Characterising how surveys of US macrovariables add value.

Michael P. Clements

Department of Economics

University of Warwick*

May 15, 2012

Abstract

We investigate two characteristics of consensus survey forecasts which are shown to contribute to their accuracy. Consensus forecasts incorporate the effects of perceived changes in the long-run outlook. They also feature departures from the path toward the long-run expectation. At the level of the individual forecasts, there is some evidence that the departures from the long-run path enhance accuracy, but to a lesser extent.

Keywords: consensus forecast, long-run expectations, forecast path.

JEL classification: C53, E37.

*An earlier version of this paper was circulated as 'Why are survey forecasts superior to model model forecasts?', Warwick Economic Research paper No. 954. I am grateful to Dean Croushore and Tom Stark for helpful comments on the way in which the SPF forecast data is collected.

1 Introduction

Survey expectations are sometimes 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.' This quotation from Ang *et al.* (2007) is rather general, and we try and unpack it. Are there specific characteristics of survey forecasts that can be shown to contribute to their accuracy? 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. The forecast of a stationary process will approach the unconditional mean of the process as the forecast horizon increases. Model-based forecasts share this characteristic to the extent that they are of the equilibrium-correction class, see e.g., Clements and Hendry (2006), who establish the generality of this class of model. Models are backward-looking in that they project forward past patterns in the data, such as the mean of the data during the estimation period. 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. Hence models fail to 'quickly adapt' to the changed circumstances. So survey forecasts might benefit from the incorporation of a forward-looking element especially at times of 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 is intended to capture perceived changes in the long-run mean and counter the tendency of the models' forecasts to 'correct' towards the long-run mean that characterized the past data. Patton and Timmermann (2010, 2011) argue that in order to explain disagreement between forecasters at longer horizons it is reasonable to assume that forecasters have different views about the long-run values of variables, and that changes in these beliefs would lead to changes in their shorter-horizon forecasts. Our first aim is to determine whether survey forecasts do incorporate perceived changes in the outlook which contribute to more accurate (shorter-term) forecasts.

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

Secondly, in the absence of useful information about future events at the time the forecasts are made, we would generally expect that the sequence of 1 to h-step ahead forecasts would approach the long-run expectation in a (more or less) monotonic fashion as the forecast horizon increases, irrespective of whether the long-run expectation later turns out to be a poor forecast. Suppose, as a simple illustration, that inflation is currently lower than expected, but that a forecaster expects higher inflation in six months time (due to an expected temporary rise in commodity prices, or a pre-election spending boom, etc) but that inflation is then expected to slow to the monetary authorities' long-run or target rate. Inflation will then be forecast to 'overshoot' its long-run expected value over the medium term. Knowledge of future developments of the sort described would contribute to the good performance of survey expectations documented by Ang *et al.* (2007). Our aim is to assess whether such departures from the path to the long-run position enhance forecast accuracy, or simply constitute noise: do they represent knowledge of future developments, or are they spurious in the sense that they are uninformative about the future course of the variable over the medium term? Thus, we analyze the term structure of forecasts from 1 to h-steps ahead made from each 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 this is determined. Whether or not such forecasts enhance accuracy is an empirical question. Forecasters might find it difficult to do better than follow 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 3), and the median forecasts of the year-ahead annualized 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 Forecasters. 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

[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.

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 the macroeconomic outlook.

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 values 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. 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 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, and occur in 'normal times' as well as periods of financial turmoil. Of interest is whether NC-forecasts generally improve forecast accuracy. In addition to real output growth and CPI inflation, we include in our study consumption and investment, the GDP deflator measure of inflation, the unemployment rate and the three-month Treasury bill rate. Hence we consider a range of important macroeconomic variables. We also consider 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

[3] Of course 'the' model forecasts could be made more adaptive by considering models with time-varying parameters and a rolling forecasting scheme - as explained in section 3, 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.

behavioural interpretation.[4]

There is a large recent literature on expectations formation, but this is nearly orthogonal to the issues considered in this paper. Two prominent strands emphasize information rigidities when agents update their forecasts: sticky information (see, e.g., Mankiw and Reis (2002), Mankiw, Reis and Wolfers (2003) and Coibion and Gorodnichenko (2011)) and noisy information (see, e.g., Woodford (2001) and Sims (2003)). Other authors consider forecasts of a fixed target as the forecast horizon shortens: Patton and Timmermann (2010, 2011). These papers consider the properties of sequences of forecasts of a fixed length as the forecast origin moves forward, or of a fixed target, as new information is acquired through time. Our interest is in the 'shape' of the forecast path as the forecaster looks further ahead from a given forecast origin.

The plan of the remainder of the paper is as follows. Section 2 describes the US Survey of Professional Forecasters (SPF) data and the reporting practices which might have a bearing on our findings. Section 3 considers 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 4 defines NC-forecasts and the impact of such behaviour on the accuracy of the consensus forecast. Section 5 analyses whether the accuracy-enhancing characteristics of the consensus forecast are also found at the level of the individual forecasts. Section 6 provides further discussion of some of the differences we find across variables. Section 7 offers some concluding remarks. Details on the pooled regression estimators of section 5 are provided in an Appendix.

2 Nature of the SPF forecast 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).[5] 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.

[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.

[5]Full details are provided by the Philadelphia Fed: 'Documentation for the Philadelphia Fed's Survey of Professional Forecasters', http://www.phil.frb.org/econ/spf/.

Our sample consists of the quarterly SPF surveys from 1981:Q3 to 2008:4. For these surveys we have, for a number of key macrovariables, individual respondents' 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 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[6], and calculate the consensus forecasts from this subset.

We analyze the forecasts of the following variables: real GDP (RGDP), consumption (RCONSUM), non-residential investment (RNRESIN), the GDP deflator (PGDP), CPI inflation, the unemployment rate (UNEMP), and the three-month Treasury bill rate (TBILL). All these variables apart from TBILL are subject to revision, so we use 'real-time data'[7] throughout to ensure that forecasters are not assumed to have information that would not have been available at the time. Full descriptions of the variables, and their identifying mnemonics in the SPF and Real-Time Dataset for Macroeconomists (RTDSM) are in table 1.

The reporting practices for the variables are as follows. From 1990:Q3 onwards, respondents have had the option of reporting their forecasts of RGDP and its components as either levels (L) or growth rates (G) (prior to 1990:3 they were reported as levels). Whether reported as L or G, they are recorded in the forecast database as levels. From the recorded levels, we analyze RGDP and its components as quarter-on-quarter percentage growth rates. The CPI forecasts are reported as annualized quarter-over-quarter percentage changes. We also analyze PGDP (reported as either L or G, as for RGDP and its components) as annualized quarter-over-quarter percentage changes, so that it is directly comparable to CPI (in terms of MSE of the forecasts). UNEMP and TBILL are both reported by the respondents, and recorded in the SPF, as percentages, and we analyze them as percentages. The relevance of the way the forecasts are reported is considered below.

3 Median SPF forecasts and the changed outlook

Our first hypothesis is that the accuracy of the survey forecasts relative to the benchmark forecasts will improve when there are changes in the long-run outlook (as measured by the survey forecasts). The

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

[7]This is taken from taken from the Real-Time Data for Macroeconomists (RTDSM) provided at http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/, see Croushore and Stark (2001).

benchmark forecasts serve to control for changes in predictability over time, e.g., for the possibility that changes in the long-run outlook are associated with greater uncertainty about the future, so that the accuracy of the shorter-term survey forecasts actually worsens during these times relative to when the long-run outlook is more stable. Our benchmark forecasts are simple autoregressive model forecasts. The models are 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 RTDSM. To make the timing conventions clear, consider the model-based forecasts corresponding to the 1981:Q3 survey forecasts. The model is 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 2 provides a comparison of the accuracy of the median survey forecasts and the benchmark forecasts on MSFE. The first three columns report the MSFEs for each forecast horizon, and the ratio of the median survey to the benchmark (the remaining columns will be explained later). The survey forecasts are markedly more accurate at the shorter horizons for the first five variables (RGDP down to CPI): at $h = 1$ the median survey forecast MSFEs range from as little as two fifths to three-quarters of the model MSFEs. For UNEMP and TBILL the $h = 1$ ratios are as little as 0.22 and 0.05, and are only around two thirds of the model MSFEs at the year ahead horizon, $h = 5$. The MSFEs are computed using estimates of the actual values published in the second quarter following the data being forecasted.[8]

The greater accuracy of the survey forecasts is unsurprising given the timeliness and breadth of the information they draw on, and is not the focus of this paper.[9] Our interest is in determining whether we

[8]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 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.

