Forecasters' ability – what do we usually assess and what would we like to assess?

By Michael K. Andersson and Ted Aranki*
by Michael K. Andersson and Ted Aranki
The authors work in the Monetary Policy Department.

In this article, we propose a method for the comparison of various forecasters' ability. One problem in comparing forecasts is that forecasts are prepared at different points in time. This means that forecasts are based on differing amounts of information. The closer one comes to the outcome date for the variable being forecast, the more information the forecaster has regarding the development of the variable. Consequently, a comparison of the accuracy of forecasts should allow adjustments to be made for such differences. We achieve this by simultaneous estimation of the forecasters' ability and the effects of the amount of information. The proposed method of comparison is applied to a body of data covering ten Swedish forecasters. This data covers the period 1999–2008. We examine the importance of the amount of information and the ability of the various forecasters for the entire period and for a specific year, namely 2008, which is the most recent year for which we have an outcome.

What do we usually assess and what would we like to assess?

In the world of sports, winning is all-important and winners are considered to be the best. But is it true that the winner is always the best athlete (or that the best man always wins)? Sometimes this assertion holds true: the 100-metre sprinter who crosses the finishing line first wins and, assuming that the underlying conditions have been fair, it would also be reasonable to describe this sprinter as 'the best'.[1] However, there are also sports in which the equipment used is important to the result achieved – possibly

* We would like to thank Stefan Palmqvist, Lars E.O. Svensson and Joanna Gerwin for suggestions and comments on previous drafts of this article. We would furthermore like to thank the National Institute of Economic Research for sharing data. Any remaining inaccuracies or shortcomings in this article are of our own making.

[1] However, considering the increasing frequency of injuries and doping in sports, it can be questioned whether the winner is always the best man.

even more important that the sportsman himself. One such sport is motor racing's Formula One. It is usually claimed that Michael Schumacher is the best driver of all time, but it could equally likely be the case that it was the car Schumacher drove (a Ferrari) that was the best of all time.

A similar statement could be made regarding forecasters. Can we be certain that the forecaster ranked highest in a traditional statistical evaluation is also the best forecaster? Or could it possibly be the case that this forecaster publishes its reports at a later date than all of the others and thus has an information advantage? It is thus not a foregone conclusion that the forecaster with the best forecasting accuracy under a (standard) evaluation also has the best ability in making forecasts.

Forecast elvaluations are important

Forecasts are perishable goods. They are interesting on their date of publication, but are replaced by newer forecasts relatively quickly. However, occasional studies of previously published forecasts are important, not least as important economic and political decisions are often based on them. A forecaster's accuracy is normally assessed with the aid of average forecast errors – that is, on the basis of calculations of the average amount by which forecasts have deviated from outcome. As the economy is constantly affected by different events that are difficult to foresee, the accuracy of forecasts varies. For instance, a large forecast error may be due to a shock that could not have been predicted. An assessment of an individual year thus provides only limited information on the forecaster's accuracy. It is therefore also informative to compare the precision of different forecasters, preferably over a longer period of time.

The Riksbank, the Ministry of Finance and the National Institute of Economic Research regularly evaluate their forecasts and compare them with those of other institutions.[2] Furthermore, Blix et al. (2001), Bergvall (2005) and Andersson et al. (2007) have published more detailed comparisons of Swedish forecasters. International studies of panels of forecasters include Bauer et al. (2003), who assess the participants of the Blue Chip panel of US forecasters. Goh and Lawrence (2006) compare the precision and ranking of a number of New Zealand forecasters.

2 The Riksbank publishes an annual forecast levaluation in "Material for Assessing Monetary Policy in Sweden". "The Ministry of Finance and the National Institute of Economic Research present similar assessments in the "Spring Fiscal Bill" and in the first issue of the report, "The Swedish Economy" each year, respectively.

The comparison of forecasts can be misleading

Forecast comparisons are based on analyses of observed forecast errors.[3] Usually, the average forecast error and the mean squared forecast error (or the mean absolute forecast error) is employed to study the degree of accuracy of forecasts. The average forecast error indicates whether there exists a systematic level error (bias) in the forecasts, while the mean square error summarises the bias and dispersion of forecast errors. These measurements can be used to compare various forecasters' accuracy, with the desired values of the calculated measurements being as low as possible. Forecasts that are always accurate have no bias and their mean square error equals zero.

Forecast comparisons based on such statistical measurements are sufficient if the compared forecasts were made at the same point in time and are thereby based on the same quantities of information. However, as different forecasters publish their forecasts at different points in time, in practice implying that the forecasters have varying quantities of information (in the form, for example, of outcome, indicators and forecasts by other agencies) when they prepare their forecasts, a straight comparison of forecast errors is not entirely fair. A forecaster that systematically publishes its forecasts after everybody else can be expected, on average, to have a better accuracy than the other forecasters.

What do we wish to assess?

One legitimate question arising in comparisons of different forecasters is whether the accuracy of the forecasters is the most interesting factor to study. Could it actually be their ability in making forecasts that forms the area of real interest? Accuracy is usually compared, even though it is ability that is being discussed.

If it is, then, ability that is of interest, how can this ability be separated from other factors affecting accuracy? This is not entirely obvious. One could, for example, dwell upon the parallel with the sporting world, in which competition and comparisons are commonplace. Michael Schumacher is considered by many to be a giant within Formula One, where he is history's most successful driver. But is Schumacher's success due to his being the most skilful driver – or could it be due to his having driven a better car?[4] In Formula One, it is (probably) not enough to be the most

---

3 The term forecast error refers to the difference between outcome and forecast. The error for a forecast made at a point in time t and which refers to an outcome of a variable at time T is defined as pf($T|t$) = outcome(T) forecast($T|t$).

4 Presumably winning (a combination of the driver's skill and the car's performance) is the most interesting element of Formula One.

skilful driver – in order to be a potential champion, a very good car is also needed. During the 2000s, Kimi Räikkönen was one of Schumacher's foremost challengers. During the years 2002–2006, Schumacher finished 1, 1, 1, 3 and 2, while Räikkönen finished 6, 2, 7, 2 and 5. Räikkönen thus only succeeded in getting a better placing than Schumacher in the total series in one of these years, namely 2005. Would it, then, be fair to say that Schumacher was a better driver than Räikkönen during this entire period?[5] During these years, we know that Schumacher drove for Ferrari, while Räikkönen drove for McLaren-Mercedes. So, has it been established, without doubt, that Schumacher was the better driver – or could it be that Ferrari cars were better than Mercedes?

