ICMA Centre

Henley Business School
UNIVERSITY OF READING

Discussion Paper

Forecasters' Disagreement about How the Economy Operates, and the Role of Long-run Relationships

November 2015

Michael P Clements
ICMA Centre, Henley Business School

Discussion Paper Number:
ICM-2015-09

AMBA ACCREDITED EFMD EQUIS ACCREDITED AACSB ACCREDITED

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

Forecasters' Disagreement About How The Economy Operates, And The Role Of Long-Run Relationships

Michael P. Clements
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA
m.p.clements@reading.ac.uk.

November 13, 2015

Abstract

Macro-variables such as consumption, investment and output are expected to move together in the long run. We consider whether survey forecasts of these quantities suggest beliefs about equilibrium relationships play a prominent role in expectations formation. Evidence is brought to bear from an analysis of multivariate measures of forecaster disagreement, as well as tests of forecast optimality. The analysis of disagreement provides little support for the proposition that equilibrium considerations play a key role. Moreover, we generally reject forecast optimality for a majority of forecasters, but there is no evidence that this is due to long-run mis-specification.

Journal of Economic Literature classification: C53.

Keywords: Expectations formation, Equilibrium, Multivariate Disagreement.

1 Introduction

In this paper we consider whether professional forecasters have similar models in mind when they formulate their expectations of consumption, investment and output, and whether those models include well-defined long-run relationships. Broad classes of economic theory suggest the existence of balanced growth paths (of the Solow-Ramsey model) and the 'great ratios' of Kosobud and Klein (1961), or two-sector models (such as, e.g., Whelan (2003)) which predict that the key NIPA aggregates grow at constant (but possibly different) rates in the long run. Of interest is whether these ideas inform the regular practice of forecast production. Equilibrium relationships have played a prominent role in terms of the econometric modelling of dynamic relationships, especially following the coupling of error-correction and cointegration,[1] and underpin DSGE modelling (see, e.g., Del Negro and Schorfheide (2013) for a recent review). The forecasts of survey respondents might be expected to bear their imprint.

We know that equilibrium between integrated variables requires cointegration, and that cointegration implies error-correction. The existence of error-correction implies that short-run disequilibria, as captured by the error-correction terms, will be useful for short-horizon forecasting of the growth rates, and in some circumstances, for longer-horizon forecasting too.[2] Although cointegration is a long-run phenomenon, the belief that cointegration only affects long-horizon forecasts is mistaken. As stressed by Clements and Hendry (1995a), the greatest impact of cointegration on forecasts of growth rates will be at the shortest horizons. Hence the influence of cointegration should be detectable using short-horizon forecasts if such relationships influence forecaster behaviour.

We tackle the question of whether the behaviour of professional forecasters is affected by the postulated equilibrium relationships through a number of related analyses. The first uses multivariate measures of disagreement to assess the extent to which agents share a common model of the economy for the set of variables of interest. The assumption that agents share a 'common model' of the economy (which informs their forecasts) is a key feature of recent implementations of models of informational rigidities.[3] These models have played a prominent role in explaining why macro-forecasters disagree. Hence our analysis will shed light on the common model assumption for a set of variables for which agents might be expected to agree about the workings of the economy. There are few studies looking at multivariate disagreement, and the studies there are arguably consider sets of variables about which there is less likely to

[1]The first paper to bring 'error-correction' to the economics literature was Sargan (1964), with Davidson, Hendry, Srba and Yeo (1978) being especially influential, whilst Engle and Granger (1987) is a key paper on cointegration.

[2]See Engle and Yoo (1987), Clements and Hendry (1995a) and Christoffersen and Diebold (1998), and Elliott (2006) for a review and synthesis.

[3]See e.g., Mankiw and Reis (2002), Woodford (2001), Mankiw, Reis and Wolfers (2003), Sims (2003), Coibion and Gorodnichenko (2012), Andrade and Le Bihan (2013) and Andrade, Crump, Eusepi and Moench (2014).

be a consensus about the relationships between the variables (discussed further below).

We consider the set of relatively prolific forecasters, defined as those who filed returns to in excess of 84 of the possible 92 surveys between 1990:4 and 2013:3. Of interest is whether these forecasters are reasonably homogeneous when a multivariate measure is used to assess the extent to which they disagree with the consensus. This measure has the following attractive property: suppose two forecasters disagree with the consensus to the same extent, when each variable is considered in isolation. If one forecaster agrees more with the consensus view about the working of the economy (in a sense explained below), then that forecaster receives a lower overall disagreement score. We consider whether these individuals do disagree with the consensus to different extents on average, and whether these differences are persistent (i.e., whether those who are a long way from, or alternatively close to, the consensus at one time remain so subsequently).

Secondly, after the analysis of the most prolific forecasters, all the forecasters are taken together, and we consider the extent to which agreement about the workings of the economy tempers overall disagreement, and how these measures vary with the business cycle, and the forecast horizon. Is there greater agreement during expansions, say, or concerning the longer-run outlook relative to short-term prospects?

Thirdly, we decompose the overall disagreement measures in to disagreement about particular variables, and the correlations between the forecasts of different variables (i.e., the extent of agreement about the relationships which exist between the specific variables). In so doing, we consider disagreement about the consumption-output and investment-output ratios, and also the dynamic relations which ought to hold in the presence of cointegration and error-correction.

In summary, our results using disagreement suggest a relatively high degree of commonality in forecasters' beliefs concerning the mechanisms that generate the growth rates of consumption, investment and output growth. That said, formal testing reveals statistically significant differences between individual forecasters when we adopt a multivariate approach to disagreement. We find that multivariate disagreement declines in the forecast horizon, and that the percentage reduction in the measure due to agreement about the workings of the economy increases between $h = 0$ and $h = 1$, and thereafter changes little. Multivariate disagreement is counter-cyclical. However, there is little evidence that forecasts are generated by individuals who believe in a common equilibrium towards which the economy is moving.

Given that the analysis of disagreement does not suggest a role for equilibrium beliefs in expectations formation, we sharpen our analysis of whether forecasts are generated by models with long-run relationships by presenting tests for the presence of such effects in the forecasts of the respondents. The tests we run include the traditional Mincer-Zarnowitz (Mincer and Zarnowitz (1969)) forecast efficiency tests, a multi-horizon extension due to Patton and Timmermann (2012), and a test constructed to have power to detect omitted error-correction behaviour. The tests are shown by Monte Carlo to have reasonable size and power properties in samples of the size considered empirically. Patton and Timmermann (2012) have recently suggested short-horizon forecasts might be used in the place of outturns (or actual values) in testing for forecast optimality, with the advantage for macro-forecasting that the investigator would not need to take a stance on the vintage of data being forecast. We assess the operating characteristics of these tests in our sample, including an analysis of the extent to which the inference concerning rationality varies with the use of actuals or short-horizon forecasts, and the length of the forecasts, for specific individuals (namely, the more prolific respondents). This provides useful information on the properties of the tests in real-time forecasting exercises. Although the application of the tests to the empirical forecast data rejects forecast optimality for a number of respondents, there is little evidence that these rejections are due to long-run mis-specification.

Our paper is related to a large literature on disagreement, [4] but the number of papers using multivariate approaches is much smaller. Banternghansa and McCracken (2009) was one of the first papers to consider multivariate approaches to forecast disagreement, and Dovern (2014) uses multivariate disagreement to assess whether forecasters disagree because they have different views about the outlook for the economy e.g., whether an expansion is likely to continue or give way to a period of slower growth or even a contraction, or whether disagreement occurs because forecasters have different views about how the economy operates (even though they might be of a similar opinion regarding the prospects for growth, for example). In the context of forecasting inflation, GDP growth and the unemployment rate, Dovern (2014) finds that the disagreement-based evidence for agents sharing a common model of the economy is weak. However, this may simply be because some of the actual data correlations between these variables are small. If the actual data series were generated independently of each other forecasters would not be expected to adopt a common model of their joint determination. On the other hand, there are good reasons to expect the variables we consider to be relatively highly correlated, so it is of interest to assess whether these correlations are present in the forecast data. Moreover, the balanced growth paths (of the Solow-Ramsey model) or two-sector models (such as, e.g., Whelan (2003)) all predict that the key National Income and Product Accounts aggregates grow at constant (possibly different) rates in the long run, which imply relationships between the short-run forecasts of these variables and of transformations of these variables, so we are able to assess the extent to which the forecasts are consistent with these postulated relationships.

More generally, our paper relates to the literature on how expectations are formed, as reviewed in e.g., Pesaran and Weale (2006). Evidence is brought to bear on the common model assumption underlying common implementations of models of information rigidities, and we also consider the role of long-run relations in expectations formation.

[4] See, for example, Zarnowitz and Lambros (1987), Bomberger (1996), Rich and Butler (1998), Capistrán and Timmermann (2009), Lahiri and Sheng (2008), Rich and Tracy (2010) and Patton and Timmermann (2010).

The plan of the remainder of the paper is as follows. In section 2 the forecast data used throughout the paper are discussed. Section 3 presents the multivariate measures of disagreement, and section 3.1 applies these measures to analyze the nature of the disagreement between each of the most prolific forecasters and the consensus. Section 3.2 considers all the forecasters together, and the extent to which the forecasters agree on the way the economy operates, including the relevance of equilibrium for the generation of forecasts. In section 4 the tests of forecast optimality are outlined, and we derive the population parameters of the test regressions when forecasts are mis-specified by the omission of relevant error-correction terms. Section 5 outlines the Monte Carlo used to investigate the small forecast-sample performance of the tests. Section 6 discusses a number of issues in applying the tests to survey data. The results using shorter-horizon forecasts in place of actual values in the tests are discussed in section 6.3, and the empirical part ends in section 6.4 with an in-depth investigation of a number of issues concerning testing for forecast optimality. Section 7 offers some concluding remarks.

2 Forecast Data: SPF Respondents' Forecasts

We use the US Survey of Professional Forecasters (SPF). The SPF is a quarterly survey of macroeconomic forecasters of the US economy that began in 1968, 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 Survey of Professional Forecasters (SPF): see Croushore (1993). We use the SPF because it is made freely available by the Philadelphia Fed, so that any results can be readily reproduced and checked by other researchers. Its constant scrutiny is likely to minimize the impact of respondent reporting errors. An academic bibliography of the large number of published papers that use SPF data is available at: http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/academic-bibliography.cfm.

The SPF provides multi-horizon forecasts of GDP, consumption and investment for around a third of a century. Specifically, forecasts of these variables are available from the 1981:3 surveys to the present. For the analysis of disagreement, we restrict the sample to 1990:4 onwards, corresponding to the period for which the Philadelphia Fed assumed responsibility. Forecasts are made of the current quarter (i.e., the quarter in which the survey takes place), and of the quarterly values of the variables in each of the next four quarters, so that the longest-horizon quarterly forecast is of the same quarter of the year in the following year. Forecasts are also provided of the current (survey-quarter) calendar-year levels of the variables, and of the levels in the following year. Hence for surveys made in the first quarter of a year, the next year forecast approximately corresponds to a 2-year or 8-quarter forecast, whereas for fourth quarter surveys these forecasts have a horizon of 5-quarters. Hence from the first-quarter surveys we obtain an annual series of 2-year ahead calendar-year forecasts (as well as the annual series of 1-year ahead calendar-year forecasts from the forecasts of the current year).

We use the 92 surveys from 1990:4 to 2013:3 inclusive. The 2013:3 survey provides a forecast of 2014:3. We use the vintage of data available two quarters after the reference quarter for outturns, which explains the forecast data endpoint. At the time of writing, the latest data vintage is 2015:1, which means we have the second-quarterly vintage estimates for 2014:3.

Table 1 provides details concerning the actual and forecast data.

The switch from a 'fixed-base-year' to chain-weighted estimates of real GDP and its components in the 1990's may potentially have an impact on our analysis, as with chain-weighting it is no longer true that the GDP components sum to GDP, or to intermediate sub-aggregates. The investment series is the sum of private non-residential investment, and residential investment, although strictly-speaking these components are not summable. However, it seems likely that any resulting distortions will be of secondary importance.

In the course of this research, a small number of aberrant observations were identified, and these observations were replaced by missing values. Appendix 1 details the small number of changes which were made to the published SPF data.

3 Multivariate Measures of Disagreement

Banternghansa and McCracken (2009) convincingly argue for a multivariate approach to the analysis of forecaster disagreement. Survey respondents are typically asked to report forecasts for a number of variables. Banternghansa and McCracken (2009) argue that one might then consider multivariate measures of disagreement, which consider the distances between the vectors of forecasts, rather than analyzing disagreement about individual variables in isolation of each other. Assuming the vector of forecasts is produced as a coherent whole and reflects the forecaster's beliefs about the inter-dependencies that exist between the variables, it is reasonable to take the correlations across variables into account in determining the extent to which forecasters disagree. They present an illustration (see their Figure 1): if two variables are positively correlated, then this would be expected to be reflected in forecasts of these two variables, and an individual who records forecasts of the two variables of different sign might could reasonably be said to disagree to a greater extent than a forecaster who produces forecast of the same sign, even if the (Euclidean) distance of the two pairs of forecasts from the consensus is the same.

To capture this idea, they first define the cross-sectional forecast covariance matrix as:

$$S_{t|t-h} = N_{t,h}^{-1} \sum_{i=1}^{N_{t,h}} \left(y_{i,t|t-h} - \overline{y}_{t|t-h} \right) \left(y_{i,t|t-h} - \overline{y}_{t|t-h} \right)' \tag{1}$$

where $y_{i,t|t-h}$ is the vector of forecasts made by i at time $t-h$ for a target y_t, $N_{t,h}$ is the number of forecasters of y_t at time $t-h$, and $\overline{y}_{t|t-h} = N_{t,h}^{-1} \sum_{i=1}^{N_{t,h}} y_{i,t|t-h}$. Then they define their multivariate disagreement measure for individual i forecasting the vector y_t at forecast origin $t-h$ as the Mahalanobis distance:

$$D_{i,t|t-h} = \sqrt{\left(y_{i,t|t-h} - \overline{y}_{t|t-h}\right)' S_{t|t-h}^{-1} \left(y_{i,t|t-h} - \overline{y}_{t|t-h}\right)}. \tag{2}$$

To illustrate, suppose y consists of just two variables, and $y_{i,t|t-h} - \overline{y}_{t|t-h} = (1,1)'$, so that a respondent's forecasts of both variables differ from the consensus forecasts by a positive amount (of 1 unit). Then the Euclidean measure of disagreement (by setting $S = I_2$) is $\sqrt{2}$. Suppose the diagonal elements of S are unity and the off-diagonal element is ρ. If $\rho = 0.9$, so the cross-sectional covariance between the other respondents' forecasts of the two variables (equivalently, forecast errors) is positive, then $D = 2/1.9$, which is less than $\sqrt{2}$, which is in turn less than the value of D when $\rho = -0.9$.

Banternghansa and McCracken (2009) consider the extent to which each individual FOMC (Federal Open Market Committee) member disagrees in terms of their forecast vectors of four key variables. Disagreement amongst FOMC members is found to be small relative to the level of disagreement between the forecasts of the SPF.