[9]In terms of timeliness, note that the survey responses for quarter t are filed around the middle of the middle month of quarter t, whilst the model forecasts use data only through quarter $t-1$ (and only on the variable being forecast). Information on the quarter being forecast would be expected to improve the model forecasts of that quarter (see, for example, Montgomery, Zarnowitz, Tsay and Tiao (1998) and Clements and Galvão (2008)). For UNEMP and TBILL the timing of the publication of data is such that the survey respondents will know the value for the first month of the survey quarter when they file their forecasts (the same is not true for CPI). Faust and Wright (2009) consider the issue of timeliness within a more general study that compares the US Greenbook inflation and output growth forecasts to large dataset methods and simple univariate forecasting methods. They find that once the univariate model forecasts are adjusted for the Greenbook's use of knowledge of the current

can identify simple characteristics of the survey forecasts which contribute to their enhanced accuracy. We know that survey forecasts are based on superior information sets to any purely model-based forecast, and will incorporate judgement. Can this superior information be shown to improve short-term forecasts by foreseeing changes in the long-run outlook (this section), or is it due to induced departures from the equilibrium path (the following section)? Better model forecasts could be obtained, but however sophisticated their forecasts would not be informed by perceived changes in the long-run outlook.[10]

Table 3 presents evidence on the association between short-term survey-forecast accuracy (relative to the benchmark) and changes in the longer-term outlook. It records the ratios of the MSFEs of the survey forecasts to the benchmark forecasts, for a particular forecast horizon, and for a subset of the forecasts. For each variable, the first row is a comparison for all survey forecasts, and the second row is calculated for the subset of forecasts for which the corresponding long-run median forecast changed relative to the previous forecast value by more than one standard deviations. For RCONSUM, RNRESIN and UNEMP we observe a sizeable improvement in the relative accuracy of the 1-step survey forecasts (compared to the benchmarks) associated with one standard deviation changes in the long-run expectation. There are also improvements to the 1-step forecasts of RGDP and PGDP of a smaller size. The exceptions are CPI and TBILL. We note that the improvements are rarely large beyond 1-step ahead, except for UNEMP. (We also considered changes in long-run expectations in excess of two standard deviations, but there were generally too few instances to draw reliable conclusions). One might object to describing the 5-quarter ahead forecasts as long-run forecasts. These are the longest horizon forecasts provided by the SPF for all variables over our sample period,[11] but actual 'long-run forecasts' may not be very different. The SPF provides median forecasts of the average rate of CPI inflation over the next ten years. Comparing these with the median five-quarter ahead forecast of the CPI inflation rate for the period for which they are available (1991:Q4 to 2010:Q4) gave a correlation of 0.93.

Table 4 present the results of a more formal analysis that by and large confirms the findings of table 3. We regress the difference in accuracy of the short-term survey forecasts and the benchmark forecasts on the change in the long-run outlook as measured by the longer-horizon survey forecasts. Because the forecasts of quarterly growth rates can be quite volatile, we take as a smoother measure the sum of

state of the economy, these simple models are competitive for forecasting output growth, but not inflation.

[10]In terms of forecasting inflation, it may be difficult to find a much better forecasting model - Stock and Watson (2010) claim that 'it is exceedingly difficult to improve systematically upon simple univariate forecasting models, such as the Atkeson-Ohanian (2001) random walk model (although that model seems to have broken down in the 2000s) or the time-varying unobserved components model in Stock and Watson (2007)'.

[11]Forecasts of the annual value for the year after the survey quarter are provided, which would provide longer forecasts for the earlier quarters of the year, but does not allow the calculation of a series of longer-horizon rolling forecasts.

the current- and next-quarters' forecasts, which corresponds approximately to the half-yearly percentage change (between $t-1$ and $t+1$). Similarly, the measure of the longer-run outlook is the change between surveys in the half-yearly growth rate one year ahead. To be exact, the regression we estimate by OLS is given by:

$$\left|y_t^{t+2} - y_{t|t-1}^{Med} + y_{t+1}^{t+3} - y_{t+1|t-1}^{Med}\right| - \left|y_t^{t+2} - y_{t|t-1}^{BM} + y_{t+1}^{t+3} - y_{t+1|t-1}^{BM}\right|$$
$$= \beta_1 + \beta_2 \left|\sum_{i=3}^{4} \left(y_{t+i|t-1}^{Med} - y_{t-1+i|t-2}^{Med}\right)\right| + \zeta_t \tag{1}$$

where t runs over the surveys from 1981:Q4 onwards, and we use HAC (heteroscedasticity and autocorrelation consistent) standard errors to account for possible heteroscedasticity and the overlapping nature of the forecasts (we set the autoregressive order correction to 4). As indicated by (1), forecast accuracy is mean absolute error (MAE) as this is more robust than mean squared error (MSE) to occasional large forecast errors. The dependent variable consists of the difference in accuracy between the survey and benchmark forecasts (y^{Med} and y^{BM}, respectively) of the current and next quarter ($y_{t|t-1}$ and $y_{t+1|t-1}$). y_t^{t+2} and y_{t+1}^{t+3} are estimates of the value of y in quarters t and $t+1$ (the subscript) taken from the data vintage available two quarters later (in quarters $t+2$ and $t+3$, respectively: the superscript). The explanatory variable is the absolute value of the difference between the longer-horizon forecasts issued from survey quarter t relative to the forecasts made in the previous survey quarter $t-1$.[12]

We also consider a variant where the dependent variable is the half-yearly growth rate between the current quarter and two quarters ahead (so the dependent variable is defined as $\left|y_{t+1}^{t+3} - y_{t+1|t-1}^{Med} + y_{t+2}^{t+4} - y_{t+2|t-1}^{Med}\right| - \left|y_{t+1}^{t+3} - y_{t+1|t-1}^{BM} + y_{t+2}^{t+4} - y_{t+2|t-1}^{BM}\right|$, with the explanatory variable unchanged).

For all the variables other than CPI we obtained negative estimates of β_2. For RCONSUM, RNRESIN and TBILL these were statistically significant at conventional significance levels, and for UNEMP at the 10% level. See table 4. A negative β_2 indicates that changes in the long-run outlook are associated with more accurate short-horizon forecasts being issued at the same time. For RGDP β_2 is large and negative, but not significant. For CPI it is positive but not significant. We have highlighted the results for (1), whereas for the dependent variable defined as the t to $t+2$ growth rate there is less evidence of β_2 being negative.

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

These results generally agree with those in table 3. For all variables other than CPI inflation the short-run consensus forecasts are more accurate (relative to the benchmark) at times of changes in long-run expectations.

4 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.

4.1 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:

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

where y_{t-1} is the value in the period prior to the survey quarter.

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:

$$y_{t-1+h|t-1} : y_{t-1} < y_{t+4|t-1} < y_{t-1+h|t-1} \text{ or: } y_{t-1} > y_{t+4|t-1} > y_{t-1+h|t-1}. \tag{3}$$

In the empirical work, we take y to be the quarter-on-quarter growth rate for RGDP, RCONSUM and RNRESIN. 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, because the growth rates will have well-defined long-run expectations. Similarly, for PGDP and CPI y refers to the annualized quarter-on-quarter growth rate, because CPI is reported in this form. For both UNEMP and TBILL we assume that y is the level (i.e., the percentage value). These transformations of the variables are also used for evaluating the accuracy of the forecasts.[13]

4.2 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 5 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 $h = 4$. For CPI inflation, about an eighth of the forecasts are characterized as btt at $h = 1$.

In terms of os, we find quite different patterns: generally 20 to 40% of the median survey forecasts are os, and the proportion does not depend upon the horizon. The exception in UNEMP - only around 10% of the shorter-horizon consensus forecasts are os.

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 degree of predictability 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,

[13] For RGDP, RCONSUM, RNRESIN and PGDP, we take the value of the quarterly growth for y_{t-1} from the survey-quarter RTDSM data vintage (see Data Appendix), so it is from the vintage of data that would have been available to the respondent to survey t. For CPI, UNEMP and TBILL the value of y_{t-1} is provided by the respondent as their 'estimate' for the quarter prior to the survey quarter.

