 1990:Q3 (see table 6). The results do not differ materially from those for the whole period, suggesting that any changes in the way the survey was administered at that time do not matter greatly for our purposes.

The statistics reported in table 5 are the result of a fairly broadbrush approach. The MSFE/MAE calculations average across different numbers of forecasts from different surveys, without making any allowance for the fact that forecast errors from a given survey will be correlated because of common macroeconomic shocks, or that the overlapping nature of the forecasts means that forecast errors will be correlated across time. To control for these aspects, and to conduct statistical inference on the effect of NC behaviour on forecast accuracy at an individual level, we estimate pooled regressions based on the approach of Keane and Runkle (1990) and Bonham and Cohen (2001). Specifically, we regressed the difference in forecast accuracy of the reported and artificial forecasts (as measured by MAE) on the absolute difference in the reported and counterfactual forecasts (which will be zero for non-NC forecasts):

$$\left|y_{t+h} - y_{i,t-1+h|t-1}\right| - \left|y_{t+h} - \widetilde{y}_{i,t-1+h|t-1}\right| = \beta_1 + \beta_2 \left|y_{i,t-1+h|t-1} - \widetilde{y}_{i,t-1+h|t-1}\right| + \varepsilon_{ith} \quad (4)$$

If $\beta = 0$, then we conclude that NC behavior has no systematic effect on MAE (once we allow for common macro shocks and the overlapping nature of the forecasts). Alternatively, $\beta > 0$ indicates that the reported forecasts have a larger MAE than the counterfactuals, so that on average NC behaviour worsens forecast accuracy (and conversely for $\beta < 0$). We estimate (4) over all i and t for a given $h = \{1, 2, 3, 4\}$. To allow for the overlapping nature of the forecasts and for the dependence in forecast errors across individuals resulting from common macro shocks, we assume the following covariance structure for the ε_{ith}, where for a given h, and with t denoting the survey quarter (t =1981:2 ..., 2007:2), then for an individual i:

$$\begin{aligned} E\left[\varepsilon_{ith}^2\right] &= \sigma_0^2 \\ E\left[\varepsilon_{ith}\varepsilon_{i,t+k,h}\right] &= \sigma_k^2 \text{ when } 0 < k \leq h, \text{ and } 0 \text{ otherwise} \end{aligned}$$

and for any pair of individuals i, j:

$$\begin{aligned} E\left[\varepsilon_{ith}\varepsilon_{jth}\right] &= \delta_0^2 \\ E\left[\varepsilon_{ith}\varepsilon_{j,t+k,h}\right] &= \delta_k^2 \text{ when } 0 < k \leq h, \text{ and } 0 \text{ otherwise.} \end{aligned}$$

In the Appendix we explain this particular structure further, including how it results from using actual values published in the second quarter following the data being forecasted. We also record how the model is estimated and how the estimated covariance matrix of the disturbances is obtained given the unbalanced nature of the panel.

The results are summarized in table 7, and broadly confirm the broadbrush approach of table 5. We find that except for the current-quarter ($h = 1$) CPI inflation forecasts, β_2 is positive and significant. CPI inflation is predictable (beyond convergence to the long-run expectation) at the shortest horizon, as we find $\beta_2 < 0$, but in all other cases the NC-behaviour at the individual level tends to worsen the accuracy of the reported forecasts.

5 Conclusions

At the level of the aggregate or consensus forecasts, we have shown that it is possible to discern two characteristics that contribute to the superiority of the survey forecasts over purely model-based forecasts. These characteristics are identifiable from forecasts of different horizons made in successive periods. Firstly, we show that short-horizon forecasts issued at times of changes in the long-run outlook tend to outperform purely model-based forecasts for GDP and most of its components. Hence we provide support for the contention that survey forecasts are able to foresee changes in the economic environment, rather than simply extrapolating past patterns, and do so in a way that leads to more accurate short-horizon forecasts. Secondly, we show that consensus forecasts do not always move monotonically to the long-run expectation as the forecast horizon increases, and moreover, that such departures from the path toward the long-run expectation on average tend to enhance forecast accuracy.

Having categorized these two characteristics of consensus forecasts, we consider whether individual forecasts also display accuracy-enhancing departures from the long-run expectation path. We find that 'smoothing out' these departures improves the average squared individual forecast error for all the variables we consider other than the CPI inflation rate. Only for forecasting CPI inflation are the reported forecasts more accurate on average than counterfactual forecasts characterised by convergence to the long-run position. There is some evidence that the average accuracy of the individual forecasts (other than for CPI inflation) is adversely affected by some large 'idiosyncratic' errors (e.g., reporting/typographical errors) which are removed when the counterfactual forecasts are switched in, because the relative improvement of the average individual forecast is lessened when accuracy is measured by absolute loss rather than squared error loss. Nevertheless, there is scant evidence at the individual level that the counterfactuals are less accurate than the reported forecasts, in stark contrast to the findings for the consensus forecast. This suggests that the consensus forecast is successful in 'the pooling of large amounts of information; the efficient aggregation of that information' (Ang *et al.* (2007)) at least to the extent that the resulting forecasts cannot be readily improved by ironing out departures from the equilibrium path.