Dovern (2014) suggests using $\sqrt{\det\left(S_{t|t-h}\right)}$ as a measure of overall disagreement. It contains disagreement due to two sources: the direction of the economy; and how the economy operates. The elements on the main diagonal of $S_{t|t-h}$ measure the disagreement about the outlook for the specific variables, whereas the covariances indicate disagreement about how the economy operates. In the context of forecasting inflation, GDP growth and the unemployment rate, an analysis of how the elements of $S_{t,h}$ vary over time reveals that the cross-sectional correlations appear very noisy. By way of contrast, the diagonal elements vary less over time and account for the tendency of disagreement to move counter-cyclically. Dovern (2014) also finds the unconditional cross-section correlations (full-sample averages) between the three variables are relatively low - the year ahead inflation and output growth forecasts exhibit a correlation of only 0.16, and the growth and unemployment rate forecasts a correlation of -0.25. [5]

3.1 Multivariate Disagreement and the Most Prolific Individuals

Forecasters are obviously heterogeneous in the sense that they do not report identical forecasts at each point in time. The more interesting question is whether differences in forecasters are systematic, in the sense that some respondents' forecasts tend to differ more or less than others. The alternative would be that individual respondents are more or less the same on

[5] The correlation between output and unemployment is known as Okun's Law (see, e.g. Ball, Jalles and Loungani (2015) for a cross-country analysis from a forecasting perspective), but there is perhaps less reason to expect a strong correlation between output growth and inflation.

average (across time), and overall disagreement at any point in time is as likely to be due to any one forecaster disagreeing with the consensus as any other forecaster.

Table 2 reports the evidence for forecaster heterogeneity based on the multivariate disagreement measure (equations (1) and (2)). The table records results for the 16 forecasters who reported forecast vectors (of the growth rates of consumption, investment and output growth, at the $h = 0$ and $h = 4$ horizons) in response to 84 or more of the 92 surveys between 1990:4 and 2013:3. The forecasters are ordered numerically by forecaster id., and the 'Ave' in the second and sixth columns give the average of eqn. (2) across all surveys to which the individual responded, for $h = 0$ and $h = 4$, respectively.

For the $h = 0$ forecasts we report formal tests of whether the population means of the $D_{i,h}$ differ across individuals, i.e., of the null that $H_0 : \mu_{i,h} = \mu_{j,h}$ versus $H_1 : \mu_{i,h} \neq \mu_{j,h}$ for two individuals i, j, where $i \neq j$, where $\mu_{i,h}$ denotes a population mean. The $\{D_{i,t|t-h}\}$ are regarded as realizations, and we report a standard t-test for two population means allowing the variances to be unequal. (An autocorrelation-consistent estimator is used). Each individual is compared against the respondent with the smallest sample average, id84, and the largest, id431. p-values are reported, so of interest are one-sided tests of whether respondents population means $\mu_{i,h}$ are significantly larger, in the former case (labelled Equal(1) in the table) or are significantly smaller (Equal(2)): see table notes. In the top panel, the entries in column (3) indicate that over half the respondents have mean values significantly larger than that of id84: there are statistically significant differences between the forecasters. The same is true when we use Euclidean distance (bottom half of the table).

In addition, two ways of assessing persistence across time are considered. The first is the correlation between adjacent forecasts, reported in columns five and seven, which indicate such correlations are typically low for most forecasters. Secondly, we compare the ranks of forecasters based on their average levels of multivariate disagreement in the first and second halves of the sample. For $h = 0$, the full sample rank and two sub-sample ranks are given in columns eight to ten. We do not undertake a formal analysis, but note that the respondents with ranks 1 and 2 in the first sample are 4 and 2 in the second, and those with ranks of 15 and 16 are 12 and 13. This degree of relative constancy for $h = 0$ is not apparent for $h = 4$: those ranked 1 and 2 on the first sample are 15 and 16 on the second.

The bottom panel addresses the question of whether the multivariate measure reported in the top panel suggests a different picture of forecaster heterogeneity compared to simply summing the (squared) differences between the forecasts (without weighting by cross-sectional covariances). A comparison of the two panels of the table indicates some differences but by and large the conclusions are unaltered. That is, there are significant differences between the average levels of disagreement of the individual forecasters howsoever measured. By simply looking at the low and high-ranking individuals, it is apparent that there is more stability for

$h = 0$ than for $h = 4$, and two individuals with ranks 1 and 2 (whole sample, columns 8 and 11) are the same in the top and bottom panels.

3.2 Application to the Forecasters *En Masse* - The Impact of (Dis)agreement about the Workings of the Economy.

In this section, disagreement for the group of forecasters taken together[6] is analyzed using the measure of overall disagreement, $\sqrt{\det\left(S_{t|t-h}\right)}$. We consider how disagreement varies over the business cycle, and the extent to which disagreement about the how the economy operates (the off-diagonal elements of $S_{t|t-h}$) affects the overall measure over the course of business cycle. We calculate $\sqrt{\det\left(S_{t|t-h}\right)}$ for the vector of forecasts of (difference of log) growth rates for consumption, investment and output, where $S_{t|t-h}$ is given by eqn. (1). Table 3 reports results for current quarter, next quarter and year ahead forecasts ($h = 0, 1$ and 4). For $h = 4$ as well as forecasts of the quarterly growth rate 4 quarters ahead, we also report the smoother 'cumulative' growth rate between the current quarter and the same quarter in the following year. If X is the level of the variable, the cumulative forecast is given by $100 \times \frac{1}{4} \ln\left(X_{t|t-4}/X_{t-4|t-4}\right)$, as opposed to $100 \ln\left(X_{t|t-4}/X_{t-1|t-4}\right)$. The table reports the averages of the disagreement measures across survey quarters. Hence the first row is given by $T^{-1}\sum_t^T \sqrt{\det\left(S_{t|t-h}\right)}$ when the number of surveys is T.

We see that the overall disagreement measure clearly declines in h, and the disagreement about the cumulative growth rate at h is less than a half of that of the forecast of the 4-quarter ahead growth rate. In order to isolate the effect of disagreement about the structure of the economy, we calculate the (square root) of the product of the terms on the leading diagonal of $S_{t|t-h}$, i.e., $\sqrt{\prod_j S^{jj}_{t|t-h}}$ where S^{jj} is the element in the j^{th} row and j^{th} column of S. This corresponds to the overall measure when S is diagonal, and so effectively sets the correlations between the forecasts of the variables to zero. This corresponds to there being no agreement about how the economy operates, and is simply the product of the cross-sectional standard deviations, i.e., the product of the individual variable disagreement measures. In the table we report the time averages of this measure, as well as the ratio of the time average of the overall measure to the time average of the 'diagonal' $S_{t|t-h}$ measure. The ratio shows that agreement about the structure of the economy reduces overall disagreement by around 30% at $h = 0$, by around 40% at $h = 4$, and nearly halves disagreement for the cumulative $h = 4$ forecasts.

Figure 1 displays the time paths for $h = 0$ (top panel) and $h = 4$ cumulative forecasts (bottom panel) for the 1990:4 to 2013:3 surveys. At both horizons disagreement increases around the 2007:4-2009:2 recession, in part emanating from higher disagreement about the

[6] All the forecasters who responded to 12 or more surveys are included.

outlook for the individual variables (as given by $\sqrt{\prod_j S^{jj}_{t|t-h}}$).[7] This is significantly reduced, especially for the $h = 4$ forecasts, once the offset for correlations between forecasts is taken into account using $\sqrt{\det\left(S_{t|t-h}\right)}$. Disagreement at the $h = 4$ horizon shows little correlation with the business cycle save for the latest recession (the 1990:3-1991:1 and 2001:1-2001:4 recessions have no apparent effect). The 1990:3-1991:1 recession inflates disagreement about the current-quarter forecasts, and there are other spikes in the product of the three variables' standard deviations which are not attributable to business cycle peaks and troughs. Some of these are reduced or even largely removed in the overall measure which takes cross-variable correlations into account.

In summary, multivariate measures of disagreement: i) decline in the forecast horizon; ii) show less excessive variation of a non-business cycle variety when an adjustment is made for cross-variable correlations in variables (reflecting a degree of commonality in beliefs about how the economy operates), and; iii) are counter-cyclical - disagreement at $h = 0$ increases during the first and last recessions in the sample, while for $h = 4$ disagreement only registers an increase during the last (2007:4-2009:2) recession.

Table 4 provides a breakdown of the components of $S_{t|t-h}$, averaged over t. The cross-sectional standard deviations for the three growth rates are declining in h. Disagreement about the 2-year ahead annual forecasts and the $h = 4$ cumulative forecasts are broadly similar. Forecasts are expected to eventually converge to the historical sample average for a stationary variable, so that this result is not surprising. [8] On the contrary, note that disagreement about the Great Ratios (expressed here in natural logarithms, $c/y = \ln(C/Y)$ and $i/y = \ln(I/Y)$) is increasing in h. However, whereas King, Plosser, Stock and Watson (1991) establish that c/y and i/y constitute cointegrating combinations in the post WWII period up to the end of their sample (1990), the one-sector balanced growth path model that implies these great ratios are stationary appears to fit the US experience less well over the last quarter of a century: see, e.g., Whelan (2003). According to Whelan (2003), the equilibrium relationships should be amended to $\ln C - 0.95 \ln Y$ and $\ln I - 1.35 \ln Y$. However the findings are little affected. If there was a common perception of the long-run savings ratio (either $\ln(C/Y)$ or $\ln\left(C/Y^{0.95}\right)$) then we would expect disagreement about the forecasts of these quantities to decline in h. There is little evidence that the forecasts of consumption, investment and output are generated by forecasters with such beliefs.

Table 5 indicates increasing contemporaneous correlations in h, at least for consumption and output: these are the off-diagonal elements that reduce the overall disagreement measure

[7] Note that the time-dating in the figures corresponds to the forecast origin, not the period being targetted. Hence the rise in disagreement around the time of the recent recession lags the onset (based on the NBER chronology), not taking place until the second half of 2008.

[8] Although Andrade *et al.* (2014) consider a model of forecaster behaviour which features 'shifting endpoints', and in which disagreement does not disappear as the forecast horizon increases.

relative to the product of the individual variables' standard deviations.

Finally, our analysis of disagreement so far has considered forecasts of a number of variables made at the same origin of the same future period (i.e., the same horizon), but we can also analyze forecasts of different horizons, and look at 'dynamic correlations'. We consider the correlations between forecasts, say, $\Delta c_{t|t-h}$ and $c_{t-1|t-h}\ /y_{t-1|t-h}$. If error correction holds in the forecasts, then other things being equal, we would expect a negative correlation. A forecaster who expected an above average value of $c_{t-1|t-h}\ /y_{t-1|t-h}$, for example, would likely have a below average $\Delta c_{t|t-h}$ (where average refers to the cross-section). All the correlations in table 6 are of the wrong sign other than in the last row (the correlation between $\Delta y_{t|t-h}$ and $i_{t-1|t-h}\ /y_{t-1|t-h}$).

In summary, our results suggest a relatively high degree of commonality in forecasters beliefs concerning the mechanisms that generate the growth rates of consumption, investment and output growth, but there is little evidence from the forecaster disagreement results for belief in a common equilibrium towards which the economy is moving.

The next step is to test the optimality of the forecasts of the individuals, using general tests, and tests designed to detect the omission of error-correction.

4 Tests of forecast optimality, and testing for omitted error-correction terms

We consider the popular test of Mincer and Zarnowitz (1969), as well as extensions due to Patton and Timmermann (2012), and a test directed toward finding omitted error correction. The tests are described in section 4.1, and their ability to detect omitted error-correction is established in section 4.2.

4.1 Tests of Forecast Optimality

The obvious general test to use is the Mincer and Zarnowitz (1969) (MZ) regression which tests forecast optimality at a given horizon. The regression is:

$$y_t = \delta_0 + \delta y_{t|t-h} + u_t \tag{3}$$

where the observations range over t for a given h, and the null of optimality is that $\delta_0 = 0$ and $\delta = 1$, and HAC standard errors are used for multi-step forecasts to account for the overlapping forecasts phenomenon. Recently Patton and Timmermann (2012) propose the univariate optimal revision regression (henceforth ORR), which is applicable when fixed-event forecasts (see, e.g., Nordhaus (1987) and Clements (1995)) are available, as here. This test can be motivated by writing a short horizon forecast (e.g., $h_1 = 1$) as:

$$y_{t|t-h_1} \equiv y_{t|t-h_H} + d_{t|h_1,h_2} + \dots d_{t|h_{H-1},h_H} \tag{4}$$

where $h_1 < h_2 < \ldots < h_H$, with h_H the longest horizon forecast of the target y_t, and $d_{t|h_j,h_{j+1}} = y_{t|t-h_j} - y_{t|t-h_{j+1}}$. Then rather than regressing y_t on $y_{t|t-h_1}$, say, as in (3), the ORR test replaces $y_{t|t-h_1}$ by (4) and estimates:

$$y_t = \delta_0 + \delta_H y_{t|t-h_H} + \sum_{i=1}^{H-1} \delta_i d_{t|h_i,h_{i+1}} + u_t. \tag{5}$$

The null hypothesis is that $\delta_0 = 0$ and $\delta_1 = \delta_2 = \ldots = \delta_H = 1$. Under the null, the error for the short-horizon forecast $y_{t|t-h_1}$ is uncorrelated with all forecasts of the target y_t made at earlier times (and hence on smaller information sets). Equation (5) becomes $y_t = y_{t|t-h_1} + u_t$. Hence the ORR test has power to reject the null against the alternative that the short-horizon forecast error is systematically related to revisions in earlier forecasts of the target value.

Neither the MZ or ORR tests are designed to detect the form of mis-specification of interest here. To these two tests we add an additional test that supplements the MZ regression (3) with additional variables known at $t - h$, e.g.,

$$y_t = \delta_0 + \delta y_{t|t-h} + \kappa' z_{t-h} + u_t$$

where (as shown) z_{t-h} may comprise a vector of such variables. The null is now that $\delta = 0$, $\delta_1 = 1$ and $\kappa = 0$. We consider a simpler version in which $\delta = 1$ in the maintained model, and the test regression is:

$$y_t - y_{t|t-h} = \delta_0 + \kappa' z_{t-h} + u_t \tag{6}$$

with $\delta_0 = 0$ and $\kappa = 0$ under the null. Given the focus on whether survey expectations embody long-run information, z_t will be the error- or equilibrium-correction term at time t, and we refer to this test as ECT.