In the example above, it cannot be identified how much of the performance depends upon the driver and how much depends upon the car. In order to make it possible to identify the most skilful driver, both drivers would have needed to have "exchanged" cars with one another (preferably with the aid of random selection). Another method of identifying a driver's skill would be to allow a third driver to alternately drive Ferrari and Mercedes. This would allow an objective comparison to be made between the cars, after which the drivers' skill could be identified, given the cars' performance.

When comparing forecasters, it is normal to state when the forecasts under comparison were made. The National Institute of Economic Research relates the publishing date of each forecaster's report to the publication date of its own report. However, this is only equivalent to saying that Schumacher won *and* that he drove a Ferrari.

In this article, we take matters one step further and propose a method that takes into account the different amounts of information held by forecasters when they make their forecasts. We use the difference between publication date and outcome date (in months) as an approximation of the value of the available amount of information. The method is based on a model in which the importance of the quantity of information and the ability of the forecaster are estimated simultaneously. Unlike the Formula One example, we have sufficient variation in the data material to separate the effects of the available information from the forecaster's ability.

One way of considering the importance of the quantity of information

Assume that $\hat{x}(h)_{it}$ is a forecast made by forecaster i for variable x, at point in time t and which is published h months before the outcome of variable x is known. This implies that forecaster i has access to information up

[5] We cannot, based on a so-called sign test, reject that both drivers were equally skilful.

to and including time t to make its forecast. The absolute forecast error, which is comprised of the difference between the outcome for point in time T (x_T) and the forecast at point in time t in absolute figures, can be expressed as follows

$$(1) \qquad \varepsilon_{it} = |x_T - \hat{x}(h)_{it}|$$

We model the absolute forecast error as a function of the distance to outcome and each forecaster's ability in making forecasts according to the following general specification

$$(2) \qquad \varepsilon_{it} = \alpha_1 h_{it} + \alpha_2 h_{it}^2 + \alpha_3 h_{it}^3 + \mu_i + \lambda_t + e_{it}$$

where h_{it} is a horizon variable approximating the information available during the period of time until and including the publication date t. The coefficients preceding the horizon variable (α_1, α_2 och α_3) measure the marginal effect on the absolute forecast error of increasing the horizon by one month. The variables h_{it}^2 and h_{it}^3 are included in the model to provide it with the functional form best resembling the empirical relation between the absolute forecast error and the available information.[6,7] The parameter μ_i describes forecaster $i{:}s$ average ability (described in the literature as individual effect), while λ_t reflects the differing levels of difficulty in forecasting for different years. This quantity is usually called a time-specific effect and is shared by all forecasters, but varies across time. The model's residual, e_{it}, is an error term that is assumed to be randomly distributed, with mean zero and constant variance.

Forecasts from ten different institutions

The analysis presented in this study is based on data gathered by the National Institute of Economic Research.[8] The forecast comparison covers ten forecasting institutions and their full-year forecasts for GDP growth, CPI and unemployment (rate) figures for the period 1999–2008. GDP and CPI are measured as the average annual percentage change, while unemployment is measured as the annual average of the number of unemployed (in relation to the size of the labour force).

6 A description of the manner in which the average forecast error can be approximated by use of the forecast horizon is presented in the Appendix.

7 Assessment of the model proceeds from equation (2), allowing the data to determine which trend components will finally be included in the specification, that is, we perform individual tests of whether $\alpha 1$, $\alpha 2$ and $\alpha 3$ can statistically be separated from zero.

8 The data supplied by the National Institute of Economic Research covers the period 1994–2007. We have complemented the data with information for 2008.

In order to evaluate each forecaster's accuracy, we study the forecasts made up to two years before the publication of outcome. This provides a maximum horizon of 24 months.[9]

Figure 1 summarises the relation between the forecasters' absolute forecast errors and the horizon. We can observe that errors are minor during the short horizons (a few months) and, in general, increase as the distance to outcome increases. This is not surprising. When forecasts are made closer to the date of outcome, more details of the approaching outcome are known (refer to the Appendix for a more detailed description of outcome effects).

A further description of the dataset is presented in Table 1. Among other information, the table indicates that the number of forecasts published differs between the various forecasters. During the period studied, 1999–2008, the Swedish Trade Union Confederation published the smallest number of forecasts (37) and the National Institute of Economic Research the largest (81).

The lower portion of each panel in Table 1 indicates the average absolute forecast error for the respective variable. The standard deviation of the data has been allocated between the various forecasters and within each individual forecaster. In general, forecast error does not vary greatly between forecasters, while the variation is greater within the individual forecasters' sets of forecast errors. For example, the variation, measured as a standard deviation, between the various participants' average forecast error is 0.08 per cent in GDP forecasts, while the respective participants' forecast error in the same forecasts indicates a variation of 0.85.[10] This indicates that forecasters regularly adjust their forecasts and, at the same time, that the forecasts do not markedly differ between the various participants. According to our interpretation, this implies that 'herd behaviour' is taking place among the studied forecasters.[11]

Table 2 presents the number of forecasts that each respective forecaster has published at various horizons. There are some regularities regarding when various forecasters publish their forecasts, and dates of publication vary between the forecasters. The final row in the table indicates the mean horizon of the various forecasters. A comparison of these mean horizons provides an indication of the manner in which a correction of the forecast error can affect the result. As most forecasters' average horizons lie relatively close to one another, adjustments of forecast errors are expected to be minor.

---

9 Forecasts prepared the year following the assessment year are excluded from the investigation as the data does not cover them.

10 Note that the variation between different forecasters and the variation for each forecaster do not add up to the total variation, as the two standard deviations are not calculated around the same mean value.

11 The occurrence of herd behaviour among forecasters is not particularly remarkable. Forecasters study approximately the same information and have access to each other's forecasts and analyses.

The mean horizon for the entire dataset is 12.1 months. The Riksbank's mean horizon of 11.3 implies that the Riksbank, on average, publishes its forecasts 0.8 months (24–25 days) later than the average forecaster. This implies, in turn, that the Riksbank, on average, has access to more information than other forecasters. An adjustment of the forecast error as regards the information set should thus increase the Riksbank's forecast error relative to the other forecasters. The reverse probably applies to those forecasters with a longer mean horizon than average (for all forecasters).

As different forecasters publish their forecasts at different points in time within and across the years, information regarding each forecaster's mean horizon is not sufficient to adjust the forecast error. In order to perform a fair adjustment, information regarding the distance to outcome for *all* forecasts must be utilised.

Forecasters' estimated ability

In this section, we present the estimated models, discuss the importance of the quantity of information and analyse the various forecasters' average forecasting ability. We estimate and compare this ability for the entire period 1999–2008, as well as for 2008 alone. Forecast evaluations usually focus on individual years. Such analyses have a certain value, but tend to have the character of a description. Consequently, in order to be able to draw conclusions regarding general forecasting ability, a broader range of years is required, particularly as accuracy can vary widely from year to year.

