

How far ahead can we forecast? Evidence from cross-country surveys

Gultekin Isiklar and Kajal Lahiri
Department of Economics
University at Albany – SUNY
Albany, NY 12222. USA
klahiri@albany.edu

Abstract:

Using monthly forecast data from Consensus Economics Inc. for 18 developed countries reported over 24 different forecast horizons during 1989-2004, we found wide diversity in the quality of forecasts across countries, and the horizons at which forecasts start becoming useful. For only seven countries (viz., Austria, Belgium, Canada, Denmark, Germany, Norway and U.S.) the initial 24-month-ahead forecasts beat the naïve no-change forecast. In terms of the worst performance, Irish, Portuguese, Swiss and Dutch forecasts beat the naïve forecast as late as at horizons 10-13 months. We also found that that generally the survey forecasts do not have much value when the horizon goes beyond 18 months. The flow of new information to annual GDP growth forecasts follows a bell-shaped curve over horizons with a peak point when the forecast horizon is around 13 months.

Corresponding Address: Professor K.Lahiri, Department of Economics, University at Albany – SUNY, Albany, NY 12222. USA. Email: klahiri@albany.edu; phone; (518 442 4758; Fax: (518) 442 4736.

1. Introduction

How far ahead into the future forecasts have value and how the information content of forecasts changes over forecast horizon have been the focus of various studies.¹ Most of these studies, however, provide measures for the information content of *optimum* forecasts over forecast horizons by modeling the actual data generating process. For example, Öller (1985) and Galbraith (2003) provide estimates of the length of the forecast horizon at which the optimum forecasts start to contain information by assuming that the actual process follows an ARIMA process. Similarly, Oke and Öller (1999) provide estimates by modeling the actual process using VARMA process.

Granger (1996) pointed out that a feature that will provide limits to how far ahead one can forecast is when the (forecastable) signal gets lost in the (unforecastable) noise. In other words, forecasts will not provide any information when the measurement errors start to make the information content of signals negligible compared to noise. In reality, the measurement errors are not only driven by the level of noise attributed to the data generating process but also to other factors. For example, delays in data releases and data revisions, not to mention structural breaks that are only detectable *ex post*, are some of the factors that may affect the information content of real-time forecasts. In these situations, a forecaster will seem like responding to information that are relevant but also to those that are not. These factors do not cause problems in *ex post* analysis of historical data but may induce significant deformation in the information content of real-time forecasts.

¹ See for example, Parzen (1982), Öller (1985), De Gooijer and Klein (1992), Diebold and Kilian (1997), Oke and Öller (1999), and Galbraith (2003).

Only a hand full of studies have used real-time survey data to estimate the information content in forecasts (e.g. Mills and Pepper, 1999; Vuchelen and Gutierrez, 2005), and no one has examined the dynamics of how the information content of forecasts change over horizons and how new information increases the information value of forecasts. However, understanding the changes in the information content of forecasts over horizons and, for example, the timing of the arrival of most important information is critical for both the forecasters and the clients. It is well known that many popular forecasting agencies like the OECD, Blue Chip, etc. produce forecasts several times a year from an initial 24-month-ahead forecast. Information on the dynamics of information accumulation over forecast horizons can provide forecasters an important parameter in their selection of the timing and the frequency of forecasting service. For the clients of forecasting firms, the information content of forecasts can be an important consideration in their decisions on how to use and when to use these forecasts.

In this study we address these issues using 15 years of monthly private sector forecast data for 18 developed countries reported over 24 different forecast horizons. We study various characteristics of real GDP growth forecasts over forecast horizons and their differences across countries, and propose two measures for the content of new information in forecasts. We find that the flow of new information to annual GDP growth forecasts follows a bell-shaped curve over horizons with a peak point when the forecast horizon is around 12-13 months.

The remaining of the paper is structured as follows. Section 2 presents data. Section 3 discusses certain stylized facts on the evolution of forecasts in a cross-country

setting, and reports estimates on the flow of new information at various horizons using two alternative approaches. Section 4 concludes.

2. Data

In this study three data sets are used. The main data of the study on real GDP forecasts come from the *Consensus Economics Inc.* The second and third data sets are for the actual data series but with different vintages. Our historical data for real GDP growth rates (to calculate the 10-year GDP growth averages and to model the actual GDP growth but not to evaluate forecasts) is constructed from the IMF's *International Financial Statistics* (February 2002 edition). Our real time data set for forecast evaluation is mainly constructed from the OECD's mid-year *Economic Outlook*, 1990 to 2004. However, as it will be explained shortly, there are several non-matching variables in the OECD reports in the early 1990's so we need to use several additional data sources such as DEA's *Survey of Business* or Bundesbank's *Monthly Economic Reports* to match the exact data for these years. The details of the data sets follow.

Since October 1989, the *Consensus Economics Inc.* has been polling more than 600 forecasters each month and recording their forecasts for principal macroeconomic variables (including GDP growth, inflation, interest rates and exchange rates) for a large number of countries. Forecasts are made for the current year (based on partial information about developments in that year) and for the following year. The number of panelists ranges from 10 to 30 for most of the countries, and for the major industrialized countries the panelists are based in countries they forecast.

We study the consensus forecasts of annual average real GDP growth. Survey respondents make their first forecasts when there are 24 months to the end of the year;

that is, they start forecasting GDP growth in January of the previous year, and their last forecast is reported in the beginning of December of the target year. So for each country and for each target year we have 24 forecasts of varying horizons. Our data set ranges from October 1989 to June 2004. The countries we study are the 18 industrialized countries for which forecasts are available from *Consensus Economics Inc.*²

There have been several major changes in the definition of forecast variable since the survey started in 1989. For example, while real GNP was being forecast in the first few years for some countries, the real GDP became the forecast variable since early 1990s. For example, this switch occurred in January 1992 for the US and in January 1993 for Germany. In our data sample, the most significant changes were for Germany. While West Germany's real GNP growth was being forecast through December 1992, after January 1993 the forecast variable became real GDP of West Germany. In addition, unified Germany's GDP growth was added to the survey and West Germany's GDP forecast was removed in May 1997.

In order to evaluate the forecast errors correctly, the forecasts should be matched with the actual data being forecast. It is well documented in the literature that data revisions may have an important impact on the perceived performance of the forecasters. Since forecasters cannot possibly be aware of data revisions after they report their forecasts, we use an early revision as the actual value, which is compiled from the mid-year reports of OECD *Economic Outlook* immediately following the target year. But because of the changes in variable definitions (i.e., GNP to GDP or West Germany to

² There are only a very few of studies that have used the *Consensus Forecasts* data set. These are Artis and Zhang (1997), Batchelor (2001), Harvey, Leybourne and Newbold (2001), Loungani (2001), Juhn and Loungani (2002), Gallo, Granger and Jeon (2002), and Isiklar, Lahiri and Loungani (2006). However, none of these studies consider the empirical findings analyzed in this paper.

Unified Germany) some of the data are not available in the June issue of OECD *Economic Outlook*. We collected these data from the original sources such May or June issues of BEA's *Survey of Business* or Bundesbank's *Monthly Economic Reports* in the year immediately following the target year.

3. Evolution of Fixed-Target Forecasts over Horizons

Figure 1 presents the reported forecasts and the realized actual values between 1991 and 2002. Each country's forecasts are divided into three sub graphs, which are located horizontally in the figures. Plots start when the forecast horizon is 24, which is reported in January of the previous year, and end when the forecast horizon is 0, which gives the actual realization. Gallo *et al.* (2002) presented this type of graphs for 1993-96 and three major countries: U.S., U.K., and Japan. We can now examine certain stylized characteristics of the forecast evolution in greater depth.

First, note that for the first six months or so (i.e., for horizons 24 to 18 months), the consensus forecasts do not seem to change very much. This empirical observation leads us to believe that over these horizons, forecasters do not receive dependable information to revise their forecasts systematically. There are important exceptions, however. For the target year 1994, forecasts for Belgium, France, Ireland, and Spain were active from the beginning.

Second, except for Ireland and Japan, the initial forecasts for all other countries seem to be starting from a relatively narrow band and then tend to diverge from these initial starting points. For example, for Austria, Belgium, Denmark and several other countries 24-month ahead forecasts are located between 2 percent and 3 percent, and as information is accumulated these forecasts tend to move towards their final destination.

One may conjecture that these initial long-term forecasts are nothing but unconditional means of the processes. While this conjecture seems to hold for most of the forecasts, the initial forecasts of Irish and Japanese GDP growth rates seem to behave differently. For Ireland, the forecasts tend to move upward and for Japan the forecasts tend to move downward as we go from the far left panel (forecasts for 1991 to 1994) to the far right panel (forecasts for 1999 to 2002). This movement of the 24-month-ahead forecasts implies that these long-term expectations have been changing for these two countries, and that recent short run forecast errors have affected the longer-run expectations. See Frenkel (1975) who hypothesized such a feedback.

Third, Gallo *et al.* (2002) noted that the consensus forecasts sometimes do not converge to the right target due to copycat behavior of non-dominant forecasters. In our expanded data set, even though we see evidence for such a behavior for certain years for some countries, evidence is not pervasive. For Ireland the one-month ahead forecasts underestimated the targets repeatedly, but this can be explained by the exceptional Irish growth during the nineties. As documented by them, and we also find in Figure 1, U.S. growth for 1995 was seriously overestimated even a month before the end of the target. This again can be explained by the fact that in the U.S., 1995 was a sudden growth slowdown year.

Fourth, Gallo *et al.* (2002) noted that there are generally large shifts in the 12-month-ahead forecasts, possibly due to the release of new figures for the previous year. A look at these graphs reveals certain undeniable regularities on how the fixed-target forecasts evolve over time. We thus proceed to examine more rigorously the timing of

the arrival of important information when forecasters break away for their initial estimates. This issue is discussed in the following sections.