$\widetilde{y}_{t|t-1}$, 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 $y_t > \frac{1}{2}\left(y_{t-1} + y_{t|t-1}\right)$. This way of constructing counterfactuals has the merit of assessing whether the NC characteristic enhances accuracy relative to the closest forecast that does not possess this property.[14]

The columns of table 2 headed '$\tilde{y}_{btt}$' and '$\tilde{y}_{os}$' 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 the adjustment to the reported forecasts markedly worsens forecast accuracy. For both UNEMP and TBILL the NC-characteristics are found to be the most beneficial: the MSFE ratios of the smoothed to reported forecasts are larger than 1.5 for btt for $h = 1$, and for TBILL for os at $h = 1$ is 2.7. The NC-characteristics are also clearly valuable for these two variables beyond 1-step ahead. The respondents' forecasts of the quarterly values of both UNEMP and TBILL will be informed by the published values of these variables for the first month of the survey quarter,[15] indicating the usefulness of partial data on the variable for the quarter being forecast in general, and that the btt and os measures capture the way in which this information enhances forecast accuracy. For UNEMP, btt is more prominent, whereas the results in table 2 point to os for TBILL.

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. The ratio is close to one with few exceptions, so that the effect of NC-behaviour on the model forecasts is neutral. (The exception is TBILL, where removing os improves the model forecasts by around 10%, but this is dwarfed by the impact of os on the consensus forecasts).

One might suppose that survey respondents would have better information about the future course of the economy around business cycle turning points, and that it would be at such times that NC-forecasts would be made. To investigate this, figure 6 plots the relative accuracy of the one-step ahead median survey forecasts of output growth over time compared to the artificial forecasts for which the btt-characteristics of the survey forecasts is suppressed.[16] In the figure, the circles are the absolute errors

[14]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 $\left|y_{t-1+h|t-1} - \widetilde{y}_{t-1+h|t-1}\right|$.

[15]Both variables are published at the monthly frequency. Unlike CPI, the value for the first month of the quarter is published before the SPF returns are made around the middle of the second month of the quarter.

[16]We only present this figure for a single variable and forecast horizon to save space. We choose output growth, and consider

of the median forecasts with the btt-characteristic removed, minus the absolute errors of the original median forecasts. Positive values indicate that the original forecasts - with the btt-characteristic - are more accurate.[17] We also indicate on the same figure the NBER business cycle peaks and troughs - 1981:3 (P), 1982:3 (T), 1990:3 (P), 1991:1 (T), 2001:1 (P), 2001:4 (T). We find the btt-characteristic 'adds value' to the $h = 1$ forecasts of the three survey quarters that follow an NBER peak, and that the first two of these (the 1981:4 survey, following the 1981:3 peak, and the 1990:4 survey, following the 1990:3 peak) give the largest absolute gains to the median forecasts relative to the 'neutral' forecasts. However, as is evident from figure 6, the btt-characteristic does not solely capture superior information at or around turning points, as there are clear gains to forecasts with this property at other times. The predominance of circles with positive values points to the accuracy-enhancing nature of the btt characteristic.

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. There is evidence that changes in the long-run outlook improve accuracy for all the variables other than CPI, whereas the NC-characteristics generally enhance the accuracy of all the short-horizon consensus forecasts.

5 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 consensus forecasts of all the variables, and especially of CPI, UNEMP and TBILL. Table 6 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, declining to less than half this fraction at $h = 4$. Roughly one third of the forecasts of the variables other than UNEMP and TBILL are os, and this fraction is largely the same across forecast horizons. These findings are comparable to the findings for the consensus forecasts.

The results regarding the impact on forecast accuracy are recorded in table 7. 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

btt-forecasts at $h = 1$, because at this horizon approximately a quarter of median forecasts have the btt characteristic.

[17] The zero values are due to the median forecasts not possessing the btt characteristic.

the MSFE of the $\widetilde{y}$ forecasts to the MSFE of the reported forecasts. There is a clear dichotomy in the findings. The adjusted forecasts are more accurate than the originals for GDP, its components, and PGDP, indicating that individuals' NC-behaviour worsens forecast performance. But for CPI, UNEMP and TBILL the 'smoothed' counterfactual forecasts are markedly less accurate. The os CPI forecasts, and the btt UNEMP forecasts 'add value' at 1 and 2-steps ahead. The results for the individuals' forecasts of CPI, UNEMP and TBILL match those for the consensus forecasts, but for the other variables there is a qualitative difference between the individual and consensus forecasts.

The statistics reported in table 7 are the result of a fairly broadbrush approach. The MSFE 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). We adopt a forecast encompassing framework (see, e.g., Clements and Harvey (2009) for a recent review and exposition) to test for predictive ability, for the reasons explained below. We consider whether one forecast encapsulates all the useful predictive information contained in a second forecast. If the artificial forecasts encompass the reported forecasts, then the NC-behaviour is irrelevant in terms of squared-error loss forecast accuracy: once we have the artificial forecasts (no-NC property by construction), it is impossible to improve the accuracy of these forecasts by combining them linearly with forecasts that do possess the NC-property. Under forecast encompassing, the optimal weight on the reported forecasts in a linear combination with the artificial forecasts would be zero.

We use the form of the test proposed by Nelson (1972) and Granger and Newbold (1973) (see also Chong and Hendry (1986)), which regresses the error from one set of forecasts on the difference between the two sets of forecasts (or equivalently, the difference between the two sets of errors), namely:

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

Relative to the Fair and Shiller (1989) specification:

$$y_{t+h} = \alpha + \alpha_1 y_{i,t-1+h|t-1} + \alpha_2 \widetilde{y}_{i,t-1+h|t-1} + v_{ith},$$

regression equation (4) imposes the restriction that $\alpha_1 + \alpha_2 = 1$. The reported forecasts encompass the artificial when $\beta_2 = 0$, and encompassing holds in the reverse direction when $\beta_2 = 1$. The forecast encompassing framework is able to detect whether NC-behaviour is beneficial to *some degree*. To see this, note that the Artificial forecast can be written as:

$$\widetilde{y}_{i,t-1+h|t-1} \equiv y_{i,t-1+h|t-1} - nc_{i,t-1+h|t-1}$$

where $nc_{i,t-1+h|t-1} = y_{i,t-1+h|t-1} - \widetilde{y}_{i,t-1+h|t-1}$ is the non-zero difference between the Reported and the Artificial forecasts when the Reported has the NC-property. (For btt, for example, $nc_{i,t-1+h|t-1} = y_{i,t-1+h|t-1} - y_{t-1}$). The Artificial is adjusted by removing the whole of the nc term from the Reported. The forecast encompassing framework allows that the optimal combined forecast may remove only some fraction of nc. From (4) the combined forecast is (ignoring β_1):

$$\begin{aligned}(1-\beta_2)y_{i,t-1+h|t-1} + \beta_2\widetilde{y}_{i,t-1+h|t-1} &= (1-\beta_2)y_{i,t-1+h|t-1} + \beta_2\left(y_{i,t-1+h|t-1} - nc_{i,t-1+h|t-1}\right) \\ &= y_{i,t-1+h|t-1} - \beta_2 nc_{i,t-1+h|t-1}\end{aligned}$$

allowing that not removing all (when $\beta_2 < 1$) the NC-property may be optimal.

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:

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

Appendix A provides further details, and records how the model is estimated, and the calculation of the estimated covariance matrix of the disturbances given the unbalanced nature of the panel.

Table 8 reports the results of testing the null hypotheses that $\beta_2 = 0$, and that $\beta_2 = 1$, as well as the estimates of β_2. Consider the forecasts of real output growth. At $h = 1$, $\beta_2 = 0.66$, suggesting that the weights on the artificial and reported are about 2:1 in the optimal combination, but we reject the null hypotheses that $\beta_2 = 0$ and that $\beta_2 = 1$, so neither forecast encompasses the other. The NC-characteristic has some value. For longer horizon forecasts the size of the estimated coefficient β_2 increases, and we continue to reject $\beta_2 = 0$ but not that $\beta_2 = 1$, suggesting that the artificial forecasts encompass the reported. We observe this pattern whether we consider forecasts which are either btt or os, or those which are btt, or those which are os. A broadly similar pattern holds for RCONSUM, RNRESIN and PGDP, except for PGDP the Artificial encompass the Reported for all horizons ($\beta_2 = 1$), while we always reject the null of forecast encompassing in the reverse direction.

For CPI at $h = 1$ we find $\beta_2 \simeq 0.1$ (taking btt and os together): the reported forecasts encompass the artificial (we do not reject $\beta_2 = 0$, but do reject $\beta_2 = 1$). At $h = 2$, the weights on the two sets of forecasts are close to a half, and neither set encompasses the other. For longer horizon forecasts the artificial forecasts are preeminent as we found for the other variables. The results for UNEMP and TBILL clearly favour the reported forecasts, and not just at $h = 1$ as for CPI. (β_2 is generally not significantly different from zero).