THE MODEL CAN BE USED TO ESTIMATE FORECAST ABILITY

Table 3 presents the estimation result for the forecasts for GDP, CPI and unemployment, respectively. For each specification, the model includes a set of constant time effects and individual effects (see equation 2). The estimated linear portion of the horizon variable (h) is positive and differs significantly from zero for each of the three variables. This means that, just as expected, the average forecast error increases the further away from the outcome date the forecast is published. The squared horizon variable (h^2), which enables the forecast error to increase or decrease at a faster rate than the linear portion, is only significantly separated from zero for the GDP specification.[12] The cubed horizon term (h^3) is not significant

[12] The linear-quadratic horizon effect does not differ significantly between the various forecasters, according to a variation test.

for any of the specifications.[13] At the same time, these estimates indicate that the linear portion of the horizon specification is most important for the approximation of the effects of the available information. The marginal effect on forecast error of publishing forecasts one month earlier is 0.11–0.004*h* for GDP, 0.036 for CPI and 0.028 for unemployment.[14] The marginal effect for (for example) CPI indicates that the absolute forecast error could be expected to decrease by 0.036 in the event that a forecast were to be published one month later. That the marginal horizon effect for GDP is a function of the horizon h is due to the fact that the squared horizon term h^2 is included in the model for GDP.[15]

The time effects are strongly significant for each of the three variables – implying that in certain years, it is easier or more difficult to forecast the outcome of the variables than it is in other years.[16] A similar test is used to investigate differences in the ability of the various forecasters. As regards both CPI and unemployment, a joint F-test indicates that there exist significant differences between the forecasters concerning the accuracy of forecasts for these variables. On the other hand, we find no significant differences in their ability in making forecasts for GDP. However, pair wise t-tests indicate differences between certain forecasters. These two tests differ from each other in that the F-test determines whether any particular forecaster's precision deviates from the average value for all forecasters, while the t-test investigates whether two individual forecasters have differing abilities. Table 4 presents significance tests for the Riksbank's forecasting ability in comparison with other forecasters. We will discuss the result of these tests later in this article.

It is important that the horizon variable and forecasting ability are not strongly correlated with each other, as this would raise doubts as to the possibility of separating ability from the horizon effect. We find no serious indications of such multicollinearity (dependency between the explanatory variables) in the respective specification.[17] Furthermore, the model diagnostics indicate that the specifications function well, which, in turn, suggests that the models can be used to analyse and compare the ability of the various forecasters.

Below, we present the estimated forecast ability (adjusted for quantity of information at publication date) for each forecaster as regards forecasts for GDP, CPI and unemployment. Panels (b) and (c) in Figures

---

13 The cubed trend term is included to allow a flexible representation of the importance of the horizon. We provide further information regarding this in the Appendix of this article.

14 The marginal effects are estimated as the derivative of the estimated relation $\alpha_1 h_{it} + \alpha_2 h_{it}^2 + \alpha_3 h_{it}^3$

15 Compare the horizon effect for GDP (panel (a) in Figure 2) with the equivalent effects for CPI and unemployment (panel (a) in Figures 3 and 4, respectively).

16 We use an F-test with the null-hypothesis that λ_t is equal for all years against the alternative that all λ_t are not equal – that is to say that outcome for all years is not as difficult to forecast.

17 For this, we have used variance inflation factors (VIF).

2, 3 and 4 indicate forecast ability for the entire period 1999–2008 and for 2008 alone.[18] In each figure, we present the estimated ability and the more traditional measure mean absolute error (MAE). Both of these measurements are stated as deviations from all forecasters' average accuracy. A positive column (value >0) implies a greater adjusted forecast error than the average forecaster and thus a lower forecast ability, while a negative column (value <0) implies the reverse. As a complement to these figures, Table 5 presents the ranking of forecasters, 1999–2008 and 2008, for the three investigation variables.

In our example from the world of Formula One, Ferrari probably had a better car, as the company invested more resources in Formula One than elsewhere. This also holds true for forecasters – the public institutions (the Ministry of Finance, National Institute of Economic Research and Riksbank) have larger forecasting organisations than, for example, the commercial banks. Even though our method does not examine this resource aspect, our results still provide a certain degree of information about it.[19]

SIZE DOES NOT MATTER – WHEN PREDICTING GDP

The ability of both the Ministry of Finance and the Riksbank in the forecasting of GDP appears to be relatively good for the entire period under examination (see panel (b) in Figure 2 and Table 5 for a ranking of forecasters). However, even smaller participants such as Nordea and Skandinaviska Enskilda Banken are included among those with the best forecasting ability. Consequently, it cannot be taken for granted that organisations with major resources produce better GDP forecasts than those with lesser resources.

The Swedish Retail Institute, Svenska Handelsbanken, the Confederation of Swedish Enterprise and Swedbank are the forecasters with a greater than average adjusted forecast error and, thus, lower ability in their GDP forecasts over time (1999–2008).

Panel (c) in Figure 2 shows that the forecast ability for an individual year, in this case 2008, can deviate greatly from that estimated on a longer sample. For example, the Confederation of Swedish Enterprise reports the most accurate forecasts for 2008. However, considered across the entire period (1999–2008), the Confederation of Swedish Enterprise is

[18] The estimate of μ_i in Equation (2) reports forecast ability during the years 1999–2008. In order to compare forecast precision for individual years, Equation (2) must include an interaction term that is only active for the year in question.

[19] We can only comment on the differences between larger and smaller forecasters. The effect of the amount of resources invested by a forecaster could be identified in a similar manner to the horizon effect with the aid of the inclusion, among other factors, of the number of employees, their educational level and wage costs. However, this lies outside the scope of this study.

placed among those forecasters with lower than average ability (see panel (b) in the same figure).

THE MAJOR ACTORS HAVE MADE THE BEST CPI FORECASTS

The test results indicate that, over time, there exist systematic differences between the levels of ability of different forecasters in the prediction of CPI. The major authorities made the most accurate CPI forecasts during the period 1999–2008 (see the ranking in Table 5 and estimated ability in panel (b) in Figure 3). These authorities also provided reliable forecasts for 2008.

THE AUTHORITIES' ABILITY IN MAKING UNEMPLOYMENT FORECASTS LIES CLOSE TO AVERAGE

Generally seen, the major forecasters made forecasts lying, in terms of accuracy, close to the average for all participants, with the exception of the National Institute of Economic Research, which had the best ability in forecasting unemployment during the period 1999–2008. The least accurate unemployment forecasts were made by the labour market organisations – the Swedish Trade Union Confederation and the Confederation of Swedish Enterprise.