4. Forecast variance and forecast horizons

Forecasts presented in Figure 1 clearly show that the initial 24-month ahead forecasts are reasonably stable over years, and as forecast horizon decreases, they tend to diverge. In other words, forecast variability increases as the forecast horizon decreases. While variability of forecast errors is usually associated with uncertainty, forecast variability is inversely related with uncertainty. This argument may look counter intuitive but can be easily rationalized by noting that

$$y_t = f_{t,h} + \varepsilon_{t,h}$$

where $f_{t,h}$ denote the h -period ahead forecast and $\varepsilon_{t,h}$ denote the error associated with this forecast. Since rational expectations imply that $Cov(f_{t,h}, \varepsilon_{t,h}) = 0$ we have $Var(y_t) = Var(f_{t,h}) + Var(\varepsilon_{t,h})$, which implies that the variations in forecasts and forecast errors move in opposite directions as forecast horizon changes (note $Var(y_t)$ does not depend on the forecast horizon).³ Therefore as forecast horizon increases the forecast error variability and therefore the uncertainty increases, but the forecast variability decreases.

This observation is confirmed more clearly in Figure 2, where we present the variances of the forecasts over our sample period at each forecast horizon. The last points in the charts, the points when the horizon is zero, give the variances of the actual values. As it is clear from these figures, as forecast horizon decreases the variance of the

³ This point is noted by several studies. See, for example, Muth (1985).

forecasts steadily increases. Another way of looking at this increasing variability of forecasts is that as forecast horizon increases more information is accumulated and as more information is accumulated in the forecasts the variation of forecasts increases. This information accumulation process can be mimicked using a simple MA model for the data generating process. Suppose that the actual process has a moving average representation of order q so that

$$y_t = \mu + \sum_{k=0}^q \theta_k \varepsilon_{t-k} \quad (1)$$

Then an optimum forecast at horizon h will be

$$f_{t,h} \equiv E(y_t | I_{t,h}) = \mu + \sum_{k=h}^q \theta_k \varepsilon_{t,k} \quad (2)$$

and the variance of the forecast is

$$\text{Var}(E(y_t | I_{t,h})) = \sigma^2 \sum_{k=h}^q \theta_k^2 \quad (3)$$

Similarly the variance of the forecast when forecast horizon is $h-1$ is

$$\text{Var}(E(y_t | I_{t,h-1})) = \sigma^2 \sum_{k=h-1}^q \theta_k^2$$

so that

$$\text{Var}(f_{t,h-1}) = \text{Var}(f_{t,h}) + \theta_{h-1}^2 \sigma^2$$

So when the forecast horizon is very long, i.e. several years, the forecasts tend to converge towards the mean of the process, and as information is accumulated, the forecasts move up and down increasing the forecast variability. It is interesting to note that Mankiw and Shapiro (1986) used the same argument to conclude that U.S. GDP revisions are “news” rather than “noise”. If successive revisions incorporate useful

information about past GDP growth, then we will expect the successive revised figures to have more variance than the initial announcement.

While the positive slope in the forecast variance graphs is clear in all figures, there are some differences across countries that are worth mentioning. First, for some countries, e.g., Japan and the USA, the positive slope is not very clear in the longer-run forecasts, especially when horizon is more than 18 months. As just shown, the forecast variability increases because of the variability of the accumulated shocks, i.e. $\theta_k \varepsilon_{t-k}$ s. Therefore, if forecast variability does not change much over several horizons as it is the case, for instance, in Japanese forecasts for horizons from 24 to 15, this may mean that the information acquired 15 months ago does not have much impact on the actual value, i.e. $|\theta_k|$ is small. Of course, it may also be related to the informational inefficiency of the forecasts too. It is possible that while the information over this period is important, the forecasters do not incorporate the information in their forecasts causing less than optimal variability in the forecasts. This will be addressed later when we present forecast evaluation measures that are based on forecast errors over forecast horizons.

It is interesting to note that for some countries, the variation of actual values is much larger than the variation of one-month ahead forecasts. This can be especially observed for Finland, Ireland, Norway, Portugal, and Switzerland. There can be several reasons for this. One-period-ahead forecasts can be written as:

$$y_t = f_{t,1} + u_{t,1}$$

where $u_{t,1}$ is the forecast error of the one-step ahead forecast. Suppose that forecasts are efficiently constructed so that $E(y_{t,h} | I_{t,h}) = f_{t,h}$ which implies that $Cov(f_{t,1}, u_{t,1}) = 0$.

Then if the variation of the actual values is very large compared to the forecasts then this

implies that the variance of $u_{t,1}$ is very large, which means that there is significant information revealed on the last month of the year.⁴ But if the forecasts are not efficient and if $Cov(f_{t,1}, u_{t,1}) \neq 0$ then we have

$$Var(y_t) = Var(f_{t,1}) + Var(u_{t,1}) + 2Cov(f_{t,1}, u_{t,1})$$

which implies that the large difference between the variation of actual and the forecasts can be due to both the variation in noise and the inefficiency $Cov(f_{t,1}, u_{t,1})$. So the large difference between the variances may occur even if the actual process is not very noisy.

Finally, it is worth noting the implications of rational expectations and implicit expectations in the graphs. As it is well known, Muth's (1961) rational expectations hypothesis requires that forecasts should be uncorrelated with the forecast errors, which also implies that the variance of the actual process should be larger than the variance of the forecasts, i.e. $Var(y_t) > Var(f_{t,h})$. On the other hand, implicit expectations, which was pioneered by Mills (1957), requires that the actual realizations should be uncorrelated with the forecast errors, and the variance of actual realizations should be lower than the variance of forecasts, i.e. $Var(y_t) < Var(f_{t,h})$.⁵ The evidence in Figure 2 clearly supports the implications of rational expectations since variances of actual realizations are larger than those of forecasts in majority of the cases. However, short-run

⁴ Let us note that large data revisions between December and the June may be also responsible for large variation in $u_{t,1}$.

⁵ Using this difference between the rational and implicit expectations, Muth (1985) claims that firm production forecasts are not consistent with rational expectations hypothesis and proposes a hybrid model of expectation formation in which rational and implicit expectations are special cases. Also see Lovell (1986) for a comparison of rational expectation and implicit expectations hypotheses. Lahiri and Lee (1979) justify additional variance in forecasts in terms of possible errors in measurement in the survey data in a rational expectations model.

forecasts of some countries, namely France, Germany, Denmark, Japan, and the UK seem to mildly violate this relation.

5. Information content of forecasts

Information value of a forecast is related with how accurate the forecasts are. In this section, we will provide statistics such as MSE, MAE and Theil's U statistic along with another statistic recently proposed by Diebold and Kilian (1997). While MSE and MAE depend on the variability of the actual process, Theil's U statistic scales the RMSE by the variability of underlying data and has the advantage of being independent of the variance of the actual process. Formally,

$$U_h(y_n) = \sqrt{\frac{\sum_{t=1}^T (y_t - f_{t,h})^2}{\sum_{t=1}^T (y_t - y_n)^2}} \quad (4)$$

which compares the forecast errors with a naive forecast y_n . If U_h is more than one, the forecasts does not beat the naive forecast. An important issue in calculating the U_h is the selection of the naive forecast. In this study, we will follow the literature and use the forecast of no change as the naive forecast, i.e., $y_n = y_{t-1}$.⁶

To see the improvement in the performance of forecasts over decreasing horizons, we also provide a R^2 measure, which is based on the idea of predictability measure proposed by Diebold and Kilian (1997). They propose a measure of predictability as

⁶ Note that when the forecast horizon is more than 12 months, the forecasters do not know y_{t-i} . So, when forecast horizon is more than 12 months, judging the forecasters for not being able to beat the benchmark forecast may not be fair due to additional information advantage of the benchmark forecast. Thus, we also use several other benchmark forecasts such as second lagged actual y_{t-2} and 10-year real GDP growth. In most of the cases the forecasters could beat both of these naive forecasts at every horizon.

$p_{s,k} = 1 - \frac{E(L(e_s))}{E(L(e_k))}$, where $E(L(e_k))$ denotes the expected loss in the long-run forecasts

and $E(L(e_s))$ denotes the expected loss in the short-run forecasts. If mean squared errors are used as the loss function, then we have⁷

$$p_{s,k} = 1 - \frac{MSE_s}{MSE_k}$$

Diebold and Kilian (1997) used this measure to compute the predictability of several macro variables using the realized data and noted that it would be interesting to use this measure on the forecast survey data. When k-period ahead survey forecasts are used as the naïve forecast, then $p_{s,k}$ will give the improvement in the forecasts as horizon decreases. To the best of our knowledge, no study has ever used this statistic on the survey data.

Table 1 presents MAE, MSE, and Theil's U statistics for 12-month and 24-month ahead forecasts. Later we also provide the Theil's U statistics for all horizons. For 24-month ahead forecasts, Theil's U statistic is less than one for only Canada, Denmark, Germany, and the U.S. The worst performers in 24-month ahead forecasts are Portugal, Ireland, and Netherlands, which have Theil's U statistics 1.45, 1.41 and 1.39 respectively. For 12-month ahead forecasts, all the countries, with the exception of Ireland and Portugal, have Theil's U statistics less than one implying that the forecasts have value over the no-change forecast.

⁷ This measure is also related to the *forecast content function* proposed by Galbraith (2003). In Galbraith's *forecast content function*, MSE of the unconditional mean forecast replaces MSE_k , which also defines the so-called *skill score* that has been used extensively in other disciplines (see, for instance, Murphy (1988)).

Figure 3 presents Diebold and Kilian's $p_{h,24}$ and Theil's $U_h(y_{t-1})$ for each forecast horizon and country. Notice that large values of Theil's U imply large forecast errors. On the other hand, large values of $p_{h,24}$ imply that forecasts improve significantly over the 24-month ahead forecast $f_{t,24}$. The right axes in the figures show $p_{h,24}$ whereas the left axes show the values of $U_h(y_{t-1})$. Since, 1.0 is the threshold value of $U_h(y_{t-1})$ for determining whether the forecast can beat the naïve forecast of no change, the plots in Figure 6 include a horizontal line through 1.0. In addition, to pinpoint the longest horizon at which the forecasts beat the naive forecast, the graphs also include a vertical line through the longest horizon at which the estimated $U_h(y_{t-1})$ is lower than one. This provides an easy way to compare the countries with each other. For all the countries, as forecast horizon decreases the quality of the forecasts increase as expected. The graphs also point out significant heterogeneity across countries.