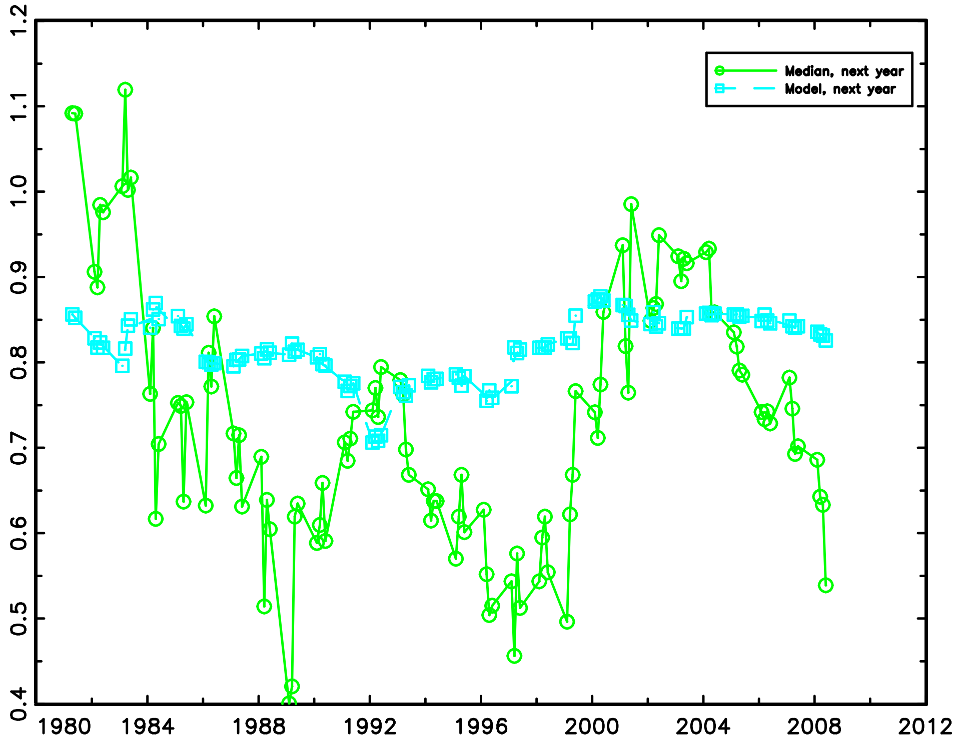

Figure 1: Real output growth year-ahead forecasts, plotted against the time the forecasts were made.

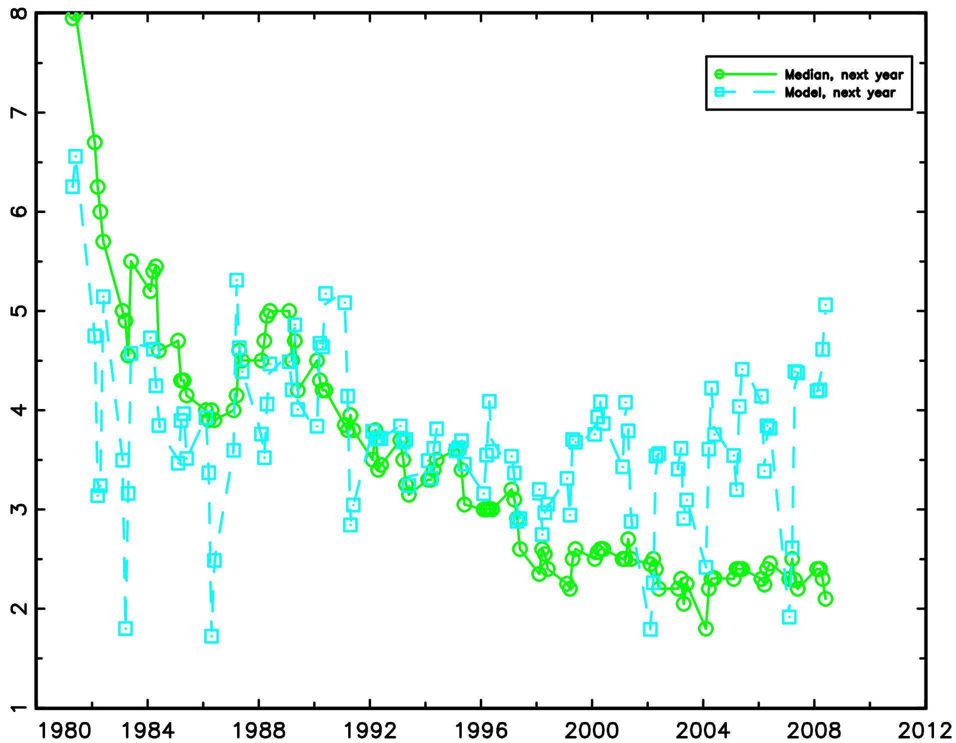

Figure 2: CPI inflation year-ahead forecasts, plotted against the time the forecasts were made.