Patton and Timmermann (2012) show that the actual value of y_t can be replaced by a short-horizon forecast, say, $y_{t|t-h_1}$, to give:

$$y_{t|t-h_1} = \delta_0 + \delta y_{t|t-h_2} + u_t \tag{7}$$

where $h_2 > h_1$, and e.g., :

$$y_{t|t-h_1} = \delta_0 + \delta_H y_{t|t-h_H} + \sum_{i=2}^{h_H-1} \delta_i d_{t|h_i,h_{i+1}} + u_t \tag{8}$$

when $h_H > h_{H-1} > \ldots > h_1$. This requires that the short-horizon forecast is a conditionally unbiased proxy for the actual value, and the interpretation of, say, (7) is that it tests the rationality of both $y_{t|t-h_1}$ and $y_{t|t-h_2}$. The practical advantage, as discussed in section 6, is to obviate the need to select the vintage(s) of data to be used as actual values. We can perform a similar substitution in (6), For example, replacing y_t by $y_{t|t-h_1}$ results in:

$$y_{t|t-h_1} - y_{t|t-h_2} = \delta_0 + \kappa' z_{t-h_2} + u_t. \tag{9}$$

That the revision between the fixed-event forecasts of y_t made at time h_1 and h_2 should be unpredictable at the time the longer horizon (h_2) forecast is made is called a 'strong-efficiency' test by Nordhaus (1987). Tests on (8) are closely related to the weak efficiency tests of Nordhaus (1987): forecast revisions should be unpredictable from earlier revisions.

In the next section we consider whether these tests are able to detect the omission of cointegration.

4.2 Power to detect the omission of cointegration in forecast generation

We assume a first-order VAR for n $I(1)$ variables with r cointegrating vectors, and specify the model that omits cointegration as a zero-order VAR in differences. Simple expressions can be obtained for the regression parameters in the test regressions for forecasts generated from models with and without error-correction terms. We explore by Monte Carlo the effect of complicating factors such as including lags in the DV model.

Following Clements and Hendry (1995a), the first-order dynamic linear system is written as:

$$x_t = \Upsilon x_{t-1} + \tau + v_t, \tag{10}$$

where $v_t \sim IN[0, \Omega]$ for $t = 1, 2, \ldots, T$. Υ is an $n \times n$ matrix of coefficients. In vector equilibrium-correction form we have:

$$\Delta x_t = \Pi x_{t-1} + \tau + v_t, \tag{11}$$

where $\Pi = \Upsilon - I_n = \alpha\beta'$ and α and β are $n \times r$ of rank $r < n$ when $x_t \sim I(1)$ and the cointegrating rank is r.[9] The subscript on the identity matrix, 'I_n' denotes its order.

We can also write the system as:

$$z_t = G z_{t-1} + \Psi_0 + \epsilon_t, \tag{12}$$

[9] In addition we assume that $\alpha'_\perp \Theta \beta_\perp$ is full rank, where Θ is the mean-lag matrix (here, simply Υ), and $\alpha_\perp$ and $\beta_\perp$ are full column rank $n \times (n-r)$ matrices such that $\alpha'\alpha_\perp = \beta'\beta_\perp = 0$ (see Johansen (1992)).

where $z_t' = (x_t'\beta : \Delta x_{b,t}')$, with β normalized such that its first r rows are the identity matrix, i.e., $\beta = (I_r : \beta_2')'$, $\epsilon_t \sim IN_n[0, \Sigma]$. In (12), $\Psi_0 = (\tau'\beta : \tau_b')' = Q\tau$ where $\tau_b = J'\tau$ when $J' = (0 : I_{n-r})$, $\Delta x_{b,t} = J'\Delta x_t$, and:

$$Q = \begin{pmatrix} \beta' \\ J' \end{pmatrix}, \; G = \begin{pmatrix} (I_r + \beta'\alpha) & 0 \\ \alpha_b & 0 \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ \alpha_b & 0 \end{pmatrix} \tag{13}$$

and:

$$\Sigma = \begin{pmatrix} \beta'\Omega\beta & \beta'\Omega J \\ J'\Omega\beta & J'\Omega J \end{pmatrix}.$$

The system in (12) determines both the conditional and unconditional means and variances of all the $I(0)$ variables. For $\alpha \neq 0$, the long-run solution for the system is defined by:

$$E[z_t] = (I_n - G)^{-1} Q\tau = \begin{pmatrix} -(\beta'\alpha)^{-1}\beta'\tau \\ \alpha_b(\beta'\alpha)^{-1}\beta'\tau + \tau_b \end{pmatrix}. \tag{14}$$

Using (14) we can show that the expectation of Δx_t is $E[\Delta x_t] = K\tau$, where $K = \left(I_r - \alpha(\beta'\alpha)^{-1}\beta'\right)$, so that $K\tau$ is the growth in the system, and $E[\beta' x_t] = -(\beta'\alpha)^{-1}\beta'\tau$ is given directly. We let $w_t = \beta' x_t$, and the cointegrating combinations follow a VAR(1):

$$w_t = \lambda w_{t-1} + \beta'\tau + \epsilon_{1t}, \tag{15}$$

where $\epsilon_{1t} = \beta v_t$.

Then, letting $x_{t+h|t}$ denote the conditional h-step ahead expectation (of $t+h$ conditional on period t) we have:

$$x_{t+h|t} = \Upsilon^h x_t + \sum_{i=0}^{h-1} \Upsilon^i \tau \tag{16}$$

and:

$$\begin{aligned} w_{t+h|t} &\equiv \beta' x_{t+h|t} = \beta'\Upsilon^h x_t + \sum_{i=0}^{h-1} \beta'\Upsilon^i\tau \\ &= \lambda^h w_t + \sum_{i=0}^{h-1} \lambda^i \beta'\tau \end{aligned} \tag{17}$$

using $\beta'\Upsilon^i = \lambda^i\beta$.[10]

[10] Note that:
$$\beta'\Upsilon^i = (\beta' + \beta'\alpha\beta')\Upsilon^{i-1} = (I_r + \beta'\alpha)\beta'\Upsilon^{i-1} = \lambda\beta'\Upsilon^{i-1} = \cdots = \lambda^i\beta'.$$

In terms of forecasting the differences, from (16):

$$\Delta x_{t+h|t} = \Upsilon^{h-1}\alpha\beta' x_t + \Upsilon^{h-1}\tau.$$

From $\Upsilon^n\alpha = \alpha\lambda^n$, and $\Upsilon^n = \Upsilon^{n-1}\left(I_n + \alpha\beta'\right) = \Upsilon^{n-1} + \alpha\lambda^{n-1}\beta = \ldots = I_n + \alpha\beta' + \alpha\lambda\beta' + \ldots + \alpha\lambda^{n-1}\beta' = I_n + \alpha\left(\sum_{i=0}^{n-1}\lambda^i\right)\beta'$:

$$\Delta x_{t+h|t} = \alpha\lambda^{h-1}\beta' x_t + \left(I_n + \alpha\left(\sum_{i=0}^{h-2}\lambda^i\right)\beta'\right)\tau.$$

For large h,

$$\Delta x_{t+h|t} \to \left(I_n - \alpha\left(\beta'\alpha\right)^{-1}\beta'\right)\tau = K\tau. \tag{18}$$

Consider now forecasts from the VAR in differences (DV), that is, the model that omits the cointegrating combinations. Denote the forecasts of the growth rates and cointegrating combinations by $\widetilde{\Delta x}_{t+h|t}$ and $\widetilde{w}_{t+h|t}$. In the first-order model for x_t, the growth rate forecasts are:

$$\widetilde{\Delta x}_{t+h|t} = K\tau \tag{19}$$

so that the variables are forecast to increase at their (population) average growth rates. The levels forecasts are then:

$$\widetilde{x}_{t+h|t} = \widetilde{x}_{t+h-1|t} + K\tau = x_t + hK\tau \tag{20}$$

and so the forecast of the cointegrating combination is:

$$\widetilde{w}_{t+h|t} = \beta'\widetilde{x}_{t+h|t} = w_t \tag{21}$$

It follows immediately that the correctly-specified and DV model forecasts of the growth rates converge in h (compare (18) and (19)), accounting for the finding of Clements and Hendry (1995b) that cointegration only improves short-horizon forecasts of growth rates. In terms of forecasting cointegrating combinations, the VECM forecasts are equivalent to using a stationary VAR(1) whereas the DV model forecasts correspond to using a no-change or random walk forecast. Some straightforward algebra establishes that the unconditional variance of the correctly-specified model is smaller by a positive definite matrix:

$$\sum_{r=0}^{h-1}\sum_{q=0}^{h-1}\beta'\alpha\lambda^r V\left[w_T\right]\lambda^{q\prime}\alpha'\beta \tag{22}$$

which does not disappear as h gets large. In the bivariate case, λ is a scalar, and (22) simplifies to:

$$\alpha V\left[w_T\right]\alpha'\left(\frac{1-\lambda^h}{1-\lambda}\right)^2 \tag{23}$$

indicating that the relative accuracy of the correctly-specified model is increasing in h (see Clements and Hendry (1995b) for details).

4.2.1 ORR test and forecasting cointegrating combinations.

Using (21), it is simple to show that the ORR test as in (8) will reject the null in population when we consider DV model forecasts of the cointegrating combination(s). [11] Suppose $H=3$ and $h_1=1$, $n=2$ and $r=1$, so there is a single cointegrating relationship, w_t, a scalar. Let $\widetilde{w}_{t|t-j}$ denote a forecast of w_t made at $t-j$ using the DV model. Then the ORR regression is given by:

$$\widetilde{w}_{t|t-1}=\delta_0+\delta_3\widetilde{w}_{t|t-3}+\delta_2\left(\widetilde{w}_{t|t-2}-\widetilde{w}_{t|t-3}\right)+u_t \tag{24}$$

Using (21), the DV forecasts are given by $\widetilde{w}_{t|t-j}=w_{t-j}$, and the ORR regression becomes:

$$w_{t-1}=\delta_0+\delta_3 w_{t-3}+\delta_2\left(w_{t-2}-w_{t-3}\right)+u_t.$$

From (15) the population values of the least squares estimators of the parameters δ_2 and δ_3 are $\delta_2=\delta_3=\lambda$, where $\lambda=1+\beta'\alpha$, and $|\lambda|<1$. The population value of δ_0 is $\delta_0=\beta'\tau\left(1-\lambda\right)^{-1}$. Suppose $\beta'=[1,-1]$. Recall that the null has $\delta_0=0$ and $\delta_2=\delta_3=1$. Hence the power of the test will be greater the further λ from 1, i.e., the larger the absolute value of the elements of $\alpha=[\alpha_1,\alpha_2]'$ (given that $\alpha_1<0$ and $\alpha_2>0$). Only when Π is the null matrix will $\delta_0=0$ and $\delta_2=\delta_3=1$, their values under the null.

4.2.2 MZ test and forecasting cointegrating combinations.

The MZ regression for a given h, when $h_2=h_1+1$, is:

$$\widetilde{w}_{t|t-h}=\delta_0+\delta\widetilde{w}_{t|t-h-1}+u_t. \tag{25}$$

Written for DV forecasts, we obtain:

$$w_{t-h}=\delta_0+\delta w_{t-h-1}+u_t.$$

From (15), $w_{t-h}=\lambda w_{t-h-1}+\beta'\tau+\varepsilon_{1t-h}$, the population values of δ_0 and δ are: $\delta_0=\lambda$

[11] We do not explicitly consider ORR tests of DV model forecasts of growth rates, given our DV model forecasts do not depend on t or h.

and $\delta = \lambda$, independent of h.

Were we to regress $\widetilde{w}_{t|t-1}$ on $\widetilde{w}_{t|t-h}$, say, then from:

$$w_{t-1} = \beta'\tau \sum_{i=0}^{h-2} \lambda^i + \lambda^{h-1} w_{t-h} + \sum_{i=0}^{h-2} \lambda^i \varepsilon_{1t-h-i}$$

we obtain $\delta_0 = \beta'\tau \sum_{i=0}^{h-2} \lambda^i$, and $\delta = \lambda^{h-1}$, so that the power will increase in h because the coefficient δ approaches 0 as h increases (compared to its hypothesized value of 1 under the null).

4.2.3 ECT test and forecasting cointegrating combinations.

Consider (9) for a given h, when $h_2 = h_1 + 1$, and setting $z_{t-h_2} = w_{t-h_2}$, then in terms of DV model forecasts $\widetilde{w}_{t|t-h}$ and $\widetilde{w}_{t|t-h-1}$ of the cointegrating combination:

$$w_{t-h} - w_{t-h-1} = \delta_0 + \kappa w_{t-h-1} + u_t. \tag{26}$$

From $w_{t-h} - w_{t-h-1} = (\lambda - 1) w_{t-h-1} + \beta'\tau + \varepsilon_{1t-h}$, $\kappa = \lambda - 1$.

When $h_1 = 1$, and $h_2 = h > 1$, for DV model forecasts we obtain:

$$w_{t-1} - w_{t-h} = \delta_0 + \kappa w_{t-h} + u_t. \tag{27}$$

From $w_{t-1} = \lambda^{h-1} w_{t-h} + \beta'\tau \sum_{i=0}^{h-2} \lambda^i + \sum_{i=0}^{h-2} \lambda^i \varepsilon_{1t-1-i}$, $w_{t-1} - w_{t-h} = \left(\lambda^{h-1} - 1\right) w_{t-h} + \beta'\tau \sum_{i=0}^{h-2} \lambda^i + \sum_{i=0}^{h-2} \lambda^i \varepsilon_{1t-1-i}$, and so $\kappa = \lambda^{h-1} - 1$, and the power of the test based on κ increases in h, as $\kappa \to -1$ as h gets large.

5 Monte Carlo

We assess the size and power properties of the tests in forecast samples of the size encountered in practice, when the form of mis-specification is the omission error-correction. We check the large forecast-sample results are as expected, and then assess the performance at empirical sample sizes. To this end, we generate data from:

$$y_t = (I_2 + \Pi) y_{t-1} + \delta + \varepsilon_t \tag{28}$$

where $y_t = [y_{1t}\ y_{2t}]'$, $\Pi = \alpha\beta'$, where α and β are 2 by 1, and $\varepsilon_t \sim IN(0, \Sigma)$. We set $\alpha = [-0.25\ 0.10]'$, $\beta = [1\ -1]'$, $\delta = [0\ 0]'$ and Σ is the identity matrix. We generate forecasts after estimating a correctly-specified model by maximum likelihood estimation with the reduced rank restriction for Π imposed, as well as from a VAR in differences (with either no, or 1, lagged differences). In all cases constant terms are estimated. We also report results for a VECM where we abstract from parameter estimation uncertainty by using the true values of the parameters (α, β and δ).

The variants of the tests that we consider in the Monte Carlo are those which use short horizon forecasts in place of the actual values, as the LHS variable, and for MZ and ECT, we consider 'adjacent forecasts', so that the RHS forecast is of length one longer than the LHS forecast. [12]Consequently the tests we analyze are:

1. For ORR, eqn. (8):

$$y_{t|t-h_1} = \delta_0 + \delta_H y_{t|t-h_H} + \sum_{i=2}^{H-1} \delta_i d_{t|h_i,h_{i+1}} + u_t$$

where the null is $\delta_0 = 0$, and $\delta_i = 1$, $i = 1, \ldots, H-1$, and we set $h_1 = 1$ and $H = 5$, and the generic variable 'y' is in turn Δx, Δy and w.

2. For MZ, eqn. (7):

$$y_{t|t-h_1} = \delta_0 + \delta y_{t|t-h_2} + u_t$$

where the null $\delta_0 = 0$, and $\delta = 1$, when $h = h_1 = h_2 - 1$ for $h = 1$ to 4.