The findings reported in table 8 broadly confirm those of table 7, but the forecast encompassing perspective softens the distinction between the NC-characteristic improving the consensus forecasts, but worsening the individual forecasts, for RGDP, RCONSUM, RNRESIN and PGDP. Finding β_2 less than 1 suggests that the NC-characteristics have some value in enhancing the individual forecasts, even though in a head-to-head comparison, the Artificial are more accurate on average than the Reported.

6 Discussion of the key findings

We consider possible explanations for the finding that btt and os enhance the accuracy of the individual respondents' forecasts for CPI, UNEMP and TBILL, but not for the expenditure variables and the GDP deflator. The obvious explanation is that CPI, UNEMP and TBILL are intrinsically more predictable - the forecasters have superior information on the likely direction and value of the variable over the short term than for the expenditure components. The forecast btt and os behaviour reflects this, and these aspects of the forecasts enhance their accuracy. The low MSFEs of the short-term forecasts of UNEMP and TBILL (both in absolute terms and relative to the benchmarks) are consistent with the predictability of these variables. However, CPI does not fit this pattern. In terms of the consensus forecasts, the MSFEs for the CPI are greater than for PGDP, so the value of btt/os does not simply depend on the predictability of the variable. Recall that the forecasts of PGDP are evaluated in terms of the annualized quarter-on-quarter growth rates, to be in comparable units to the reported CPI forecasts. Figure 7 gives the time series plots of (annualized quarter-on-quarter) CPI and PGDP inflation over the last 60 years. The two inflation rates display the same broad movements, but markedly different quarter-on-quarter changes. The greater variability of the CPI over the later period is consistent with the higher MSFE.[18]

An alternative explanation turns on the reporting practices for the expenditure variables (including the GDP) deflator. Respondents may formulate their projections in terms of growth rates, or in terms of billions of dollars of expenditure. Prior to 1990, forecasts had to be reported as levels. As of 1990:Q3, respondents may report either levels or growth rates. In either case, the SPF forecast database records the forecast as a level, and the form in which the forecast was reported is not stored. A majority tend to report levels.[19] A reasonable conjecture is that a forecast of the value of RGDP (say) in the next four quarters formulated in terms of growth rates is less likely to display spurious or unintended btt/os movements than a forecast of the levels of the variable for the same periods. Unfortunately we are unable to test this conjecture formally without knowledge of how the forecast were reported. That btt/os is of much more value to the individual CPI forecasts, reported as growth rates, than for the PGDP forecasts (reported as levels or growth rates) is consistent with the conjecture, but there may be other reasons for the observed differences. Both UNEMP and TBILL are reported in their ‘natural units’ in which individuals are likely to frame their forecasts.

A second issue is the finding that changes in the long-run outlook are associated with relatively more accurate short-horizon forecasts for all variables other than the CPI. A concern is that CPI inflation is close to having a unit root, so there may be less of a tendency for inflation to move toward the equilibrium long-run inflation rate (or for the short-horizon forecasts to be influenced much by perceived changes in that equilibrium). Some authors model inflation as having a unit root process (see, for example, Stock and Watson (2008)). Although formally, we reject a unit root in CPI inflation, CPI inflation is clearly more persistent than the growth rates of GDP and its components.[20] However, the PGDP deflator displays similar persistence, as do UNEMP and TBILL. There does not seem to be simple explanation in terms of the time-series properties of the variables.

[18] Note however that the figure plots the values of the series taken from the latest vintage. These will differ from the real-time values available at each forecast origin, particularly for the GDP deflator.

[19] I am grateful to Dean Croushore and Tom Stark for drawing this to may attention.

[20] Regressing the change in the annualised quarter-on-quarter change in CPI on four lags, a ‘lag level’ and a constant gave a coefficient on the lag level of -0.16, and an augmented Dickey-Fuller test of the null of a unit root of -3.70, which rejects the null at conventional levels. Whether the *level* of real US GDP is integrated of order one or is stationary about a trend, possibly with a break, is a contentious issue (see, e.g., Nelson and Plosser (1982), Perron (1989)), but either way there is no suggestion that differencing the level of GDP more than once is required.

7 Conclusions

At the level of the aggregate or consensus forecasts, we have shown that it is possible to discern two accuracy-enhancing characteristics of the survey 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 enjoy a relative advantage over purely model-based forecasts for GDP, consumption, investment, the GDP deflator, the unemployment rate and the Treasury Bill rate. Hence we provide support for the contention that survey respondents 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. By way of contrast, such departures are neutral for the model forecasts, as expected: this attribute is unique to the survey forecasts.

We find that survey respondents possess better information about the future course of output growth around business cycle turning points, but that departures from the path toward the long-run expectation occur at other times too, and to good effect in terms of forecast accuracy.

One of our most emphatic empirical findings is that departures from the equilibrium path are associated with large improvements in the accuracy of the CPI inflation forecasts, and of the unemployment rate and Treasury Bill rate. The findings for CPI inflation complement those of Faust and Wright (2009), who find that the US Greenbook inflation forecasts remain superior to univariate time-series model forecasts even when the latter are adjusted to draw on the Greenbook's knowledge of the current state of the economy. Our approach allows us to explain the reason for the superiority of survey forecasts in terms of readily understood attributes of those forecasts. The SPF consensus forecasters successfully foresee future movements beyond what could be predicted by a simple-time series forecasting model. In the case of the SPF consensus forecasts, the removal of the overshooting attribute leads to a worsening of the one-quarter ahead forecasts by around 50% on MSFE. The findings for the unemployment rate and Treasury Bill rate indicate similarly large effects.

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 the real expenditure variables and the PGDP deflator, but for CPI, UNEMP and TBILL the individual respondents' forecasts are more accurate on average than the counterfactual forecasts which imply no departures from convergence to the long-run position. The forecast encompassing framework indicates that for the shortest-horizon individual forecasts it is not optimal to completely remove these departures even for the expenditure variables. The average accuracy of the counterfactual forecasts of these variables can be improved by tilting them toward the reported forecasts (which display departures from the long-run expectation path).

Hence 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.

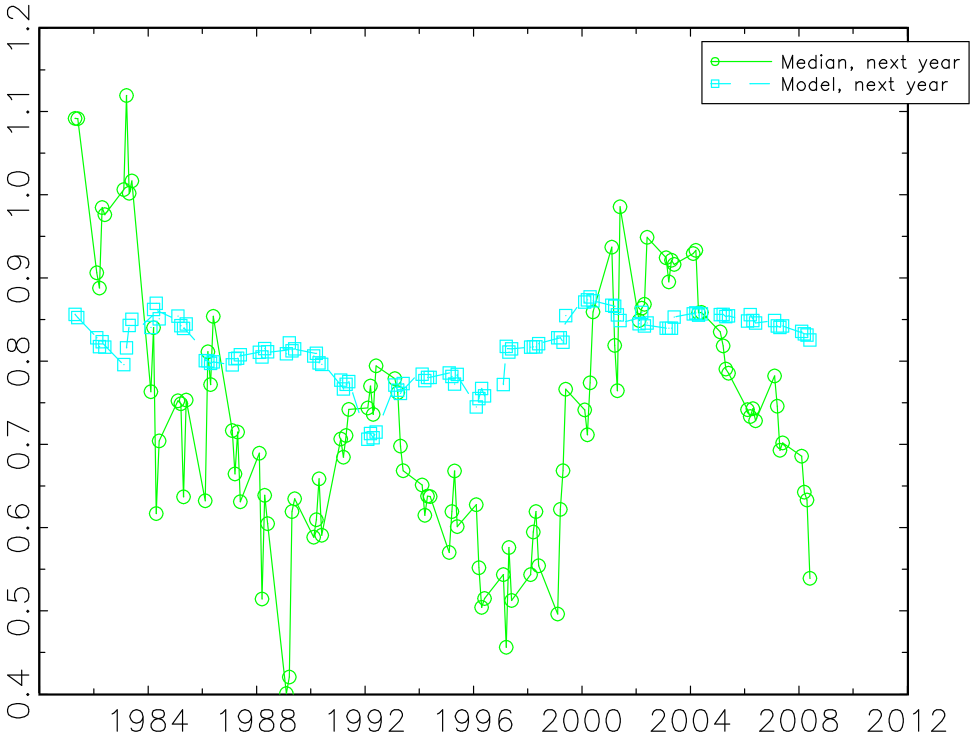

Figure 1: Real output growth year-ahead forecasts, plotted against the time the forecasts were made.