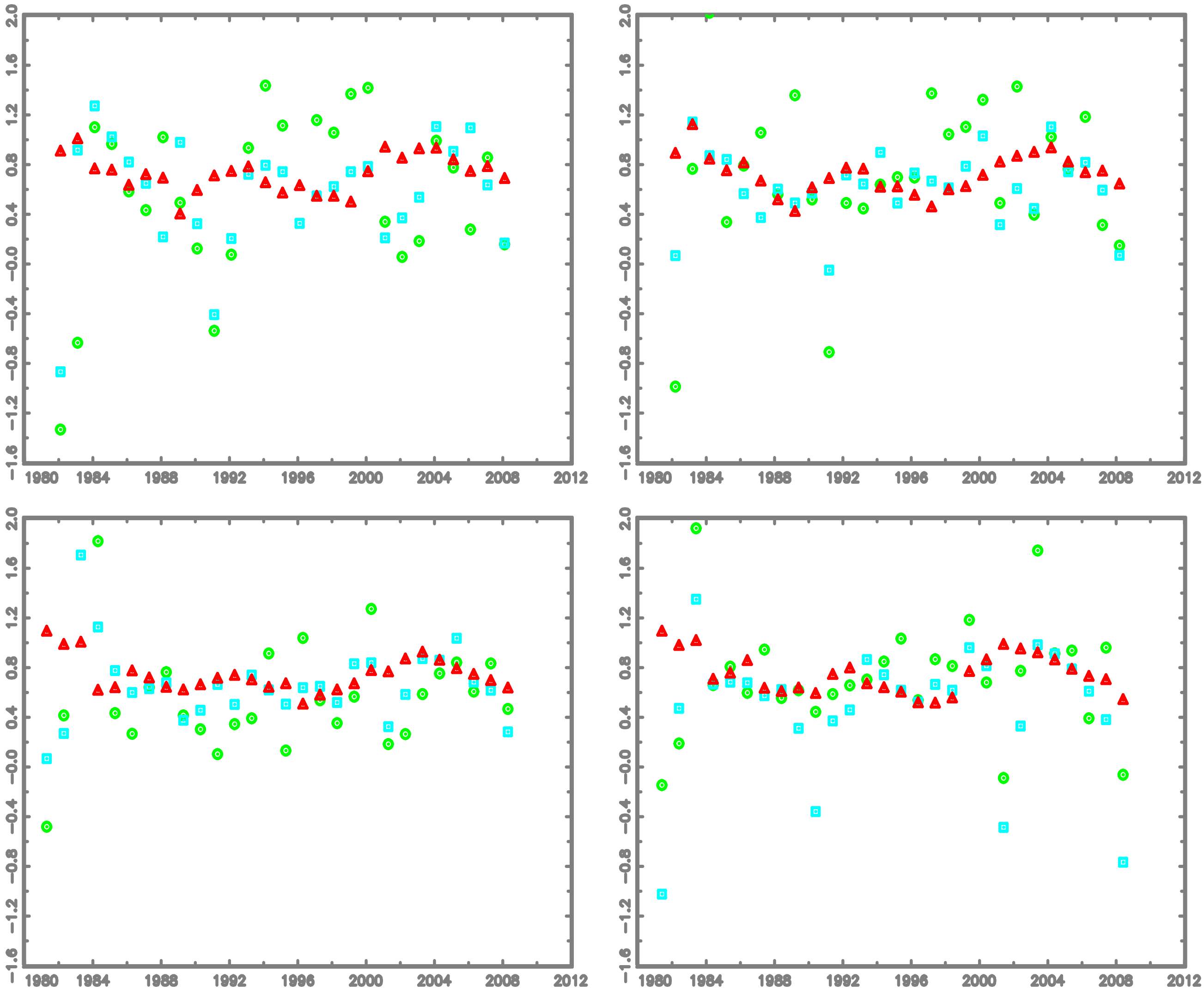

Figure 3: Median survey forecasts of quarter-on-quarter percentage output growth, for the survey quarter (square), and the same quarter a year ahead (triangle). The circles denote the first estimates of the actual value for the previous quarter.

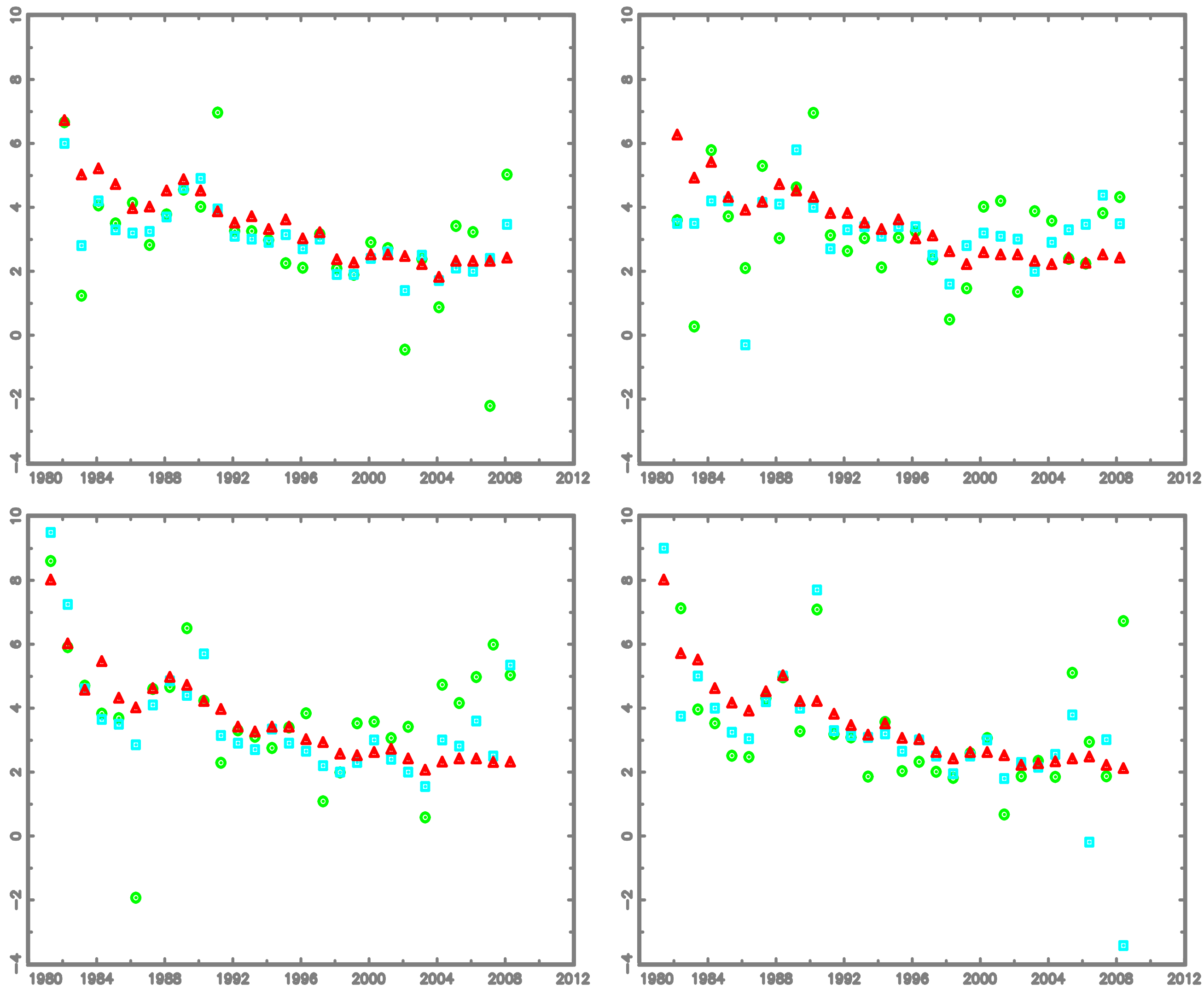

Figure 4: Median survey forecasts of quarter-on-quarter percentage CPI inflation at an annualised rate, for the survey quarter (square), and the same quarter a year ahead (triangle). The circles denote the first estimates of the actual value for the previous quarter.

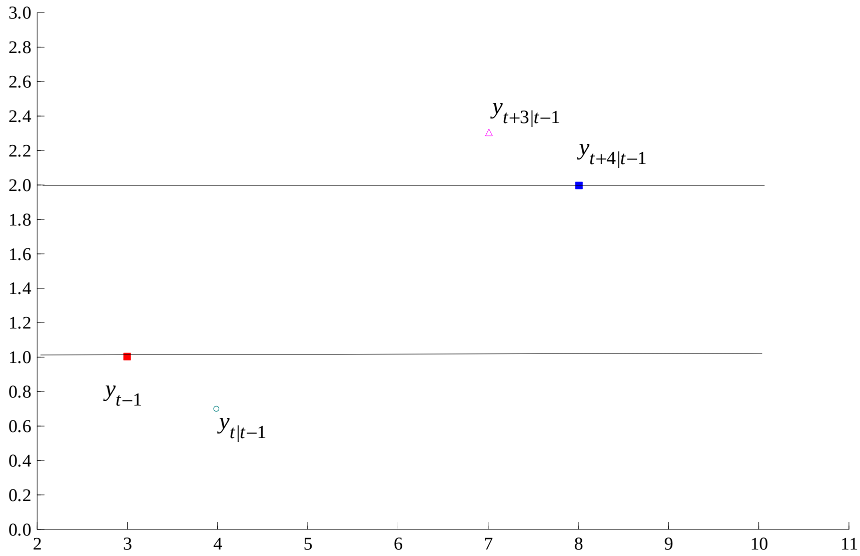

Figure 5: Graphical depiction of 'bucking the trend' and 'overshooting' forecasts, $y_{t|t-1}$ and $y_{t+3|t-1}$.

