Work in progress

Are revisions to *Consensus Forecasts* predictable?

Anna Kulhavy
Economics Department, Reserve Bank of New Zealand

Christie Smith
Economics Department, Reserve Bank of New Zealand

ABSTRACT

Simple time series analysis of changes to *Consensus* forecasts of annual GDP growth point to forecast revisions being serially correlated, suggesting that forecasts can be improved upon by anticipating future revisions. Our approach examines whether revisions to *Consensus* GDP growth forecasts for individual countries exhibit distinctive properties and therefore, contain extra information that would be lost through aggregation. ARMA models are applied to *Consensus* forecasts revisions of annual average GDP growth in the contemporaneous and following year sub-samples for 14 different countries. The individual country forecasts are then combined to obtain next period's expected forecast revision to an aggregate 14-country GDP series (GDP14). Forecast performance is empirically evaluated by deriving out-of-sample forecasts and comparing these to an ARMA forecast based on the aggregate GDP14 revisions. Although individual country forecast revisions exhibit distinctive time series properties, we do not find compelling statistical evidence to unequivocally favour a combination ARMA forecast over a GDP14 ARMA forecast. However, the adjustment each model implies for GDP14 out to a six-steps ahead in response to a range of impulses (stylised revisions) can differ considerably over some draws of revision vectors. Although we cannot draw any strong conclusions about the significance of these differences, this exercise is particularly useful for tracing the differences between the propagation mechanisms underlying each model. In response to a 0.2 percentage point forecast revision to the United States and Japanese contemporaneous year forecasts at time t, the combination forecast implies an initial adjustment to GDP14 at time t+1 of approximately 0.03 percentage points. For the same impulse, the GDP14 ARMA implies an adjustment to GDP14 of approximately 0.05 percentage points.

1 INTRODUCTION AND MOTIVATION

Every month *Consensus Forecasts Inc.*[1] survey macroeconomic forecasts in various countries from prominent forecasters in the respective countries. These forecasts – for a given variable and country – are then aggregated into a mean forecast, which is referred to as the *Consensus* forecast.[2] A number of macroeconomic variables are forecast by the organisations surveyed by *Consensus*, including gross domestic product (GDP) for the current and following year.

The Reserve Bank of New Zealand forms a measure of foreign demand for New Zealand goods and services by calculating a weighted average of the gross domestic products of 14 of New Zealand's trading partners.[3] Forecasts of this 'GDP14' aggregate are derived by aggregating *Consensus* forecasts for the individual countries. A filtered version of the GDP14 aggregate is used as an exogenous input into the Reserve Bank's Forecasting and Policy System (FPS). FPS is the Reserve Bank of New Zealand's primary macroeconomic model.

The Reserve Bank of New Zealand adopts *Consensus* forecasts in forming its view about the global outlook for growth because of the resource constraints imposed by being a small institution. Zarnowitz and Braun's (1993) analysis supports the use of '*Consensus*' forecasts, since they show that averaging a variety of forecasts – as *Consensus Inc.* does – almost invariably leads to improvements in forecast accuracy relative to the individual forecasts. Of the methods they examine, Clemen and Winkler (1986) also find that simple averages or Bayesian methods are the best means of combining forecasts.

In this paper we concentrate on the *Consensus* GDP forecasts which the Reserve Bank uses to forecast demand for New Zealand exports. A key issue for the Reserve Bank is to understand whether *Consensus* forecasts represent the 'best guess' about future foreign demand for New Zealand goods and services. One way of addressing this issue is to consider the time series properties of the forecast *revisions*.

Consensus Forecasts surveys forecasters monthly, asking for forecasts of various variables. With respect to GDP, forecasters are asked to provide a forecast of annual average GDP growth for the current year (which we will term the contemporaneous-year forecast) and the following year (which we will term the following-year forecast).[4]

[1] www.consensuseconomics.com

[2] *Consensus* forecasts should not to be confused with the concept in Gregory et al. (2001): they examine whether forecasters do hold some underlying 'consensus' view in common.

[3] GDP14 is currently an aggregate of the gross domestic products of Australia, United States, Japan, United Kingdom, Germany, France, Italy, Canada, China, Hong Kong, Malaysia, Singapore, South Korea and Taiwan weighted by each country's normalised share of New Zealand's exports. These 14 countries account for just over 90 per cent of New Zealand's exports. The composition of GDP14 is reviewed periodically.

[4] At a quarterly frequency forecasters are also asked to provide a quarterly GDP forecast track for the current and following year. At a half-yearly frequency forecasters are asked to provide long-run forecasts, which extend out approximately ten years into the future. This paper does not analyse these less frequent data.

Consensus forecasts of annual average GDP growth are regularly revised. The monthly forecast revisions for the aggregate annual average growth in GDP14 series range from –0.9 to 0.3, and the mean of the absolute revisions is around 0.1 percentage points per month.[5] Figures 1 and 2 plot the forecast revisions to GDP14 for the sample period February 1995 to February 2002. The data do not appear to be white noise: the data appear to exhibit serial correlation, such that positive revisions tend to follow positives, and negative revisions tend to be followed by more negatives.

In previous internal work at the Reserve Bank of New Zealand simple AR(1) models were applied to forecast revisions of the aggregate GDP14 time series, and to some individual countries, to assess the degree of serial correlation in the forecast revisions. This work indicates that the forecast revisions are indeed serially correlated. For the United States, for example, the AR(1) coefficient is about 0.36, and is found to be highly significant.

In the current research, the pre-eminent goal is to see whether much – if any – improvement can be made on the *Consensus* forecasts that form the basis of the Reserve Bank's exogenous world forecast track. For both theoretical and econometric reasons, we apply a more formal time series methodology to the forecast revisions. We specifically develop parsimonious empirical models of *Consensus* forecast revisions, to see if these are predictable from <u>previous</u> forecast revisions. This is a less ambitious goal than trying to reconcile the relationship between forecast revisions and the totality of the information set that might affect the forecasts.

In our approach we develop models for the GDP forecast revisions (henceforth simply 'revisions') for the <u>individual countries</u> included in GDP14. Zarnowitz and Tobias (2000) suggest that disaggregation can lead to substantial improvements in forecasting an aggregate (loosely 'world') measure of GDP. Similarly, the Reserve Bank of New Zealand takes a disaggregated (or components) approach to forecasting New Zealand GDP.

An important theme in such an investigation is to assess whether the predicted revisions that result from the application of time series models are in fact <u>material</u>. It is often easy to devote a lot of resources to a topic, and then find that the improvements, while they may even be highly significant in a statistical sense, are not very important for practical decision-making.

There are a number of associated questions with the main empirical objective:

1.) Do forecast revisions from different countries exhibit different properties?
2.) How does a forecast of GDP14 that is built up from disaggregated models fare in comparison to a time series model estimated directly on the sample data (a 'GDP14 model')?
3.) Are there any systematic differences in the models developed for the forecast revisions of the various different countries?
4.) How do time series forecasts of revisions perform in comparison to 'naïve' estimates of revisions?

[5] Average annual growth forecasts for GDP14 were revised by 0.4 percentage points in March 2002 on the basis of revisions to *Consensus* forecasts.

5.) Are there differences between the models of contemporaneous- and following-year forecasts?
6.) Within the Reserve Bank it has also been suggested that forecasters may pay relatively less attention to the following-year forecast, or conversely that they focus more closely on the forecasts of data for the current year. Is there evidence that this conjecture is true? And what implications would this have for the usefulness of the forecasts?