Regarding the Riksbank's forecasts

In this article, our primary aim has been to describe our method of evaluating forecasts and comparing the forecasts of ten Swedish forecasters. Consequently, we have tried not to focus specifically on the Riksbank's forecasting ability. However, in this section we take matters one step further and analyse the Riksbank's own forecasts, comparing them with those of the other forecasters.

In comparison with those prepared by other forecast institutions, the Riksbank's forecasts for GDP appear to be relatively accurate (see Table 5 for a ranking of forecasters). According to our ranking, the Riksbank has made the second most accurate GDP forecasts for the entire period, but the quantitative difference between the forecasters with the best forecast precision is small. This is shown by the point estimations presented in panel (b) in Figure 2. Paired significance tests of the Riksbank and other forecasters indicate that the Riksbank, over time, has been significantly better at predicting GDP than have the Swedish Retail Institute, Svenska Handelsbanken, the Confederation of Swedish Enterprise and Swedbank (see Table 4). The accuracy of the Riksbank's forecasts for GDP growth was also relatively high in 2008.

The Riksbank also belongs to the group of forecasters making the most accurate CPI forecasts over time. Paired significance tests indicate that the Riksbank's CPI forecasts have been significantly better than similar forecasts by the Swedish Retail Institute, Nordea, Svenska Handelsbanken and the Confederation of Swedish Enterprise (see Table 4). The Riksbank is also included among the best forecasters of CPI in the individual year of 2008. However, the differences within the group of forecasters demonstrating better than average accuracy are minor. The Riksbank's ability for 2008 was significantly better than that of the Swedish Trade Union Confederation and Nordea.

Over time, the Riksbank's unemployment forecast was only marginally better than average. In contrast, the Riksbank's forecasts for 2008 have been among the most accurate and significantly better than those of the Ministry of Finance, the National Institute of Economic Research, the Swedish Trade Union Confederation and the Confederation of Swedish Enterprise. These forecasters belong to the group of forecasters with a greater than average adjusted forecast error for this individual year. Over the longer period of time the forecasts prepared by the Riksbank have been significantly more accurate than those prepared by Nordea and Svenska Handelsbank (see Table 4).

THE RIKSBANK'S FORECASTS HOLD UP WELL

Our assessment of mean ranking indicates that the Riksbank's forecast accuracy, in terms of each of the three variables, has been the second best of the investigated forecasters (see Table 5). The mean ranking is calculated as the mean value of each forecaster's ranking for the individual variables (GDP, CPI and unemployment). The Riksbank is ranked as second best for GDP, third best for CPI and fourth best for unemployment. Consequently, the Riksbank's mean rank is 3.0 (= (2+3+4)/3).[20] One interesting observation is that the three forecasters with the most resources are included among the four best forecasters, according to the mean ranking. The best mean ranking over the entire period was attained by the National Institute of Economic Research (2.7), while the worst mean ranking was attained by Svenska Handelsbanken (9.0).

[20] The mean ranking provides a better way of aggregating the variables than calculating the sum (or mean value) of the mean absolute error for the variables. The mean absolute error cannot be compared across variables as different levels of difficulty apply to the forecast of different variables. Mean ranking is not affected by this problem. However, mean ranking does not consider the extent of the differences between the levels of forecasting ability of the various participants.

The equivalent mean ranking calculation for 2008 indicates that, all in all, the Riksbank's and Svenska Handelsbanken's forecast precision was best among all investigated forecasters. Their mean ranking for 2008 is 3.0. The third most precise forecasts were made by the Swedish Retail Institute and SEB, which both attained mean rankings of 5.0. The least accurate forecaster for this year was the National Institute of Economic Research, with a mean ranking of 8.3.

A comparison of our measurement of precision and a traditional measurement

What are we actually interested in? Do we want to know who won or who was most skilful? In Formula One, it is, of course, most important to win, and winning often entails that the driver in question is considered to be the best (even if he has also probably driven the best car). However, when we evaluate forecasters, it is the second option we are looking for – we want to know who is most skilful. So far, we have used our method to compare the ability of the investigated forecasters. In this section, we compare the precision measurements we propose in this study (that is, the measurements answering the question: who is most skilful?) with a standard MAE evaluation (which seeks to answer the question: who won?). The results are presented in panels (b) and (c), respectively, in Figures 2, 3 and 4. In addition, Table 5 illustrates a comparison of the rankings of forecasters.

Unlike a study of mean absolute errors, our precision measurement indicates, for example, that the Riksbank, on the whole, has a worse place in the ranking. This is due to the fact that the Riksbank often publishes its forecasts at a later date than the other investigated forecasters. The opposite effect can be seen for Swedbank, which, on average, publishes earlier than other forecasters. Swedbank improves its ranking considerably using this method, as compared with an assessment of mean absolute error.

This empirical result also indicates the importance of the manner in which an investigation is conducted, as well as the importance of considering the importance of the quantity of information, in order to obtain a fairer comparison of different forecasters.

As a concluding observation, we would like to mention that Kimi Räikkönen changed to Ferrari (Schumacher's old team) for the 2007 season. That year, Räikkönen won the entire Formula One series.

Summary

In this article, we introduce a method for the comparison of different forecasters which considers that they publish their forecasts at different points in time. This method is applied to a body of data including forecasts made by ten Swedish forecasters. The result of the method, in the form of the ranking of forecasters, can deviate from the result provided by more traditional statistical assessment measurements. It is thus meaningful to adjust for differences in publication date when comparing forecasters.

References

Andersson, M.K., G. Karlsson and J. Svensson (2007), "The Riksbank's Forecasting Performance", Sveriges Riksbank Working Paper Series No. 218.

Baltagi, B.H. (2001), "Econometric Analysis of Panel Data", 2nd ed., John Wiley and Sons, Great Britain.

Baltagi, B.H. and P. Wu (1999), "Unequally Spaced Panel Data Regression with AR(1) Disturbances", *Econometric Theory* 15(6): 814–823.

Bauer, A., R.A. Eisenbeis, D.F. Waggoner and T. Zha (2003), "Forecast Evaluation with Cross-sectional Data: The Blue Chip Surveys", *Federal Reserve Bank of Atlanta Economic Review*, Second Quarter 2003.

Bergvall, A. (2005), "Utvärdering av Konjunkturinstitutets prognoser", Occasional Study No 5, March 2005, National Institute of Economic Research.

Blix, M., J. Wadefjord, U. Wienecke and M. Ådahl (2001), "How Good is the Forecasting Performance of Major Institutions?", Sveriges Riksbank Economic Review 2001:3, .

Goh, K.L. and D. Lawrence (2006), "Treasury's Forecasting Performance: A Head-to-Head Comparison", New Zealand Treasury Working Paper 06/10.