Table 1: Forecast accuracy (MSFE) of the consensus and model forecasts

	h	Median Survey	Model Forecasts	Median / Model	$\tilde{y}_{btt}$	$\tilde{Model}$/ Model	$\tilde{y}_{os}$	$\tilde{Model}$/ Model
	1	0.20	0.29	0.67	1.13	1.00	1.04	1.00
RGDP	2	0.28	0.35	0.81	1.05	1.00	0.97	1.00
	3	0.31	0.34	0.92	1.01	1.00	0.97	1.00
	4	0.30	0.30	0.97	1.00	1.00	0.96	1.00
	5	0.28	0.30	0.93	.	.	.	.
	1	0.23	0.31	0.74	1.11	1.00	1.19	1.01
RCONSUM	2	0.30	0.30	0.98	0.99	1.00	1.04	1.01
	3	0.30	0.28	1.05	1.00	1.00	1.00	1.00
	4	0.31	0.30	1.04	1.00	1.00	1.00	1.00
	5	0.30	0.30	1.01	.	.	.	.
	1	2.98	4.41	0.67	1.04	1.00	1.10	1.00
RNRESIN	2	3.77	4.79	0.79	1.02	1.00	1.03	1.00
	3	4.34	5.07	0.86	1.01	1.00	1.00	1.00
	4	4.75	5.27	0.90	1.00	1.00	1.01	1.00
	5	4.77	5.25	0.91	.	.	.	.
	1	5.72	8.22	0.70	1.00	1.00	1.07	1.00
RRESINV	2	8.10	11.09	0.73	0.99	1.00	1.07	1.00
	3	9.89	11.89	0.83	1.00	1.00	1.04	1.00
	4	10.58	12.60	0.84	1.00	1.00	1.04	1.00
	5	9.94	13.07	0.76	.	.	.	.
	1	5.38	12.29	0.44	1.00	1.00	1.28	1.01
RFEDGOV	2	6.83	9.67	0.71	1.00	1.00	1.12	0.99
	3	7.02	8.13	0.86	1.00	1.00	1.03	1.00
	4	6.92	7.41	0.93	1.00	1.00	1.01	1.00
	5	6.84	7.94	0.86	.	.	.	.
	1	0.35	0.47	0.73	1.02	1.00	1.03	1.02
RSLGOV	2	0.32	0.43	0.74	1.00	1.00	1.02	1.00
	3	0.34	0.44	0.77	1.00	1.02	1.00	1.02
	4	0.36	0.49	0.74	1.00	1.00	0.98	1.00
	5	0.36	0.54	0.67	.	.	.	.
	1	0.05	0.08	0.66	1.07	1.02	1.06	0.99
PGDP	2	0.06	0.09	0.71	1.01	1.00	1.12	1.00
	3	0.09	0.08	1.13	1.01	1.00	1.00	1.00
	4	0.09	0.08	1.15	1.00	1.00	1.01	1.00
	5	0.11	0.09	1.11	.	.	.	.
	1	1.14	2.76	0.41	1.32	1.01	1.45	0.99
CPI_1	2	2.50	3.47	0.72	1.01	0.98	1.10	1.02
	3	2.91	3.20	0.91	1.00	1.00	0.98	1.00
	4	3.07	3.34	0.92	1.01	1.00	0.99	1.00
	5	3.42	3.37	1.02	.	.	.	.
	1	1.14	2.76	0.41	1.14	1.01	1.57	0.99
CPI_2	2	2.50	3.47	0.72	1.00	0.98	1.10	1.02
	3	2.91	3.20	0.91	1.00	1.00	0.98	1.00
	4	3.07	3.34	0.92	1.00	1.00	0.99	1.00
	5	3.42	3.37	1.02	.	.	.	.

Notes: The column '$\tilde{y}_{btt}$' is the ratio of the MSFE when the btt-forecasts are replaced by non-btt forecasts to the MSFE of the reported median forecasts. The column '$\tilde{y}_{os}$' is the same when the os-forecasts are replaced by non-os forecasts. The columns $\tilde{Model}$/Model are the MSFE of the btt and os-AR forecasts to the AR forecasts.

Table 2: Relative median survey and model short-horizon forecast accuracy as a function of the change in the long-run outlook

	$\hat{\beta}_1$	$\hat{\beta}_2$	Spearman	p-value
RGDP	0.0092 (0.9070)	-0.6010 (0.2408)	2.1542	0.0156
RCONSUM	0.0958 (0.1112)	-1.4700 (0.0004)	1.6806	0.0464
RNRESIN	-0.4182 (0.0375)	-0.3137 (0.4066)	0.0386	0.4846
RRESINV	-0.3926 (0.2310)	-0.2789 (0.1057)	1.3163	0.0940
RFEDGOV	-0.7308 (0.0275)	-1.6091 (0.0486)	2.2196	0.0132
RSLGOV	-0.3261 (0.0000)	0.5319 (0.0510)	-1.1804	0.8811
PGDP	-0.0579 (0.0806)	0.0404 (0.8090)	0.2512	0.4008
CPI	-0.6887 (0.0175)	0.0279 (0.9148)	0.2265	0.4104

The table displays the estimates from regression equation (1) with HAC p-values of the null that the corresponding coefficient is zero (in parenthesis), and the Spearman rank correlation test statistic and p-value of the null that the dependent and slope variables are unrelated.