When we look at the performance rankings based on Theil's $U_h(y_{t-1})$, we observe that in addition to the four country forecasts that beat the naive forecast when horizon is 24 months, (i.e., American, Danish, German, and Canadian forecasts given in Table 1), we now see that Austrian and Norwegian forecasts beat the naive forecast when horizon is 23 months. In terms of the worst performance in beating the naive forecast, we find that Irish and Portuguese forecasts beat the naïve forecast at horizons 10 and 11 months respectively. These are followed by Switzerland and the Netherlands that beat the naive forecast when the horizon is 13 months.

The Diebold-Kilian measure of predictability $p_{h,24}$ shows the improvement in the information content of the forecasts as measured by the decrease in the mean squared

errors over the MSE of the 24-month ahead forecasts. As shown in Figure 3, the predictive ability of GDP forecasts for some countries (e.g., France, Canada, Denmark, Finland, Japan, USA) does not improve over the 24-month ahead forecasts when the horizon is relatively long, but for some other countries (e.g., Germany, Ireland, Spain), each additional month increases the information content of the forecasts over the previous month even in longer-run forecasts. For most of the countries, we see that MSE substantially decreases in the short-run forecasts causing $p_{h,24}$ to be close to 100% when the forecast horizon is one month. Two exceptions are Norwegian and Irish GDP growth forecasts, where the final values of $p_{h,24}$ are less than 80%.

6. Timing of the most valuable information

The slope of the plots in Figure 3, which can be interpreted as a measure of the improvement in forecast quality over horizons, is found to be quite different from country to country. For example, Norwegian statistics do not have a steep slope, which implies that Norwegian forecasts do not improve much with the decreasing horizon, but the Japanese statistics have very steep slope implying that the forecast quality increases sharply as new information is acquired.

This last point brings us to another important query: Around what horizon does the most valuable information is received or, in other words, at what horizon the forecasts improve the most? The answers to these questions are related to the slope of $p_{h,24}$ or U_h curves and are addressed in the next sections, where we provide alternative approaches to measure the content of new information at a particular horizon.

The first measure is based on the performance of forecasts and makes use of both the actual and the forecast data. The second measure is based only on the forecast revisions data and can be seen as the content of new information in the forecasts as perceived by the forecasters. Following the previous literature cited before, another measure for the content of new information will be constructed from the “optimum forecasts” using the time series representation of the actual quarterly GDP growth. We do this for the purpose of comparison.

6.1. New information based on forecast errors

The first difference in the MSE_h will give an estimate for the new information content in forecasts when the horizon is h . From equation (2) an optimum forecast $f_{t,h}$ satisfies

$$\Delta MSE(f_{t,h}) \equiv MSE(f_{t,h+1}) - MSE(f_{t,h}) = \theta_h^2 \sigma^2 \quad (5)$$

which is equivalent to the information content of the new information in the actual process. Now suppose that $\tilde{f}_{t,h}$ is not an optimum forecast and is generated according to

$$\tilde{f}_{t,h} \equiv E(y_t | \tilde{I}_{t,h}) = \tilde{\mu} + \sum_{k=h}^q \tilde{\theta}_k \tilde{\varepsilon}_{t,k} \quad (6)$$

where q denotes the largest forecast horizon at which the first forecast is reported. It defines the conditional mean of the actual process when the horizon is q , i.e.

$\tilde{\mu} = E(y_t | \tilde{I}_{t-q})$, $\tilde{\varepsilon}_{t,h}$ denotes the 'news' component utilized by the forecaster, and $\tilde{\theta}_h$

denotes the impact of this news component as perceived by the forecaster.

For convenience, let us assume that the forecasters observe the news $\varepsilon_{t,h}$ correctly but their utilization of news is not optimum, so that $\tilde{\theta}_h \neq \theta_h$ and $\tilde{\varepsilon}_{t,h} = \varepsilon_{t,h}$. From equation (6), we see that the forecast errors follow:

$$y_t - \tilde{f}_{t,h} = (\mu - \tilde{\mu}) + \sum_{k=h}^H (\theta_k - \tilde{\theta}_k) \varepsilon_{t,k} + \sum_{k=0}^{h-1} \theta_k \varepsilon_{t,k} \quad (7)$$

where the first component on the RHS denotes the bias in the forecast, the second component denotes the errors due to inefficiency, and the third component denotes the errors due to unforecastable events after the forecast is reported. Calculating mean squared errors and assuming that sample estimates converge to their population values, we get

$$MSE_h = (\mu - \tilde{\mu})^2 + \sum_{k=h}^H (\theta_k - \tilde{\theta}_k)^2 \sigma^2 + \sum_{k=0}^{h-1} \theta_k^2 \sigma^2 \quad (8)$$

Similarly calculating MSE_{h+1} and taking the first difference we find that

$$\Delta MSE_h \equiv MSE_{h+1} - MSE_h \text{ is}$$

$$\Delta MSE_h = \theta_h^2 \sigma^2 - (\theta_h - \tilde{\theta}_h)^2 \sigma^2 \quad (9)$$

which gives the improvement in the content of the forecasts with the new information. The first element on the RHS represents the maximum improvement in the quality of forecasts if the information is used efficiently, but the second component represents the mistakes in the utilization of the new information. If the usage of the most recent information $\tilde{\theta}_h$ differs from its optimum value θ_h , then the gain from the utilization of new information decreases. In the special case when $\tilde{\theta}_h = \theta_h$, the equation (9) is

equivalent to equation (5). In this case, ΔMSE_h will be an estimate for the content of the new information in the actual process $\theta_h^2 \sigma^2$.

6.2. New information based on forecast revisions

While the use of ΔMSE_h provides the improvement in forecasting performance at horizon h and therefore it gives the information content of the news in terms of forecasting ability, a similar measure can be constructed based solely on forecast data without using the actual data on GDP growth. Notice that, based on equation (2), the optimum forecast revisions $r_{t,h} \equiv f_{t,h} - f_{t,h+1}$ is nothing but

$$r_{t,h} = \theta_h \varepsilon_{t,h} \quad (10)$$

In the not-optimum case of equation (6), we have the forecast revision process

$$r_{t,h} = \tilde{\theta}_h \varepsilon_{t,h} \quad (11)$$

Calculating the mean squared revisions and taking the probability limit we get

$$MSR_h = p \lim_T \frac{1}{T} \sum_{t=1}^T r_{t,h}^2 = \tilde{\theta}_h^2 \sigma^2$$

which gives a measure for the reaction of the forecasters to news. But since forecasters react to the news based on their perception of the importance of the news, this measure can be seen as the content of the new information as perceived by the forecasters. Note the clear difference between ΔMSE_h and MSR_h . While the first one is driven by the forecast errors, the latter has nothing to do with the actual process. But both of the measures should give the same values if the survey forecasts are optimum.

The difference between MSR_h and ΔMSE_h may provide important behavioral characteristics of the forecasters such as over or under reaction to the news at a specific

forecast horizon. MSR_h can be seen as a measure of how forecasters interpret the importance of news at a specific horizon, and ΔMSE_h can be seen as the “prize” they get as a result of revising their forecasts. Suppose that forecasters make large revisions at horizon h^* but the performance of the forecasts do not improve much at that horizon, then one may conjecture that the forecasters react excessively to the news. To see this more clearly calculate the difference between these two measures

$$MSR_h - \Delta MSE_h = 2(\tilde{\theta}_h^2 - \theta_h \tilde{\theta}_h) \sigma^2$$

which is positive when $\tilde{\theta}_h^2 > \theta_h \tilde{\theta}_h$, which is the same as the condition $|\tilde{\theta}_h| > |\theta_h|$. But $|\tilde{\theta}_h| > |\theta_h|$ is equivalent to overreaction to the news when the horizon is h .

6.3. Empirical comparisons

Before providing the graphs of the ΔMSE_h and MSR_h , let us try to determine the shape of them conceptually. As shown earlier, we expect to see forecast variability to increase as new information is accumulated. If the information content at a particular period is significantly larger than the previous period, we expect to see an increase in forecasting performance in that period.

When the forecast horizon is very short, we expect the new information content in forecasts to be small for two reasons. First, the impact of a shock is determined by the length of the time for the shock to be totally absorbed by the economy. When there is not enough time for the transmission mechanisms to create a significant impact on the output, the impact will be small. This implies that as horizon gets smaller the impact on the GDP growth of a typical shock will be correspondingly smaller. Second, since the forecast variable is annual average real GDP growth, current year forecasts are highly driven by

the quarterly real GDP announcements and data revisions during the year. So as we approach the end of the target year, a lot of information will be already known and it is expected that in the last few months the impact of the information will be very small. Because of these, we expect that the new information content will be small when the forecast horizon is short. Similar to the first reasoning above, when the horizon is long the final impact is expected to be small since most of the impact will be consumed before the target year even starts. These two observations suggest that when the horizon is too short or too long the information content of news is expected to be small so that we expect to see a bell curve for the impact of shocks with respect to the forecast horizons.

Figure 4 presents ΔMSE_h (dots) and MSR_h (squares) values along with fitted quadratic polynomial curves. The quadratic curves fitted for ΔMSE_h are shown in bold lines and the quadratic curves fitted for MSR_h are shown in dashed lines. As it is clear from the figure, both the bold line and dashed lines in charts (with an exception of Finland) present a significant quadratic polynomial as expected. It is also worth noting several important differences and similarities across countries. First of all, for most of the countries the peak of quadratic line is when forecast horizon is close to 12 months and usually when the horizon is between 10 months and 13 months. The exceptions to this argument are (when ΔMSE_h is considered) Finnish and Irish forecasts. It is interesting to note that both of these countries have observed significant unusual movements in their real GDP growth rates in 1990s.

In majority of the cases, ΔMSE_h and MSR_h contours look similar, which implies that forecast revisions are mostly consistent with the improvement in forecasting performance. Although forecast revision based estimates do not use any actual values and

they do not depend on the traditional forecast error measure (as the ΔMSE_h graphs), peak points of the two measures mostly match. But it may be also worthwhile to note that when the peak points do not coincide, MSR_h peaks a few periods later than ΔMSE_h . For most countries ΔMSE_h is larger than MSR_h at all horizons, which implies that forecast revisions are sticky, and forecasters stagger their reactions to news.⁸

The finding that for most of the countries the largest information content is revealed when the horizon is close to 12 months, and especially, between December of the previous year and March of the current year (when the horizon is 10 months to 13 months) begs another interesting question: What is the source of information that is revealed during this time? Clearly the GDP announcements can be regarded as one possible information source that may have a large impact on the forecasting performance. During January to March, however, the first quarter GDP figures are not yet released. The U.S. and U.K. GDP figures for the first quarter are released during the last week of April and in most other countries GDP figures are released in the second half of May. Therefore for most of the countries the improvement in the forecasts cannot be attributed to the release of the first quarter GDP figure. However, it is possible that relevant monthly indicators such as employment, industrial production, manufacturing index data that are known to be correlated with GDP, may be used to revise forecasts.