Tables and figures

In the tables and figures below, the ten investigated forecasters are designated as follows:
FD – Ministry of Finance
HUI – Swedish Retail Institute
KI – National Institute of Economic Research
LO – Swedish Trade Union Confederation
NORDEA
RB – Riksbank
SEB – Skandinaviska Enskilda Banken
SHB – Svenska Handelsbanken
SN – Confederationof Swedish Enterprise
SWED – Swedbank.

Table 1: Descriptive statistics for absolute forecast error

		Number of forecasts	Mean value	Standard deviation	Min	Max
			Panel (a) GDP			
FD		41	0.97	0.82	0	3.50
HUI		79	0.96	0.87	0	3.70
KI		81	0.94	0.90	0	4.00
LO		37	0.98	0.85	0	3.20
NORDEA		73	0.93	0.75	0	3.00
RB		80	0.91	0.83	0	3.20
SEB		80	0.93	0.84	0	3.60
SHB		70	1.14	0.95	0	3.80
SN		73	1.02	0.84	0	3.20
SWED		42	1.12	0.95	0	3.30
Total variation	N	656	0.98	0.86	0	4.00
Between	n	10		0.08	0.91	1.14
Within	N/n	65.6		0.85	-0.16	4.05
			Panel (b) CPI			
FD		41	0.38	0.38	0	1.30
HUI		79	0.43	0.36	0	1.50
KI		81	0.33	0.38	0	1.20
LO		37	0.43	0.41	0	1.40
NORDEA		73	0.50	0.52	0	2.10
RB		80	0.35	0.39	0	1.80
SEB		80	0.42	0.40	0	1.70
SHB		70	0.48	0.53	0	2.30
SN		73	0.51	0.46	0	1.80
SWED		42	0.45	0.45	0	1.70
Total variation	N	656	0.43	0.43	0	2.30
Between	n	10		0.06	0.33	0.51
Within	N/n	65.6		0.43	-0.08	2.25
			Panel (c) Unemployment			
FD		41	0.38	0.35	0	1.20
HUI		79	0.35	0.31	0	1.80
KI		81	0.32	0.31	0	1.20
LO		37	0.45	0.42	0	1.50
NORDEA		73	0.45	0.37	0	1.40
RB		78	0.35	0.34	0	1.60
SEB		80	0.33	0.29	0	1.20
SHB		70	0.47	0.44	0	1.90
SN		73	0.42	0.43	0	2.00
SWED		42	0.36	0.34	0	1.60
Total variation	N	654	0.38	0.36	0	2.00
Mellan	n	10		0.06	0.32	0.47
Within	N/n	65.4		0.36	-0.09	1.96

Note. The first row in the table describes the absolute error in the GDP forecasts published by the Ministry of Finance between 1999 and 2008. The Ministry of Finance made 41 forecasts and the absolute error averaged 0.97, with a standard deviation of 0.82. The smallest absolute error registered for the Ministry of Finance is 0 and the greatest is 3.50.

In total, for the entire body of data, 656 forecasts by the ten forecasters have been analysed. The average of all of these GDP forecast errors is 0.98 and the standard deviation is 0.86.

"Between" indicates the spread between the various forecasters' mean absolute error, while "Within" refers to the degree to which each forecaster's absolute error deviates from its mean absolute error. As regards the forecasts for GDP, we find the smallest average forecast error with the Riksbank (0.91) and the greatest with Svenska Handelsbanken (1.14); these are the figures presented in the columns "Min" and "Max" on the row "Between" in panel (a).

Table 2: Number of forecasts by horizon (time in months to outcome date) and forecaster

Horizon	FD	HUI	KI	LO	NORDEA	RB	SEB	SHB	SN	SWED	Total
1	1	10	6			8	2	4	8	1	40
2			3	1	7	1	8	3		1	24
3	2	7	2	5	1	10				1	28
4	8	2		1	6	1	5	5	5		33
5			10	2	3		5	4	5	6	35
6						1					1
7		10	9			9		2	7	3	40
8					7		10	6			23
9	8	5	1	3	3	1		3	2	1	27
10	2	5	9	6		7	2		7	1	39
11				1	4	3	8	3	3	1	23
12			1		6			5		7	19
13		10	5			8	2	3	7	1	36
14			4	1	7	1	8	4		1	26
15	3	7	2	4	1	10				1	28
16	7	3		1	5		6	6	5		33
17			10	2	3		4	3	5	5	32
18											0
19		10	9	1		10		1	6	3	40
20					7		10	7			24
21	8	5		2	3			2	3	1	24
22	2	5	10	6		8	4		6	1	42
23				1	5	2	6	4	4	1	23
24					5			5		6	16
Total	41	79	81	37	73	80	80	70	73	42	656
Mean horizon	12.2	11.3	11.8	12.7	12.8	11.3	12.2	12.7	12.1	13.5	12.1

Note. Horizon 1 signifies that the forecast was published one month before publication of outcome, while Horizon 24 signifies that the forecast was published two years before outcome became known. The information in the other columns indicates the number of forecasts each institution made at each horizon. For example, in the years under analysis, the Ministry of Finance made one forecast in December of the year referred to by the forecast, and 41 forecasts in total. All forecasters published a total of 40 forecasts for the forecast year in December of that same year. Furthermore, it can be noted that the Ministry of Finance, on average, published its forecasts 12.2 months before the end of the forecast year. The equivalent figure for the entire body of data is 12.1 months. The column "Total" indicates that the 656 forecasts are spread relatively evenly across the year, with the exception of June.

Table 3: Estimation result

	GDP	CPI	Unemployment
Horizon	0.111	0.036	0.028
	(10.3)**	(19.3)**	(16.8)**
Horizon^2	–0.002	– – –	– – –
	(–5.7)**		
Horizon^3	– – –	– – –	– – –
Time effects	Yes**	Yes**	Yes**
Forecaster effects	Yes	Yes*	Yes*
R^2	0.88	0.73	0.74
Number of observations	656	656	654

Note. The upper part of the table presents the estimated coefficients for the horizon components of equation (2). The *t*-value (based on White's robust estimations of standard error) for the estimated coefficients are presented in parentheses. ** indicates that the parameter, or effect, is significantly different from zero at the one-percent level and * that it is statistically significant at the five-percent level.