Table 3: Proportion of forecasts that 'buck the trend' and 'overshoot'

	'Bucking the tend' forecasts								'Overshooting' forecasts							
	Median survey forecasts				Model forecasts				Median survey forecasts				Model forecasts			
$h =$	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
RGDP	0.25	0.17	0.10	0.02	0.02	0.00	0.00	0.00	0.20	0.34	0.32	0.42	0.00	0.00	0.00	0.00
RCONSUM	0.14	0.09	0.05	0.03	0.05	0.00	0.00	0.00	0.33	0.36	0.37	0.39	0.30	0.04	0.25	0.07
RNRESIN	0.18	0.11	0.04	0.04	0.00	0.00	0.00	0.00	0.29	0.25	0.30	0.27	0.03	0.01	0.01	0.02
RRESINV	0.08	0.05	0.05	0.03	0.00	0.00	0.00	0.00	0.29	0.36	0.47	0.39	0.00	0.01	0.02	0.04
RFEDGOV	0.05	0.04	0.00	0.01	0.03	0.00	0.01	0.01	0.72	0.55	0.59	0.41	0.14	0.14	0.08	0.04
RSLGOV	0.09	0.05	0.03	0.03	0.05	0.01	0.01	0.01	0.38	0.30	0.35	0.42	0.19	0.15	0.18	0.18
PGDP	0.10	0.08	0.05	0.02	0.17	0.08	0.02	0.00	0.25	0.34	0.39	0.35	0.09	0.05	0.06	0.05
CPI_1	0.24	0.12	0.06	0.05	0.31	0.15	0.05	0.08	0.25	0.27	0.28	0.30	0.21	0.26	0.14	0.28
CPI_2	0.13	0.03	0.01	0.00	0.31	0.15	0.05	0.08	0.28	0.30	0.30	0.27	0.21	0.26	0.14	0.28

Notes: The table reports the proportion of the forecasts from the 100 survey quarters 1981:3 to 2008:4 that are 'btt' and 'os' at each horizon h, for both the median surveys and the model forecasts.

Table 4: Individual NC proportions

	btt-forecasts				os-forecasts			
$h =$	1	2	3	4	1	2	3	4
RGDP	0.26	0.20	0.15	0.10	0.28	0.34	0.37	0.41
RCONSUM	0.17	0.14	0.10	0.08	0.40	0.39	0.41	0.42
RNRESIN	0.21	0.15	0.10	0.07	0.30	0.30	0.35	0.37
RRESINV	0.21	0.14	0.10	0.07	0.26	0.33	0.38	0.40
RFEDGOV	0.12	0.12	0.10	0.08	0.53	0.43	0.44	0.43
RSLGOV	0.17	0.14	0.11	0.10	0.40	0.39	0.40	0.43
PGDP	0.21	0.16	0.13	0.10	0.31	0.37	0.39	0.42
CPI_1	0.22	0.14	0.09	0.07	0.27	0.32	0.32	0.30
CPI_2	0.18	0.10	0.07	0.05	0.26	0.31	0.30	0.29

Notes: The table records the proportion of all the forecasts across individuals and surveys which are each type of NC forecast.
The CPI_1 calculations are based on the RTDSM value of CPI inflation in $t-1$ (as for the other variables). The CPI_2 calculations are based on the individual reported survey values. The variable mnemonics are defined in the Data Appendix.

Table 5: Individual accuracy, 1981:Q3 – 2008:Q4

	h	MSFE	# btt	# non -btt	btt-ratio MSFE	btt-ratio MAE	# os	# non -os	os-ratio MSFE	os-ratio MAE
	1	0.27	694	2124	0.92	1.00	771	2047	0.96	1.00
RGDP	2	0.36	534	2284	0.90	0.97	958	1860	0.88	0.93
	3	0.41	420	2398	0.91	0.97	1041	1777	0.84	0.95
	4	0.36	275	2543	0.94	0.98	1165	1653	0.87	0.96
	1	0.35	415	2282	0.93	0.99	1122	1575	0.91	1.01
RCONSUM	2	0.37	355	2342	0.93	0.98	1055	1642	0.88	0.97
	3	0.36	250	2447	0.94	0.97	1135	1562	0.90	0.94
	4	0.38	205	2492	0.96	0.98	1159	1538	0.88	0.97
	1	5.87	549	2117	0.80	0.96	786	1880	0.81	0.99
RNRESIN	2	4.48	376	2290	0.98	1.00	777	1889	0.94	0.99
	3	5.87	253	2413	0.81	0.97	928	1738	0.95	0.96
	4	5.47	182	2484	0.99	1.00	993	1673	0.90	0.98
	1	7.78	532	2138	0.97	0.99	757	1913	0.89	0.95
RRESINV	2	9.97	381	2289	0.94	0.98	939	1731	0.93	0.96
	3	11.25	284	2386	0.98	0.99	1083	1587	0.92	0.95
	4	12.10	208	2462	0.99	0.99	1132	1538	0.96	0.98
	1	6.49	327	2257	0.87	0.98	1364	1220	1.08	1.04
RFEDGOV	2	6.26	315	2269	0.99	0.99	1125	1459	1.00	0.98
	3	6.52	270	2314	1.01	1.00	1130	1454	0.96	0.98
	4	6.33	217	2367	0.99	0.99	1103	1481	0.96	0.98
	1	0.61	436	2158	0.84	0.94	1043	1551	0.85	0.94
RSLGOV	2	0.54	361	2233	0.86	0.95	1009	1585	0.84	0.96
	3	0.59	279	2315	0.76	0.95	1044	1550	0.91	0.96
	4	0.52	256	2338	0.93	0.97	1113	1481	0.90	0.96
	1	0.11	601	2200	0.89	0.96	844	1957	0.78	0.93
PGDP	2	0.13	450	2351	0.83	0.97	1000	1801	0.88	0.98
	3	0.14	371	2430	0.92	0.97	1069	1732	0.80	0.94
	4	0.14	274	2527	0.91	0.97	1167	1634	0.82	0.94
	1	1.87	616	2119	1.10	1.06	720	2015	1.14	1.07
CPI_1	2	2.99	404	2331	1.00	1.00	818	1917	1.01	0.99
	3	3.21	272	2463	0.99	0.99	837	1898	0.96	0.98
	4	3.36	182	2553	1.00	1.00	798	1937	0.98	0.99
	1	1.87	486	2249	1.07	1.05	669	2066	1.16	1.07
CPI_2	2	2.99	274	2461	1.00	1.00	785	1950	1.01	0.99
	3	3.21	190	2545	1.00	1.00	801	1934	0.95	0.97
	4	3.36	119	2616	1.00	1.00	769	1966	0.99	0.99

The column 'btt-ratio MSFE' is the ratio of the MSFE when the btt-forecasts are replaced by counterfactual non-btt forecasts (as explained in the text) to the MSFE of the reported individual forecasts. The column 'btt-ratio MAE' is the same but using MAEs, rather than MSFEs. The columns 'os-ratio MSFE' and 'os-ratio MAE' are the same when the os-forecasts are replaced by non-os forecasts.