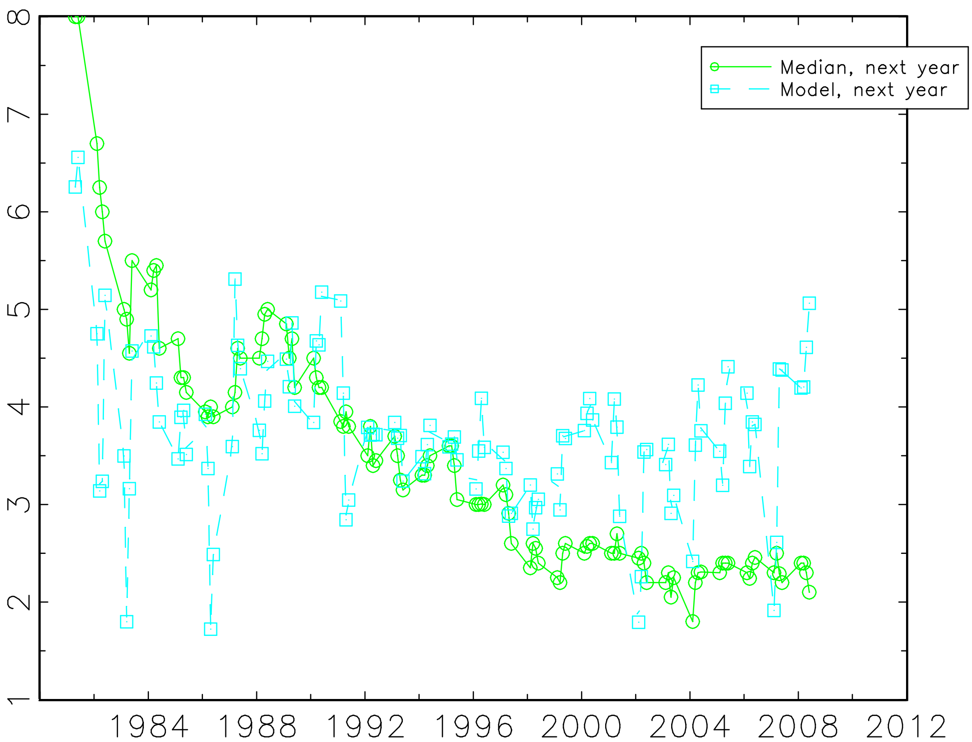

Figure 2: CPI inflation year-ahead forecasts, plotted against the time the forecasts were made.

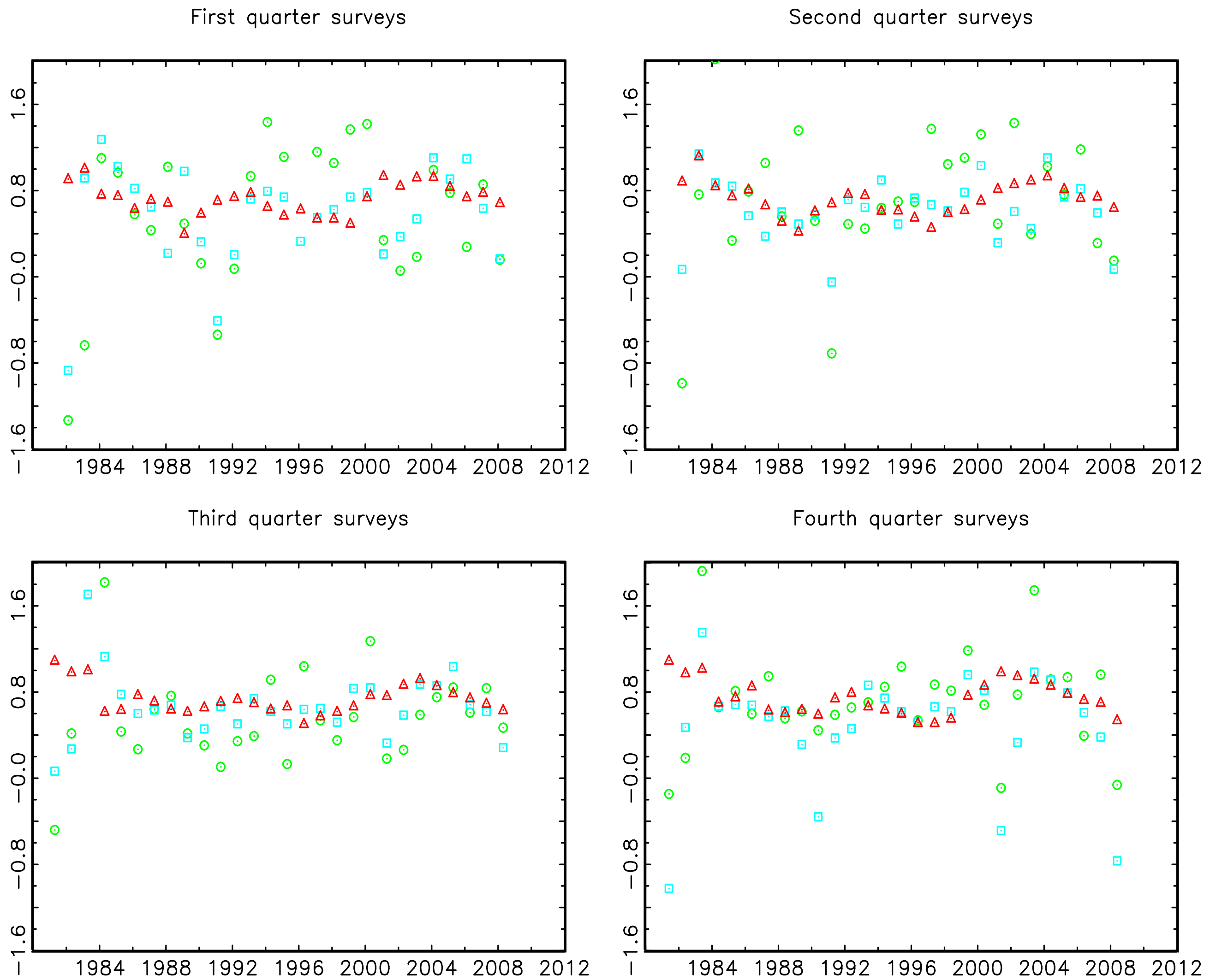

Figure 3: Median survey forecasts of quarter-on-quarter percentage output growth, for the survey quarter (square), and the same quarter a year ahead (triangle). The circles denote the first estimates of the actual value for the previous quarter.