TABLE 4: PAIRED SIGNIFICANCE TESTS OF THE ABILITY OF THE RIKSBANK AND OTHER FORECASTERS

	GDP	CPI	Unemployment
	Panel 1: 1999–2008		
RB vs FD	0.47	0.53	0.42
RB vs HUI	**0.05**	**0.02**	0.48
RB vs KI	0.27	0.79	0.90
RB vs LO	0.38	0.29	0.14
RB vs NORDEA	0.59	**0.03**	**0.06**
RB vs SEB	0.47	0.20	0.88
RB vs SHB	**0.02**	**0.08**	**0.02**
RB vs SN	**0.06**	**0.00**	0.14
RB vs SWED	**0.09**	0.29	0.70
	Panel 2: 2008		
RB vs FD	0.37	0.41	**0.01**
RB vs HUI	0.15	0.35	0.70
RB vs KI	**0.04**	0.19	**0.03**
RB vs LO	0.53	**0.03**	**0.05**
RB vs NORDEA	0.87	**0.04**	0.35
RB vs SEB	0.24	0.53	0.13
RB vs SHB	0.44	0.57	0.41
RB vs SN	0.90	0.12	**0.02**
RB vs SWED	0.18	0.37	0.31

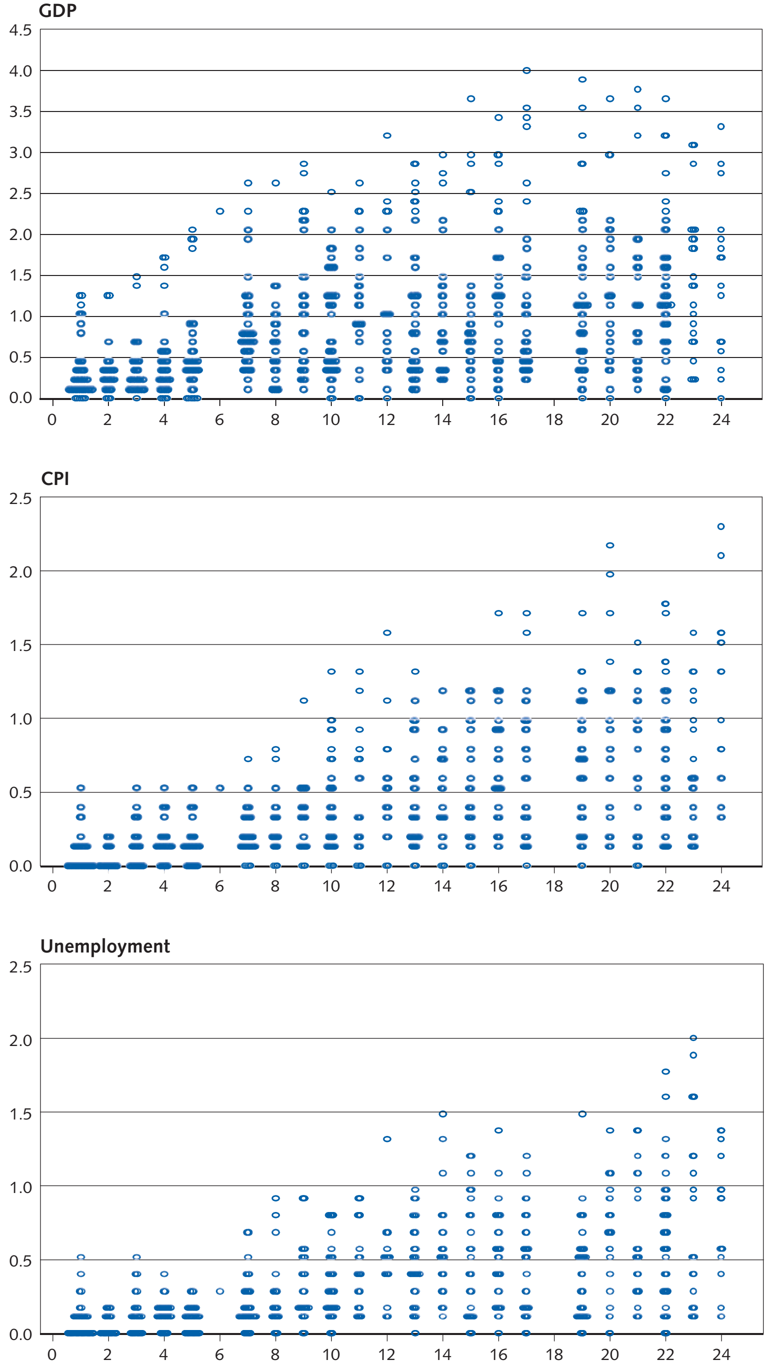
Note. The table presents the p-values from a test of the null-hypothesis that the Riksbank's ability is equivalent to the ability of the other forecaster, against the alternative that the Riksbank's ability is superior. A p-value lower than 0.1 (at the selected significance level of 10 per cent) indicates that the Riksbank has made statistically proven more precise forecasts (see the figures in bold).

TABLE 5: RANKING BASED ON FORECASTING ABILITY AND MAE, 1999–2008 AND 2008.

	GDP		CPI		Unemployment		Mean ranking	
	Ability	MAE	Ability	MAE	Ability	MAE	Ability	MAE
				Panel 1: 1999–2008				
FD	4	6	2	3	6	6	4.0	5.0
HUI	7	5	8	5	5	3	6.7	4.3
KI	6	4	1	1	1	1	2.7	2.0
LO	5	7	5	6	9	8	6.3	7.0
NORDEA	1	2	9	9	8	9	6.0	6.7
RB	2	1	3	2	4	4	3.0	2.3
SEB	3	3	6	4	2	2	3.7	3.0
SHB	10	10	7	8	10	10	9.0	9.3
SN	8	8	10	10	7	7	8.3	8.3
SWED	9	9	4	7	3	5	5.3	7.0
				Panel 2: 2008				
FD	6	4	4	2	9	5	6.3	3.7
HUI	9	8	5	5	1	2	5.0	5.0
KI	10	10	7	6	8	6	8.3	7.3
LO	3	5	9	9	7	7	6.3	7.0
NORDEA	2	2	10	10	4	4	5.3	5.3
RB	4	3	3	1	2	1	3.0	1.7
SEB	7	7	2	4	6	8	5.0	6.3
SHB	5	6	1	3	3	3	3.0	4.0
SN	1	1	8	7	10	10	6.3	6.0
SWED	8	9	6	8	5	9	6.3	8.7

Note. Ability is an estimated individual effect according to Equation (2) and MAE is a mean absolute error. The mean ranking is calculated as the mean value of each forecaster's ranking for the three variables of GDP, CPI and unemployment. The rankings (based upon ability according to our assessment and MAE) are separated in order to allow comparisons between both approaches. Comparisons are presented for 1999–2008 (panel 1) and for 2008 (panel 2).