As a precursor to our main results and some preliminary answers to the above, this paper finds that revisions to *Consensus* country growth forecasts exhibit distinctive time series properties that can be characterised by *(stable?)* parsimonious ARMA representations. Although the revisions of some country groupings share some broad features, the individual models are distinguished on the whole more by their differences than their similarities. Compared to an ARMA forecast based on the aggregate GDP14 revisions, a combination of the individual forecasts perform analogously at the one step ahead and marginally better as the forecast horizon lengthens. However, the adjustment each model implies for GDP14 out to a six-steps ahead in response to a range of impulses (stylised revisions) can differ considerably over some draws of revision vectors. Although we cannot draw any strong conclusions about the significance of these differences, this exercise is particularly useful for outlining the different propagation mechanisms underlying each model. Results from testing the conjecture that forecasters pay less attention to the following-year forecasts do not seem to favour this view.

The rest of the paper is structured as follows. Section 2 describes the data, section 3 describes the methodology used to select the time series models, and section 4 describes the results. Conclusions and directions for future results are discussed in the final section.

2 DESCRIPTION OF THE DATA

We focus on the *Consensus* GDP forecasts for the fourteen individual countries. Technically it is possible to disaggregate further, to look at the forecasts from individual forecasters, but we have not chosen to do so in this paper. The individual forecasts are a rich source of information, and at the end of this paper we suggest ways in which the individual forecast data might be used.

Consensus Forecasts Inc survey forecasters monthly, asking forecasters to provide forecasts of annual average GDP growth for the current and following calendar years. Because the forecasters are forecasting annual average growth there is a great deal of inertia in the underlying variable. Technically, developments in the previous 12 to 23 months can have an impact on the current year's annual average growth.

In order to develop policy projections the Reserve Bank is interested in obtaining an exogenous forecast of the future path of GDP14. The monthly *Consensus* surveys provide an abbreviated summary of the outlook for GDP14, since the forecasts cover only the next two calendar years. As mentioned in footnote 4, the forecasters provide a rather more detailed perspective on the likely future path of GDP14 every three months.

In practice, the forecast horizon – the window of time being forecast – is not fixed at two years. Rather, it is two years in January and shrinks to 13 months in December. For example, in January 2002 the forecasters provide forecasts for calendar 2002 and calendar 2003. However, in December 2002 forecasters are still forecasting growth for the same two calendar years.

Figure 3 provides a stylised illustration of the forecasts and forecast revisions that form the basis of the analysis. Columns (3) and (4) illustrate the monthly annual-average GDP forecasts. Columns (5) and (6) illustrate the forecast revisions – the changes in forecasts. As is evident from the table, the following-year forecast is provided for the first time in January, so it is not possible to assess the change in forecast (since it was not forecast previously).

Another thing to note is that the GDP for a given year – 1991 for example – is only forecast 24 times; 12 times in 1990 and 12 times in 1991.

The sample of data used in this paper varies depending on the country being investigated. The forecast revision data for the G7 countries, with the exception of Japan, is from February 1990 to February 2002. The forecast revision data for Japan extends from January 1994 through to February 2002. The data for the remaining Asia-Pacific countries start in January 1995 and run through to February 2002. Further summary statistics of the data are presented in table 1.

3 RATIONALE FOR MORE COMPLEX MODELS

A typical assumption in theoretical macroeconomics is that expectations are rational, i.e. unbiased and efficient. This assumption has spawned a large literature testing whether forecasts exhibit these properties. The assumption of rationality is useful for analytical purposes, but looks rather more doubtful empirically (see Keane and Runkle (1990) and Bonham and Cohen (1995) for contrasting views).

Let $F^i_{t+h|t}$ denote the *Consensus* forecast for country i, made at time t, of GDP in time t+h (denoted GDP_{t+h}). If expectations (and similarly forecasts) are rational then revisions or changes in the forecasts should be uncorrelated with previously available information. Formally, $F^i_{t+h|t} - F^i_{t+h|t-1}$ should be uncorrelated with all previous information available at time t, I_t. As mentioned in the introduction, we focus on a subset of the previously available information. We examine whether, for j≥0, $F^i_{t+h|t-j} - F^i_{t+h|t-j-1}$ has explanatory power for $F^i_{t+h|t} - F^i_{t+h|t-1}$. We will use the following notation to refer to the time t forecast revision: $F^i_{t+h|t} - F^i_{t+h|t-1} = Y^i_{t+h|t}$.

There are a variety of reasons why empirical forecasts may not have the desired properties implied by rational expectations. Batchelor and Dua (1992) identify two psychological motives that may cause forecasts to appear biased or inefficient. First, forecasters may suffer from peer pressure, and may be unduly influenced by the (recently revised) *Consensus* forecast. Secondly, they may be 'conservative' in their forecast revisions. In other words, they may not fully revise their forecasts on receipt of new information. Spencer and Huston (1993), also appealing to the psychology literature, highlight the importance of ‘ambiguity’ – when individuals are more uncertain about some ‘opinion’ they are more likely to be swayed by others’ opinions.

If these conjectures are true, then changes to the *Consensus* forecast would also influence future individual forecasts, which in turn would affect the future *Consensus* forecast. Spencer and Huston (1993) and Gallo et al. (1999) find that there is evidence that the *Consensus* view of the world affects individuals’ forecasts, whereas Batchelor and Dua (1992) find that forecasters appear to be ‘variety-seeking’, i.e. they appear to give too little weight to the known forecasts of others.

Laster et al. (1999) try to formally model the incentives and constraints faced by forecasters. They argue that a forecaster accrues benefits from having <u>accurate</u> forecasts, but also from having <u>divergent</u> forecasts, since extreme views often garner publicity. This latter benefit is consistent with the cliché that there is no such thing as bad publicity.

It is also worth highlighting the fact that our analysis is conducted on *Consensus* (i.e. averaged) forecasts. It is possible that, due to the resource costs of forecasting, some forecasters may not update their individual forecasts every month – they may have a quarterly forecasting cycle for example. If this is the case then the *Consensus* forecast revisions would exhibit a moving average term, since new information would be incorporated into individual forecasts across several months.

Tests of rationality are best directed at testing the properties of forecast revisions from individual forecasters, rather than the aggregation of forecasts considered here (Bonham and Cohen, 2001). Additionally, in evaluating forecasts from individual forecasters, one would ideally assess these revisions on the basis of the information set available to the particular forecaster at the point in time when the forecast was being made, rather than just a subset of the information set.

The current paper should not really be thought of as yet another effort to test the rationality forecasts. Rather, the goal is a pragmatic, empirical one: can we improve on the exogenous forecasts that are used in the Reserve Bank’s model of the economy?

We apply autoregressive moving average models, ARMA(p,q) models, to obtain parsimonious parameterisations of the forecast revision data. These models and the model selection criteria that are used in this paper are described in the next section. We justify the moving average terms on the basis of periodic forecast updating described above (where the updating frequency for forecasts may be lower than the frequency at which the forecasts are reported). ARMA specifications can also arise from the summation of autoregressive processes.

Stationary ARMA models have an AR(∞) representation. This poses something of a problem in that there is not an infinite history of forecast revisions for a given year. (There are, in fact, only 24 monthly forecasts of annual GDP growth for any particular year.) In our regressions we have simply stacked the forecasts from different years. Forecasts from previous years are clearly in the decision-maker’s information set, but they are not necessarily forecasts of the same object. Thus, our use of ARMA models (and the implied AR(∞) models) are testing whether prior information is informative about future revisions, but the prior information is not necessarily a forecast revision of the same underlying variable (e.g. the prior information available to explain a revision to the contemporaneous year forecast in January 2002, that is the forecast revision for growth in calendar year 2002, are the previous monthly revisions to contemporaneous year growth forecasts for 2001, 2000, 1999 and so on).