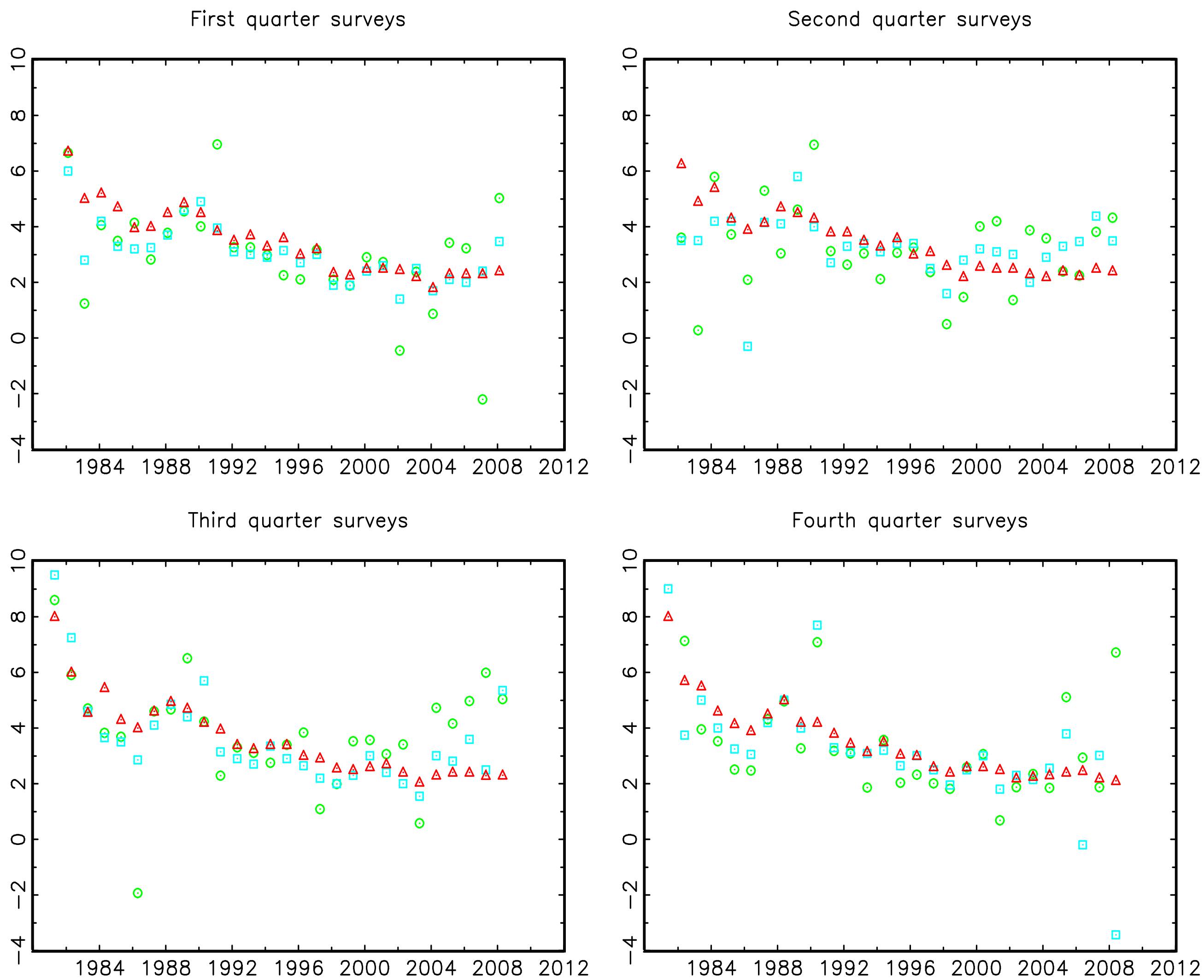

Figure 4: Median survey forecasts of quarter-on-quarter percentage CPI inflation at an annualised rate, for the survey quarter (square), and the same quarter a year ahead (triangle). The circles denote the first estimates of the actual value for the previous quarter.

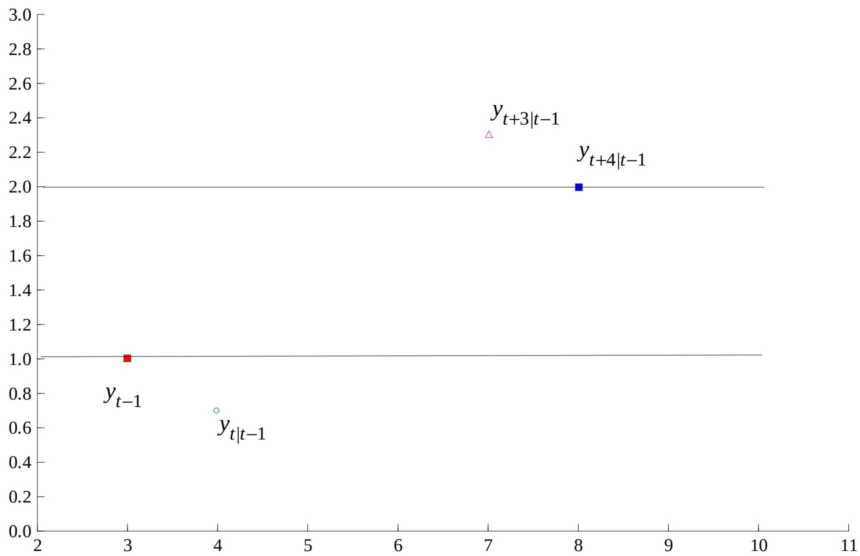

Figure 5: Graphical depiction of 'bucking the trend' and 'overshooting' forecasts, $y_{t|t-1}$ and $y_{t+3|t-1}$.

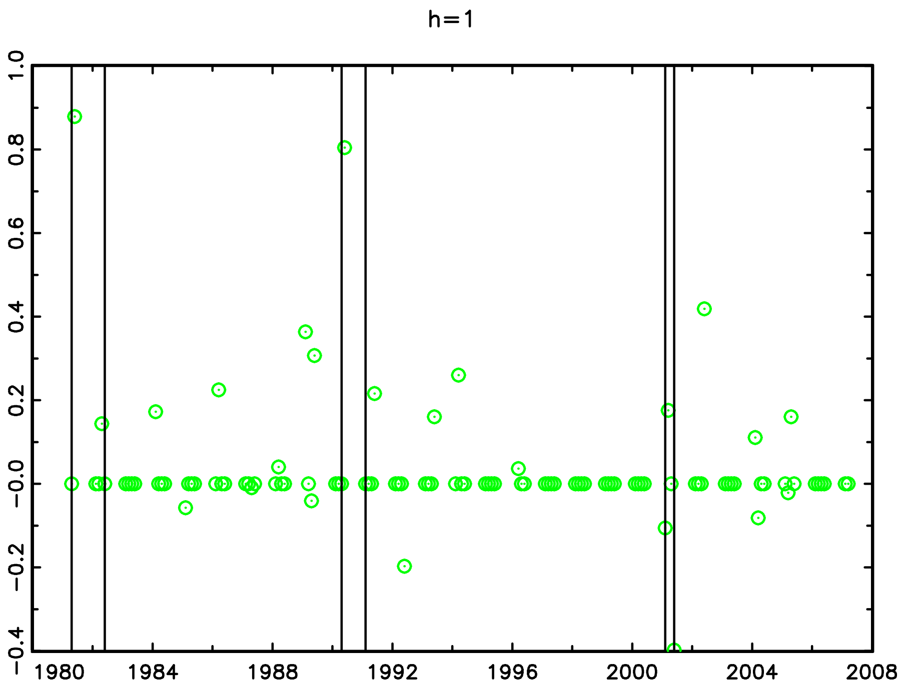

Figure 6: The one-step ahead median survey forecasts of output growth: the circles are the absolute errors of the median forecasts with the btt-characteristic removed, minus the absolute errors of the original median forecasts. The solid lines denote the business cycle peaks and troughs - 1981:3 (P), 1982:3 (T), 1990:3 (P), 1991:1 (T), 2001:1 (P), 2001:4 (T).

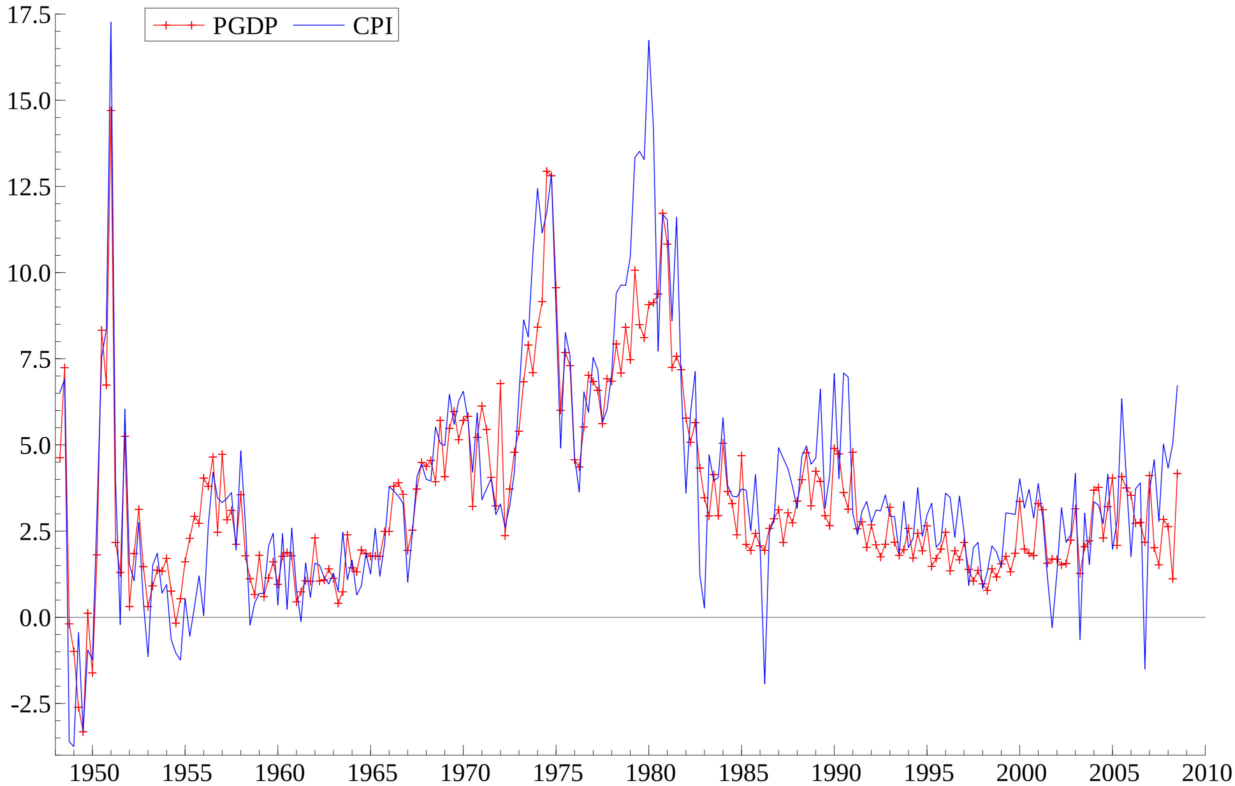

Figure 7: Annualized quarter-on-quarter PGDP and CPI inflation.