Table 6: Individual accuracy, 1990:Q3 – 2008:Q4

	h	MSFE	# btt	# non-btt	btt-ratio MSFE	# os	# non-os	os-ratio MSFE
	1	0.22	504	1663	0.94	593	1574	1.03
RGDP	2	0.25	387	1780	0.98	734	1433	0.93
	3	0.27	293	1874	0.96	807	1360	0.96
	4	0.27	201	1966	0.97	901	1266	0.92
	1	0.26	315	1809	0.89	874	1250	0.99
RCONSUM	2	0.28	266	1858	0.95	839	1285	0.91
	3	0.27	177	1947	0.94	916	1208	0.89
	4	0.26	151	1973	0.98	945	1179	0.94
	1	3.23	389	1698	0.98	612	1475	0.98
RNRESIN	2	3.79	268	1819	0.98	606	1481	0.96
	3	4.13	177	1910	0.97	733	1354	0.91
	4	4.29	138	1949	0.98	760	1327	0.96
	1	5.54	397	1695	0.98	591	1501	0.89
RRESINV	2	7.40	260	1832	0.95	741	1351	0.92
	3	8.69	189	1903	0.99	872	1220	0.91
	4	9.57	138	1954	0.99	913	1179	0.94
	1	3.51	280	1739	0.95	1027	992	1.00
RFEDGOV	2	3.74	249	1770	0.94	854	1165	0.89
	3	3.73	210	1809	0.99	886	1133	0.93
	4	3.61	160	1859	0.98	869	1150	0.96
	1	0.50	308	1709	0.86	853	1164	0.87
RSLGOV	2	0.47	257	1760	0.88	816	1201	0.84
	3	0.48	185	1832	0.78	865	1152	0.94
	4	0.40	163	1854	0.97	908	1109	0.95
	1	0.08	403	1696	0.91	653	1446	0.71
PGDP	2	0.07	304	1795	0.98	785	1314	0.92
	3	0.08	255	1844	0.95	810	1289	0.83
	4	0.09	186	1913	0.95	865	1234	0.81
	1	1.55	448	1655	1.08	571	1532	1.17
CPI_1	2	2.46	297	1806	1.00	645	1458	0.97
	3	2.52	207	1896	0.99	673	1430	0.93
	4	2.65	138	1965	1.00	634	1469	0.98
	1	1.55	361	1742	1.07	552	1551	1.17
CPI_2	2	2.46	203	1900	1.00	650	1453	0.98
	3	2.52	141	1962	0.99	661	1442	0.93
	4	2.65	88	2015	0.99	623	1480	0.98

See notes to table 5.

Table 7: $\left|y_{t+h} - y_{i,t-1+h|t-1}\right| - \left|y_{t+h} - \widetilde{y}_{i,t-1+h|t-1}\right|$ on $\left|y_{i,t-1+h|t-1} - \widetilde{y}_{i,t-1+h|t-1}\right|$

	h	btt and os		btt		os	
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
RGDP	1	-0.04	0.32	-0.02	0.24	-0.02	0.35
		0.00	0.00	0.02	0.00	0.02	0.00
	2	-0.03	0.59	-0.01	0.47	-0.02	0.67
		0.00	0.00	0.07	0.00	0.00	0.00
	3	-0.04	0.75	-0.01	0.68	-0.03	0.76
		0.00	0.00	0.02	0.00	0.00	0.00
	4	-0.03	0.69	0.00	0.70	-0.02	0.66
		0.00	0.00	0.00	0.00	0.00	0.00
RCONSUM	1	-0.07	0.34	-0.01	0.33	-0.05	0.29
		0.00	0.00	0.01	0.00	0.01	0.00
	2	-0.04	0.52	-0.01	0.44	-0.03	0.52
		0.00	0.00	0.03	0.00	0.00	0.00
	3	-0.02	0.58	0.00	0.59	-0.01	0.54
		0.01	0.00	0.01	0.00	0.07	0.00
	4	-0.03	0.64	0.00	0.56	-0.02	0.65
		0.00	0.00	0.00	0.00	0.00	0.00
RNRESIN	1	-0.32	0.77	-0.11	0.73	-0.20	0.75
		0.00	0.00	0.00	0.00	0.00	0.00
	2	-0.08	0.31	-0.04	0.25	-0.04	0.30
		0.01	0.00	0.02	0.00	0.12	0.00
	3	-0.16	0.81	-0.06	0.90	0.00	0.21
		0.00	0.00	0.00	0.00	0.79	0.00
	4	-0.09	0.63	-0.01	0.30	-0.07	0.68
		0.00	0.00	0.01	0.00	0.00	0.00
RRESINV	1	-0.05	0.25	-0.01	0.09	-0.04	0.39
		0.29	0.00	0.65	0.00	0.15	0.00
	2	-0.07	0.26	-0.02	0.25	-0.04	0.25
		0.31	0.00	0.33	0.00	0.45	0.00
	3	-0.04	0.28	-0.02	0.23	-0.03	0.30
		0.52	0.00	0.22	0.00	0.64	0.00
	4	-0.02	0.20	-0.01	0.14	-0.02	0.23
		0.75	0.00	0.58	0.00	0.77	0.00
RFEDGOV	1	-0.22	0.21	-0.08	0.71	0.05	-0.23
		0.01	0.00	0.00	0.00	0.39	0.00
	2	-0.02	0.10	0.02	-0.09	-0.04	0.19
		0.65	0.00	0.36	0.00	0.18	0.00
	3	-0.04	0.20	0.00	-0.05	-0.03	0.26
		0.12	0.00	.	.	0.16	0.00
	4	-0.07	0.50	-0.02	0.52	-0.05	0.46
		0.00	0.00	0.00	0.00	0.02	0.00