In addition, various leading indicators (e.g., stock market index, interest rate spread, building permits, unemployment insurance claims, etc.) which have predictive power up to a year and which usually become available promptly can be valuable information sources too.

⁸ Isiklar *et al.* (2006) find similar evidence for G7 countries using the same data source, but a different methodology. See Mankiw and Reis (2001) and Sims (2003) for alternative explanations.

In addition to these variables, it is also possible that the GDP figure for the previous year, which is released as early as late January for some countries and in February for the majority of the countries, can provide important information about the current year GDP growth. But clearly there is a difference in content between the information provided by macro variables for the first few months of the year and the GDP data release for the previous year. While macro variables such as industrial production provide information about the events in the first few months of the year being forecast, the released GDP figure for the previous year has actually no information value in this respect.

But, previous year's GDP figure may increase the information content of the forecasts significantly for several reasons. First of all, information of the previous year's GDP level determines the base of the GDP growth forecast for the following year, which may have substantial effect on the current year forecasts. In addition, if GDP growth has a large serial correlation and forecasters employ extrapolative expectations to capture this, the release of last year's GDP figure may initiate large revisions and increase the forecasting performance. It is also possible that previous year's GDP growth will have a large impact on the forecasting performance if the forecasters can learn from their previous mistakes and employ an error correction model as it is usually posited in the adaptive expectations model.

6.4. Content of new information implied in the actual process

In this section we provide a measure of new information in an “optimum” forecast, which is based on modeling the actual process. The content of the new information in the actual process can be calculated by estimating equation (1) using the

actual quarterly real GDP growth data. For example, one may think of fitting an MA model on the real GDP growth series, and then treating the estimated MA coefficients as estimates for θ_h coefficients. As pointed before, this approach is the main idea behind several studies in the calculations of the information content of optimum forecasts, e.g., Öller (1985).

But, in this study the forecasts are what are called “fixed-event” forecasts. So the target variable represented by y_t is not quarterly real GDP growths but annual real GDP growths. This implies that we have to make a transformation on the MA coefficients estimated using the quarterly real GDP growth series to be comparable with the annual real GDP growth forecasts.

Suppose that y_t denotes the annual real GDP growth as before and $\tilde{y}_{t,q}$ denotes the annualized quarterly real GDP growth q quarters before the end of the year t . For example, $\tilde{y}_{t,1}$ is the GDP growth rate in the last quarter of year t and $\tilde{y}_{t,4}$ is the GDP growth rate in the first quarter of the year t . Note that with this notation we have $\tilde{y}_{t,k} = \tilde{y}_{t-1,k+4}$. Then by definition:

$$y_t = \frac{1}{4} \sum_{k=1}^4 \tilde{y}_{t,k} \quad (12)$$

Now suppose that $\tilde{y}_{t,q}$ has the following $MA(\infty)$ representation:

$$\tilde{y}_{t,q} = \gamma_0 \varepsilon_{t,q} + \gamma_1 \varepsilon_{t,q+1} + \gamma_2 \varepsilon_{t,q+2} + \dots + \gamma_k \varepsilon_{t,q+8} + \dots \quad (13)$$

Then substituting this MA process in the equation (12) gives the MA representation for the actual process:

$$\begin{aligned}
y_t = & \frac{1}{4} [\gamma_0 \varepsilon_{t,1} \\
& + (\gamma_0 + \gamma_1) \varepsilon_{t,2} \\
& + (\gamma_0 + \gamma_1 + \gamma_2) \varepsilon_{t,3} \\
& + (\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) \varepsilon_{t,4} \\
& + (\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4) \varepsilon_{t,5} \\
& + (\gamma_2 + \gamma_3 + \gamma_4 + \gamma_5) \varepsilon_{t,6} \\
& + (\gamma_3 + \gamma_4 + \gamma_5 + \gamma_6) \varepsilon_{t,7} \\
& \vdots \\
&]
\end{aligned}$$

More specifically, the MA form can be represented as follows:

$$y_t = \sum_{k=0}^{\infty} \delta_k \varepsilon_{t,k+1} \quad (14)$$

where

$$\delta_k = \sum_{i=\max(k-3,0)}^k \gamma_i \quad (15)$$

The intuition behind this representation is clear. While last quarter shocks have only a unique chance of having an impact on the annual GDP growth, third quarter shocks will have the chance two times: contemporaneous effect on the third quarter GDP (via γ_0) and then a secondary effect on the last quarter GDP growth (via γ_1). Similarly, first quarter shocks will have four impact coefficients. When the horizon is larger than 4 quarters, the shocks will have 4 chances to have an effect on the current year GDP growth. But in this case, there will not be any contemporaneous impact since the effect will be seen on the previous year's GDP growth.

To estimate the γ_k s, we use the seasonally adjusted quarterly real GDP growth and estimate an AR(4) model. Then we transformed this AR(4) model into an $MA(\infty)$

representation which gives us γ_k s. This is a 'safe' way to get reasonable a MA representation. Direct MA modeling is an alternative way but the models may not converge under certain conditions. After getting the MA coefficients of the quarterly model, we construct the MA coefficients of the annual model using equation (15) and then calculate the optimum percentage of variation at horizon k as

$$100 \times \frac{\delta_k^2}{\sum_{i=1}^8 \delta_i^2}$$

which are presented in Figure 5.

To be comparable with the survey forecast data we use 1990-2001 period to estimate AR (4) models for each country. Note that the longest horizon we are interested is two years and we use quarterly GDP growth rates to estimate the γ_k s, we have only eight observations to plot for each country.⁹ For the sake of brevity and in order to see a stable “flow of information” curve, Figure 5 presents the percentage shares of 18 countries aggregated and separately only for UK and US. Since we use quarterly data to generate the shares, we plot each quarter's value in the center of the quarter. So, for example, the first estimated share of contemporaneous shock is plotted when horizon is 2 months. The plots show that the impact of the new information in the optimum forecast follows a bell-shape curve with a peak around 13 months as expected. With a few exceptions, a similar relationship was found for each of the sample countries.¹⁰

⁹ There were outliers in the data too: Germany on 1991:1 (8.32%), Portugal on 1988:1 (15.6%), and Norway on 1997:2 (6.9%). In addition, the GDP growth rates of Spain behave abnormally during 2000:2 to 2001:1 having growth rates of 3.8, -3.04, 5.3 and -2.2 percent respectively. With these data points, the model failed to converge so we used the data until 2000:1 for Spain. Except the Spanish case, the results were, however, not affected by the control of the outliers.

¹⁰ These countries are: Austria, Denmark, Japan, Norway, and Portugal. The anomaly for these countries can possibly be explained by the estimated AR coefficients due to small samples and their robustness.

We do not expect these graphs based on the actual data generating process of the target series to be exactly same as those based on survey forecasts in Fig. 4. As we have pointed out at the beginning, in real life, forecast errors are driven not only by the level of “well behaved” noise attributed to the data generating process but also by numerous other factors. For example, randomness in data revisions, structural breaks, model misspecifications, outliers, etc. that are only detectable *ex post*, are some of the factors that may affect the information content of real-time forecasts. These factors do not cause problems in *ex post* analysis of historical data of the target variable, but may induce significant difference in the information content of real-time survey forecasts. That is why a comparison between the three approaches is important.

7. Conclusion

In this paper we study the characteristics of fixed-target monthly GDP growth forecasts for 18 developed countries during 1989-2004. We study how forecasting performance improves as forecast horizon decreases, and at what horizons forecasts start to become informative. Since there are many forecasting organizations around the world that provide forecasts for many macroeconomic variables with horizons up to 12 quarters or more, it is interesting to explore the value of these forecasts, and thereby try to understand the limits to how far ahead today’s professional forecasters can reasonably forecast. Since the panel of forecasters in *Consensus Economics* are all private market professionals, the limits to forecasting that these forecasters exhibit can safely be taken as indicative of the current state of economic forecasting. However, the answer from our exhaustive data analysis did not turn out to be a “single-liner”. We have found wide diversity in the quality of the forecasts across countries, and the horizons at which

forecasts start becoming useful, possibly reflecting the forecast difficulty of the underlying series.

We used Theil's U statistic with the lagged GDP growth as the benchmark, and another measure of predictability recently suggested by Diebold and Kilian (2001) with the 2-year ahead forecast as the benchmark. In terms of Theil's U, for only seven of the eighteen countries (*viz.*, Austria, Belgium, Canada, Denmark, Germany, Norway and U.S.) the initial 24-month-ahead forecasts beat the naïve forecast. In terms of the worst performance, Irish, Portuguese, Swiss and Dutch forecasts beat the naïve forecast as late as at horizons 10-13 months.

In terms of the Diebold-Kilian skill measure and the variance functions, we found for majority of the countries the long-term forecasts for up to 18 months are as good as the initial 24-month-ahead forecasts. That is, over these longer horizons, forecasters do not receive dependable information to adjust their forecasts systematically. We also observed a similar pattern when we looked the horizons at which the survey forecasts beat the naïve no-change forecast. These findings imply that the survey forecasts do not have much value when the horizon goes beyond 18 months or so.

In this paper we have proposed two alternative approaches to measure the content of new information in survey forecasts. The first measure is based on the improvement in the forecasting performance over horizons, and the second measure is based on forecast revisions and can be considered as a measure of content of new information as perceived by the forecasters. Whereas the latter can be interpreted as a measure of how forecasters interpret the importance of news in real time, the former is the *ex post* "prize" they get as a result of revising their forecasts. Under rationality and without too much unforeseen

errors in the sample period, these two approaches will yield similar results. We observe that for most of the countries the most significant improvement in forecasting performance comes when the forecast horizon is between 10 months and 13 months. The estimates based on forecast revisions, though not exactly the same, are very similar.