3. For ECT, eqn. (9):

$$y_{t|t-h_1} - y_{t|t-h_2} = \delta_0 + \kappa w_{t-h_2} + u_t$$

where the null is $\delta_0 = 0$, and $\kappa = 0$, when $h = h_1 = h_2 - 1$ for $h = 1$ to 4.

Forecast data are generated such that there are a number of forecasts of length $h = 1, \ldots, H$ (with $H = 5$), for each of $P - H + 1$ forecast targets. Writing the first target as t, then for the longest-horizon forecast $y_{t|t-H}$ the model estimation sample is 1 to $t-H$ (after discarding startup observations), for the $H-1$ length forecast $y_{t|t-(H-1)}$ the estimation sample is augmented by one observation (1 to $t - H + 1$), and so on. That is, we adopt a recursive forecasting scheme. We then consider forecasts of $t + 1$, and so on.

Then, for each replication of the Monte Carlo we have forecast data $\mathbf{F}$, organized as shown in (29) to show how it is used in the tests.

$$\overbrace{\begin{bmatrix} y_{t|t-H} & y_{t|t-(H-1)} & \cdots & y_{t|t-2} & y_{t|t-1} \\ y_{t+1|t+1-H} & y_{t+1|t+1-(H-1)} & & y_{t+1|t-1} & y_{t+1|t} \\ \vdots & & & & \\ y_{t+P-H|t+P-2H} & y_{t+P-H|t+P-2H+1} & & y_{t+P-H|t+P-H-2} & y_{t+P-H|t+P-H-1} \end{bmatrix}}^{\mathbf{F}} \quad (29)$$

[12] As established in section 4.2, using adjacent forecasts, as opposed to fixing h_1 and increasing h_2, would be expected to result in some power loss.

As an illustration, for the MZ test with $h = 1$ forecasts as the dependent variable, and $h_2 = 2$, the last column of (29) is regressed on the penultimate column. The ORR tests can be obtained in a straightforward way using the columns of $\mathbf{F}$.

For the ECT test, the explanatory variable is $[w_{t-H},\ w_{t+1-H},\ \ldots,\ w_{t+P-H}]'$ when $h_2 = H$ (with $h_1 = H-1$), and the dependent variable is the second column of $\mathbf{F}$ minus the first column.

The way in which data are generated means that as we increase T, the initial (or minimum) estimation window size increases, as does the average window size. Increasing P increases the the number of observations ('targets') for the tests, and also the size of the average estimation samples.

5.1 Simulation results

To focus on how the properties of the tests depend on the forecast sample, we assume large estimation samples (in excess of 500 observations) throughout. We also report results for a VECM using the population values of the parameters to gauge whether the large-estimation sample results are close to the known parameter results for the VECM.

The left side of table 7 reports the rejection rates of the tests when there are 500 sequences of 1 to 5-step ahead forecasts (columns headed $P = 500$). Consider first the ORR rejections frequencies (first panel). The null is that the revision between the 1 and 2-step ahead forecasts is unpredictable from revisions to earlier forecasts of the same target, so that the regression error is iid under the null and correction to the estimated variance-covariance matrix of the parameter estimates is not required (see section 6.2). As expected, the actual sizes are indistinguishable from the 5% nominal for the Known-parameter VECM, and for the estimated VECM for forecasting the growth rates (Δx and Δy), but a little inflated for forecasting the cointegrated combination w. Results are reported for two models that eschew long-run relations, the DV(0), which simply includes a constant, and DV(1), which includes a lagged difference (as well as the constant). The rejection frequencies are a little lower for Δx and Δy, for the DV(1), but for both models the rejection rate is 1 when the ORR test is applied to w.

For forecast samples of the size available in the SPF (i.e., around 100 forecasts), the right side of table 7 indicates lower rejection frequencies for Δx and Δy, especially when a lagged difference term is included to mop up the serial correlation resulting from the omission of the ECM (as in the DV(1)). However, the rejection frequencies remain close to 1 for forecasting w (and the results for the estimated VECM indicate only minor size distortions at this forecast sample size).

Next, for the MZ test, we can again show that no autocorrelation correction is required when h-step forecasts are regressed on $h+1$-step ahead forecasts. The middle panel of table 7 shows MZ tests of the VECM forecasts were approximately correctly sized at all h for Δx and Δy for $P = 500$, with a slight tendency to being over-sized when $P = 100$, and for forecasting

w when the VECM is estimated. However, the powers to reject the null for Δx and Δy for the DV(1) are low, suggesting MZ tests for the growth rates are unlikely to detect the omission of ECM terms in practice, unless we test forecasts of the cointegrating combination.

For the ECT test (bottom panel) similar considerations apply as for the MZ test, in that no auto-correlation correction of the parameter-estimator covariance matrix is required. The MZ and ECT results are similar for the VECM with estimated parameters, and identical by construction when parameters are known.[13] However, unlike for the MZ tests, we obtain relatively high rejection rates for the DV(1) for both Δx and Δy, and not just for w. As for MZ, tests based on w also have rejection rates close to 1 for all $h = 1$.

In summary, the Monte Carlo evidence suggests that the tests most likely to reliably detect long-run mis-specification are: *i*) the ORR test applied to w; *ii*) the MZ test applied to w; and *iii*) the ECT test applied to Δx, Δy and w.

6 Application to Survey Expectations

6.1 The effects of data revisions

When analyzing survey expectations data, such as the US SPF, a number of complicating factors arise. Firstly, the actual data are subject to revisions (see, e.g., the review articles by Croushore (2011a, 2011b) as well as Landefeld, Seskin and Fraumeni (2008) and Fixler, Greenaway-McGrevy and Grimm (2014)). This raises two potential difficulties: the question of which vintage of data to use as actual values; and possible distortionary effects from periodic rebasing of the data. Generally, researchers have preferred to use a vintage released soon after the reference quarter, rather than the latest-available vintage at the time of the investigation. The latter will typically include benchmark revisions, rebasings, and other methodological changes to the way the data are collected and measured.[14] This issue can be side-stepped by using short-horizon forecasts in place of actuals, as in section 5. Even so, the effects of the regular rebasings of the data remain problematic for the following reason. When the actual series are rebased the levels of the variables are shifted. The SPF forecasts are rebased in tandem. For example, following the 1992Q1 rebasing, the forecast returns to the 1992Q1 (and subsequent)

[13] The equivalence between the MZ and ECT tests for the known parameter VECM holds for the following reason. The MZ test can be parameterized as regressing $\Delta x_{t|t-h} - \Delta x_{t|t-h-1}$ on a constant and $\Delta x_{t|t-h-1}$, and testing whether these two regressors jointly have any explanatory power. ECT regresses $\Delta x_{t|t-h} - \Delta x_{t|t-h-1}$ on w_{t-h-1} and a constant. But the two sets of regressors span the same space, so the tests are equivalent. In addition, the VECM known parameter tests are identical for Δx, Δy and w for both MZ and ECT. Consider the MZ tests. The LHS variables $\Delta x_{t|t-h}$ and $\Delta y_{t|t-h}$ are both multiples of w_{t-h}, and the RHS variables, $\Delta x_{t|t-h-1}$ and $\Delta y_{t|t-h-1}$, respectively, are multiples of w_{t-h-1}, so the test statistic values are identical.

[14] The Real Time Data Set for Macroeconomists (RTDSM) maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)) has greatly facilitated the use of real-time data in macro analysis and forecasting research.

surveys are on the new base, while the returns to earlier surveys are on the old base. Rebasings may have relatively small effects on the growth rates of the series, but the same is not true of the levels of the series (or combinations of the levels of different series).[15]

Consider for example the forecasts of the levels of consumption (say) from the 1992Q1 survey. Consider the simple MZ forecast revision regression, where the forecast $y_{t|t-h_1}$ is related to $y_{t|t-h_2}$, and suppose we restrict attention to adjacent forecasts, namely, set $h_2 = h_1+1$. Then the regression becomes:

$$y_{t|t-h} = \delta + \delta_1 y_{t|t-(h+1)} + u_t.$$

Problems may arise when the forecast origin $(t-h)$ of the shorter-horizon forecast is 1992Q1 (or any other quarter in which the national accounts are rebased). Then the forecast $y_{t|t-h}$ is on the new base, but $y_{t|t-(h+1)}$ is on the old base. A simple remedy which we adopt is to omit all the pairs of observations $\{y_{t|t-h}, y_{t|t-(h+1}\}$ for which $t-h$ corresponds to a rebasing quarter.[16] In our sample period there are 8 quarters in which level shifts occurred,[17] so that we lose 8 of the 129 survey quarters (1981:3 to 2013:3). Were actual values used (instead of shorter horizon forecasts), a solution would be to create a series of real-time actuals purged of the effects of rebasings,[18] but we do not pursue that option here.

In the case of the ORR regressions, the use of forecasts with horizons up to 4 quarters means that the surveys corresponding to the quarters of the shifts and the each of the subsequent 3 quarters need to be discarded, resulting in the loss of approximately nearly four times as many survey quarters.[19] As intimated above, in empirical work the effect of rebasing on growth rates is likely to be small. For growth rates we report results for all surveys. For the cointegrating combinations we remove observations from surveys which span rebasings.

A common problem with surveys such as the SPF is that there are missing observations for each respondent. Individuals do not file a response to every survey. As in much of the literature, we assume that data are missing 'at random' so that the sample is representative of the population.[20]

[15] However, such shifts may be benign, if for example C and Y are altered by the same multiplicative factor following a rebasing, the log of the ratio of C to Y would be unaffected.

[16] This simple strategy will remove all the observations where the LHS and RHS variables straddle a rebasing. For example, when $t-h$ corresponds to 1992Q2, the two forecast origins are 1992Q2 and 1992Q1, and similarly if $t-h$ is earlier than 1992Q1.

[17] The rebasing quarters are 1986Q1, 1992Q1, 1996Q1, 1999Q4, 2000Q2, 2004Q1, 2009Q3 and 2013Q3.

[18] See Clements and Galvão (2012) for a treatment of this problem in the context of estimating and forecasting with vintage-based vector autoregressions.

[19] For example, consider forecasts of the target 1992Q4. The longest horizon forecast is the $h=4$ forecast made in 1991Q4, which will be on a different basis from the forecasts of 1992Q4 made in 1992Q1, 1992Q2, 1992Q3 and (the current quarter forecast of) 1992Q4. When we roll the target forward to 1993Q1, all the forecasts are on the post 1992Q1 base.

[20] An exception is López-Pérez (2015) who considers whether the decision to contribute is related to perceived uncertainty about the outlook.

6.2 Actual values versus short(er) horizon forecasts

The use of adjacent horizon forecasts in the tests of forecast optimality serves to circumvent the need for autocorrelation corrections. Consider the MZ regression given by:

$$y_{t|t-h_1} = \delta + \delta_1 y_{t|t-h_2} + u_t \tag{30}$$

where $h_2 > h_1$, and the null is that $\delta = 0$ and $\delta_1 = 1$.

Provided $h_2 = h_1 + 1$, i.e., that the forecasts are adjacent, then under the null the error term in the regression will be serially uncorrelated. To understand why this is the case, consider for example the two rows of the regression corresponding to the targets t and $t+1$. Under the null, $y_{t|t-h_1} - y_{t|t-h_2} = u_t$ and $y_{t+1|t+1-h_1} - y_{t+1|t+1-h_2} = u_{t+1}$. Suppose the time series y_t is written as an infinite-order moving average:

$$y_t = \psi(L)\varepsilon_t = \sum_{j=1}^{h_1} \psi_{h_1-j}\varepsilon_{t-h_1+j} + \psi_{h_1}\varepsilon_{t-h_1} + \ldots + \psi_{h_2-1}\varepsilon_{t-h_2+1} + \sum_{j=0}^{\infty} \psi_{h_2+j}\varepsilon_{t-h_2-j},$$

which collapses to:

$$y_t = \psi(L)\varepsilon_t = \sum_{j=1}^{h} \psi_{h-j}\varepsilon_{t-h+j} + \psi_h\varepsilon_{t-h} + \sum_{j=0}^{\infty} \psi_{h+1+j}\varepsilon_{t-h-1-j},$$

when $h_2 - 1 = h_1 \equiv h$. The lag polynomials are written as above to clearly show the past, present and future (relative to the forecast origin) components.

In the general case,

$$\begin{aligned} y_{t|t-h_1} &= E(y_t \mid \mathcal{I}_{t-h_1}) = \psi_{h_1}\varepsilon_{t-h_1} + \ldots + \psi_{h_2-1}\varepsilon_{t-h_2+1} + \textstyle\sum_{j=0}^{\infty} \psi_{h_2+j}\varepsilon_{t-h_2-j} \\ y_{t|t-h_2} &= E(y_t \mid \mathcal{I}_{t-h_1}) = \textstyle\sum_{j=0}^{\infty} \psi_{h_2+j}\varepsilon_{t-h_2-j}, \end{aligned}$$

so that $y_{t|t-h_1} - y_{t|t-h_2} = \psi_{h_1}\varepsilon_{t-h_1} + \ldots + \psi_{h_2-1}\varepsilon_{t-h_2+1}$. (Here $\mathcal{I}_{t-h}$ denotes the information set at time $t-h$, and consists of $\varepsilon_{t-h}, \varepsilon_{t-h-1}, \ldots$.) For the target $t+1$, we have

$$y_{t+1|t+1-h_1} - y_{t+1|t+1-h_2} = \psi_{h_1}\varepsilon_{t+1-h_1} + \ldots + \psi_{h_2-1}\varepsilon_{t+1-h_2+1}$$

(by simply replacing 't' by 't+1'). In general, then $Cov\left(y_{t|t-h_1} - y_{t|t-h_2}, y_{t+1|t+1-h_1} - y_{t+1|t+1-h_2}\right) \neq 0$ since the two revisions have common ε's. But when $h_2 - 1 = h_1$,

$$Cov\left(y_{t|t-h_1} - y_{t|t-h_2}, y_{t+1|t+1-h_1} - y_{t+1|t+1-h_2}\right) = Cov\left(\psi_h\varepsilon_{t-h}\psi_h\varepsilon_{t+1-h}\right) = 0,$$

and no autocorrelation-correction is needed.

The advantage of using short-horizon forecasts in place of actuals is immediately apparent. It is a simple matter to check that the regression errors would be correlated for, say:

$$y_t = \delta + \delta_1 y_{t|t-h} + u_t$$

whenever $h > 1$.[21]

The upshot is that for simple MZ regressions we estimate:

$$y_{t|t-h} = \delta + \delta_1 y_{t|t-(h+1)} + u_t$$

for $h = 1, 2, \ldots$.

Note that the ORR regressions do not require autocorrelation-consistent estimation of the covariance matrix of the regression parameter estimates. Using the $h_1 = 0$ forecast as the actual value, provided the next-shortest horizon forecast $h_2 = h_1 + 1 = 1$, then under the null the regression error is $y_{t|t-h_1} - y_{t|t-h_2} = u_t$, which is serially uncorrelated by the above results.