Table 1: Description of Macroeconomic Variables

Variable	SPF code	Forecast reported as:	RTDSM or FRED code
Real GDP (GNP)	RGDP	L/G	ROUTPUT
Real personal consumption	RCONSUM	L/G	RCON
Real nonresidential fixed investment	RNRESIN	L/G	RINVBF
GDP price index	PGDP	L/G	P
(implicit deflator, GNP / GDP deflator)		L/G	
CPI inflation rate	CPI	q-o-q annualized	CPI
Civilian unemployment rate	UNEMP	%	RUC
3-month Treasury Bill	TBILL	%	TB3MS
(Secondary Market Rate)			

Notes: The CPI forecasts are reported as annualized quarter-over-quarter (q-o-q) percentage changes. The forecasts of RGDP and its components are reported by respondents as either either levels (L) or growth rates (G) (from 1990:3 onwards; prior to 1990:3 they were reported as levels). Whether reported as L or G they are recorded in the forecast database as levels.

All data except for the CPI, UNEMP and TBILL were reported as quarterly vintages of quarterly observations. For these three variables, quarterly vintages of monthly observations are provided, and we averaged the months to obtain quarterly series. Note that TBILL is not revised, and prior to 1994:Q3 the CPI data were not revised.

The SPF data are from the Philadelphia Fed, http://www.phil.frb.org/econ/spf/. The forecast data were downloaded from the SPF web page on May 12th 2012.

Data from the RTDSM (http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/) for all variables other than TB3MS, which is taken from FRED (Federal Reserve Economic Data, http://research.stlouisfed.org).

Table 2: 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	2.99	4.41	0.68	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	0.87	1.32	0.66	1.07	1.02	1.06	0.99
PGDP	2	1.09	1.54	0.70	1.01	1.00	1.12	1.00
	3	1.44	1.28	1.13	1.01	1.00	1.00	1.00
	4	1.56	1.36	1.14	1.00	1.00	1.01	1.00
	5	1.78	1.61	1.10	.	.	.	.
	1	1.14	2.76	0.41	1.14	1.01	1.57	0.99
CPI	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	.	.	.	.
	1	0.02	0.08	0.22	1.60	1.00	1.32	0.99
UNEMP	2	0.09	0.22	0.42	1.09	1.06	1.24	1.00
	3	0.21	0.44	0.49	1.03	1.03	1.16	1.00
	4	0.40	0.66	0.60	1.00	1.00	1.06	1.00
	5	0.58	0.88	0.66	.	.	.	.
	1	0.03	0.57	0.05	1.75	0.98	2.68	0.89
TBILL	2	0.43	1.53	0.28	1.06	0.99	1.18	0.94
	3	1.00	2.18	0.46	1.01	1.00	1.07	0.99
	4	1.94	3.37	0.58	1.00	1.00	1.02	0.99
	5	2.85	4.90	0.58	.	.	.	.

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 3: Relative accuracy of shorter-term survey forecasts by magnitude of change in the longer-term outlook

		1	2	3	4	#
RGDP	All	0.67	0.81	0.91	0.97	103
	1 s.d.	0.62	0.96	1.08	0.95	26
RCONSUM	All	0.75	0.98	1.05	1.04	103
	1 s.d.	0.64	1.67	0.94	1.18	20
RNRESIN	All	0.68	0.79	0.86	0.89	103
	1 s.d.	0.48	0.77	0.78	0.79	20
PGDP	All	0.69	0.76	1.02	1.07	103
	1 s.d.	0.65	0.62	1.31	1.18	26
CPI	All	0.40	0.72	0.88	0.91	103
	1 s.d.	0.76	0.64	1.13	1.65	22
UNEMP	All	0.22	0.42	0.50	0.61	103
	1 s.d.	0.14	0.24	0.38	0.48	20
TBILL	All	0.05	0.30	0.51	0.63	103
	1 s.d.	0.05	0.23	0.49	0.62	24

Each element in the columns 3 to 6 is the ratio of the MSFE of the survey-median forecasts to the MSFE of the corresponding benchmark forecasts, for a given horizon ($h = 1$ corresponds to a current-quarter forecast). For each variable, we give the ratio for all the forecasts for that horizon, and the second row is calculated for the subsets of forecasts for which the corresponding long-run median forecast changed relative to the previous forecast by more than one standard deviations (1 s.d.). The last column ('#') gives the number of forecasts in each case.

Table 4: Relative median survey short-horizon forecast accuracy as a function of the change in the long-run outlook

	growth $t-1$ to $t+1$		growth t to $t+2$	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
RGDP	0.01	-0.89	-0.01	-0.26
	0.91	0.30	0.91	0.53
RCONSUM	0.08	-1.61	0.03	-0.20
	0.14	0.00	0.58	0.34
RNRESIN	-0.30	-0.83	-0.32	-0.67
	0.06	0.01	0.03	0.02
PGDP	-0.19	-0.09	-0.24	0.50
	0.18	0.77	0.13	0.18
CPI	-0.73	0.15	-0.24	-0.32
	0.01	0.69	0.36	0.39
UNEMP	-0.07	-0.36	-0.05	-0.41
	0.23	0.10	0.57	0.17
TBILL	-0.19	-0.56	-0.10	-0.65
	0.18	0.00	0.43	0.00

The table displays the estimates from regression equation (1), with p-values of the significance of the coefficient below in smaller font (based on HAC standard errors). For the left side, the dependent variable is the absolute value of the error of the survey forecast of growth between $t+1$ and $t-1$, less the absolute value of the model forecast, and for the right side, the same but for the growth rate between $t+2$ and t. (In both cases the forecast origin of $t-1$). The slope regressor is the absolute change in the long-run survey expectation made at $t-2$ and $t-1$. The $t-2$ long-run expectation is of the growth rate between $t+2$ and $t+3$, and the $t-1$ forecast is of the growth rate between $t+3$ and $t+4$.

Table 5: Proportion of the Median survey forecasts that 'buck the trend' and 'overshoot'

	'Bucking the tend' btt-forecasts				'Overshooting' os-forecasts			
$h =$	1	2	3	4	1	2	3	4
RGDP	0.25	0.17	0.10	0.02	0.20	0.34	0.32	0.42
RCONSUM	0.14	0.09	0.05	0.03	0.33	0.36	0.37	0.39
RNRESIN	0.18	0.11	0.04	0.04	0.29	0.25	0.30	0.27
PGDP	0.10	0.08	0.05	0.02	0.25	0.34	0.39	0.35
CPI	0.13	0.03	0.01	0.00	0.28	0.30	0.30	0.27
UNEMP	0.25	0.20	0.11	0.03	0.07	0.10	0.15	0.17
TBILL	0.22	0.19	0.11	0.05	0.18	0.30	0.34	0.30

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

Table 6: Individual NC proportions

	'Bucking the tend' btt-forecasts				'Overshooting' 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
PGDP	0.21	0.16	0.13	0.10	0.31	0.37	0.39	0.42
CPI	0.18	0.10	0.07	0.05	0.26	0.31	0.30	0.29
UNEMP	0.18	0.13	0.06	0.02	0.05	0.10	0.14	0.15
TBILL	0.26	0.20	0.13	0.06	0.13	0.21	0.25	0.25

Notes: The table records the proportion of all the forecasts across individuals and surveys which are each type of NC forecast.