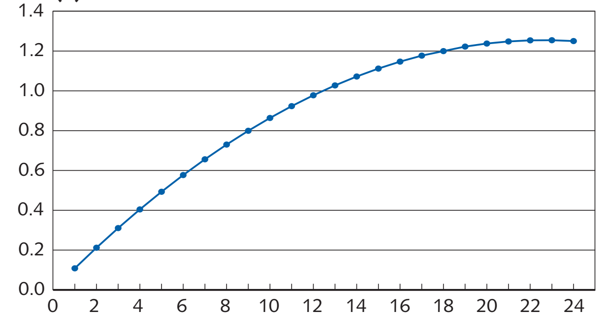
Figure 1. Forecast error (in absolute figures) for various forecast horizons, percentage points



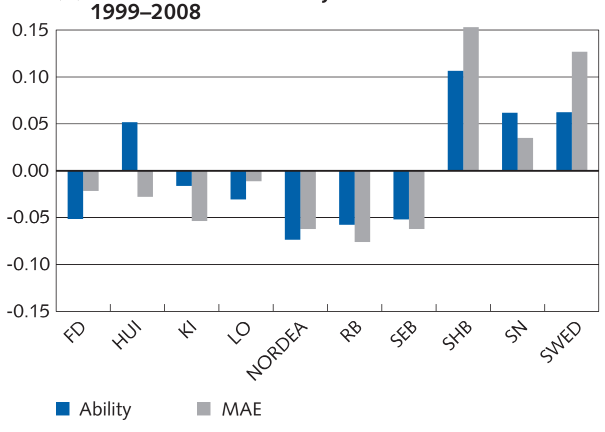
Note. The figure presents all forecasters' absolute forecast errors for GDP, CPI and unemployment in relation to the horizon. Darker and wider points indicate those cases in which there is more than one observation.

Figure 2. Result of model estimates, GDP

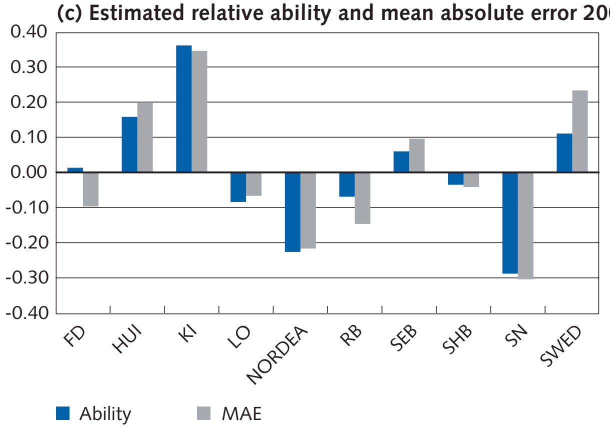
(a) Horizon effect



(b) Estimated relative ability and mean absolute error 1999–2008



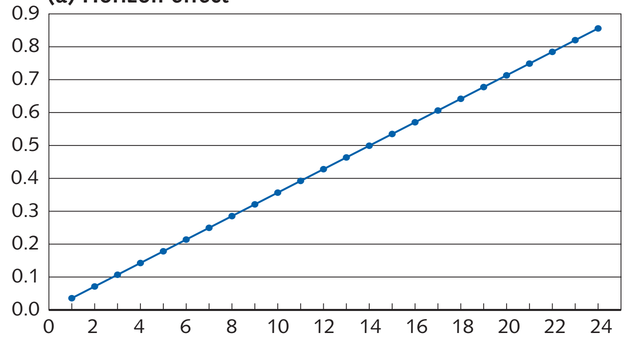
(c) Estimated relative ability and mean absolute error 2008



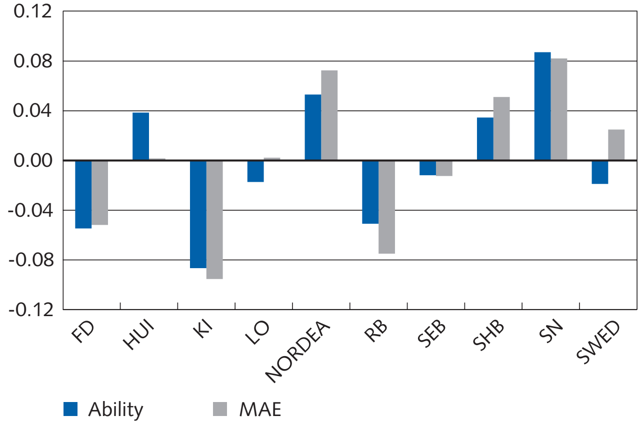
Note. Panel (a) in the figure indicates the marginal effect of the horizon, while panel (b) shows the estimated ability of each forecaster for the entire sample and panel (c) shows the corresponding quantity for 2008. The estimated ability in the figures is adjusted for the average ability of all forecasters. The zero line can thus be interpreted as the average ability.

Figure 3. Result of model estimates, CPI

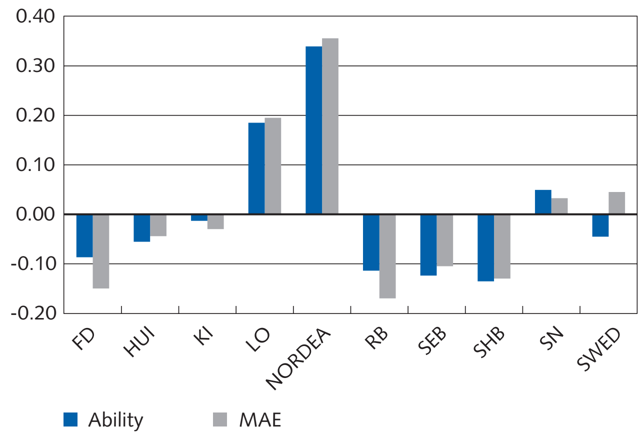
(a) Horizon effect



(b) Estimated relative ability and mean absolute error 1999–2008



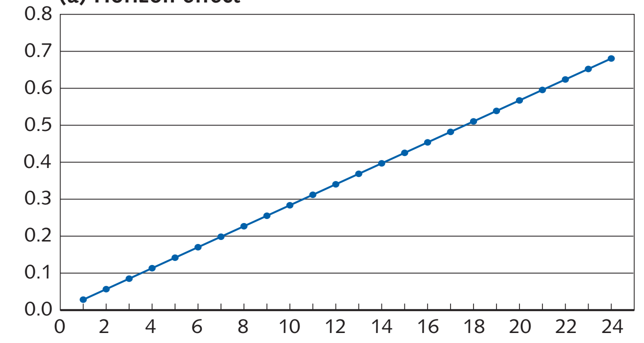
(c) Estimated relative ability and mean absolute error 2008



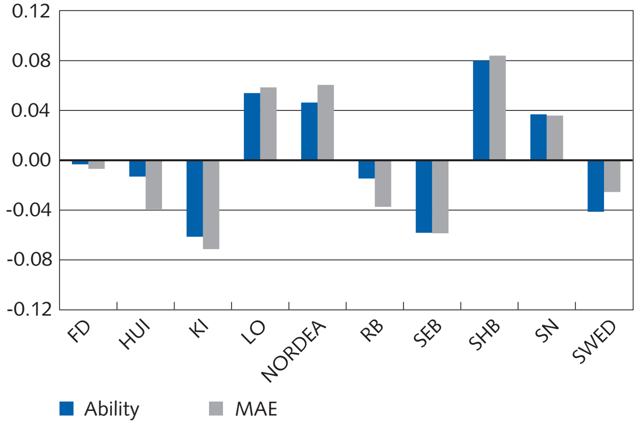
Note. See note to Figure 2.