4 MODEL SELECTION METHODOLOGY

For the purpose of forecasting a time series solely in terms of its past values AR(P) models are in general the most practical. However, the possibility that some forecasters update their forecasts at a lower frequency than others do suggests that forecast revisions may be more parsimoniously represented by models that include moving average error terms. For this reason we have elected to estimate ARMA(p,q) forecast models that take the general form given by equation [1] where revisions at time t, Y_t, are expressed in terms of past revisions (Y_{t-1}, Y_{t-2},..., etc) and current and past errors (ε_t , ε_{t-1},..., etc).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ...+ \theta_q \varepsilon_{t-q} \qquad [1]$$

ϕ and θ are unknown parameters and the ε are independent and identically distributed errors with zero mean.

To begin our approach, the ARMA specification above requires that our dependent variable Y is stationary. As one would expect, formal unit root tests reported in Appendix 1 suggest that, in all cases, monthly revisions to *Consensus* country forecasts and the aggregate GDP14 series do not have unit roots.

In the next stage we identify the appropriate lag structure of the ARMA(p,q) models. To guide our choice on the order of p and q, we follow a three step numerical procedure suggested by Hannan and Rissanen (1982). In the first step we estimate by ordinary least squares some pure AR processes. By comparing the Akaike information criterion (AIC) we select the regression with the smallest value.[6] In step two, the residuals from the first regression are taken as estimates of the unknown innovations in the ARMA model and a number of ARMA models are fitted. To determine the optimal lag structure we select the model with the smallest value of the Schwartz information criterion (SC) in step three. The estimated models are then checked for adequacy by looking at the properties of the respective estimated residuals.

Once the appropriate specification of the ARMA forecast models has been identified we proceed to construct four competing forecasts of future revisions to GDP14 growth in the contemporaneous and leading year. The first forecast is given by a weighted combination of ARMA forecasts of future revisions for the individual countries included in GDP14. We choose to model the forecast revisions for each of

[6] As cited in Granger and Newbold (1986), Shibata(1976) showed that the AIC criterion is not consistent but rather overestimates, asymptotically the true order of the model. However, the objective at this stage is not to estimate the order of the true model but rather to approximate a mixed model by an AR of sufficiently high order. It is best to use a liberal criterion for choosing this order and use the approximating regression to obtain estimates of the innovations.

the 14 countries because the time series properties of the various forecast revisions may be distinctive between countries. If one considers the range of countries included in GDP14, along with the possible reasons for serial correlation discussed above, it seems at least plausible that the time series properties of forecast revisions could manifest significant differences. The second forecast is an ARMA forecast based on revisions to the aggregate GDP14 series. This approach implicitly assumes that there are no material differences between the time series properties of forecast revisions among the countries included in GDP14. The third and fourth are naïve forecasts: namely, a no change assumption (i.e. assumes that revisions are white noise) and a forecast based on the mean historical revision. The latter forecast suggests that forecasts are systematically optimistic (if the mean revision is negative) or pessimistic (if the mean revision is positive) but systematic bias of this sort seems unlikely and we do not wish to emphasise it.

5 EMPIRICAL RESULTS

Parameter estimates

Table 2 presents the parameter estimates for the ARMA models estimated on the forecast revisions for each individual country and also on revisions to the aggregate GDP14 forecasts. The estimated models were arrived at by the methodology described in the previous section.

Clearly from table 2 no single ARMA specification of the forecast revisions has emerged across the different countries or across the two samples. In fact, the selected models are distinguished by their diversity. Despite this variation, a couple of broad characterisations can be drawn. It is striking that moving average terms are present in all the contemporaneous year models for revisions to the Asia Pacific countries. This is consistent with there being some delay or inertia in the updating of forecasts. One reason for this may be that a lot of conditioning data used in formulating forecasts is only available on a quarterly cycle. In contrast, the models for the G7 nations over the same period are all represented by pure autoregressive specifications. Not surprisingly, given the near 50/50 weighting of these two country groups in GDP14, both AR and MA features are significant in the representation of the contemporaneous year aggregate GDP14 data.

Looking more closely at the parameter estimates, we note some interesting features. With the exception of the Australian model for the following-year revisions, the parameters estimated for the various countries are highly significant. The fact that none of the other parameter estimates on the contemporaneous- or following-year data for Australia are significant, bar the MA(3) term estimated at 0.39 on the contemporaneous-year data, suggests that Australian forecast residuals may be white noise – consistent with forecasts being rational. This is not true for the autoregressive models estimated for Japan and the United States. For these two countries, the previous month's revision is a significant explanatory variable of current revisions in the contemporaneous- and following-year models. (In the contemporaneous year model for Japan, the third lagged revision is also significant.)

The degree of correlation indicated by the coefficients on the lagged revisions also differs between the two countries. In the United States the magnitude is 0.46 in the contemporaneous year model, falling to 0.4 in the following year. On the basis of these coefficients a revision in time t, of say 0.2 percentage points, would imply an adjustment to the US GDP forecast by the expected future revision in time t+1 of just under 0.1 percentage points. In the case of Japan, the coefficient estimates on the first and third lagged revisions are 0.23 and 0.25 respectively in the contemporaneous AR(3) model and 0.27 in the following year AR(1) model. A 0.2 percentage point revision to the contemporaneous year forecast in time t would imply a cumulative adjustment to the Japanese GDP forecast of around 0.1 percentage points by time t+3.

The contemporaneous year GDP14 model contains both an AR(1) and a MA(3) term, estimated at 0.65 and 0.32 respectively. This forecast places a lot of weight on the previous month's revision, which is not too surprising given that ten out of the fourteen individual countries' models include an AR(1) term. A 0.2 percentage point revision to the contemporaneous year forecast in time t would imply an adjustment to the GDP14 forecast in time t+1 of around 0.1 percentage points.

Put side by side, the United States is the only country to have the same model across the contemporaneous-year and following-year samples (for the other 13 countries different models are selected for the contemporaneous- and following-year data). For the following-year data a pattern is more noticeable across countries – a simple AR(1) specification is selected for eight out of fourteen countries.

The explanatory power of the individual country models varies considerably. The models of Australian forecast revisions explain less than 10 per cent of the variation in forecast revisions, and the Japanese and United States models are similarly uninformative. However, the models of contemporaneous-year revisions in China, Malaysia, South Korea, and the United Kingdom explain 40 per cent or more of the variation. Interestingly, the explanatory power for the following-year revisions is much lower than for the contemporaneous-year data. The contemporaneous year GDP14 ARMA explains around 45 per cent of the variation in forecast revisions.

While the Hannan and Rissanen (1982) model selection procedure looks to have estimated the majority of country-specific models in a plausible and parsimonious fashion, we do have some concerns about the more elaborate models selected for China, Malaysia and South Korea. The magnitudes and erratic signs of the estimated terms may portend that some of these contemporaneous year models are prone to spurious over fitting, which may have implications for forecasting.

XXX TO ADD XXS Stability tests for parameter estimates TO ADD XXXXXXX

Forecast comparison and model evaluation

[Include a chart of the competing forecasts]

We evaluate the combination model by comparing its out-of-sample forecasting with the GDP14 ARMA model, and a couple of naïve alternatives. The models are estimated on the in-sample period, and then 10 one-step-ahead forecasts are performed over the out-of-sample data. (The parameters are not re-estimated after each additional out-of-sample forecast revision.) Results for the contemporaneous year forecasts and following year forecasts are given in tables 3 and 4 respectively.

Consider first the comparison of forecasts from the ARMA model estimated on GDP14 data and the forecasts from the 'combination' model (built up from the ARMA models for each individual country). We find that the one-step ahead out-of-sample forecasts for each model imply virtually indistinguishable root mean squared errors (RMSE). The RMSE for contemporaneous-year and following-year forecasts were 0.17 and 0.29 respectively for both models (to 2 decimal places). However, as the forecast horizon increases the combination model performs slightly better than the GDP14 ARMA.

The first naïve model is simply that the revisions are white noise (so that the mean revision is expected to be zero). The second naïve forecast is that forecast revisions are independent, identically distributed (iid) process with a mean equal to the mean historical revision. As shown in table 3 and 4 both the combination and GDP14 models offer some improvements over the two naïve, but these improvements are not large.