Table 7: Individual accuracy, 1981:Q3 – 2008:Q4

	h	MSFE	# btt	# non -btt	btt-ratio MSFE	# os	# non -os	os-ratio MSFE
	1	0.275	694	2124	0.92	771	2047	0.96
RGDP	2	0.362	534	2284	0.90	958	1860	0.88
	3	0.408	420	2398	0.91	1041	1777	0.84
	4	0.361	275	2543	0.95	1165	1653	0.87
	1	0.345	414	2283	0.95	1122	1575	0.91
RCONSUM	2	0.368	355	2342	0.93	1055	1642	0.89
	3	0.364	250	2447	0.94	1135	1562	0.90
	4	0.379	205	2492	0.96	1159	1538	0.88
	1	3.749	549	2117	0.95	784	1882	1.00
RNRESIN	2	4.404	375	2291	1.00	778	1888	0.94
	3	4.854	252	2414	0.98	928	1738	0.94
	4	5.085	182	2484	0.99	993	1673	0.97
	1	1.917	601	2197	0.89	844	1954	0.77
PGDP	2	2.167	450	2348	0.81	997	1801	0.89
	3	2.293	371	2427	0.92	1066	1732	0.82
	4	2.324	274	2524	0.91	1164	1634	0.84
	1	1.866	486	2249	1.07	669	2066	1.16
CPI	2	2.989	274	2461	1.00	785	1950	1.01
	3	3.210	190	2545	1.00	801	1934	0.95
	4	3.353	119	2616	1.00	768	1967	0.99
	1	0.032	534	2333	1.24	138	2729	1.05
UNEMP	2	0.121	384	2483	1.06	282	2585	1.03
	3	0.250	187	2680	1.02	395	2472	1.04
	4	0.411	55	2812	1.00	408	2459	1.02
	1	0.085	705	2015	1.42	335	2385	1.25
TBILL	2	0.478	545	2175	1.09	531	2189	1.06
	3	1.075	337	2383	1.02	645	2075	1.03
	4	1.915	157	2563	1.01	651	2069	1.01

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 'os-ratio MSFE' is the equivalent for the os-forecasts.

Table 8: Tests of Forecast Encompassing of the individual Reported and Artificial Forecasts

	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	0.077	0.656	0.656	0.079	0.661	0.661	0.092	0.643	0.643
		0.075	0.000	0.000	0.079	0.000	0.000	0.028	0.000	0.003
	2	0.041	0.939	0.939	0.036	0.891	0.891	0.047	0.993	0.993
		0.495	0.000	0.369	0.571	0.000	0.191	0.427	0.000	0.946
	3	0.044	1.025	1.025	0.031	1.017	1.017	0.050	1.033	1.033
		0.530	0.000	0.642	0.672	0.000	0.825	0.481	0.000	0.675
RCONSUM	1	0.122	0.616	0.616	0.147	0.733	0.733	0.125	0.582	0.582
		0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
	2	0.136	0.839	0.839	0.156	0.866	0.866	0.144	0.816	0.816
		0.001	0.000	0.039	0.000	0.000	0.187	0.000	0.000	0.058
	3	0.151	0.971	0.971	0.165	0.963	0.963	0.153	0.975	0.975
		0.003	0.000	0.720	0.002	0.000	0.662	0.003	0.000	0.829
RNRESIN	1	0.327	0.559	0.559	0.305	0.646	0.646	0.349	0.482	0.482
		0.047	0.000	0.000	0.076	0.000	0.005	0.034	0.001	0.000
	2	0.312	0.769	0.769	0.307	0.527	0.527	0.325	0.904	0.904
		0.217	0.000	0.078	0.238	0.009	0.019	0.192	0.000	0.513
	3	0.254	0.869	0.869	0.231	0.631	0.631	0.245	1.006	1.006
		0.442	0.000	0.295	0.491	0.000	0.014	0.454	0.000	0.970
PGDP	1	-0.146	0.935	0.935	-0.167	0.931	0.931	-0.107	0.941	0.941
		0.124	0.000	0.098	0.100	0.000	0.205	0.257	0.000	0.250
	2	-0.245	0.997	0.997	-0.262	1.071	1.071	-0.243	0.905	0.905
		0.038	0.000	0.943	0.040	0.000	0.093	0.038	0.000	0.183
	3	-0.369	1.048	1.048	-0.412	0.970	0.970	-0.360	1.089	1.089
		0.029	0.000	0.262	0.023	0.000	0.673	0.031	0.000	0.142
CPI	1	-0.017	0.096	0.096	-0.014	0.192	0.192	-0.015	0.036	0.036
		0.874	0.358	0.000	0.899	0.153	0.000	0.891	0.791	0.000
	2	-0.105	0.440	0.440	-0.083	0.408	0.408	-0.100	0.448	0.448
		0.478	0.006	0.001	0.589	0.046	0.004	0.495	0.023	0.005
	3	-0.162	1.168	1.168	-0.134	0.770	0.770	-0.161	1.305	1.305
		0.455	0.000	0.368	0.558	0.002	0.351	0.451	0.000	0.190
UNEMP	1	-0.046	0.035	0.035	-0.046	-0.001	-0.001	-0.046	0.149	0.149
		0.001	0.657	0.000	0.000	0.991	0.000	0.001	0.165	0.000
	2	-0.074	0.061	0.061	-0.077	-0.140	-0.140	-0.072	0.226	0.226
		0.067	0.716	0.000	0.048	0.464	0.000	0.076	0.210	0.000
	3	-0.089	-0.261	-0.261	-0.087	-0.582	-0.582	-0.087	-0.161	-0.161
		0.212	0.375	0.000	0.228	0.133	0.000	0.231	0.591	0.000
TBILL	1	-0.071	0.124	0.124	-0.068	0.088	0.088	-0.068	0.169	0.169
		0.000	0.001	0.000	0.000	0.027	0.000	0.000	0.000	0.000
	2	-0.201	0.263	0.263	-0.191	0.122	0.122	-0.196	0.352	0.352
		0.007	0.017	0.000	0.006	0.352	0.000	0.007	0.002	0.000
	3	-0.349	0.270	0.270	-0.345	0.285	0.285	-0.346	0.259	0.259
		0.023	0.246	0.002	0.021	0.260	0.005	0.022	0.313	0.004

The table reports the estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ in (4), where $\beta_2 = 0$ is the null hypothesis that the Reported forecast encompass the Artificial, and $\beta_2 = 1$ is the null hypothesis that the Artificial forecast encompass the Reported. The corresponding p-values of these two null hypotheses are directly below the (repeated) $\hat{\beta}_2$ estimates (calculated for a two-sided alternative).

8 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 5 we assume that for an individual i:

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

and for any pair of individuals i, j:

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

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 $E\left[\varepsilon_{it}\varepsilon_{it+2}\right] = 0$ 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 $k = 5$, $E\left[\varepsilon_{it}\varepsilon_{it+k}\right] = 0$, 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:

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

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

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

We can then construct the estimator $\hat{\Sigma}$ of $\Sigma = E\left(\varepsilon\varepsilon'\right)$, where $\varepsilon = \left[\varepsilon_{11}\ \varepsilon_{12}\ \ldots \varepsilon_{1T}; \ldots; \varepsilon_{N1}\ \varepsilon_{N2}\ \ldots \varepsilon_{NT}\right]'$, 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.

Atkeson, A., and Ohanian, L. (2001). Are Phillips Curves useful for forecasting inflation?. *Federal Reserve Bank of Minneapolis Quarterly Review*, **25**, 2–11. (1).

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.

Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

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 Harvey, D. I. (2009). Forecasting combination and encompassing. In Mills, T. C., and Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 169–198: Palgrave MacMillan.

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.

Coibion, O., and Gorodnichenko, Y. (2011). What can survey forecasts tell us about information rigidities?. College of William and Mary.

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 threedimensional panel data. *Journal of Econometrics*, **68**, 205–227.

Fair, R. C., and Shiller, R. J. (1989). The informational content of *ex ante* forecasts. *Review of Economics and Statistics*, **71**, 325–331.

Faust, J., and Wright, J. H. (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics*, **27 (4)**, 468–479.

Granger, C. W. J., and Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, **5**, 35–47. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

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.

Mankiw, N. G., and Reis, R. (2002). Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips Curve. *Quarterly Journal of Economics*, **117**, 1295–1328.

Mankiw, N. G., Reis, R., and Wolfers, J. (2003). Disagreement about inflation expectations. mimeo, National Bureau of Economic Research, Cambridge MA.

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.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, **62**, 902–917. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Nelson, C. R., and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, **10**, 139–162.

Patton, A. J., and Timmermann, A. (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, **57**, 803–820.

Patton, A. J., and Timmermann, A. (2011). Predictability of output growth and inflation: A multi-horizon survey approach. *Journal of Business and Economic Statistics*, **29**, 397–410.

Perron, P. (1989). The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.

Romer, C. D., and Romer, D. H. (2000). Federal Reserve information and the behaviour of interest rates.

American Economic Review, **90**, 429–457.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50**, 665–690.

Stock, J. H., and Watson, M. W. (2007). Why has U.S. Inflation Become Harder to Forecast?. *Journal of Money, Credit and Banking*, **Supplement to Vol. 39**, 3–33.

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

Stock, J. H., and Watson, M. W. (2010). Modelling Inflation after the Crisis. *NBER Working Paper Series*, **16488**.

Woodford, M. (2001). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M. (eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In honor of Edmund Phelps*: Princeton University Press.

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.