Acknowledgements:

We thank the participants of the Leipzig conference on the “Future of Forecasting” for many helpful and encouraging comments. In particular, comments from Roy Batchelor, Anthony Davies, Fushang Liu, Lars-Erik Öller, Franz Seitz, Xuguang Shen, Herman Stekler, and Kenneth Wallis are gratefully acknowledged. However, we are responsible for any errors and omissions.

References

- Artis, M. and W. Zhang, (1997), "International business cycles and the ERM: Is there a European business cycle?" *International Journal of Finance and Economics*, 2, 1-16.
- Batchelor, R. A. (2001), "How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus," *The Review of Economics and Statistics* 33, 225-35.
- De Gooijer, J. G. and A. Klein (1992), "On the cumulated multi-step-ahead predictions of vector autoregressive moving average processes", *International Journal of Forecasting* 7, 501-513.
- Diebold, F. X. and L. Kilian (2001), "Measuring predictability: theory and macroeconomic applications", *Journal of Applied Econometrics*, 16 (6), 657-669.
- Frenkel, J. A. (1975), "Inflation and the formation of expectations", *Journal of Monetary Economics* 1, 403-433.
- Galbraith, J. W. (2003), "Content horizons for univariate time-series forecasts", *International Journal of Forecasting* 19, 43-55.
- Gallo, G.M., C. W. J. Granger, and Y. Joen, "Copycats and common swings: The impact of the use of the forecasts in information sets," *IMF staff Papers*, 49, 1, 4-21 (2002).
- Granger, C. W. J. (1996), "Can we improve the perceived quality of economic forecasts?", *Journal of Applied Econometrics* 11(5), 455-473.
- Harvey, D.I., S.J. Leybourne, and P. Newbold, (2001), "Analysis of a panel of UK macroeconomic forecasts," *Econometrics Journal* 4, 37-55.
- Isiklar, G., K. Lahiri, and P. Loungani (2006), "How quickly do forecasters incorporate news? Evidence from cross-country surveys", Forthcoming in *Journal of Applied Econometrics*.

- Juhn, G. and Loungani, P. (2002), "Further cross-country evidence on the accuracy of the private sector output forecasts", *IMF Staff Papers*, 49, 49-64.
- Lahiri, K. and J.S. Lee (1979), "Tests of Rational Expectations and the Fisher Effect," *Southern Economic Journal*, October 1979
- Loungani, P. "How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth," *International Journal of Forecasting* 17, 419-432 (2001).
- Lovell, M. C. (1986), "Tests of the rational expectations hypothesis", *American Economic Review* 76, 110-124.
- Mankiw, N.G. and M.D. Shapiro (1986), "News or noise? An analysis of GNP revisions", *Survey of Current Business*, 66, 20-25.
- Mankiw, N. G. and R. Reis (2002), "Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve", *Quarterly Journal of Economics* 117(4), 1295-1328.
- Mills, E.S. (1957), "The theory of inventory decisions", *Econometrica* 25, 222-238.
- Mills, T. C. and G. T. Pepper (1999), "Assessing the forecasters: an analysis of the forecasting records of the treasury, the London Business School and the National Institute", *International Journal of Forecasting* 15, 247-257.
- Mincer, J.A. (1969), "Models of adaptive forecasting", in J. Mincer, ed., *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, Vol. 6, NBER, New York, pp. 83-111.
- Murphy, A. H. (1988), "Skill scores based on the mean square error and their relationships to the correlation coefficient", *Monthly Weather Review*, 116, 2417-2424.
- Muth, J. F. (1961), "Rational expectations and the theory of price movements", *Econometrica* 29, 315-335.
- Muth, J. F. (1985), "Properties of some short-run business forecasts", *Eastern Economic Journal* 11(3), 200-210.

- Oke, T. and L.-E. Öller (1999), "Testing for short memory in a VARMA process", *Journal of Forecasting* 18, 477-487.
- Öller, L.-E. (1985), "How far can changes in business activity be forecasted?", *International Journal of Forecasting* 1, 135-141.
- Parzen, E. (1982), "ARARMA models for time series analysis and forecasting", *Journal of Forecasting* 1, 67-82.
- Sims, C. A. (2003), "Implications of rational inattention", *Journal of Monetary Economics* 50(3), 665-690.
- Vuchelen, J. and M.-I. Gutierrez (2005), "A direct test of the information content of the OECD growth forecasts", *International Journal of Forecasting* 21, 103-117.

Table 1 Goodness of fit of forecasts (MAE, RMSE, and Theil's U)

Country	12-month ahead forecasts			24-month ahead forecasts		
	MAE	RMSE	Theil U	MAE	RMSE	Theil U
Austria	0.98	1.16	0.81	1.24	1.48	1.01
Belgium	0.99	1.15	0.72	1.28	1.68	1.02
Canada	1.21	1.36	0.7	1.44	1.7	0.88
Denmark	0.72	0.99	0.75	0.96	1.14	0.84
Finland	2.24	2.89	0.87	2.7	3.37	1.07
France	0.79	0.99	0.73	1.15	1.5	1.08
Germany	0.79	1.03	0.6	1.49	1.96	0.99
Ireland	2.35	2.76	1.08	2.98	3.67	1.41
Italy	0.77	0.87	0.65	1.39	1.61	1.19
Japan	1.41	1.58	0.74	1.9	2.3	1.04
Netherlands	0.89	1.06	0.88	1.38	1.72	1.39
Norway	0.92	1.13	0.73	1.14	1.33	1
Portugal	0.98	1.31	1.02	1.4	1.89	1.45
Spain	0.61	0.86	0.66	1.18	1.58	1.2
Sweden	0.9	1.13	0.69	1.46	1.84	1.13
Switzerland	1.22	1.45	0.92	1.71	2.04	1.26
UK	0.77	1.02	0.7	1.08	1.62	1.12
USA	0.96	1.09	0.64	1.28	1.59	0.96

Fig. 1. Evolution of fixed-target forecasts over horizons

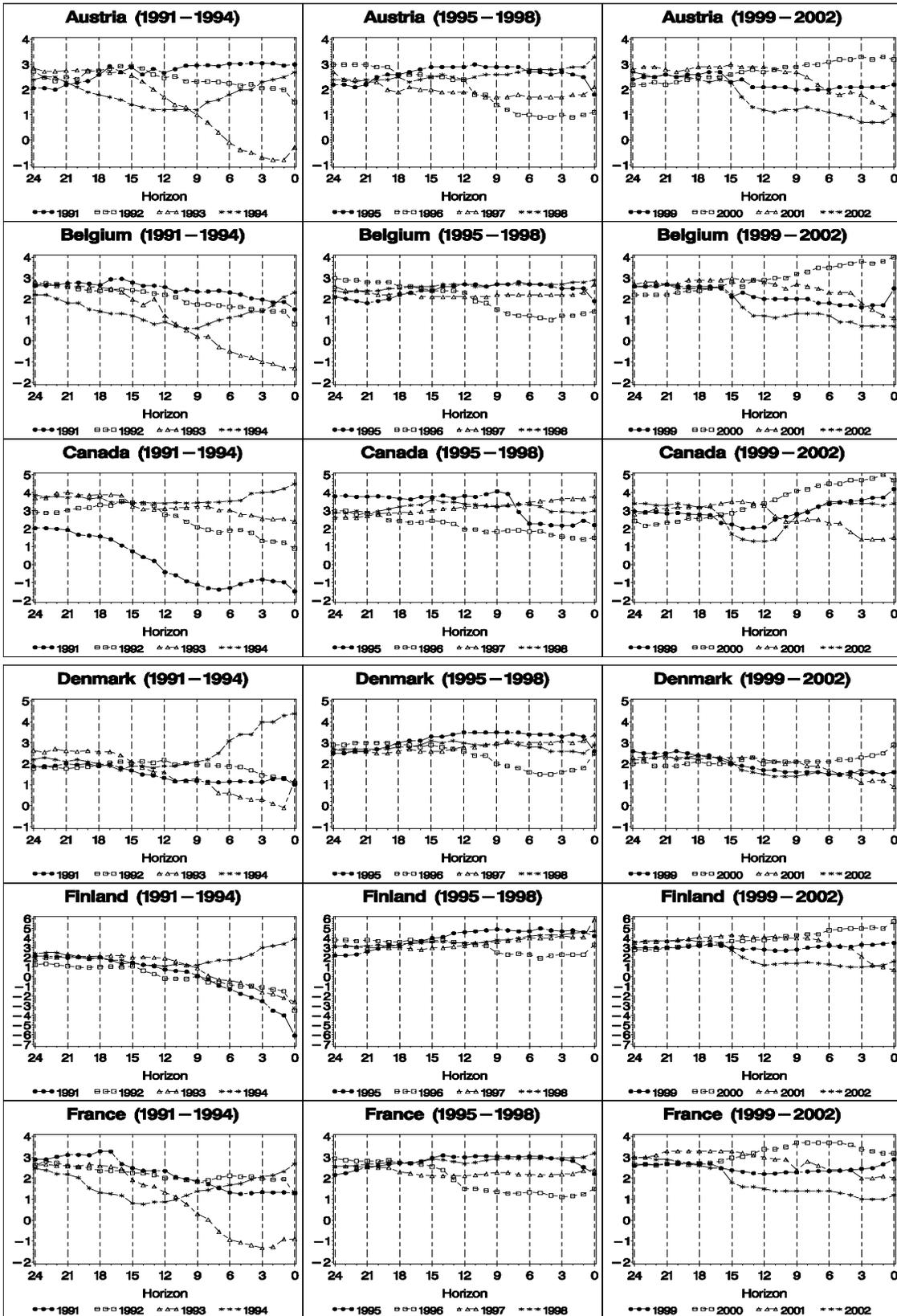


Fig. 1. Cont.