Does disaggregation materially affect forecasts?

An important question that arises is whether it is important to build up a forecast of the future revisions by disaggregating – estimating models for individual countries, developing forecasts for the revisions from the individual countries, and then aggregating the forecasts. Or is it equally effective, and more parsimonious, to concentrate on the revisions in the aggregated data – the GDP14 time series.

Clearly there is a loss of information involved in collapsing a vector of revisions into a single (weighted-aggregate) revision. Are, however, the individual country models that have been estimated sufficiently different, such that aggregation has a material impact on GDP14 forecast revisions? If we get a large revision for a GDP forecast for Australia, does this provide the same information about future revisions to forecasts of GDP14 as a large revision to, say, a forecast of China's GDP?

To address this question and examine the effects of disaggregation we have introduced a range of different 'revision vectors'[7] or impulses and applied these to both the combination model and the GDP14 ARMA model. This exercise allows us to consider what these impulses would imply for GDP14 forecast revisions up to six periods out from the time of the impulse, and also to get a better understanding of the propagation mechanism that underlies each forecast model.

In total we consider six impulse vectors. Each vector covers forecast revisions from different geographical areas, and impulses of different magnitudes are systematically applied to provide a sense of scale. Tables 5 to 10 compare the results of applying each impulse vector to both the combination and GDP14 ARMA models. Table 5 presents the results of applying a revision vector that introduces the same impulse (forecast revision) for all of the 14 countries simultaneously. The other tables report the implied adjustments to GDP14 from impulse vectors that introduced a forecast revision to a single country or a subset of countries, whilst holding revisions to the forecasts of all other countries at zero. For example, in table 6 the impulse vector has Australian GDP forecasts being subjected to a revision, but all other countries' GDP forecasts are assumed to remain unchanged. Tables 7 to 10 summarise the results of

7 Each element of the $[14 \times 1]$ impulse vector corresponds to the revision from the jth country.

impulse vectors introduced for the United States, Japan, Europe, and Non-Japan Asia. In the Europe vector, forecasts for the UK, Germany, Italy, and France are all subjected to the same impulse, whereas the forecasts for the remaining 10 countries are left unchanged.

The magnitude of the first impulse applied to all countries' GDP forecasts at time t is listed in the first column, row one in table 5. The next six rows in the first column of each section report the successive impact of the impulse on future revisions of GDP14 implied by GDP14 ARMA and combination forecast models. The last row in column one reports the *cumulative* impact implied by each model six periods after the impulse was applied.

The first result we note from table 5 is that the successive adjustments implied by the GDP14 ARMA are almost double that implied by the combination ARMA. In addition, the combination ARMA's forecasts of future revisions dissipate more quickly over the forecast horizon. Concentrating on the cumulative impact, we can see that the total impact implied by the combination ARMA for future revisions for GDP14 is approximately the same as the initial impulse. In contrast, the cumulative effect implied by the GDP14 ARMA forecast model is almost double the initial impulse.

Results from applying the all-forecasts-revised impulse vector illustrate that combination model and the GDP14 ARMA can have quite different implications for predicted forecast revisions. In this example, disaggregation appears to have had a material impact on the implied future revisions. Intriguingly, tables 6 to 10 show that impulses to Europe and non-Japan Asia yield similar adjustments irrespective of whether one is using the GDP14 ARMA or the combination model. In contrast, impulses to Australia, the United States or Japan imply quite different adjustments depending on whether one uses the combination model or GDP14 ARMA model. Generally, the forecast adjustment implied by a forecast revision in Australia is about one quarter the size when using the combination model, as compared with the GDP14 model. Comparable results are obtained for Japan and the United States, though the factors are one third and one half respectively.

Taken together, results from the impulse analysis tend to suggest that the divergence between the adjustments implied by the two models arise primarily from differences in the time series properties of some individual country revisions compared with the aggregate GDP14 revisions. Individual ARMA forecasts for the United States, Japan and Australia in particular attach far less weight to past revisions in forecasts of future revisions than the average model, as represented by the GDP14 ARMA.[8] It is possible of course that the contribution from the individual country forecasts of future revisions to GDP14 future revisions may be under estimated in the combination forecast. This would be the case to the extent that the GDP14 ARMA forecast model captures some interrelationships between different countries growth outlook, such as the relationship between US growth and growth in other economies.

[8] The country specific forecast models for Australia and the United States have the largest weighting in the combination ARMA.

But how can the cumulative adjustments implied by the two models be so divergent when the forecasts are relatively comparable in the out-of-sample forecasting tests? When considering this question it should first be noted that in the out-of-sample forecasting tests we are comparing the forecasting performance of models estimated from contemporaneous year data against <u>empirical</u> revisions. From this we saw that the combination and GDP14 models estimated from the contemporaneous- and following-year data perform relatively comparably. In the impulse analysis it is not clear where the stylised revisions lie in relation to the empirical revision vectors.[9] It is very likely that we are getting the models to forecast future revisions in response to a range of impulses that the models are not used to. Vectors that, in practice, may not occur very often.

Although we cannot say anything about the likelihood of the impulse vectors we have applied or draw any conclusions about which of the two models is preferred, we believe that the impulse analysis is constructive for at least two reasons. First, it serves as a convenient illustration of how a revision at time t evolves. That is, it clearly traces the propagation mechanisms that underlie each model. Table 6 starkly shows how the different time series features in the Australian and the GDP14 models arrive at quite different adjustment profiles for GDP14 when an impulse is applied to Australian forecast revisions. A similar distinction is apparent in table 7 where the impulse is applied to the forecast revisions of the United States. The second motive for including the impulse analysis is because it shows that over some draws of revision vectors the implied adjustment can vary considerably depending on whether the forecast is derived from the combination model or the GDP14 ARMA model.

The cumulation of forecast revisions in response to an impulse seen in the last row of table 5 is also somewhat troubling. In practice, of course, a forecast $F_{t+h|t-j}$ of GDP_{t+h} only goes back as far as a maximum of j=24. Thus, there can be at most 23 revisions. The impulses that we have described above have been applied to the models of the contemporaneous-year forecasts, to which there are only 11 revisions. Thus, it would never seem to make sense to cumulate beyond 11 periods because there are only a finite number of possible future revisions to the forecast object in question, contemporaneous year growth (data revisions – as opposed to forecast revisions – might add another few months). For example, suppose a contemporaneous-year forecast is revised in January. Then such an impulse could lead to revisions in February, March, and so on through to December.

The ARMA models that have been estimated are stationary **[check that the roots of the AR polynomials do indeed lie outside the unit circle]**, thus implying that the impact of any shock to forecast revisions dies out over time. However, the <u>cumulative</u> impact of a forecast revision shock may be positive, and may be greater in magnitude than the original shock – indeed much greater. This is particularly troubling, since it suggests that we should pay much greater attention to forecast revisions than intuitively seems reasonable.

Suppose one considered cumulating over the first three periods, and suppose for example that the estimated model is an AR(1). This implies the cumulated revisions

[9] Unfortunately, due to the data requirements for undertaking such an exercise, we cannot compute where the stylised impulses we have applied would fall in the distribution of empirical revision vectors.

from an impulse of K sum to $\sum_{j=1}^{3} K\varphi^{j}$ over three periods. The cumulative impact summing over three periods will be larger than the <u>initial</u> K impulse when ϕ>0.55.

Do forecasters concentrate on contemporaneous-year forecasts?