In terms of the regression (9) using equilibrium-correction terms, no correction is needed either, provided $h_2 = h_1 + 1$. Hence we can use:

$$y_{t|t-h} - y_{t|t-(h+1)} = \delta_0 + \kappa' z_{t-(h+1)} + u_t \tag{31}$$

for $h = 1, 2, \ldots$. However, as noted previously, the use of shorter-horizon forecasts on the LHS requires that such forecasts differ from the actual by an unpredictable error. Further, the use of adjacent forecasts may have lower power than, say, fixing h_1 and increasing h_2. We explore the empirical import of these issues in section 6.4.

6.3 Results

We calculate ORR, MZ and ECT tests for all respondents who made 30 or more forecasts over the period 1981:3 to 2013:3, for $h = 0, 1, 2, 3$, where $h = 0$ denotes a forecast of the current quarter. We use shorter-horizon forecasts on the LHS, and the forecasts are adjacent (for MZ and ECT). The proportion of forecasters for whom the null was rejected for each of the tests is recorded in table 8. The first panel reports ORR tests for four different values of h_H in eqn. (8). The ECT test results are for two variants of the test: *i*) regressing the forecast revision onto a forecast of the ECM, $w_{t|t-h-1}$, and *ii*) regressing the forecast revision onto w_{t-h-1}. (The latter is the version analyzed in section 5)). Here w is either the consumption-output,

[21] Under the null, $y_t - y_{t|t-h} = \sum_{j=1}^{h} \psi_{h-j} \varepsilon_{t-h+j}$, and $y_{t+1} - y_{t+1|t+1-h} = \sum_{j=1}^{h} \psi_{h-j} \varepsilon_{t+1-h+j}$, and so $Cov\left(y_t - y_{t|t-h}, y_{t+1} - y_{t+1|t+1-h}\right) \neq 0$ when $h > 1$.

or investment-output, log ratio. We found the results were largely unchanged if instead we used the equilibria suggested by the two-sector model, and these results are not reported. To illustrate the timings, suppose $h = 1$ (in table 8). The regression using actual data is:

$$y_{t|t-1} - y_{t|t-2} = \delta + \delta_1 w_{t-2} + u_t.$$

If t is 2010:1, say, the the dependent variable is the 1-step ahead forecast of y in 2010:1 from the 2009:4 survey, less the 2-step ahead forecast (of 2010:1) from the 2009:3 survey, and the right-hand-side variable is the value of the (log) consumption-income ratio at the time of the 2009:3 survey. This is calculated from the 2009:3 vintage of data for observation quarter 2009:2 (the data for 2009:3 will not be available at the time the forecast is made).

Finally, the results reported here drop forecast observations that straddle rebasings when the tests are applied to, or include, levels terms (i.e., the consumption and investment to output ratios), but otherwise no observations are dropped.

Table 8 indicates that the tests of forecast optimality reject for a quarter to a three-quarters of all respondents, depending on the test, and the variable being forecast, and the horizon. However, the pattern of results we obtain does not suggest that the rejections are primarily due to a neglect of equilibrium relationships. The neglect of equilibrium relationships would be expected to show in high rejection rates for the growth rates for ECT compared to the ORR and MZ tests of the forecasts of growth rates. But the non-directed tests do not reject any less often than ECT. The test outcomes for ECT are similar for the two versions of the test we report. Moreover there is no indication that the forecasts of the cointegrating combinations are rejected more often than forecasts of growth rates, which would have been expected (based on the simulation results of section 5 for the forecasts generated by models without long-run equilibria).

So although the ECT results using w (actual and forecast) indicate a systematic relationship between forecast revisions and EC terms for around a third of the forecasters at short horizons ($h = 0, 1$), and fewer at longer horizons, factors other than incorrect long-run specification have a role to play. The ORR and MZ tests are of course uninformative about the reasons for the rejection of forecast optimality. Finally, we ran ECT tests including the growth rates of consumption and income (or investment and income) at the time the longer-horizon forecast was made. In terms of (31), $z_{t-(h+1)}$ consists of Δc_{t-h-2}^{t-h-1} and Δc_{t-h-2}^{t-h-1} (when we consider consumption of output growth forecasts), where the scripts formalize the argument rehearsed above: the superscript indicates that the data vintage is that available at the time the longer horizon (length $h + 1$) forecast was made, and the subscript indicates this is for the values of the growth rates in the previous quarter.

The rejection rates for ECT tests in the direction of lagged growth rates are similar to those when ECM terms are used. Hence the directed tests using either information set (ECMs, or growth rates) are lower than the ORR and MZ tests (with $h_H = 4$ and $h = 3$, respectively), indicating that the general tests reject optimality in directions that are not solely attributable to either omitted ECM, or to dynamic mis-specification.

6.4 Robustness check: Using Actual Values

In section 6.3 we use short-horizon forecasts in place of actuals in the regression-based tests of forecast optimality. The advantages of doing so include not having to choose which vintage of data to use as actual values, and that for short-horizon forecasts of length one less than the RHS forecasts corrections for auto-correlated errors are not required. Nevertheless, the substitution of the short-horizon forecasts for the actual values requires the optimality of the former, otherwise tests of revisions may have no power to detect mis-specification, a situation described by Nordhaus (1987).[22]

In this section we consider departures from full-information rational expectations forecasts which are not detectable using short-horizon forecasts, but are readily detectable using actual values. Hence the rejection rates of forecast optimality recorded in section 6.3 may be an under-estimate if cases of non-optimality are masked by the use of short-horizon forecasts as actual values. The results of re-running the tests using actual values on the LHS are described in section 6.4.2.

6.4.1 Mis-specified forecasts

A possible alternative to full-information rational expectations forecasts, especially at 'large' h, is that the updated forecast differs from the forecast made in the previous period by a random error, unrelated to the target variable, and hence does not incorporate new information. That is:

$$y_{t|t-h} = y_{t|t-h-1} + u_{t|t-h} \tag{32}$$

where $u_{t|t-h}$ is orthogonal to y_t and $y_{t|t-h-1}$. Then consider an MZ regression of $y_{t|t-h_1}$ on a constant and $y_{t|t-h_2}$, where $h_1 < h_2$, as in (7). Then the population values of the regression parameters are:

$$\delta = \frac{Cov\left(y_{t|t-h_1}, y_{t|t-h_2}\right)}{Var\left(y_{t|t-h_1}\right)} = 1,\ \delta_0 = E\left(y_{t|t-h_1}\right) - \delta E\left(y_{t|t-h_2}\right) = 0$$

[22] Nordhaus (1987, p. 673) described the possibility of 'A baboon could generate a series of weakly efficient forecasts by simply wiring himself to a random-number generator, but such a series of forecasts would be completely useless.'

since $y_{t|t-h_1} = y_{t|t-h_2} + \sum_{s=t-h_2+1}^{t-h_1} u_{t|s}$ and $Cov\left(\sum_{s=t-h_2+1}^{t-h_1}, y_{t|t-h_2}\right) = 0$. Forecast revisions which do not add news will nevertheless be detectable using actual values, since:

$$\delta = \frac{Cov\left(y_t, y_{t|t-h_1}\right)}{Var\left(y_{t|t-h_1}\right)} = 0,\ \delta_0 = E\left(y_t\right) - \delta E\left(y_{t|t-h_1}\right) = E\left(y_t\right)$$

since neither $\delta = 1$ nor $\delta_0 = 0$ (unless $E\left(y_t\right)$ happens to equal zero).

An alternative form of mis-specification (considered by Patton and Timmermann (2012)) is to assume that the reported forecast is a linear transformation of the optimal, e.g.,:

$$y_{t|t-h} = \gamma_h + \lambda_h y_{t|t-h}^* + w_{t|t-h},\quad w_{t|t-h} \sim D\left(0, \sigma_{w_h}^2\right) \tag{33}$$

where $y_{t|t-h}^* = E\left(y_{t|t-h} \mid \mathcal{I}_{t-h}\right)$ is the conditional expectation (and so differs from $y_{t|t-h}$ by a random error with a conditional mean of zero on $\mathcal{I}_{t-h}$). Clearly $\gamma_h \neq 0$ (with $\lambda_h = 1$) would produce biased forecasts which are detectable, but Patton and Timmermann (2012) show that there are values of the vector $\left(\gamma_h, \lambda_h, \sigma_{w_h}^2\right)$ other than $(1, 0, 0)$ which constitute non-detectable mis-specification, i.e., for which $\delta_0 = 0$ and $\delta = 1$ does not hold. A case of interest suggested by the recent literature is when the forecasts differ from the optimal by a random (reporting or measurement) error,[23] corresponding to $\gamma_h = 0$ and $\lambda_h = 1$, but $\sigma_{w_h}^2 \neq 0$. Then the population values of the MZ regression (7) are given by:

$$\begin{aligned}
\delta &= \frac{Cov\left(y_{t|t-h_1}^* + w_{t|t-h_1}, y_{t|t-h_2}^* + w_{t|t-h_2}\right)}{Var\left(y_{t|t-h_2}^* + w_{t|t-h_2}\right)} = \frac{Var\left(y_{t|t-h_2}^*\right)}{Var\left(y_{t|t-h_2}^*\right) + Var\left(w_{t|t-h_2}\right)} < 1, \\
\delta_0 &= E\left(y_{t|t-h_1}\right) - \delta E\left(y_{t|t-h_2}\right) \neq 0
\end{aligned}$$

since $Cov\left(y_{t|t-h_1}^* - y_{t|t-h_2}^*, y_{t|t-h_2}^*\right) = 0$ and so $Cov\left(y_{t|t-h_1}^*, y_{t|t-h_2}^*\right) = Var\left(y_{t|t-h_2}^*\right)$. Clearly, such a form of mis-specification is detectable.

In summary, some forms of mis-specification will be detectable using actual values on the RHS of regression equations such as the MZ test, but not using short-horizon forecasts. Specifically, ‘updating’ the revised forecast but not embodying additional information is a case in point.

In the next section we investigate the consequences of using forecasts in place of actual values in the context of forecasting Δc, Δi and Δy. We look at the concordance of the test outcomes for individual respondents.

[23] The assumption forecasters make idiosyncratic errors is sometimes included in studies of forecaster behaviour, e.g., Davies and Lahiri (1995).

6.4.2 Results

We again consider the most prolific forecasters for the period 1990:4 and 2013:3. There were generally fewer forecasts per forecaster than in table 2, because of the nature of the tests.[24] We consider results by forecaster using short(er)-horizon forecasts on the LHS of the test regressions, and using actual values. Directly comparing the use of shorter horizon forecasts and actual values for the same reported forecasts allows an assessment of the practical importance of the choice of the LHS variable for assessing forecast optimality.

For ORR we report results using eqn. (8) with $h_1 = 0$ (current quarter forecasts) and $h_H = 1, 2, 3, 4$, i.e., using short-horizon forecasts as the LHS variable. We also report ORR tests based on eqn. (5) with $h_H = 1, 2, 3, 4$, that is, with actual outcomes as the LHS variable.

For MZ regressions, we report tests of eqn. (7) with: *i*) $h_1 = 0, 1, 2, 3$ and $h_2 = h_1 + 1$ (adjacent forecasts); *ii*) $h_1 = 0$ and $h_2 = 1, 2, 3, 4$; and of eqn. (3) with $h = 1, 2, 3, 4$.

For the MZ results using actual values (3) the results are based on HAC corrections, as are the results for (7) *i*). [25] We use the actual values available two quarters after the reference quarter. We consider only forecasts of growth rates, and no adjustment is made for the potential effects of the rebasings.

Of interest are: 1) the potential loss of power from using short-horizon forecasts in place of actual values; 2) the value of increasing the number of revisions included in the ORR test; 3) the potential loss of power of considering adjacent forecasts in the MZ tests, i.e., MZ (*i*) above versus MZ (*ii*), and these questions are used to focus the discussion of the results.

We consider the results for consumption growth, then investment growth, and finally output growth.

For consumption growth, table 9 shows that the proportion of forecasters for whom the ORR test rejects is lower when the (current-quarter) forecast replaces the actual values, for example, 40% of forecasters rather than 60%, and that increasing h_H (the number of revisions in the test) does not have much effect. However, although the average results across respondents are broadly similar whether forecasts or actuals are used on the LHS, at the individual level the choice is often vital, and the inference made depends on the value of h_H. For example, for the first two respondents in the table there is no evidence against rationality using the forecast as the LHS variable, but the null hypotheses return p-values of zero when actual values are used. Moreover, for the fourth respondent (id407) the rejection occurs only when the forecast value is used. The smaller sub-table shows that the choice of using forecasts or actuals on the

[24] For example, consider the ORR test with $h_H = 4$. If a respondent failed to respond to one survey, say 2005:Q1, then the ORR test regression for observation for 2005:Q1 would be missing, but so would the test regression observations for 2005:Q2, 2005:Q3, 2005:Q4 and 2006:1Q (because we are missing, respectively, the 1, 2, 3 and 4 step forecasts from the 2005:Q1 survey).

[25] A current-quarter forecast ($h = 0$) corresponds to a 1-step ahead forecast, so that what we have termed $h = 1$ forecasts are overlapping and require a correction when the equation standard error is calculated.

LHS only yields the same inference (at the 5% level) for a given h_H around half the time. For example, for $h_H = 1$, the null is rejected for 20% of respondents for tests using both forecasts and actuals, and not rejected for 20% of respondents, so there is agreement 40% of the time. We conclude that in practical assessments of individual forecaster rationality the use of short-horizon forecasts in place of actuals may lead to different inferences. The tests may lack power against certain forms of mis-specification when short-horizon forecasts are used, but equally may falsely reject rationality for vintages of actuals not targetted by the forecaster.

By way of contrast, table 10 shows the MZ test rejection rates increase in 'h' for forecasts of consumption growth, and are higher than those of the ORR test for high h. Moreover, the rejection rates in the middle panel using the current-quarter forecasts on the LHS are higher than using actual values: at the longest horizon, we reject for 91% of respondents on the former, compared to 74% on the latter.

The sub-table indicates the inferences we make will tend to be the same for the majority of individuals when we use forecasts on the LHS whether the forecast is adjacent to the RHS forecast, or the current-quarter forecast (the agreement rates are around 80% except at horizon 2), although the choice between forecasts and actuals will lead to differences as often as not for individuals at horizon 1.

Table 11 shows that the ORR tests for forecasts of investment growth do reject rationality for a larger number of respondents as h_H increases: for $h_H = 4$ rationality is rejected for around half the individuals whether forecasts or actuals are used on the LHS. The MZ tests of the investment growth forecasts (table 12) also indicate that rationality is rejected more often as the horizon increases, and occurs for 4 in every 5 respondents at $h = 4$ using current-quarter forecasts on the LHS. At horizons of 2 or more the MZ tests using forecasts reject more often than using actuals (matching the findings for consumption growth), and reject more often than the ORR tests (again, matching the findings for the consumption forecasts).

Table 13 shows the ORR forecasts of output growth are rejected for around half the respondents using the forecast as the LHS variable irrespective of h_H, but that increasing numbers are rejected as h_H increases using actual values on the LHS. As for consumption and investment, MZ tests are increasing in the horizon, and rejection rates using actuals are far higher than for the ORR tests, and rationality is rejected for nearly all respondents at horizons of 3 and 4.