Table continued on next page

Table 8: Table continued

Table continued

	h	btt and os		btt		os	
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
	1	-0.06	0.64	-0.02	0.70	-0.03	0.51
		0.00	0.00	0.00	0.00	0.00	0.00
RSLGOV	2	-0.06	0.75	-0.02	0.76	-0.04	0.71
		0.00	0.00	0.00	0.00	0.00	0.00
	3	-0.05	0.84	-0.02	0.91	-0.01	0.53
		0.00	0.00	0.00	0.00	0.01	0.00
	4	-0.02	0.66	-0.01	0.64	-0.01	0.64
		0.00	0.00	0.03	0.00	0.00	0.00
	1	-0.03	0.66	-0.01	0.61	-0.01	0.64
		0.00	0.00	0.00	0.00	0.00	0.00
PGDP	2	-0.03	0.71	-0.01	0.76	-0.02	0.59
		0.00	0.00	0.00	0.00	0.00	0.00
	3	-0.02	0.78	-0.01	0.70	-0.02	0.79
		0.00	0.00	0.00	0.00	0.00	0.00
	4	-0.01	0.70	0.00	0.64	-0.01	0.73
		0.00	0.00	0.00	0.00	0.00	0.00
	1	0.07	-0.56	0.01	-0.48	0.05	-0.58
		0.02	0.00	0.50	0.00	0.01	0.00
CPI_1	2	0.00	0.09	0.00	0.02	0.00	0.11
		0.96	0.00	0.77	0.41	0.89	0.00
	3	0.00	0.26	0.00	0.26	0.00	0.26
		1.00	0.00	0.74	0.00	0.88	0.00
	4	0.00	0.17	0.00	0.01	0.00	0.22
		0.71	0.00	0.97	0.69	0.80	0.00
	1	0.06	-0.54	0.00	-0.42	0.05	-0.60
		0.05	0.00	0.83	0.00	0.02	0.00
CPI_2	2	0.00	0.06	0.00	-0.06	0.00	0.11
		1.00	0.00	0.97	0.00	0.99	0.00
	3	0.00	0.23	0.00	0.18	0.00	0.25
		0.80	0.00	0.69	0.00	0.67	0.00
	4	0.01	0.05	0.00	0.08	0.01	0.05
		0.11	0.00	0.79	0.00	0.06	0.01

The table reports the parameter estimates, with the corresponding p-values of the null that $\beta = 0$ (against a two-sided alternative) directly below.

6 Data Appendix

6.1 SPF data

The Survey of Professional Forecasters (SPF) began in 1968 as the ASA-NBER Survey of Forecasts by Economic Statisticians, administered by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER). Since June 1990 it has been run by the Philadelphia Fed, renamed as the SPF: see Zarnowitz (1969), Zarnowitz and Braun (1993) and Croushore (1993). Because it is a survey of professional forecasters, authors such as Keane and Runkle (1990) have argued that one can reasonably assume that the reported forecasts reflect the forecasters' expectations, which might not be true when ordinary individuals and firms are surveyed.

We use the point forecasts of a number of macro variables from the surveys from 1981:3 to 2008:4. For these surveys we have, for a number of key macrovariables, individual respondents' point forecasts for the previous quarter, the current quarter, and each of the next four quarters: see the online documentation provided by the Philadelphia Feb: 'Documentation for the Philadelphia Fed's Survey of Professional Forecasters', http://www.phil.frb.org/econ/spf/. The forecast data were downloaded from the SPF web page in January 2009.

6.2 Real-time datasets

The real-time data were taken from the Real-Time Data for Macroeconomists (RTDSM) provided on http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/, see Croushore and Stark (2001). All data except for the CPI were reported as quarterly vintages of quarterly observations. For the CPI, quarterly vintages of monthly observations are provided. We averaged the months to obtain quarterly series, and took the quarterly vintage two quarters after the quarter being forecast (as for the quarterly-vintage data) as the measure of the actual, although prior to 1994:Q3 the CPI data were not revised.

The variables we use, their SPF mnemonics, and their names in the RTDSMs are listed in table 9.

Table 9: Macroeconomic Variables in the SPF

Variable	SPF code	RTDSM code
Real GDP (GNP)	RGDP	ROUTPUT
Real personal consumption	RCONSUM	RCON
Real nonresidential fixed investment	RNRESIN	RINVBF
Real residential fixed investment	RRESINV	RINVRESID
Real federal government expenditure	RFEDGOV	RGF
Real state and local government	RSLGOV	RGSL
GDP price index (implicit deflator, GNP deflator)	PGDP	P
CPI inflation rate	CPI	CPI

7 Appendix: The estimation of the pooled regression.

To get the 'correct' standard errors for the regression (4) that pools over i and t (for a given h) we adapt the approach of Keane and Runkle (1990) and Bonham and Cohen (2001). Specifically, we allow for the overlapping nature of forecasts and for the dependence in forecast errors across individuals resulting from common macro shocks. From section 4 we assume that for an individual i:

$$\begin{aligned} E\left[\varepsilon_{it}^2\right] &= \sigma_0^2 \\ E\left[\varepsilon_{it}\varepsilon_{i,t+k}\right] &= \sigma_k^2 \text{ when } 0 < k \leq h, \text{ and } 0 \text{ otherwise} \end{aligned}$$

and for any pair of individuals i, j:

$$\begin{aligned} E\left[\varepsilon_{it}\varepsilon_{jt}\right] &= \delta_0^2 \\ E\left[\varepsilon_{it}\varepsilon_{j,t+k}\right] &= \delta_k^2 \text{ when } 0 < k \leq h, \text{ and } 0 \text{ otherwise.} \end{aligned}$$

The disturbances depend on the horizon h but this is left implicit to simplify the notation. In (4), ε_{it} will be correlated with $\varepsilon_{i,t+1}$ even for $h = 1$ step forecasts, i.e., even adjacent period forecast errors will be correlated. This is because we use revised actuals (e.g., y_t^{t+2}). So the error in forecasting

y_t^{t+2} when the forecast is made in the quarter t survey $(y_t^{t+2} - y_{t|t-1})$ will be correlated with the next survey's forecast error $(y_{t+1}^{t+3} - y_{t+1|t})$, because next period's forecast will be conditioned on (say) y_t^{t+1} (not the outcome y_t^{t+2}). But for the following forecast $E[\varepsilon_{it}\varepsilon_{it+2}] = 0$ because the forecast errors being compared are the original $(y_t^{t+2} - y_{t|t-1})$ and $(y_{t+2}^{t+4} - y_{t+2|t+1})$, such that y_t^{t+2} will be known before the forecast $y_{t+2|t+1}$ is made.