As mentioned in the introduction, it has been suggested that forecasters pay comparatively less attention to the following-year forecast and rather more attention to the near-term prospects of the economy. Unfortunately, it is not self-evident what testable implications might be associated with this conjecture. It is well-known that forecast accuracy declines with the horizon being forecast, so obviously forecast accuracy cannot be considered an indicator that forecasts of the following year are 'focussed less closely upon' (rather it is simply an indication that we know less about things further out into the future). The descriptive statistics indicate that following-year forecasts <u>are</u> revised over time, indicating that such forecasts are not totally ignored. And the mean of the absolute revisions is about the same for contemporaneous- and leading-year forecasts. The ranges are a little different, but it is not clear that you would expect the ranges to be the same, given that these are for forecasts of different horizons.

Obviously, contemporaneous- and following-year revisions are not exactly the same (since different models have been estimated for the two sub-samples of data). Still, it is not clear that this reflects a lack of attention on the part of forecasters.

One way to address this concern is to see whether forecasters tend to revise their following-year forecasts more in the latter six months of the year – since the following year is six or more months closer relatively more attention might be paid to it later in the year. If the arrival process for information is not seasonal (for example if it is a Poisson process), then one might expect (absolute) following-year revisions in the first part of the year to exhibit much the same mean as (absolute) following-year revisions in the latter half of the year.

To investigate this possibility we have split the data into two further sub-samples: February-June and July-December revisions. Table 11 presents descriptive statistics for the <u>absolute</u> revisions for these two sub-samples. We are interested in testing the hypothesis that the means of the sub-samples are equal. However, we are reluctant to assume that the variance is equal across the two sub-samples, since the horizons are different. Mood et al. (1974) outline a likelihood ratio statistic λ that has an asymptotic χ^2 distribution with one degree of freedom. This statistic is the ratio of the supremum of the likelihood in the restricted parameter space divided by the supremum of the likelihood in the unrestricted space. (The former will always be smaller than the latter.) Thus, the hypothesis will be rejected if the restricted likelihood is far from the unrestricted (i.e. if λ is small).[10] Unfortunately, the estimated supremum of the likelihood for the restricted parameter space (i.e. $\mu_1 = \mu_2 = \mu$) cannot be found analytically when the variances are allowed to vary across the two

[10] The $\chi^2(1)$ statistic that is used is actually –2ln(LR), where LR is the likelihood ratio. Thus, because of the negative sign, the null hypothesis is rejected for large values of this statistic.

sub-samples. Consequently, we have used numerical methods to estimate the supremum of the restricted likelihood.

The likelihood ratio statistics for each of the various countries are presented in Table 12. The test results are mixed. For the United States, the United Kingdom, Italy, Canada and Taiwan the null hypothesis of equal means can be rejected; for the remaining 8 countries the hypothesis is accepted. In other words, for the 6 countries identified above the mean absolute change in the following-year forecast revisions appears to be different in the second half of the year when compared to the mean absolute revisions in the first half of the year. However, there is some possibility that this might be a sample-specific result, or might (perhaps) reflect seasonality in the arrival process for information. When one compares the mean forecast revisions for the contemporaneous- and following-year forecasts from the second-half of the year we find that there is no mean difference in the revision. A comparison of the first-half contemporaneous- and following-year forecasts shows that in eight of the fourteen countries, one cannot reject the null hypothesis that the means are the same. Overall, therefore, our results do not provide strong support the view that the following-year forecasts are ignored or down-weighted.

6 CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Conclusions

This paper finds that revisions to *Consensus* country GDP growth forecasts exhibit distinctive time series properties that can be characterised by parsimonious ARMA representations. As we expected, the estimated ARMA models vary substantially across the different countries and also across the contemporaneous and following year samples. The lack of a standard model across the country-specific ARMA specifications suggests that there are good grounds for using a combination of individual forecasts to predict future revisions to GDP14. Chiefly, this approach makes use of extra information about the time series properties of country-specific revisions that would be lost through aggregation. Compared to an ARMA forecast based on the aggregate GDP14 revisions, a combination of the individual forecasts makes minor gains in forecast accuracy over the GDP14 ARMA forecast as the forecast horizon extends beyond the first-step ahead.

Although there does not appear to be much difference between the models in terms of their out of sample forecasting performance, impulses (stylised revisions) applied to the contemporaneous year ARMA forecasts show some substantial differences between the combination and GDP14 implied forecasts of future GDP14 revisions. Although we cannot draw any strong conclusions about the significance of these differences, this exercise is useful for outlining the different propagation mechanisms underlying each model.

Lastly, our results do not provide much support for the view that the following-year forecasts are ignored or down-weighted, though it has to be noted that this conjecture is difficult to test.

Suggestions for future research

This research project has highlighted a number of interesting areas for future research. Most of these projects involve using the disaggregated forecasts from individual forecasters.

Zarnowitz and Braun (1993) provide considerable evidence that – in terms of forecast accuracy – using *Consensus* (i.e. mean) forecasts is desirable relative to most individual forecasts. However, there are some caveats to that result. Zarnowitz and Braun show that around 25 per cent of the individual forecasters consistently perform better than the *Consensus* mean. A potentially important direction for future work may therefore be to try to establish which (if any) individual forecasters consistently out-perform the *Consensus* forecast. One future direction is thus to consider the optimal way of combining the 'individual-person' forecasts for a given country. Along these lines, one could further investigate whether there are systematic differences in the quality of forecasts from different 'types' of forecasters. For example, do domestic Japanese forecasters perform better or worse than foreign forecasters with offices in Japan? Similarly, are financial forecasters and big businesses better than 'independent forecasters', as is suggested by Laster et al. (1999)? Interestingly, this empirical result runs counter to a National Association of Business Economists' survey (reported in Zarnowitz and Braun, 1993), which indicated that there were no systematic differences in forecast accuracy.

Spencer and Huston (1993) raise another research direction. It would be interesting to disentangle the effect of 'own-revisions' from *Consensus* revisions on the individual forecasters. Alternatively, once could envisage a situation where there is a 'super-star' or 'guru' forecaster, so that other 'follower' forecasters take into account the forecast revisions of the guru. Do forecasters appear to take particular note of a guru forecaster?

The analysis presented in this paper essentially assumed that forecast revisions are independent across different countries. And yet over the last business cycle it has become increasingly obvious that the world is highly integrated, such that a sneeze in America can cause the rest of the world to exhibit similar symptoms. A further avenue of investigation is thus to consider the covariances of forecast revisions across different countries.

REFERENCES

Bonham, Carl S and Richard H Cohen (2001), "To aggregate, pool or neither: Testing the rational-expectations hypothesis using survey data", *Journal of Business and Economic Statistics*, 19(3): 278-291.

Clemen Robert T and Robert L Winkler (1986), "Combining economic forecasts", *Journal of Business and Economic Statistics*, 4(1): 39-46.

Gallo, Giampiero M, Clive W J Granger, and Yongil Jeon (1999), "The impact of the use of forecasts in information sets', University of California, San Diego Discussion Paper 99-18.

Gregory, Allan W, Gregor W Smith, and James Yetman (2001), "Testing for forecast consensus", *Journal of Business and Economic Statistics*, 19(1): 34-43.

Hannan E J and J Rissanen (1982), "Recursive Estimation of Mixed Autoregressive-Moving Average Order," *Biometrika*, 69: 81-9.

Laster, David, Paul Bennett and In Sun Geoum (1999), "Rational bias in macroeconomic forecasts", *Quarterly Journal of Economics*, 114(1): 293-318.

Spencer, Roger W and John H Huston (1993), "Rational forecasts: On confirming ambiguity as the mother of conformity", *Journal of Economic Psychology,* 14(4): 697-709.

Zarnowitz, Victor and Phillip Braun (1993), "Twenty-two years of the NBER-ASA quarterly economic outlook surveys: Aspects and comparisons of forecasting performance", In James H Stock and Mark W Watson (eds.) Business Cycles, Indicators, and Forecasting, NBER Studies in Business Cycles, Volume 28, The University of Chicago Press: Chicago.