In summary, for all three variables, the MZ tests using current-period forecasts on the LHS reject more often than using actuals. Moreover, the MZ tests tend to reject more often than using ORR tests, raising the possibility that the estimation of the additional parameters in the ORR tests adversely affects the power.

That more rejections are observed for MZ using forecasts as actuals suggests that concerns that this approach may lack power are unfounded.

7 Conclusions

In response to the question: Are professional forecasters guided by long-run relationships when they form their forecasts, the answer would appear to be no. We have considered three variables which economic theory would suggest *ought* to be cointegrated. We have argued that although cointegration is often regarded as a 'long-run' phenomenon, because cointegration implies an error-correction model, then at least in terms of the growth rates of the variables, the largest influence of cointegration/error-correction should be on the short-horizon forecasts. Hence considering year-ahead forecasts, as opposed to 10 or 20-year ahead forecasts, is not a reason for our failure to detect the effects of equilibrium relationships in the forecasts.

Our analysis of disagreement suggests a reasonable degree of commonality in beliefs concerning the mechanisms that generate the growth rates of consumption, investment and output growth, but little evidence of belief in a common equilibrium towards which the economy is moving. Further, the pattern of test outcomes from testing for forecast optimality does not suggest long-run mis-specification is the reason for roundly rejecting optimality for the majority of forecasters.

Hence we have failed to find evidence that forecasts of these three variables are driven by belief in long-run relationships, or that the rejections of forecaster optimality are primarily due to the neglect of such terms in expectation formation. It may be that despite the theory-based support for balanced growth rates, the evidence for such is weak in the actual data, and their influence on expectations formation is minor.[26]

On a positive note, we unearth a number of findings about forecast behaviour, and testing for optimality.

In terms of forecaster behaviour. Considering the forecast vectors of those individuals who respond often over the forecast period, we find that there are statistically significant differences between forecasters in terms of the degree to which they (dis)agree with the vector of consensus forecasts whether we weight those differences by the cross-sectional covariance matrices of forecasts or not. Secondly, there is persistence in terms of the extent to which individuals disagree with the consensus over two separate time periods, for the shortest-horizon forecasts, but less so for the year-ahead forecasts. Thirdly, cross-sectional covariances between individuals' forecasts, which measure agreement about the relationships between the variables being forecast, are relatively large, in the sense that an overall disagreement measure which includes these covariances is 30 to 40% smaller than a measure based on the indvidual-variable cross-sectional standard deviations alone (specifically, the product of these standard deviations). Finally, as mentioned above, there is agreement about growth rates, but not about the dynamic correlations between

[26] There might be measurement issues, or heterogeneity amongst forecasters in their beliefs about steady-state growth rates and long-run relationships, etc.

forecasts of growth rates and error-correction terms, which would signal the influence of equilibrium beliefs, nor is there any indication of agreement regarding the equilibrium quantities themselves.

In terms of testing for forecaster optimality.

We find that the extension to the MZ testing framework given by the ORR test does not tend to produce more rejections on average than simply using the MZ test with (say) year-ahead forecasts. That is, allowing for a systematic influence from a sequence of revisions to the target does not tend to result in more rejections. Secondly, using a short-horizon forecast in place of the actual value in the MZ test increases rejection rates. One might have expected a reduction, as the de-coupling of the forecasts and actuals in the testing procedure means that internally-consistent forecasts will not be flagged as non-optimal. That is, provided the short-horizon forecast is unpredictable from the longer-horizon forecast, the tests will not reject even if the forecasts are unrelated to actual values.

References

Andrade, P., Crump, R., Eusepi, S., and Moench, E. (2014). Fundamental disagreement. Working papers 524, Banque de France.

Andrade, P., and Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, *60*(8), 967–982.

Ball, L., Jalles, J. T., and Loungani, P. (2015). Do forecasters believe in Okun's Law? An assessment of unemployment and output forecasts. *International Journal of Forecasting*, *31*(1), 176–184.

Banternghansa, C., and McCracken, M. W. (2009). Forecast disagreement among FOMC members. Working papers 2009-059, Federal Reserve Bank of St. Louis.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, **28**, 381–392.

Capistrán, C., and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, **41**, 365–396.

Christoffersen, P. F., and Diebold, F. X. (1998). Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics*, **16**, 450–458.

Clements, M. P. (1995). Rationality and the role of judgement in macroeconomic forecasting. *Economic Journal*, **105**, 410–420.

Clements, M. P., and Galvão, A. B. (2012). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business & Economic Statistics*, **30(4)**, 554–562. DOI: 10.1080/07350015.2012/707588.

Clements, M. P., and Hendry, D. F. (1995a). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146. Reprinted in T.C. Mills (ed.), *Economic Forecasting.* Edward Elgar, 1999.

Clements, M. P., and Hendry, D. F. (1995b). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.

Coibion, O., and Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities?. *Journal of Political Economy*, *120*(1), 116 – 159.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, **"November"**, 3–15.

Croushore, D. (2011a). Forecasting with real-time data vintages, chapter 9. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 247–267: Oxford University Press.

Croushore, D. (2011b). Frontiers of real-time data analysis. *Journal of Economic Literature*, **49**, 72–100.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(1)**, 111–130.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.

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

Del Negro, M., and Schorfheide, F. (2013). DSGE model-based forecasting. In Elliott, G., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 2.*, pp. 57–140: Amsterdam, Horth-Holland.

Dovern, J. (2014). A Multivariate Analysis of Forecast Disagreement: Confronting Models of Disagreement with SPF Data. Working papers 0571, University of Heidelberg, Department of Economics.

Elliott, G. (2006). Forecasting with trending data. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 555–604: Elsevier, Horth-Holland.

Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.

Engle, R. F., and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, **35**, 143–159.

Fixler, D. J., Greenaway-McGrevy, R., and Grimm, B. T. (2014). The revisions to GDP, GDI, and their major components. *Survey of Current Business*, **August**, 1–23.

Johansen, S. (1992). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.

King, R. G., Plosser, C. I., Stock, J. H., and Watson, M. W. (1991). Stochastic trends and economic fluctuations. *American Economic Review*, **81**, 819–840.

Kosobud, R., and Klein, L. (1961). Some econometrics of growth: Great ratios of economics. *Quarterly Journal of Economics*, **25**, 173–198.

Lahiri, K., and Sheng, X. (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics*, **144(2)**, 325–340.

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

López-Pérez, V. (2015). Does uncertainty affect participation in the European Central BankâĂŹs Survey of Professional Forecasters?. Working paper series no. 1807, European Central Bank.

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.

Mincer, J., and Zarnowitz, V. (1969). The evaluation of economic forecasts. In Mincer, J. (ed.), *Economic Forecasts and Expectations*, pp. 3–46. New York: National Bureau of Economic Research.

Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. *Review of Economics and Statistics*, **69**, 667–674.

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(7)**, 803–820.

Patton, A. J., and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, *30*(1), 1–17.

Pesaran, M. H., and Weale, M. (2006). Survey expectations. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 715–776: Elsevier, Horth-Holland.

Rich, R., and Tracy, J. (2010). The relationships among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts. *Review of Economics and Statistics*, **92(1)**, 200–207.

Rich, R. W., and Butler, J. S. (1998). Disagreement as a measure of uncertainty: A comment on Bomberger. *Journal of Money, Credit and Banking*, **30**, 411–419.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co.

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

Whelan, K. (2003). A Two-Sector Approach to Modeling U.S. NIPA Data. *Journal of Money, Credit and Banking*, *35*(4), 627–56.

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*, pp. 25–58: Princeton University Press.

Zarnowitz, V., and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, **95(3)**, 591–621.

Table 1: Description of Forecast Data and Real-Time Data

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

The SPF data are from the Philadelphia Fed website (http://www.phil.frb.org/econ/spf/). For the investment series we used RNRESIN + RRESINV.
The real-time data were downloaded from http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/.

Table 2: Multivariate Disagreement Statistics for Individuals with Most Responses

1	2	3	4	5	6	7	8	9	10	11	12	13
	$h=0$				$h=4$		Ranking for $h=0$			Ranking for $h=4$		
id.	Ave	Equal(1)	Equal(2)	Corr.	Ave	Corr.	All	1st half	2nd half	All	1st half	2nd half
Weighting by cross-section covariances												
20	1.40	0.92	0.04	0.08	1.39	0.07	6	4	5	6	1	15
65	1.34	0.86	0.01	-0.11	1.42	0.10	2	2	4	2	9	6
84	1.23	0.50	0.00	0.02	1.42	0.05	1	1	2	1	3	11
411	1.39	0.95	0.03	-0.19	1.55	-0.04	5	3	8	3	16	4
420	1.55	1.00	0.28	0.00	1.54	-0.01	12	8	14	12	13	12
423	1.55	0.99	0.31	0.06	1.46	0.04	13	5	16	13	7	14
429	1.44	0.97	0.08	0.06	1.43	-0.06	7	9	6	7	4	13
431	1.63	1.00	0.50	0.05	1.59	-0.09	16	16	13	16	15	7
433	1.49	0.99	0.15	0.06	1.37	0.09	11	14	7	11	11	1
446	1.45	0.97	0.10	0.07	1.52	0.07	8	7	10	8	14	5
456	1.48	0.98	0.14	0.09	1.48	0.12	10	11	9	10	2	16
463	1.38	0.92	0.03	-0.03	1.36	-0.14	3	10	3	4	8	2
472	1.47	0.99	0.10	-0.09	1.42	-0.06	9	6	11	9	5	10
483	1.59	1.00	0.36	0.07	1.44	-0.02	15	12	15	15	12	3
484	1.39	0.92	0.03	0.09	1.40	0.17	4	13	1	5	6	8
504	1.57	1.00	0.33	-0.01	1.47	0.12	14	15	12	14	10	9
Not weighting by cross-section covariances												
20	0.64	0.91	0.12	0.05	0.50	-0.02	5	4	6	6	1	11
65	0.54	0.53	0.01	0.03	0.46	0.22	2	1	2	2	5	3
84	0.54	0.50	0.00	-0.06	0.60	0.12	1	2	1	1	7	14
411	0.73	0.98	0.36	0.06	0.67	0.08	11	11	10	11	16	10
420	0.77	1.00	0.51	0.00	0.58	-0.08	16	13	15	16	14	7
423	0.69	0.98	0.22	-0.02	0.53	0.19	8	5	9	8	2	12
429	0.75	0.99	0.44	0.28	0.62	0.05	12	16	8	12	10	16
431	0.77	1.00	0.50	0.02	0.57	0.10	15	12	16	15	13	5
433	0.76	0.98	0.47	0.20	0.52	0.21	13	15	12	13	12	1
446	0.67	0.97	0.15	0.15	0.59	0.11	7	7	7	7	15	2
456	0.70	0.98	0.24	0.11	0.57	0.24	9	3	14	9	6	13
463	0.64	0.93	0.10	-0.06	0.55	0.20	6	9	5	5	8	9
472	0.76	0.99	0.48	0.06	0.65	-0.03	14	14	13	14	11	15
483	0.71	0.99	0.26	0.26	0.53	0.01	10	8	11	10	9	8
484	0.61	0.81	0.06	0.17	0.48	0.01	3	6	3	3	3	4
504	0.64	0.94	0.08	0.03	0.48	0.18	4	10	4	4	4	6

The 2nd and 6th columns denote the mean (across surveys) value of $D_{i,t|t-h}$ (eqn. 2) for $h=0$ and $h=4$. Columns (5) and (7) are the first-order correlation for $h=0$ and $h=4$ respectively of $D_{i,t|t-h}$, all for forecasts of Δc, Δi and Δy. For $h=0$ we also report the p-values of testing the equality of means of each individual against forecaster id84 (Equal(1)) and against id456 (Equal(2)). The tests are based on the sample mean of id84 or id431 minus the other forecaster. Hence a p-value greater than 0.95 indicates the other forecaster has a population D_i significantly larger than that of the smallest (id84, Equal(1)). A p-value less that 0.05 suggests a D_i significantly less than the largest (id 431, Equal(2)) at the 5% level.
Column 8 ranks the forecasters in terms of average $D_{i,t|t-h}$ (i.e., column 2) across all surveys for $h=0$, and columns 9 and 10 show the rankings if instead the averages are calculated on the first half of the sample or the second half of the sample (again for $h=0$). Columnns 11 to 13 given the ranking for the whole sample, and the two sub-samples, when $h=4$.
The top panel reports results for the multivariate disagreement measure, where $S_{t|t-h}$ is calculated as in eqn. (2), and the bottom panel sets $S_{t|t-h}$ equal to the identity matrix, so that $D_{i,t|t-h}$ corresponds to the Euclidean distance between the forecast vector and consensus forecast vector at each point in time.

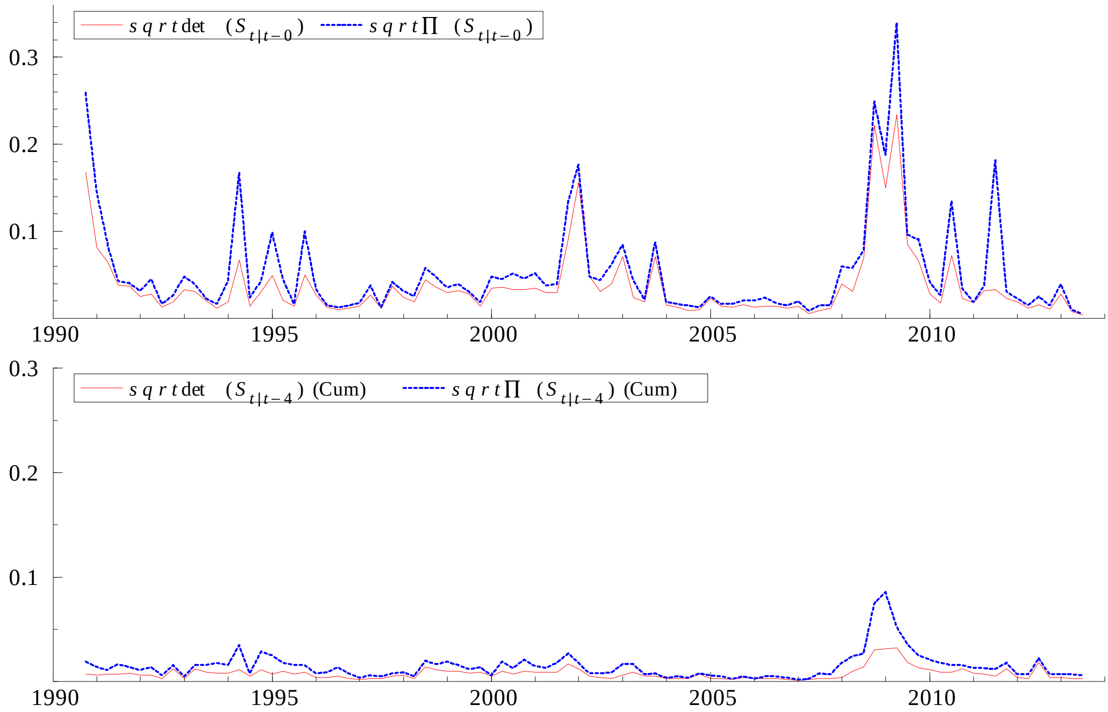

Figure 1: Measures of disagreement (overall, and for diagonal $S_{t|t-h}$) for $h = 0$, and for $h = 4$ (cumulated forecasts).