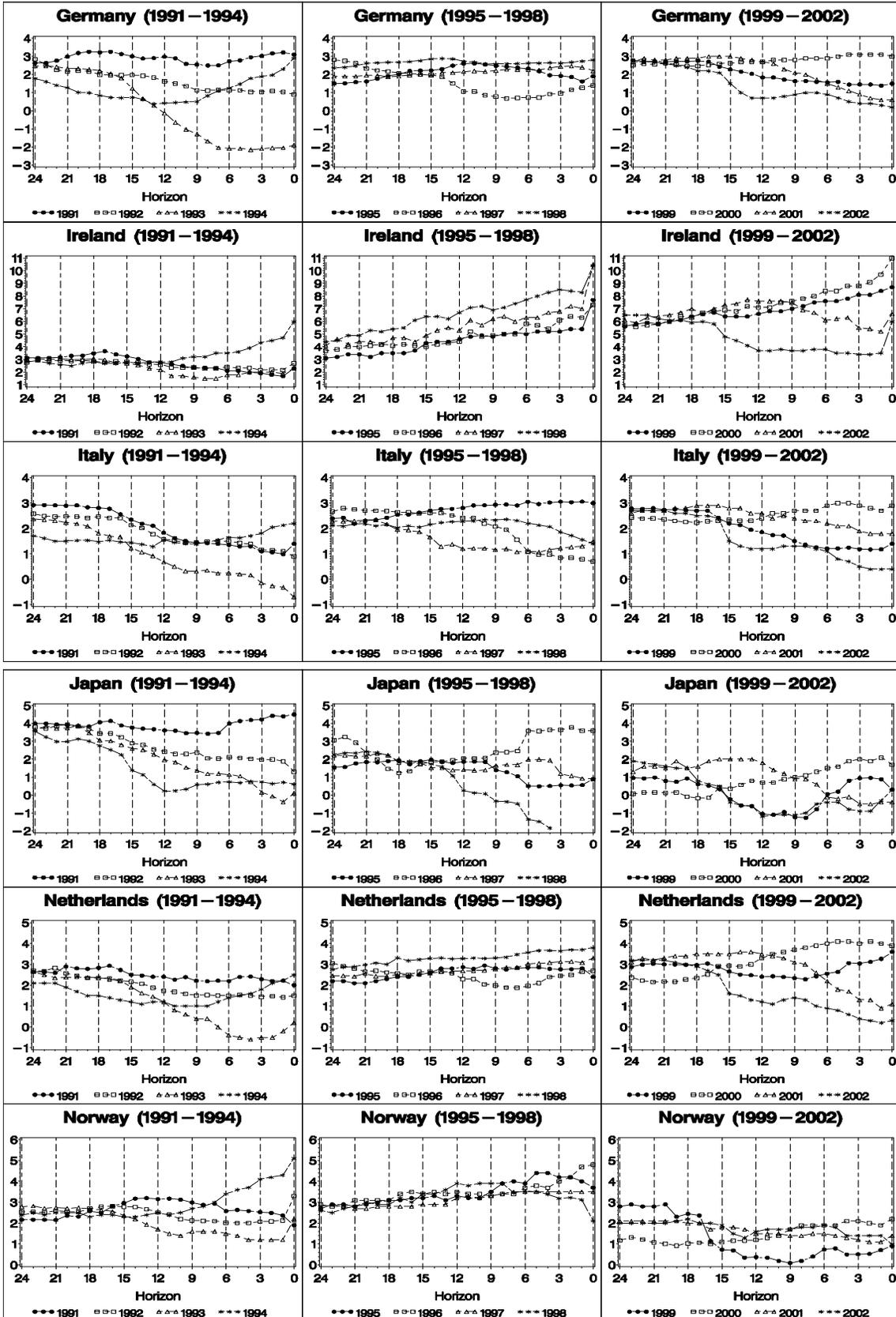


Fig. 1. Cont.

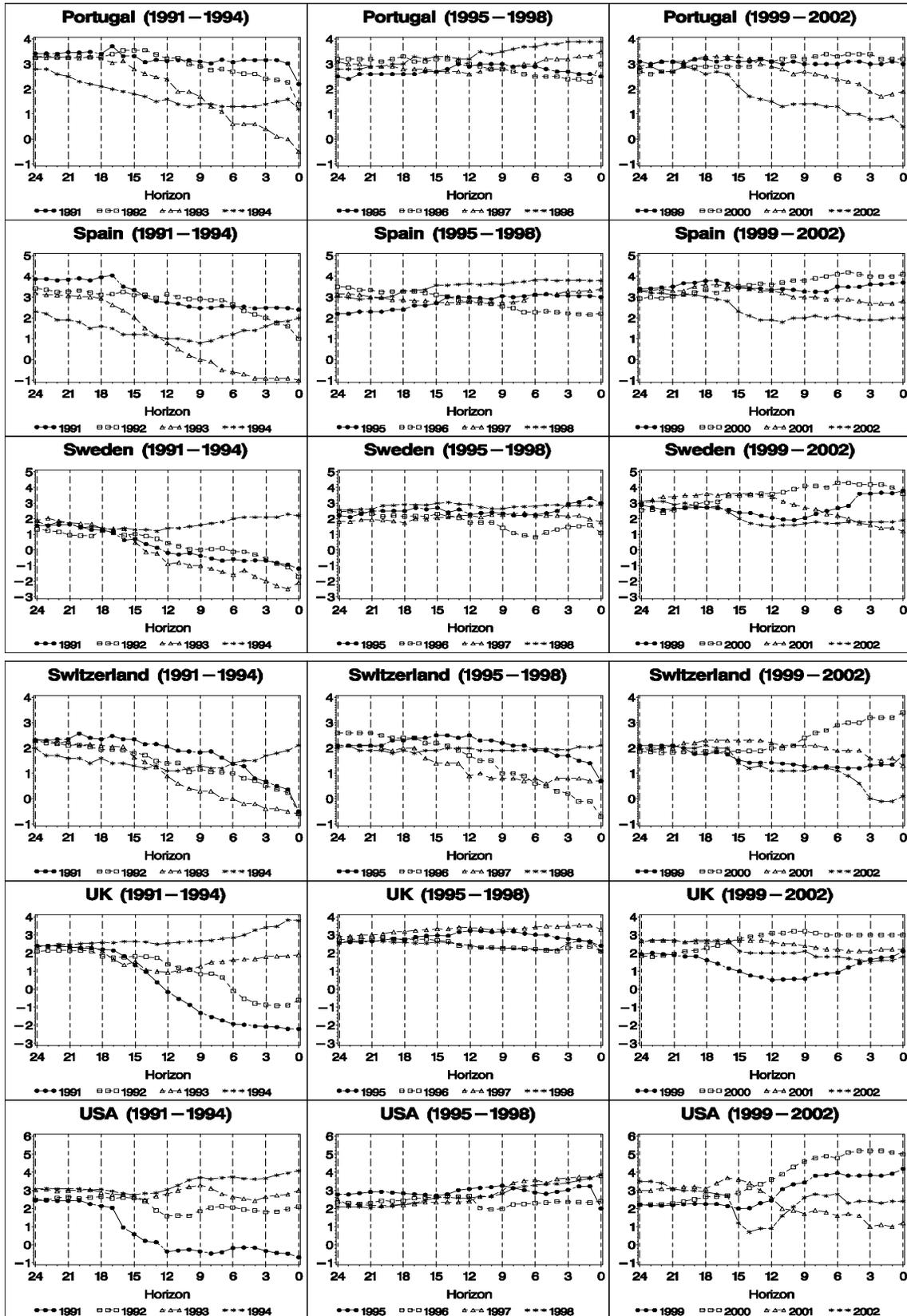


Fig. 2. Forecast variance over forecast horizons, 1989:10 – 2004:06

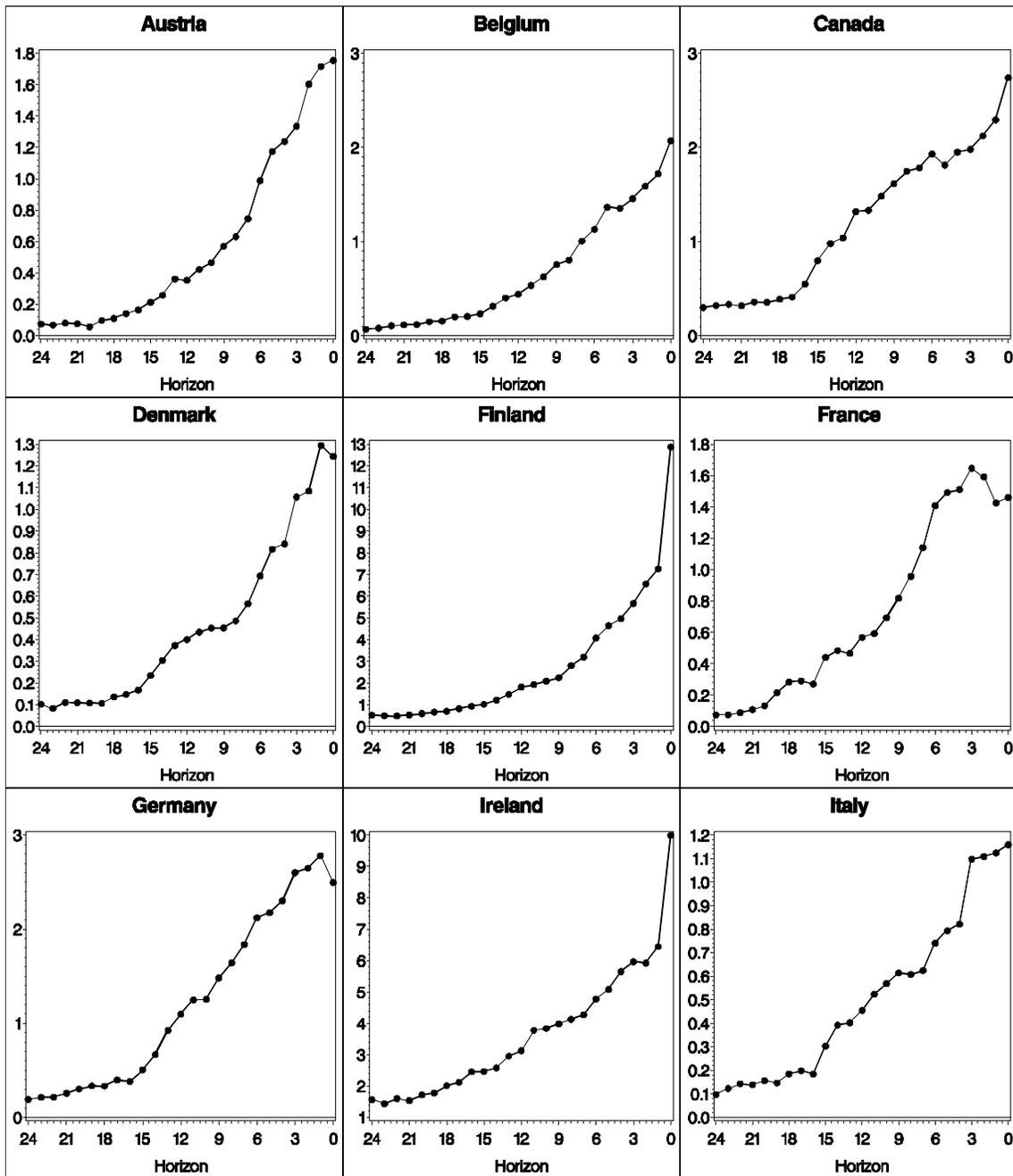


Fig. 2. Cont.