Zarnowitz, Arnold and Justin Tobias (2000), "A note on aggregation, disaggregation and forecasting performance", *Journal of Forecasting*, 19:457-469.

FIGURES AND TABLES

Figure 1
Forecast revisions to GDP14 annual average growth in the contemporaneous year, January 1995 – February 2002

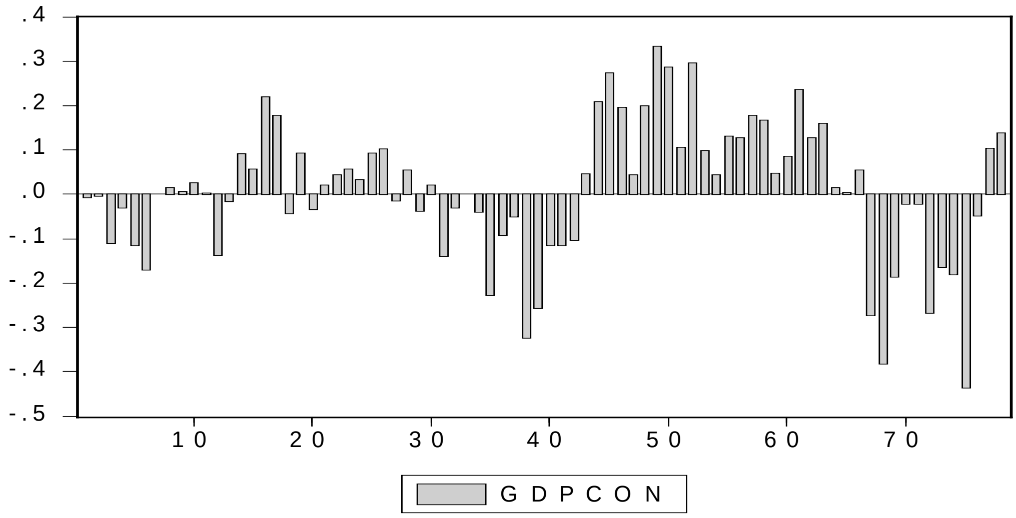

Figure 2
Forecast revisions to GDP14 annual average growth in the following year, February 1995 to February 2002

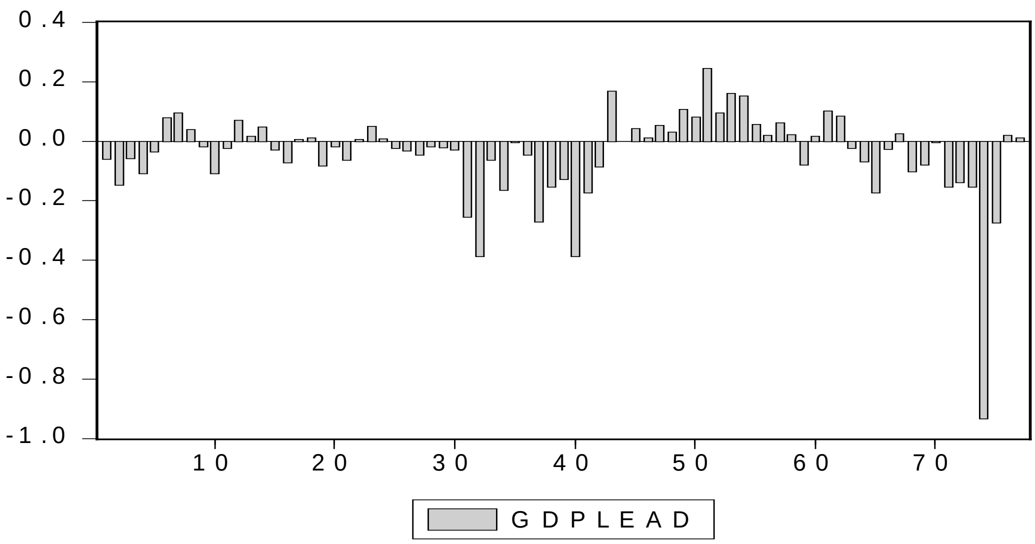

Figure 3
Stylised data

Stylised Consensus Forecast Revisions							
		Forecasts		Forecast Revisions			
Date forecast published		Contempo-raneous-Year	Following-Year	Contempo-raneous-Year	Following-Year	Years being Forecast	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1990	Jan	X13	Y1	X13 - X12	N/A		
	Feb	X14	Y2	X14 - X13	Y2 - Y1		
	Mar	X15	Y3	X15 - X14	Y3 - Y2		
	Apr	X16	Y4	X16 - X15	Y4 - Y3		
	May	X17	Y5	X17 - X16	Y5 - Y4		
	Jun	X18	Y6	X18 - X17	Y6 - Y7		
	Jul	X19	Y7	X19 - X18	Y7 - Y6	1990	1991
	Aug	X20	Y8	X20 - X19	Y8 - Y7		
	Sep	X21	Y9	X21 - X20	Y9 - Y8		
	Oct	X22	Y10	X22 - X21	Y10 - Y9		
	Nov	X23	Y11	X23 - X22	Y11 - Y10		
	Dec	X24	Y12	X24 - X23	Y12 - Y11		
1991	Jan	Y13	Z1	Y13 - Y12	N/A		
	Feb	Y14	Z2	Y14 - Y13	Z2 - Z1		
	Mar	Y15	Z3	Y15 - Y14	Z3 - Z2		
	Apr	Y16	Z4	Y16 - Y15	Z4 - Z3		
	May	Y17	Z5	Y17 - Y16	Z5 - Z4		
	Jun	Y18	Z6	Y18 - Y17	Z6 - Z5		
	Jul	Y19	Z7	Y19 - Y18	Z7 - Z6	1991	1992
	Aug	Y20	Z8	Y20 - Y19	Z8 - Z7		
	Sep	Y21	Z9	Y21 - Y20	Z9 - Z8		
	Oct	Y22	Z10	Y22 - Y21	Z10 - Z9		
	Nov	Y23	Z11	Y23 - Y22	Z11 - Z10		
	Dec	Y24	Z12	Y24 - Y23	Z12 - Z11		

* Note: The following-year forecast (e.g. of 1992 in 1991) is provided for the first time in January, hence there is no forecast revision in that month.

Table 1
Descriptive statistics for forecast revisions to annual average growth forecasts (sample: January 1995 to February 2002)

	Contemporaneous year	Following year
Mean	0.01	-0.04
Mean of absolutes	0.12	0.09
Maximum	0.33	0.24
Minimum	-0.44	-0.93
Standard deviation	0.15	0.15
Number of observations	78	77

Table 2
Parameter estimates for contemporaneous and following year forecast revisions ARMA forecast models