When any two forecasts are made by the same individual i, the covariances are σ_k^2; when by any two different individuals, by δ_k^2.

When $h = 4$, forecasts up to one year apart will still be correlated. For example, two forecasts made in the same quarter of the year in adjacent years would be $\left(y_{t+3}^{t+5} - y_{t+3|t-1}\right)$ and $\left(y_{t+7}^{t+9} - y_{t+7|t+3}\right)$. The later forecast $y_{t+7|t+3}$ contains data up to y_{t+3}^{t+4}, which does not include the original actual (y_{t+3}^{t+5}), so these two forecasts 'overlap'. When $k = 5$, $E[\varepsilon_{it}\varepsilon_{it+k}] = 0$, as e.g., $\left(y_{t+3}^{t+5} - y_{t+3|t-1}\right)$ and $\left(y_{t+8}^{t+10} - y_{t+8|t+4}\right)$ are non-overlapping. $(y_{t+8|t+4}$ conditioned on $y_{t+4}^{t+5})$.

Richer assumptions are possible, allowing σ_k^2 to be individual specific, and putting some structure on how σ_k^2 and δ_k^2 vary over k (see, for example, Davies and Lahiri (1995)) but the above makes for a relatively simple covariance structure given the highly unbalanced nature of our panel.

We follow Keane and Runkle (1990) and estimate σ_k^2 and δ_k^2, $k = 0, \ldots, h$, from the residuals of the pooled OLS regression (which imposes microhomogeneity: the same intercepts and slope parameters over all individuals), whereas Bonham and Cohen (2001) use the residuals from separate regressions for each individual. Hence:

$$\hat{\sigma}_0^2 = \frac{1}{\overline{T}} \sum_{i=1}^{N} \sum_{t_i} \hat{\varepsilon}_{it_i}^2$$

where t_i runs over all the surveys to which i responded, T_i is the number of forecasts made by i, $\overline{T} = \sum_{i=1}^{N} T_i$. Similarly:

$$\hat{\sigma}_k^2 = \frac{1}{\overline{T}} \sum_{i=1}^{N} \sum_{t_i} \hat{\varepsilon}_{it_i}\hat{\varepsilon}_{it_i-k}, \quad k = 1, \ldots, h$$

where now t_i indexes all the surveys for i for which responses were made to two surveys k-periods apart. (T_i and hence $\overline{T}$ will typically depend on k, but this is suppressed for notational convenience).

Further:

$$\hat{\delta}_0^2 = \frac{1}{\overline{\overline{T}}} \sum_{i=1}^{N} \sum_{\substack{j=1, \\ j \neq i}}^{N} \sum_{t_{ij}} \hat{\varepsilon}_{it_{ij}} \hat{\varepsilon}_{it_{ij}}$$

where t_{ij} runs over all the surveys to which i and j responded, T_{ij} is the number of such forecasts, and $\overline{\overline{T}} = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} T_{ij}$. Then finally, in obvious notation:

$$\hat{\delta}_k^2 = \frac{1}{\overline{\overline{T}}} \sum_{i=1}^{N} \sum_{\substack{j=1, \\ j \neq i}}^{N} \sum_{t_{ij}} \hat{\varepsilon}_{it_{ij}} \hat{\varepsilon}_{it_{ij}+k},\ k = 1, \ldots, h$$

We can then construct the estimator $\hat{\Sigma}$ of $\Sigma = E\left(\varepsilon\varepsilon'\right)$, where $\varepsilon = [\varepsilon_{11}\ \varepsilon_{12}\ \ \ldots \varepsilon_{1T}; \ldots; \varepsilon_{N1}\ \varepsilon_{N2}\ \ \ldots \varepsilon_{NT}]'$, using $\hat{\sigma}_k^2$ and $\hat{\delta}_k^2$, $k = 0, 1, \ldots, h$. Note that Σ (and the estimator $\hat{\Sigma}$) correspond to a balanced panel of forecasters. Write the model as:

$$Y = X\gamma + \varepsilon$$

where Y and X are ordered conformably with ε (all the time observations on individual 1, then on individual 2 etc.) and where X has two columns, the first being the intercept, and $\gamma = (\alpha\ \beta)'$. $\hat{\gamma}$ is obtained by deleting the rows of Y and X corresponding to missing observations (as in the calculation of the $\hat{\varepsilon}_{it}$ residuals). The covariance matrix for $\hat{\gamma}$ is given by the usual formula $\left(X'X\right)^{-1} X'\hat{\Sigma}X \left(X'X\right)^{-1}$ where X is again compressed to eliminate missing values, and the corresponding rows (and equivalent columns) are deleted from $\hat{\Sigma}$.

References

Ang, A., Bekaert, G., and Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better?. *Journal of Monetary Economics*, **54**, 1163–1212.

Bonham, C., and Cohen, R. (2001). To aggregate, pool, or neither: Testing the rational expectations hypothesis using survey data. *Journal of Business and Economic Statistics*, **190**, 278–291.

Clark, T. E., and McCracken, M. W. (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. In Rapach, D. E., and Wohar, M. E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty. Frontiers of Economics and Globalization. Volume 3*, pp. 93–147: Emerald.

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, **26**, 546–554. No. 4.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, **November/December**, 3–13.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105**, 111–130.

Davies, A., and Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, **68**, 205–227.

Keane, M. P., and Runkle, D. E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, **80**, 714–735.

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal of Economic Perspectives*, **22**, 193–216.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998). Forecasting the U.S.

unemployment rate. *Journal of the American Statistical Association*, **93**, 478–493.

Romer, C. D., and Romer, D. H. (2000). Federal Reserve information and the behaviour of interest rates. *American Economic Review*, **90**, 429–457.

Stock, J. H., and Watson, M. W. (2008). Phillips Curve Inflation Forecasts. Working paper 14322, NBER, Cambridge, MA.

Zarnowitz, V. (1969). The new ASA-NBER Survey of Forecasts by Economic Statisticians. *The American Statistician*, **23, No. 1**, 12–16.

Zarnowitz, V., and Braun, P. (1993). Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In Stock, J., and Watson, M. (eds.), *Business Cycles, Indicators, and Forecasting*, pp. 11–84: Chicago: University of Chicago Press and NBER.