Table 3: Overall disagreement

	$h=0$	$h=1$	$h=4$	$h=4$ (Cum.)
1. $\sqrt{\det\left(S_{t\|t-h}\right)}$	0.037	0.025	0.019	0.008
2. $\sqrt{\prod_j S^{jj}_{t\|t-h}}$	0.054	0.040	0.031	0.015
3. Ratio 1. to 2.	0.695	0.616	0.600	0.521

The entries in the first two rows are the averages (over t, 1990:4 to 2013:3, for the specified h) of the disagreement measures given in the first column. The third row records the ratios of the entries in the first two rows.

Table 4: Individual Variables - Disagreement

	$h=0$	$h=1$	$h=4$	$h=4$ (Cum.)	$h=2$ years
$\Delta c_{t\|t-h}$	0.246	0.217	0.201	0.160	0.171
$\Delta i_{t\|t-h}$	0.854	0.736	0.678	0.557	0.518
$\Delta y_{t\|t-h}$	0.219	0.214	0.201	0.152	0.150
$c_{t\|t-h} - y_{t\|t-h}$	0.229	0.337	0.608	.	0.751
$i_{t\|t-h} - y_{t\|t-h}$	0.811	1.259	2.442	.	2.881
$c_{t\|t-h} - 0.95y_{t\|t-h}$	0.224	0.331	0.598	.	0.740
$i_{t\|t-h} - 1.35y_{t\|t-h}$	0.810	1.244	2.401	.	2.867

The table reports the averages of the cross-sectional deviations over 1990:4 to 2013:3.

Table 5: Contemporaneous Correlations in Growth Rate Forecasts

	$h=0$	$h=1$	$h=4$	$h=4$ (Cum.)
$\Delta c_{t\|t-h}, \Delta y_{t\|t-h}$	0.510	0.592	0.613	0.673
$\Delta i_{t\|t-h}, \Delta y_{t\|t-h}$	0.330	0.417	0.414	0.483

Table 6: Dynamic Correlations between h-step Growth Rate Forecasts and $h-1$-step ECM Forecasts

	$h=1$	$h=2$	$h=3$	$h=4$
$\Delta c_{t\|t-h}, c_{t-1\|t-h}/y_{t-1\|t-h}$	0.129	0.199	0.199	0.207
$\Delta y_{t\|t-h}, c_{t-1\|t-h}/y_{t-1\|t-h}$	-0.101	-0.092	-0.066	-0.094
$\Delta i_{t\|t-h}, i_{t-1\|t-h}/y_{t-1\|t-h}$	0.425	0.501	0.492	0.524
$\Delta y_{t\|t-h}, i_{t-1\|t-h}/y_{t-1\|t-h}$	0.135	0.151	0.138	0.106

Table 7: Monte Carlo Estimates of Rejection Fequencies

ORR, $H = 5$

	$P = 500$			$P = 100$		
	Δx	Δy	w	Δx	Δy	w
VECM	0.05	0.05	0.09	0.06	0.06	0.08
VECM_K	0.05	0.05	0.05	0.06	0.06	0.06
DV(0)	0.64	0.78	1.00	0.25	0.28	1.00
DV(1)	0.57	0.53	1.00	0.16	0.09	0.99

MZ, h on $h+1$

	$P = 500$				$P = 100$			
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
VECM, Δx	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
VECM, Δy	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
VECM, w	0.06	0.07	0.09	0.11	0.07	0.07	0.08	0.09
VECM_K, Δx	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
VECM_K, Δy	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
VECM_K, w	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
DV(0), Δx	0.45	0.44	0.44	0.44	0.32	0.32	0.32	0.32
DV(0), Δy	0.50	0.50	0.50	0.50	0.37	0.37	0.37	0.37
DV(0), w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DV(1), Δx	0.02	0.27	0.13	0.39	0.03	0.19	0.12	0.29
DV(1), Δy	0.02	0.01	0.09	0.21	0.01	0.02	0.10	0.19
DV(1), w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ECT, h on $h+1$

	$P = 500$				$P = 100$			
	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
VECM, Δx	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06
VECM, Δy	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
VECM, w	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
VECM_K, Δx	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
VECM_K, Δy	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
VECM_K, w	0.05	0.05	0.05	0.05	0.07	0.06	0.07	0.06
DV(0), Δx	0.98	0.98	0.98	0.98	0.36	0.36	0.36	0.36
DV(0), Δy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DV(0), w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DV(1), Δx	0.89	0.91	0.85	0.94	0.64	0.60	0.59	0.31
DV(1), Δy	0.99	0.98	0.77	1.00	0.72	0.78	0.40	0.97
DV(1), w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The top panel gives the rejection frequencies for (8), reproduced here as $y_{t|t-h_1} = \delta_0 + \delta_H y_{t|t-h_H} + \sum_{i=2}^{H-1} \delta_i d_{t|h_i,h_{i+1}} + u_t$, of the null $\delta_0 = 0$, and $\delta_i = 1$, $i = 1, \ldots, H-1$, for $h_1 = 1$, $H = 5$, for Δx, Δy and w. VECM_K denotes the Known-parameter VECM. The middle panel is for (7), i.e., $y_{t|t-h_1} = \delta_0 + \delta y_{t|t-h_2} + u_t$, of the null $\delta_0 = 0$, and $\delta = 1$, when $h = h_1 = h_2 - 1$ for $h = 1$ to 4. The final panel is of (9), $y_{t|t-h_1} - y_{t|t-h_2} = \delta_0 + \kappa w_{t-h_2} + u_t$, of the null $\delta_0 = 0$, and $\kappa = 0$, when $h = h_1 = h_2 - 1$ for $h = 1$ to 4.
The left side gives results for a forecast sample size of 500, the right side for a sample of 100.

Table 8: Empirical Results: Proportion of Forecasters for which the Null is Rejected

		ORR		
	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
Δc	0.38	0.41	0.42	0.50
Δi	0.28	0.38	0.41	0.60
Δy	0.61	0.52	0.44	0.50
c/y	0.41	0.27	0.25	0.55
i/y	0.33	0.25	0.14	0.11
		MZ		
	$h = 0$	$h = 1$	$h = 2$	$h = 3$
Δc	0.38	0.57	0.69	0.75
Δi	0.28	0.44	0.54	0.59
Δy	0.61	0.56	0.63	0.64
c/y	0.44	0.33	0.48	0.56
i/y	0.33	0.34	0.26	0.32
		ECT		
Using forecasts of the error-correction term				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$
Δc	0.29	0.24	0.21	0.13
Δi	0.33	0.50	0.26	0.16
$\Delta y, (c/y)$	0.35	0.32	0.15	0.13
$\Delta y, (i/y)$	0.44	0.33	0.28	0.06
Using the actual value of the error-correction term at the forecast origin				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$
Δc	0.26	0.21	0.24	0.13
Δi	0.27	0.50	0.26	0.19
$\Delta y, (c/y)$	0.38	0.32	0.18	0.13
$\Delta y, (i/y)$	0.44	0.29	0.21	0.09
Using the actual value of Δc and Δy, or Δi and Δy, at the forecast origin				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$
Δc	0.21	0.24	0.21	0.09
Δi	0.36	0.34	0.26	0.16
$\Delta y, (c/y)$	0.24	0.32	0.21	0.19
$\Delta y, (i/y)$	0.47	0.41	0.15	0.03

In the table $h = 0$ refers to a forecast of the current (i.e., survey) quarter. The rejection frequencies are calculated for a nominal size of 5%.

Table 9: Consumption growth: ORR tests with Short-Horizon Forecasts and Actual Values

	ORR - LHS variable is a Forecast					ORR - LHS variable is an Actual Value				
id.	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
426	67	0.72	0.61	0.81	0.30	73	0.00	0.00	0.00	0.00
428	65	0.68	0.15	0.23	0.05	71	0.00	0.00	0.00	0.00
421	58	0.97	0.03	0.04	0.00	65	0.25	0.29	0.12	0.14
407	53	0.00	0.00	0.00	0.01	61	0.44	0.35	0.09	0.72
446	64	0.49	0.58	0.11	0.01	68	0.43	0.30	0.17	0.15
433	64	0.04	0.05	0.30	0.66	67	0.00	0.00	0.00	0.00
411	47	0.01	0.04	0.15	0.16	55	0.58	0.80	0.23	0.08
84	46	0.06	0.20	0.35	0.27	54	0.04	0.03	0.01	0.25
463	54	0.02	0.04	0.31	0.33	58	0.02	0.02	0.08	0.16
484	48	0.15	0.33	0.16	0.02	54	0.00	0.02	0.08	0.16
65	56	0.00	0.00	0.00	0.00	60	0.00	0.00	0.00	0.00
20	40	0.29	0.00	0.01	0.46	48	0.00	0.00	0.00	0.00
429	44	0.40	0.16	0.27	0.36	50	0.09	0.01	0.00	0.00
472	41	0.06	0.24	0.35	0.00	47	0.08	0.19	0.35	0.16
483	37	0.40	0.56	0.56	0.62	43	0.05	0.01	0.02	0.08
510	49	0.11	0.48	0.58	0.76	51	0.04	0.09	0.08	0.30
504	45	0.00	0.00	0.00	0.00	48	0.00	0.00	0.00	0.00
507	40	0.00	0.00	0.00	0.00	44	0.30	0.00	0.00	0.00
518	37	0.21	0.15	0.00	0.00	40	0.00	0.00	0.00	0.00
512	36	0.02	0.03	0.10	0.28	38	0.57	0.76	0.29	0.00
Rejns.		0.40	0.50	0.35	0.45		0.60	0.65	0.55	0.50

	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
(i) Both reject	0.20	0.30	0.25	0.20
(ii) Neither rejects	0.20	0.15	0.35	0.25
(iii) Agree	0.40	0.45	0.60	0.45

The table displays p-values of the tests of ORR forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level.
The last 3 rows give the proportion of respondents for which the tests in the left and right panels (for a given h_H) (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

Table 10: Consumption growth: MZ tests with Short-Horizon Forecasts, Adjacent and Non-Adjacent, and Actual Values

	MZ - LHS variable is a Forecast, Adjacent					MZ - LHS variable is a Forecast, Non-Adjacent					MZ - LHS variable is an Actual Value				
id.	No. forecasts	$h_1 = 1$	$h_1 = 2$	$h_1 = 3$	$h_1 = 4$	No. forecasts	$h_2 = 1$	$h_2 = 2$	$h_2 = 3$	$h_2 = 4$	No. forecasts	$h = 1$	$h = 2$	$h = 3$	$h = 4$
		$h_2 = h_1 + 1$					$h_1 = 0$								
426	76	0.72	0.00	0.00	0.02	75	0.70	0.04	0.00	0.00	82	0.00	0.01	0.00	0.00
428	74	0.68	0.01	0.02	0.00	73	0.61	0.00	0.00	0.00	81	0.01	0.01	0.01	0.01
421	70	0.97	0.47	0.00	0.43	70	0.96	0.00	0.00	0.00	79	0.30	0.24	0.00	0.00
407	64	0.00	0.00	0.01	0.00	62	0.00	0.00	0.00	0.00	73	0.40	0.00	0.00	0.00
446	70	0.49	0.06	0.05	0.05	70	0.52	0.00	0.00	0.00	75	0.44	0.03	0.00	0.05
433	68	0.04	0.00	0.00	0.00	67	0.00	0.00	0.00	0.00	74	0.00	0.00	0.02	0.03
420	49	0.62	0.19	0.00	0.00	50	0.47	0.25	0.00	0.00	67	0.01	0.14	0.72	0.03
411	58	0.01	0.01	0.45	0.00	58	0.00	0.00	0.12	0.00	72	0.60	0.16	0.69	0.03
431	53	0.21	0.18	0.16	0.13	50	0.08	0.01	0.00	0.01	67	0.35	0.27	0.46	0.09
84	58	0.06	0.05	0.03	0.09	58	0.01	0.00	0.00	0.53	67	0.04	0.08	0.10	0.78
463	61	0.02	0.56	0.83	0.01	61	0.01	0.10	0.27	0.00	66	0.01	0.14	0.02	0.00
484	57	0.15	0.03	0.10	0.10	57	0.07	0.00	0.00	0.00	64	0.00	0.08	0.00	0.00
65	61	0.00	0.00	0.00	0.00	60	0.00	0.00	0.00	0.00	65	0.00	0.00	0.00	0.00
20	51	0.29	0.16	0.05	0.03	52	0.07	0.01	0.00	0.00	64	0.00	0.00	0.00	0.00
429	53	0.40	0.18	0.00	0.00	52	0.29	0.19	0.00	0.00	62	0.08	0.33	0.02	0.00
472	48	0.06	0.06	0.00	0.00	50	0.01	0.00	0.00	0.00	59	0.07	0.36	0.04	0.00
456	47	0.00	0.05	0.00	0.00	46	0.00	0.00	0.00	0.00	58	0.00	0.00	0.03	0.00
483	45	0.40	0.35	0.06	0.00	46	0.20	0.19	0.03	0.00	57	0.11	0.64	0.24	0.18
510	52	0.11	0.00	0.07	0.02	52	0.19	0.26	0.21	0.11	54	0.06	0.00	0.07	0.04
504	49	0.00	0.06	0.00	0.00	49	0.00	0.01	0.00	0.00	52	0.00	0.03	0.02	0.12
507	46	0.00	0.00	0.01	0.00	46	0.00	0.00	0.00	0.00	51	0.26	0.43	0.08	0.00
518	42	0.21	0.00	0.04	0.01	42	0.31	0.05	0.00	0.00	45	0.01	0.03	0.14	0.11
512	40	0.02	0.00	0.23	0.01	39	0.02	0.32	0.23	0.01	42	0.57	0.00	0.00	0.00
Rejns.		0.39	0.52	0.65	0.78		0.48	0.74	0.83	0.91		0.52	0.52	0.65	0.74

Forecast Horizon	1	2	3	4
LHS variable is an Adjacent Forecast versus Non-Adjacent Forecast				
(i) Adjacent and Non Adjacent Both Reject	0.39	0.43	0.65	0.74
(ii) Neither rejects	0.52	0.17	0.17	0.04
(iii) Agree	0.91	0.60	0.82	0.78
LHS variable is Actual versus Non-Adjacent Forecast				
(i) Actual and (Non Adjacent) Forecasts Both Reject	0.26	0.43	0.57	0.70
(ii) Neither rejects	0.26	0.17	0.09	0.04
(iii) Agree	0.52	0.60	0.66	0.74

The table displays p-values of the tests of MZ forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level. The bottom table gives the proportion of respondents for which the tests in the left and middle panels (for a given h_H), and in the middle and right panels (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

Table 11: Investment growth: ORR tests with Short-Horizon Forecasts and Actual Values