	Australia		China		Hong Kong		Malaysia		Singapore		South Korea		Taiwan		Canada		France		Germany		Italy		Japan		UK		US		GDP14	
Dep var	Aucon	Aulead	Chcon	Chlead	Hkcon	Hklead	Mycon	Mylead	Spcon	Splead	Kocon	Kolead	Twcon	Twlead	Cacon	Calead	Frcon	Frlead	Gecon	Gelead	Itcon	Itlead	Jpcon	Jplead	Ukcon	Uklead	Uscon	Uslead	gdpcon	gdplead
No obs	68	67	68	67	68	67	68	67	68	67	68	67	68	67	123	122	123	122	123	122	123	122	79	78	123	122	133	122	68	67
Expl. vars																														
Dep var(-1)		0.0778					0.8880		0.4020		1.0807	0.2633		0.2213	0.4409	0.9441	0.5043	0.4246	0.4597	0.6411	0.4148	1.1802	0.2296	0.2730	0.5673	0.4766	0.4566	0.3959	0.6517	0.5733
		0.6175					6.4678		3.4197		12.1716	2.2005		1.8277	5.8017	25.9764	5.5998	5.1555	6.0672	9.2076	4.5944	21.1407	2.0571	2.4774	7.2157	5.9322	5.6285	4.7205	6.0768	5.6549
Dep var(-2)			0.4970					0.4574		0.3961							0.1848				0.2251									
			6.5328					4.0033		3.4029							2.0488				2.4700									
Dep var(-3)			-0.2694	0.4958	0.4625		0.6102				-0.2631		-0.4702						0.2019			-0.1974	0.2535		0.1544					
			-3.1176	4.5138	3.8797		4.9387				-3.0995		-4.3973						2.7180			-3.6661	2.2402		1.9630					
Dep var(-4)							-0.6481																							
							-5.8585																							
ma(1)			0.2964		0.2383	0.3449	-0.5478				-0.9345					-0.7331						-0.9885								
			1.9972		1.9230	5.7028	-2.7673				-19.5380					-8.7699						-56.4190								
ma(2)			-0.5990			0.7885																								
			-3.9674			15.9119																								
ma(3)	0.3870		0.6732					0.3093	0.3499	0.3345	0.5900		0.9457			-0.2470													0.3163	
	2.8845		4.3518					2.4965	2.6730	2.7650	17.9502		30.0858			-3.4359													2.4757	
R-squared adj.	0.0863	-0.0098	0.4322	0.1961	0.2577	0.2136	0.6082	0.3154	0.3034	0.2392	0.5184	0.0512	0.2204	0.0130	0.1873	0.1642	0.3881	0.1287	0.3383	0.3646	0.2166	0.1555	0.1336	0.0367	0.4192	0.2023	0.1661	0.1516	0.4669	0.3029
AIC	-0.0540	-1.1662	-1.5459	-0.7690	1.7477	1.4926	0.5908	0.6608	0.9972	0.8194	0.8888	1.5616	-1.3009	-0.4346	-0.6853	-1.2959	-1.6173	-1.3585	-1.5365	-1.6532	-1.8903	-1.8707	0.2569	-0.1723	-1.4850	-1.3697	-0.7953	-1.0026	-1.6261	-1.8446
Schwarz criterion	-0.0214	-1.1330	-1.3786	-0.7352	1.8146	1.5584	0.7257	0.7277	1.0630	0.8863	1.0226	1.5948	-1.2340	-0.4014	-0.6631	-1.2266	-1.5711	-1.3354	-1.4901	-1.6301	-1.8441	-1.8006	0.3183	-0.1407	-1.4386	-1.3465	-0.7723	-0.9795	-1.5603	-1.8114

Table 3
Forecast performance of forecast models of contemporaneous year forecast revisions

Root mean squared errors

Forecast model	*1-step ahead*	*2-step ahead*	*3-step ahead*	*4-step ahead*
Combination ARMA	0.1732	0.1596	0.1715	0.1809
GDP14 ARMA	0.1700	0.1717	0.1928	0.2014
No change	0.1943	0.1848	0.1946	0.2036
Sample mean	0.2122	0.1781	0.1976	0.2158

***** Table 3 and table 4 compare the out-of-sample forecast performance of the four forecast models in predicting the contemporaneous year and following year forecast revisions as measured by the root mean squared error (RMSE).

Table 4
Forecast performance of forecast models of following year forecast revisions

Root mean squared errors

Forecast model	*1-step ahead*	*2-step ahead*	*3-step ahead*	*4-step ahead*
Combination ARMA	0.2885	0.2213	0.2668	0.3173
GDP14 ARMA	0.2970	0.2375	0.2810	0.3110
No change	0.3424	0.2641	0.3047	0.3206
Sample mean	0.3085	0.2208	0.2684	0.3117

Table 5
Stylised impulses applied to all forecasts

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.03	0.06	0.11	0.17	0.22	0.28
t+2	2-step ahead	0.02	0.03	0.06	0.09	0.12	0.15
t+3	3-step ahead	0.03	0.05	0.10	0.15	0.20	0.25
t+4	4-step ahead	0.01	0.03	0.06	0.08	0.11	0.14
t+5	5-step ahead	0.01	0.02	0.03	0.05	0.06	0.08
t+6	6-step ahead	0.00	0.01	0.02	0.03	0.03	0.04
	Cumulative effect	0.09	0.19	0.37	0.56	0.75	0.94

Combination ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.02	0.03	0.06	0.09	0.12	0.15
t+2	2-step ahead	0.01	0.02	0.03	0.05	0.06	0.08
t+3	3-step ahead	0.02	0.03	0.07	0.10	0.14	0.17
t+4	4-step ahead	0.01	0.01	0.02	0.03	0.04	0.06
t+5	5-step ahead	0.00	0.01	0.02	0.02	0.03	0.04
t+6	6-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
	Cumulative effect	0.05	0.11	0.21	0.32	0.42	0.53

Table 6
Stylised impulses applied to Australian forecasts only

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.01	0.01	0.03	0.04	0.06	0.07
t+2	2-step ahead	0.00	0.01	0.02	0.02	0.03	0.04
t+3	3-step ahead	0.01	0.01	0.03	0.04	0.05	0.07
t+4	4-step ahead	0.00	0.01	0.01	0.02	0.03	0.04
t+5	5-step ahead	0.00	0.00	0.01	0.01	0.02	0.02
t+6	6-step ahead	0.00	0.00	0.00	0.01	0.01	0.01
	Cumulative effect	0.02	0.05	0.10	0.15	0.19	0.24

Combination ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.00	0.00	0.00	0.00	0.00	0.00
t+2	2-step ahead	0.00	0.00	0.00	0.00	0.00	0.00
t+3	3-step ahead	0.00	0.01	0.02	0.03	0.04	0.05
t+4	4-step ahead	0.00	0.00	0.00	0.00	0.00	0.00
t+5	5-step ahead	0.00	0.00	0.00	0.00	0.00	0.00
t+6	6-step ahead	0.00	0.00	0.00	0.00	0.00	0.00
	Cumulative effect	0.00	0.01	0.02	0.03	0.04	0.05

Table 7
Stylised impulses applied to United States forecasts only

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.01	0.01	0.02	0.03	0.04	0.05
t+2	2-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
t+3	3-step ahead	0.00	0.01	0.02	0.03	0.04	0.05
t+4	4-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
t+5	5-step ahead	0.00	0.00	0.01	0.01	0.01	0.01
t+6	6-step ahead	0.00	0.00	0.00	0.00	0.01	0.01
	Cumulative effect	0.02	0.04	0.07	0.11	0.14	0.18

Combination ARMA

t+1	1-step ahead	0.00	0.01	0.02	0.03	0.03	0.042
t+2	2-step ahead	0.00	0.00	0.01	0.01	0.01	0.018
t+3	3-step ahead	0.00	0.00	0.00	0.00	0.01	0.008
t+4	4-step ahead	0.00	0.00	0.00	0.00	0.00	0.004
t+5	5-step ahead	0.00	0.00	0.00	0.00	0.00	0.002
t+6	6-step ahead	0.00	0.00	0.00	0.00	0.00	0.001
	Cumulative effect	0.01	0.01	0.03	0.04	0.06	0.074

Table 8
Stylised impulses applied to Japanese forecasts only

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead		0.01		0.03		0.05
t+2	2-step ahead		0.01		0.02		0.03
t+3	3-step ahead		0.01		0.03		0.04
t+4	4-step ahead		0.00		0.01		0.02
t+5	5-step ahead		0.00		0.01		0.01
t+6	6-step ahead		0.00		0.00		0.01
	Cumulative effect	0.00	0.03	0.00	0.08	0.00	0.14

Combination ARMA

t+1	1-step ahead		0.00		0.01		0.02
t+2	2-step ahead		0.00		0.00		0.00
t+3	3-step ahead		0.00		0.01		0.02
t+4	4-step ahead		0.00		0.01		0.01
t+5	5-step ahead		0.00		0.00		0.00
t+6	6-step ahead		0.00		0.00		0.01
	Cumulative effect	0.00	0.01	0.00	0.03	0.00	0.05