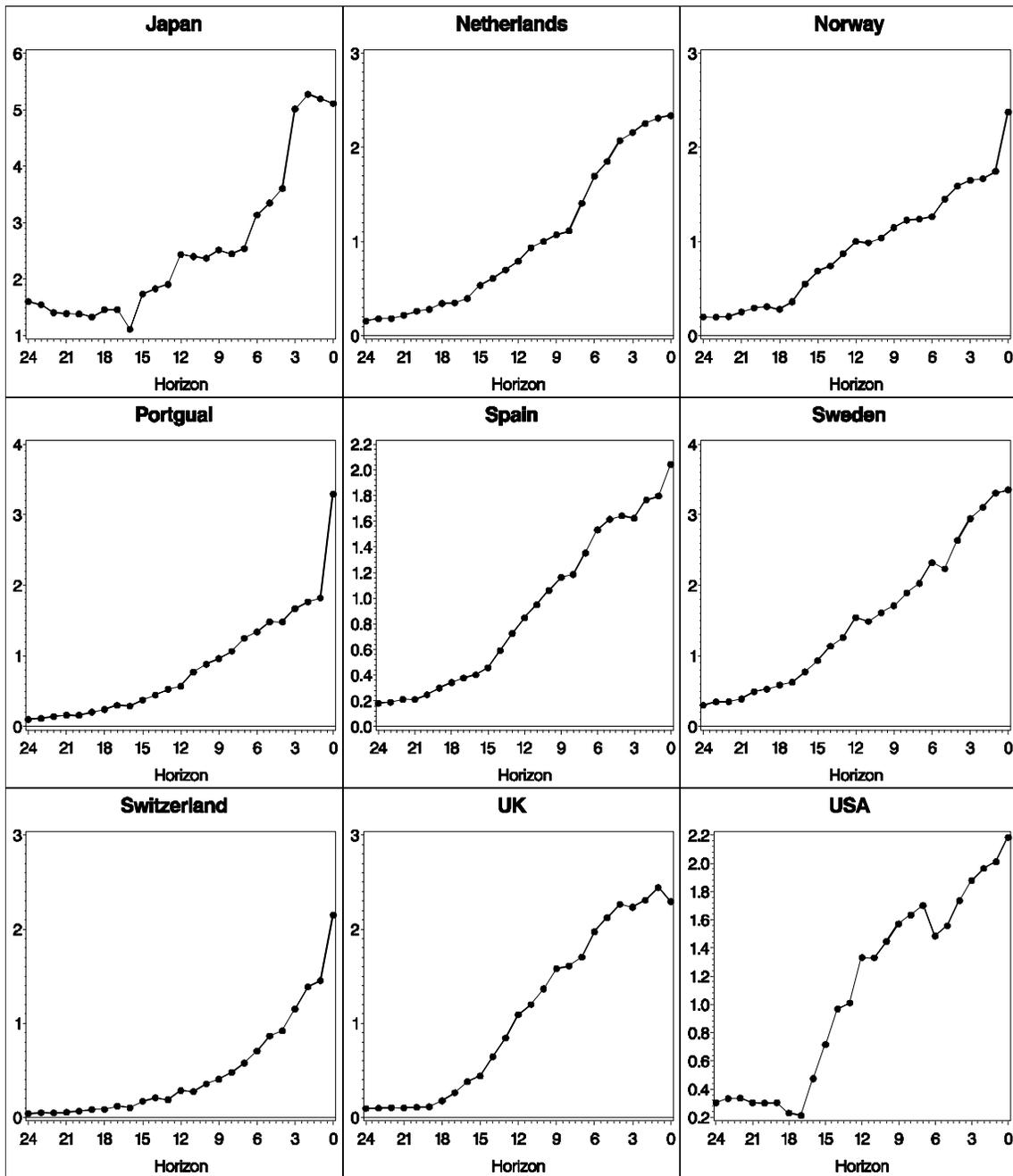


Fig. 3. Information content of forecasts over horizons, 1989:10 – 2004:06
 $U(h)$ = Theil's U; $P(h,24)$ = Diebold-Kilian predictability statistic

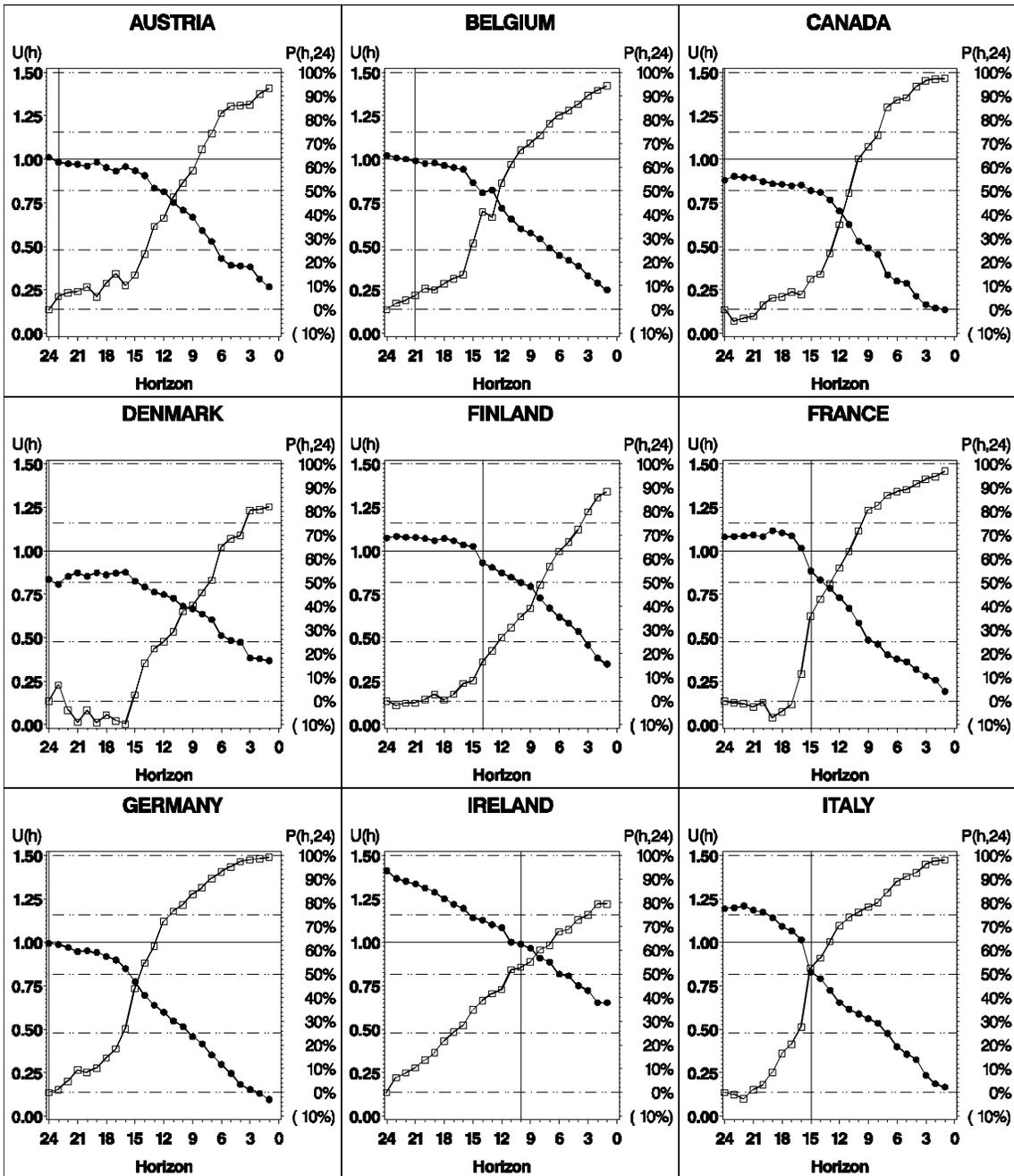


Fig. 3. Cont.