	ORR - LHS variable is a Forecast					ORR - LHS variable is an Actual Value				
id.	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
426	64	0.14	0.24	0.26	0.14	70	0.23	0.39	0.09	0.00
428	64	0.14	0.01	0.01	0.02	71	0.39	0.02	0.05	0.04
421	60	0.20	0.19	0.05	0.02	68	0.24	0.14	0.19	0.00
446	64	0.17	0.23	0.15	0.03	68	0.72	0.86	0.93	0.94
433	64	0.85	0.97	0.26	0.16	67	0.54	0.75	0.90	0.68
411	44	0.33	0.44	0.49	0.69	53	0.50	0.29	0.37	0.61
84	47	0.14	0.59	0.61	0.80	54	0.97	0.61	0.65	0.05
463	50	0.09	0.14	0.07	0.06	56	0.07	0.14	0.08	0.11
484	48	0.37	0.11	0.12	0.25	54	0.50	0.32	0.70	0.76
20	42	0.24	0.62	0.89	0.04	50	0.05	0.26	0.00	0.01
429	44	0.19	0.33	0.21	0.00	50	0.03	0.04	0.10	0.01
472	41	0.28	0.00	0.00	0.08	47	0.78	0.15	0.11	0.48
483	37	0.01	0.00	0.00	0.00	43	0.31	0.61	0.83	0.62
510	49	0.12	0.02	0.03	0.02	51	0.40	0.13	0.15	0.21
504	44	0.13	0.08	0.00	0.00	48	0.00	0.00	0.00	0.00
507	40	0.00	0.00	0.00	0.00	44	0.46	0.26	0.24	0.53
518	37	0.95	0.17	0.00	0.00	40	0.41	0.50	0.00	0.00
512	36	0.20	0.30	0.21	0.31	38	0.11	0.05	0.00	0.00
Rejns.		0.11	0.28	0.39	0.56		0.11	0.22	0.22	0.50

	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
(i) Both reject	0.00	0.06	0.11	0.33
(ii) Neither rejects	0.78	0.56	0.50	0.28
(iii) Agree	0.78	0.61	0.61	0.61

The table displays p-values of the tests of ORR forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level.
The last 3 rows give the proportion of respondents for which the tests in the left and right panels (for a given h_H) (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

Table 12: Investment growth: MZ tests with Short-Horizon Forecasts, Adjacent and Non-Adjacent, and Actual Values

	MZ - LHS variable is a Forecast, Adjacent					MZ - LHS variable is a Forecast, Non-Adjacent					MZ - LHS variable is an Actual Value				
id.	No. forecasts	$h_1 = 1$	$h_1 = 2$	$h_1 = 3$	$h_1 = 4$	No. forecasts	$h_2 = 1$	$h_2 = 2$	$h_2 = 3$	$h_2 = 4$	No. forecasts	$h = 1$	$h = 2$	$h = 3$	$h = 4$
		$h_2 = h_1 + 1$					$h_1 = 0$								
426	73	0.14	0.00	0.03	0.00	73	0.12	0.47	0.17	0.00	80	0.34	0.31	0.81	0.00
428	73	0.14	0.75	0.02	0.35	72	0.08	0.73	0.00	0.00	81	0.38	0.57	0.01	0.11
421	72	0.20	0.00	0.00	0.00	71	0.22	0.00	0.00	0.00	80	0.23	0.04	0.00	0.00
407	41	0.98	0.07	0.01	0.00	40	0.97	0.74	0.00	0.00	47	0.01	0.21	0.02	0.00
446	70	0.17	0.00	0.00	0.00	70	0.23	0.03	0.00	0.00	75	0.75	0.65	0.09	0.00
433	68	0.85	0.12	0.00	0.00	67	0.75	0.88	0.01	0.00	74	0.58	0.86	0.19	0.22
420	49	0.15	0.02	0.13	0.01	50	0.03	0.65	0.07	0.00	67	0.31	0.39	0.10	0.19
411	55	0.33	0.01	0.00	0.25	55	0.25	0.01	0.00	0.00	70	0.50	0.51	0.00	0.00
431	53	0.00	0.02	0.01	0.23	51	0.00	0.01	0.00	0.00	67	0.03	0.81	0.18	0.14
84	58	0.14	0.27	0.15	0.20	59	0.09	0.55	0.21	0.59	67	0.97	0.97	0.20	0.03
463	59	0.09	0.89	0.66	0.67	59	0.05	0.23	0.20	0.36	65	0.10	0.21	0.08	0.25
484	57	0.37	0.31	0.26	0.01	57	0.32	0.24	0.03	0.01	64	0.53	0.88	0.25	0.00
20	53	0.24	0.00	0.02	0.00	53	0.06	0.00	0.00	0.00	65	0.10	0.00	0.00	0.00
429	53	0.19	0.08	0.15	0.19	52	0.11	0.00	0.00	0.00	62	0.04	0.52	0.86	0.02
472	48	0.28	0.01	0.18	0.00	50	0.31	0.11	0.01	0.00	59	0.74	0.22	0.13	0.02
456	47	0.73	0.09	0.94	0.35	46	0.58	0.10	0.02	0.00	58	0.04	0.01	0.13	0.09
483	45	0.01	0.29	0.17	0.14	46	0.00	0.02	0.01	0.03	57	0.34	0.13	0.04	0.03
510	52	0.12	0.45	0.13	0.00	52	0.12	0.26	0.25	0.05	54	0.40	0.59	0.61	0.28
504	49	0.13	0.51	0.00	0.04	48	0.18	0.34	0.00	0.15	52	0.01	0.01	0.01	0.00
507	46	0.00	0.00	0.00	0.01	46	0.00	0.00	0.00	0.00	51	0.51	0.15	0.00	0.00
518	42	0.95	0.11	0.00	0.00	42	0.96	0.40	0.00	0.00	45	0.48	0.84	0.04	0.38
512	40	0.20	0.01	0.00	0.01	39	0.12	0.39	0.00	0.06	42	0.22	0.01	0.00	0.00
Rejns.		0.14	0.45	0.59	0.64		0.23	0.36	0.77	0.82		0.23	0.23	0.45	0.64

Forecast Horizon	1	2	3	4
LHS variable is an Adjacent Forecast versus Non-Adjacent Forecast				
(i) Adjacent and Non Adjacent Both Reject	0.14	0.27	0.55	0.55
(ii) Neither rejects	0.77	0.45	0.18	0.09
(iii) Agree	0.91	0.72	0.73	0.64
LHS variable is Actual versus Non-Adjacent Forecast				
(i) Actual and (Non Adjacent) Forecasts Both Reject	0.05	0.09	0.45	0.50
(ii) Neither rejects	0.59	0.50	0.23	0.05
(iii) Agree	0.64	0.59	0.68	0.55

The table displays p-values of the tests of MZ forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level. The bottom table gives the proportion of respondents for which the tests in the left and middle panels (for a given h_H), and in the middle and right panels (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

Table 13: Output growth: ORR tests with Short-Horizon Forecasts and Actual Values

	ORR - LHS variable is a Forecast					ORR - LHS variable is an Actual Value				
id.	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$	No. forecasts	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
426	67	0.02	0.03	0.06	0.11	73	0.02	0.00	0.00	0.00
428	65	0.02	0.02	0.09	0.14	71	0.95	0.00	0.01	0.00
421	60	0.66	0.39	0.24	0.38	68	0.73	0.89	0.09	0.00
407	57	0.00	0.01	0.03	0.03	64	0.85	0.67	0.15	0.06
446	64	0.06	0.00	0.00	0.00	68	0.98	0.01	0.01	0.01
433	64	0.00	0.00	0.00	0.01	67	0.54	0.00	0.01	0.00
411	47	0.00	0.02	0.08	0.11	55	0.58	0.04	0.00	0.01
84	47	0.87	0.80	0.44	0.42	54	0.36	0.20	0.00	0.02
463	54	0.02	0.03	0.09	0.08	58	0.01	0.01	0.20	0.05
484	46	0.02	0.06	0.21	0.08	53	0.33	0.70	0.95	0.97
65	57	0.09	0.08	0.01	0.00	60	0.00	0.00	0.00	0.00
20	40	0.32	0.62	0.98	0.96	48	0.00	0.00	0.00	0.00
429	44	0.00	0.01	0.02	0.00	50	0.63	0.01	0.00	0.00
472	44	0.46	0.71	0.48	0.04	50	0.64	0.17	0.19	0.00
483	37	0.02	0.19	0.08	0.15	43	0.15	0.66	0.25	0.21
510	49	0.00	0.00	0.00	0.00	51	0.03	0.06	0.03	0.01
504	45	0.03	0.09	0.27	0.29	48	0.00	0.01	0.00	0.00
507	40	0.00	0.00	0.00	0.00	44	0.14	0.04	0.02	0.01
518	37	0.70	0.85	0.12	0.00	40	0.14	0.00	0.00	0.00
512	36	0.08	0.02	0.02	0.04	38	0.90	0.69	0.70	0.85
Rejns.		0.60	0.55	0.40	0.50		0.30	0.60	0.65	0.80

	$h_H = 1$	$h_H = 2$	$h_H = 3$	$h_H = 4$
(i) Both reject	0.20	0.40	0.30	0.40
(ii) Neither rejects	0.30	0.25	0.25	0.10
(iii) Agree	0.50	0.65	0.55	0.50

The table displays p-values of the tests of ORR forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level.

The last 3 rows give the proportion of respondents for which the tests in the left and right panels (for a given h_H) (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

Table 14: Output growth: MZ tests with Short-Horizon Forecasts, Adjacent and Non-Adjacent, and Actual Values

	MZ - LHS variable is a Forecast, Adjacent					MZ - LHS variable is a Forecast, Non-Adjacent					MZ - LHS variable is an Actual Value				
id.	No. forecasts	$h_1 = 1$	$h_1 = 2$	$h_1 = 3$	$h_1 = 4$	No. forecasts	$h_2 = 1$	$h_2 = 2$	$h_2 = 3$	$h_2 = 4$	No. forecasts	$h = 1$	$h = 2$	$h = 3$	$h = 4$
		$h_2 = h_1 + 1$					$h_1 = 0$								
426	76	0.02	0.03	0.00	0.16	75	0.02	0.00	0.00	0.00	82	0.06	0.00	0.00	0.01
428	74	0.02	0.06	0.00	0.00	73	0.02	0.00	0.00	0.00	81	0.95	0.00	0.00	0.00
421	72	0.66	0.24	0.00	0.12	71	0.56	0.01	0.00	0.00	80	0.76	0.57	0.00	0.00
407	66	0.00	0.00	0.00	0.48	66	0.00	0.00	0.00	0.00	76	0.86	0.61	0.00	0.00
446	70	0.06	0.05	0.09	0.09	70	0.04	0.00	0.01	0.02	75	0.99	0.38	0.09	0.02
433	68	0.00	0.00	0.01	0.03	67	0.00	0.03	0.00	0.00	74	0.56	0.00	0.00	0.02
420	57	0.68	0.31	0.00	0.00	59	0.59	0.75	0.00	0.02	73	0.32	0.80	0.19	0.08
411	58	0.00	0.02	0.05	0.00	58	0.00	0.00	0.00	0.00	72	0.67	0.06	0.00	0.00
431	54	0.02	0.00	0.11	0.01	53	0.01	0.00	0.00	0.01	69	0.93	0.43	0.10	0.01
84	59	0.87	0.24	0.00	0.16	59	0.84	0.61	0.00	0.05	68	0.38	0.91	0.04	0.02
463	61	0.02	0.00	0.00	0.00	61	0.02	0.00	0.00	0.00	66	0.01	0.00	0.00	0.00
484	56	0.02	0.00	0.06	0.28	56	0.00	0.00	0.00	0.00	63	0.42	0.41	0.02	0.00
65	62	0.09	0.01	0.00	0.00	61	0.15	0.06	0.00	0.00	66	0.00	0.04	0.00	0.00
20	51	0.32	0.10	0.08	0.01	52	0.09	0.00	0.03	0.11	64	0.00	0.00	0.00	0.00
429	53	0.00	0.03	0.07	0.01	52	0.00	0.01	0.00	0.04	62	0.65	0.08	0.01	0.00
472	50	0.46	0.25	0.11	0.00	51	0.38	0.04	0.00	0.00	60	0.64	0.00	0.00	0.00
456	47	0.00	0.31	0.00	0.00	46	0.00	0.00	0.00	0.00	58	0.04	0.00	0.00	0.00
483	45	0.02	0.11	0.01	0.01	46	0.00	0.00	0.00	0.00	57	0.18	0.23	0.00	0.00
510	52	0.00	0.76	0.32	0.00	52	0.00	0.00	0.01	0.00	54	0.07	0.10	0.03	0.07
504	49	0.03	0.00	0.02	0.01	49	0.02	0.00	0.00	0.00	52	0.01	0.00	0.00	0.00
507	46	0.00	0.01	0.08	0.00	46	0.00	0.00	0.00	0.00	51	0.17	0.00	0.00	0.01
518	42	0.70	0.02	0.00	0.05	42	0.72	0.13	0.00	0.00	45	0.23	0.54	0.07	0.11
512	40	0.08	0.00	0.00	0.69	39	0.11	0.02	0.00	0.00	42	0.92	0.06	0.00	0.00
Rejns.		0.61	0.61	0.65	0.70		0.65	0.83	1.00	0.91		0.22	0.43	0.83	0.87

Forecast Horizon	1	2	3	4
LHS variable is an Adjacent Forecast versus Non-Adjacent Forecast				
(i) Adjacent and Non Adjacent Both Reject	0.61	0.52	0.65	0.65
(ii) Neither rejects	0.35	0.09	0.00	0.04
(iii) Agree	0.96	0.61	0.65	0.69
LHS variable is Actual versus Non-Adjacent Forecast				
(i) Actual and (Non Adjacent) Forecasts Both Reject	0.13	0.39	0.83	0.78
(ii) Neither rejects	0.26	0.13	0.00	0.00
(iii) Agree	0.39	0.52	0.83	0.78

The table displays p-values of the tests of MZ forecast optimality for each respondent, as well as the (average across tests) number of forecast observations for each respondent ('No. Forecasts'). The row 'Rejns' gives the proportion of respondents for which the null of optimality is rejected at the 5% level. The bottom table gives the proportion of respondents for which the tests in the left and middle panels (for a given h_H), and in the middle and right panels (i) both reject the null, (ii) both fail to reject the null, (iii) are in agreement (the sum of (i) and (ii)).

8 Appendix

8.1 Appendix 1. Data Cleaning.

We considered a number of apparently extreme observations, but ended up discarding only 3 forecasts. In the interests of transparency and replicability, these are briefly described below.

Survey quarter 1991:4, forecaster id. 422. Real output. The 4-quarter ahead is 4135.1. This gives a marked drop from the forecasts of the previous quarters. Were it 4315 instead, the forecasts of the four quarters of 1992 would equal the reported forecast for calendar 1992. Given this additional corroboration, we replace the forecast of the 1992:4 quarter by a missing value.

Survey quarter 1993:4, forecasters id 421. Real consumption. The forecast of 1994:1 is for 3418.3, which is out of kilter with the forecasts of the other quarters and with the reported forecast for 1994. The forecast of 1994:1 is replaced by a missing.

Survey quarter 1992:4, forecaster id 414. Real non-residential investment (component of total fixed investment). Forecast of 1993:3 shows a sharp drop, inconsistent with the forecast for 1993, and is replaced by a missing value.