Table 9
Stylised impulses applied to European forecasts only

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead		0.01		0.02		0.04
t+2	2-step ahead		0.00		0.01		0.02
t+3	3-step ahead		0.01		0.02		0.04
t+4	4-step ahead		0.00		0.01		0.02
t+5	5-step ahead		0.00		0.01		0.01
t+6	6-step ahead		0.00		0.00		0.01
	Cumulative effect	0.00	0.02	0.00	0.07	0.00	0.12

Combination ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead		0.01		0.02		0.04
t+2	2-step ahead		0.00		0.01		0.02
t+3	3-step ahead		0.01		0.02		0.03
t+4	4-step ahead		0.00		0.01		0.02
t+5	5-step ahead		0.00		0.01		0.01
t+6	6-step ahead		0.00		0.00		0.01
	Cumulative effect	0.00	0.02	0.00	0.06	0.00	0.10

Table 10
Stylised impulses applied to non-Japan Asia forecasts only

GDP14 ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.01	0.01	0.02	0.04	0.05	0.06
t+2	2-step ahead	0.00	0.01	0.01	0.02	0.03	0.03
t+3	3-step ahead	0.01	0.01	0.02	0.03	0.04	0.05
t+4	4-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
t+5	5-step ahead	0.00	0.00	0.01	0.01	0.01	0.02
t+6	6-step ahead	0.00	0.00	0.00	0.01	0.01	0.01
	Cumulative effect	0.02	0.04	0.07	0.11	0.14	0.18

Combination ARMA

t	Impulse magnitude	0.05	0.1	0.2	0.3	0.4	0.5
t+1	1-step ahead	0.01	0.01	0.02	0.03	0.04	0.05
t+2	2-step ahead	0.00	0.01	0.01	0.02	0.03	0.03
t+3	3-step ahead	0.01	0.01	0.03	0.04	0.06	0.07
t+4	4-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
t+5	5-step ahead	0.00	0.01	0.01	0.02	0.02	0.03
t+6	6-step ahead	0.00	0.00	0.00	0.01	0.01	0.01
	Cumulative effect	0.02	0.04	0.07	0.11	0.14	0.18

Table 11
Descriptive statistics of Sub-samples split by Year-Half

		United States	Germany	France	United Kingdom	Italy	Canada	Japan	Australia	China	Hong Kong	Malaysia	Singapore	Taiwan	Korea
Following-Year 1st-half Forecasts	Mean	0.055	0.082	0.068	0.045	0.048	0.077	0.144	0.113	0.123	0.211	0.192	0.213	0.113	0.181
	Std Dev.	0.06	0.08	0.08	0.03	0.04	0.07	0.14	0.11	0.17	0.46	0.28	0.33	0.10	0.16
	No. of Obs.	60	60	60	60	60	60	46	35	35	35	35	35	35	35
Following-Year 2nd-half Forecasts	Mean	0.149	0.116	0.112	0.123	0.109	0.130	0.187	0.100	0.121	0.316	0.291	0.364	0.212	0.341
	Std Dev.	0.22	0.13	0.13	0.13	0.13	0.16	0.20	0.12	0.13	0.53	0.33	0.36	0.30	0.59
	No. of Obs.	72	72	72	72	72	72	48	42	42	42	42	42	42	42
Contemp. 1st-half Forecasts	Mean	0.182	0.118	0.096	0.089	0.072	0.135	0.153	0.176	0.094	0.332	0.326	0.363	0.147	0.358
	Std Dev.	0.15	0.14	0.11	0.09	0.07	0.16	0.15	0.22	0.10	0.64	0.40	0.36	0.22	0.42
	No. of Obs.	61	61	61	61	61	61	42	37	37	37	37	37	37	37
Contemp. 2nd-half Forecasts	Mean	0.100	0.084	0.084	0.087	0.085	0.113	0.218	0.128	0.083	0.351	0.343	0.438	0.207	0.311
	Std Dev.	0.10	0.08	0.10	0.09	0.08	0.13	0.28	0.17	0.09	0.45	0.39	0.48	0.43	0.34
	No. of Obs.	69	69	69	69	69	69	47	42	42	42	42	42	42	42

Table 12
Chi-squared test that absolute revisions have equal means in H1 and H2

	Following-Year 1st-half vs Following-Year 2nd-half			Following-year 2nd-half vs Contemp.-year 2nd-half		Following-year 1st-half vs Contemp.-year 1st-half	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
United States	11.22	**	United States	2.95		11.22	**
Germany	3.25		Germany	2.96		3.25	
France	5.93	*	France	2.15		5.93	*
United Kingdom	20.82	**	United Kingdom	3.53		20.82	**
Italy	12.47	**	Italy	1.72		12.47	**
Canada	6.38	*	Canada	0.52		6.38	**
Japan	1.48		Japan	0.41		1.48	
Australia	0.27		Australia	0.82		0.27	
China	0.01		China	2.46		0.01	
Hong Kong	0.89		Hong Kong	0.11		0.89	
Malaysia	2.06		Malaysia	0.45		2.06	
Singapore	3.66		Singapore	0.65		3.66	
Taiwan	3.97	*	Taiwan	0.00		3.97	*
Korea	2.81		Korea	0.09		2.81	

Critical value for Chi-Squared(1) statistic: 2.71, 3.84, 6.63 for 10, 5, and 1% significance levels respectively.
* Significant at 5%; ** Significant at 1%.

Appendix 1
Unit root tests of forecast revisions

Contemporaneous year forecast revisions	**Augmented Dickey-Fuller test (3 dependent var lags)**	**Phillips-Perron Z test**
	no trend	**no trend**
Australia	-3.424307 ***	-7.640483 ***
China	-2.126746 **	-7.249914 ***
Hong Kong	-2.830579 *	-5.944263 ***
Malaysia	-2.334036 **	-4.426365 ***
Singapore	-2.569941 **	-4.702771 ***
South Korea	-2.380648 **	-4.669291 **
Taiwan	-3.176331 **	-4.798730 ***
Japan	-2.933725 ***	-6.561121***
United States	-5.315665 ***	-5.339722 ***
United Kingdom	-2.899754 ***	-3.824298 ***
Germany	-2.930756 ***	-5.679633 ***
France	-2.447017 **	-4.104105 ***
Italy	-2.253799 **	-4.558730 ***
Canada	-3.623983 ***	-5.097959 ***

*** Null of a unit root is rejected at the 1 percent level
** Null of a unit root is rejected at the 5 percent level
* Null of a unit root is rejected at the 10 percent level

Following year forecast revisions	**Augmented Dickey-Fuller test (3 dependent var lags)**	**Phillips-Perron Z test**
	no trend	**no trend**
Australia	-3.515866 ***	-7.750993 ***
China	-2.024874 **	-7.727188 ***
Hong Kong	-3.378021 ***	-8.052099 ***
Malaysia	-2.377447 **	-5.736780 ***
Singapore	-2.255673 **	-5.795319 ***
South Korea	-4.009499 ***	-6.769044 ***
Taiwan	-2.143898 **	-5.354350 ***
Japan	-3.476900 ***	-6.656212 ***
United States	-3.479813 ***	-6.045584 ***
United Kingdom	-2.819250 ***	-5.624719 ***
Germany	-2.264830 **	-3.715437 ***
France	-3.012638 ***	-5.462222 ***
Italy	-2.033155 **	-5.631792 ***
Canada	-3.287624 ***	-6.180684 ***

*** Null of a unit root is rejected at the 1 percent level
** Null of a unit root is rejected at the 5 percent level
* Null of a unit root is rejected at the 10 percent level