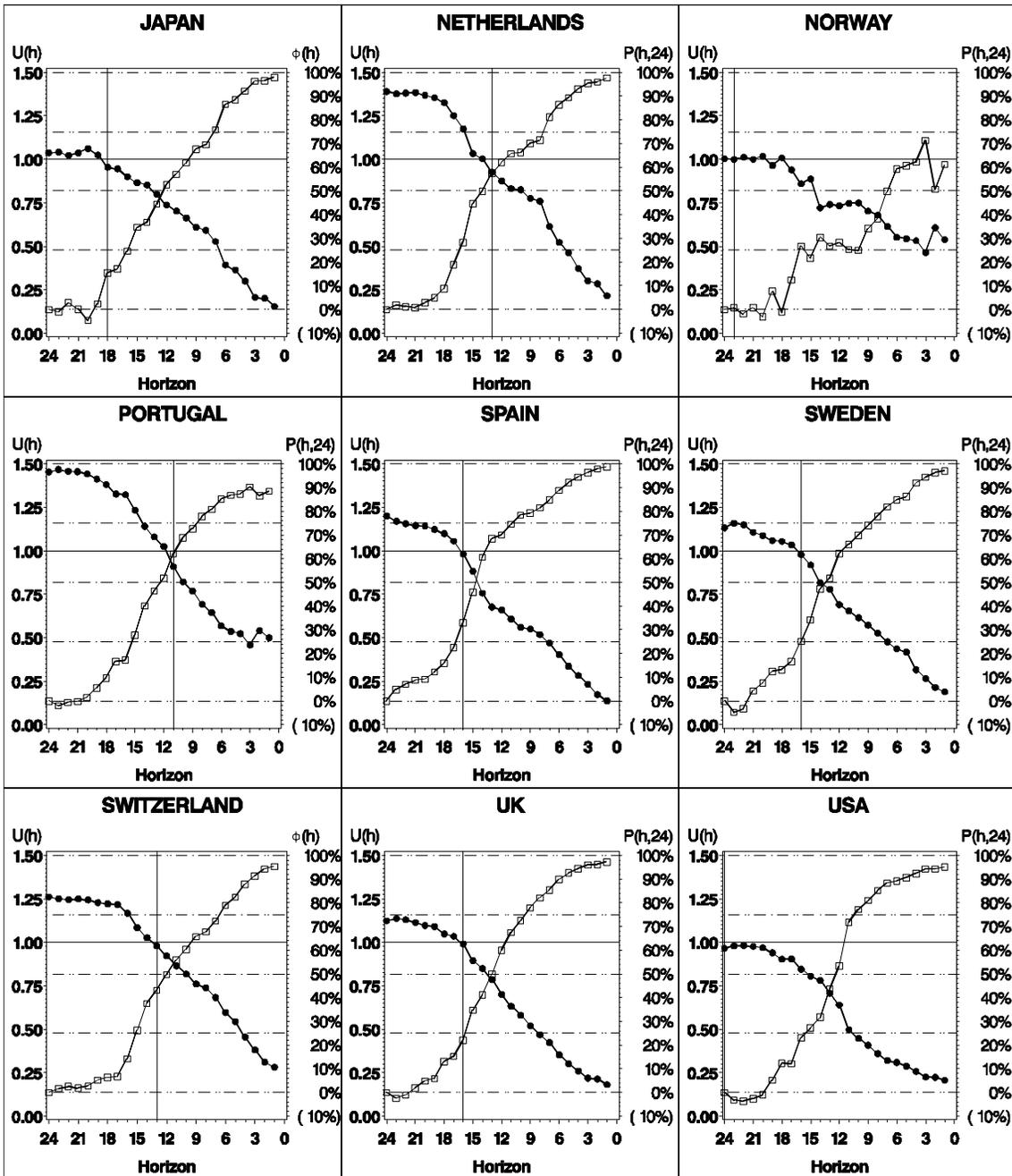


Fig. 4. Flow of information arrival over horizons, 1989:10 – 2004:06
 $\Delta MSE = \text{dots}$; Mean Square Revision (MSR) = squares

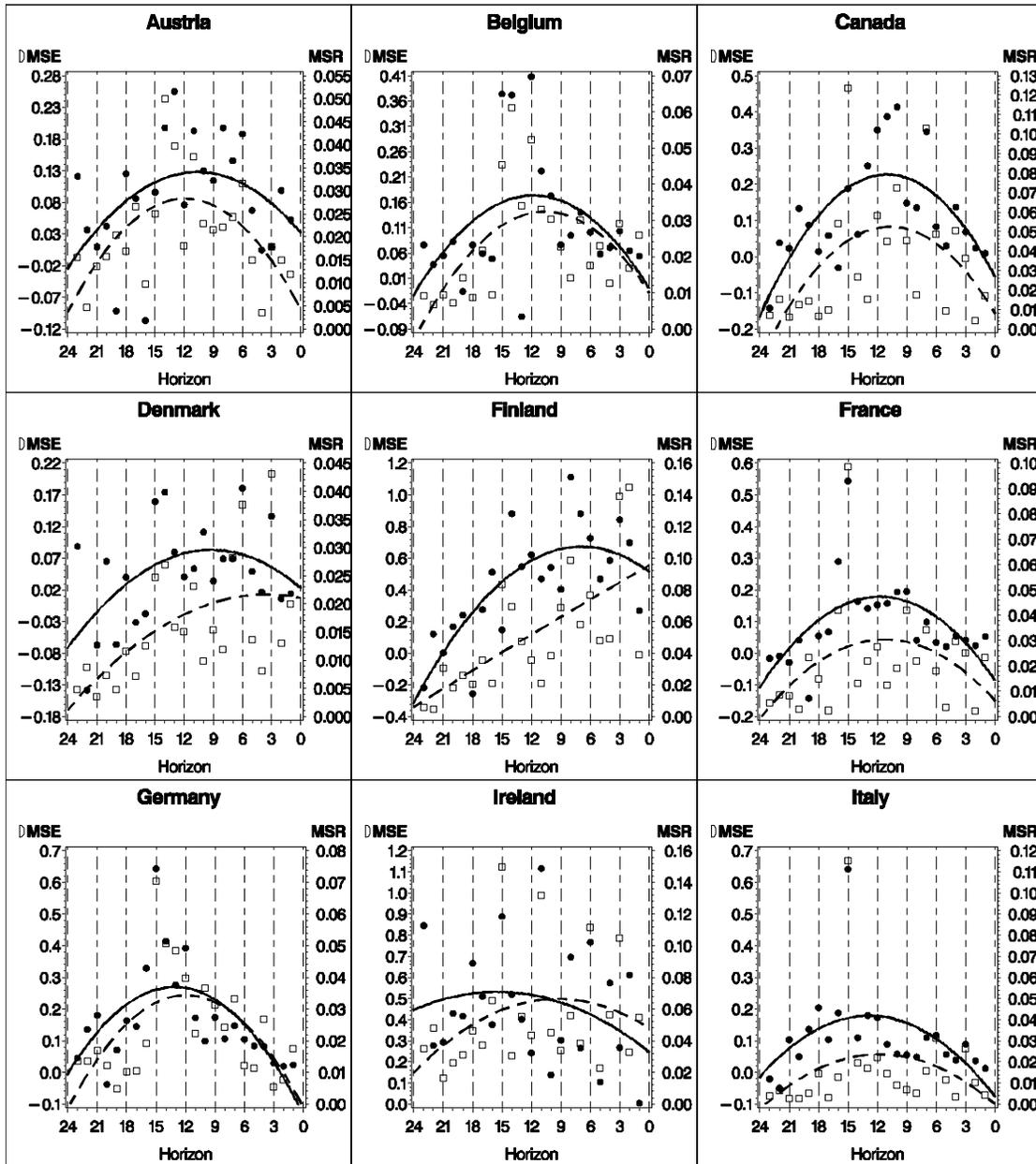


Fig. 4. Cont.

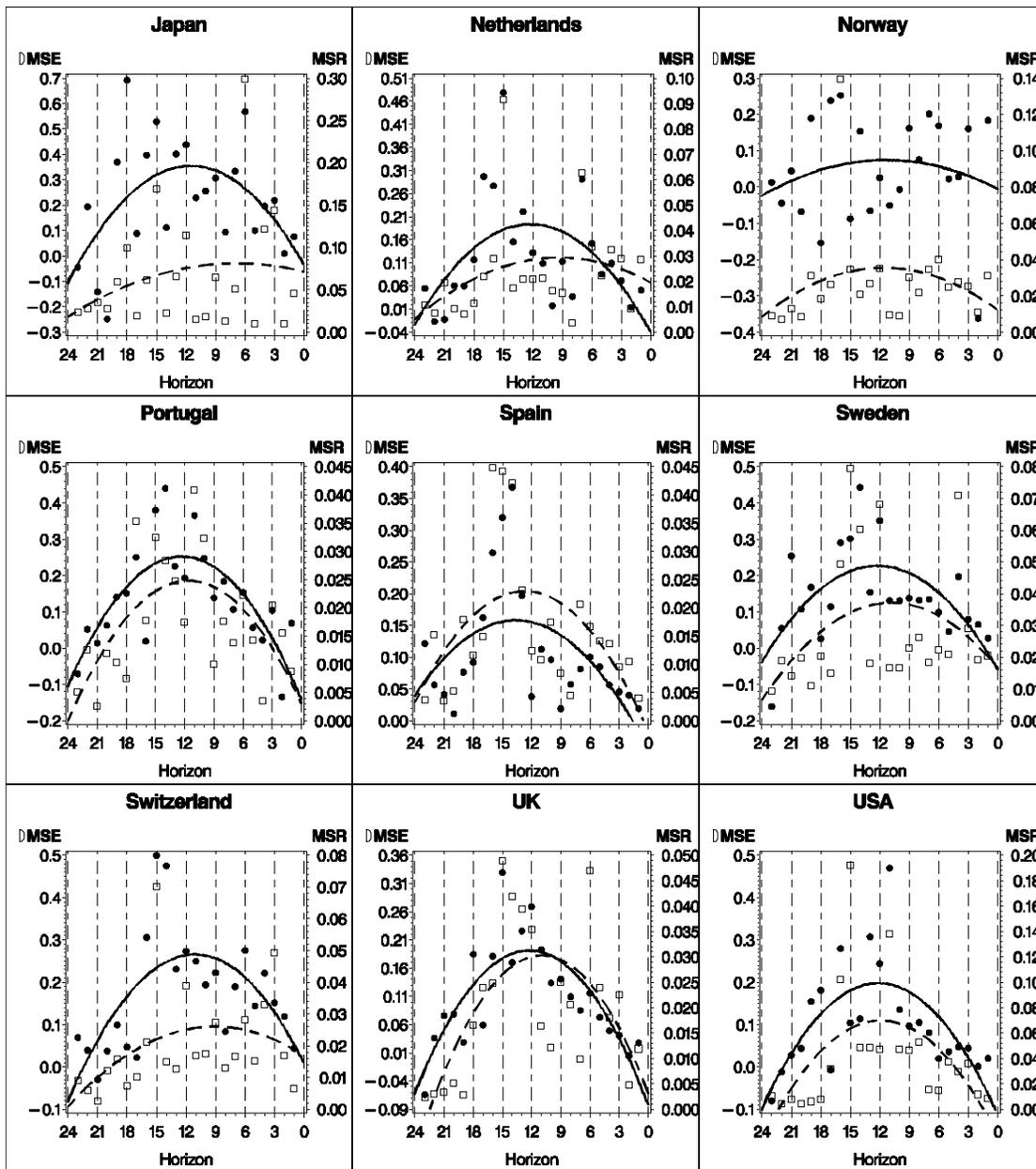


Fig. 5. Information flow based on ARMA model of GDP growth, 1989:10 – 2004:06