Figur 4. Result of model estimates, unemployment

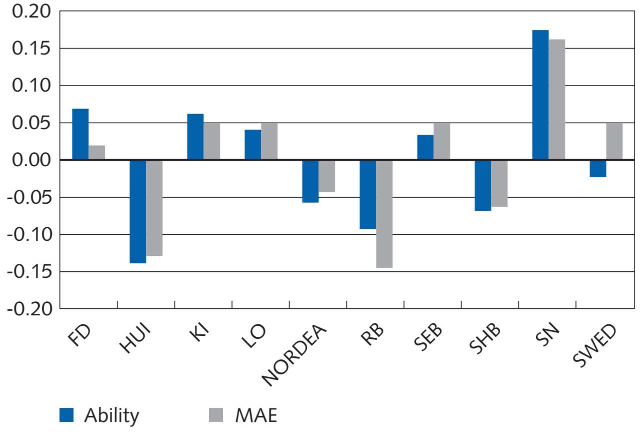
(a) Horizon effect



(b) Estimated relative ability and mean absolute error 1999–2008



(c) Estimated relative ability and mean absolute error 2008



Note. See note to Figure 2.

Appendix: The importance to the forecast error of the quantity of information

It is easily understandable that forecasts made at a point in time close to the outcome date will, on average, be more accurate than those made earlier.[21] The most important reason for this is that more 'components' of the year's outcome become known as time passes. In this appendix, we present an example showing the manner in which the annual growth rate depends on quarterly growth rates. In addition, we demonstrate the fashion in which a forecast's dependence on information on the outcome of the forecast variable (and the time aspect of this), as well as other information, can be utilised to improve (average) forecast precision. We call these three effects outcome effect, own effect and information effect, respectively. Together, these effects provide an understanding of how the distance to outcome impacts the expected forecast errors.[22]

In order to be able to understand how the information included in quarterly outcomes impacts the precision of annual forecasts, we start by studying the manner in which annual changes and quarterly changes are related. The annual percentage change in a variable is defined as the change in the variable in relation to that variable's value in the same period of the previous year. For a variable measured per quarter[23] , this means that the annual percentage change during the year's first quarter can be divided up into the four most recent quarters' percentage change as follows

$$\text{(A1)} \quad \begin{aligned} \frac{y_{T,1}-y_{T-1,1}}{y_{T-1,1}} &= \frac{y_{T,1}}{y_{T-1,1}}-1 \\ &= \frac{y_{T,1}}{y_{T-1,4}}\times\frac{y_{T`-1,4}}{y_{T-1,3}}\times\frac{y_{T`-1,3}}{y_{T-1,2}}\times\frac{y_{T`-1,2}}{y_{T-1,1}}-1, \end{aligned}$$

in which $y_{T,q}$ is the level of the variable y in quarter q of the year T. Let $g^4_{T,q}$ and $g^1_{T,q}$ be the annual and quarterly percentage change, respectively, for quarter q of year T. Equation (A1) can thereby be expressed as

$$\text{(A2)} \quad 1+g^4_{T,1} = (1+g^1_{T,1}) \times (1+g^1_{T-1,4}) \times (1+g^1_{T-1,3}) \times (1+g^1_{T-1,2}).$$

21 For example, one might imagine that it would be easy to forecast average unemployment rate (the level of unemployment as a percentage of the labour force) for one year if monthly outcome until the end of November is already known. In this case, only a forecast for December would need to be made, and this forecast would have a weight of 1/12. However, calculations become more complicated when a variable is measured by growth rates.

22 Here, we will only discuss expected, or average, forecast errors. In reality, the actual forecast errors deviate from the expected errors, for reasons including the exposure of the economy to (unexpected) disruptions.

23 Equivalent calculations can easily be generalised for arbitrary data frequencies.

The annual percentage change during quarters 2, 3 and 4 are calculated in the same fashion as equation (A2). In this case, yearly growth for the year T, $\Delta_4 y_T$, can be stated as the average of the annual percentage change in the year's various quarters, as follows

$$\text{(A3)} \qquad \Delta_4 y_T = \left(\frac{1}{4} \sum_{q=1}^{4} 1 + g_{T,q}^{4} \right) - 1$$

Equation (A3) holds exactly for CPI as measured by Statistics Sweden, and the equation holds approximately for GDP.[24] If we use Equation (A2) in Equation (A3), we obtain the following relation between quarterly growth and annual growth.

$$\text{(A4)} \qquad \begin{aligned} \Delta_4 y_T = \frac{1}{4}\Big[& \left(1+g_{T,1}^{1}\right)\times\left(1+g_{T-1,4}^{1}\right)\times\left(1+g_{T-1,3}^{1}\right)\times\left(1+g_{T-1,2}^{1}\right)+ \\ & \left(1+g_{T,2}^{1}\right)\times\left(1+g_{T,1}^{1}\right)\times\left(1+g_{T-1,4}^{1}\right)\times\left(1+g_{T-1,3}^{1}\right)+ \\ & \left(1+g_{T,3}^{1}\right)\times\left(1+g_{T,2}^{1}\right)\times\left(1+g_{T,1}^{1}\right)\times\left(1+g_{T-1,4}^{1}\right)+ \\ & \left(1+g_{T,4}^{1}\right)\times\left(1+g_{T,3}^{1}\right)\times\left(1+g_{T,2}^{1}\right)\times\left(1+g_{T,1}^{1}\right)\Big]-1. \end{aligned}$$

Equation (A4) indicates that annual growth is a function of all quarterly growth from quarter 2 of the previous year until the final quarter of the year to which the calculation applies. The equation also indicates that the opening quarter of the forecast year weighs differently in terms of yearly growth. The different weights are presented in Table A1.

Table A1: The relative importance of each quarter in annual average calculations.

Year		T-1			T		
Quarter	2	3	4	1	2	3	4
Weight	1/16	2/16	3/16	4/16	3/16	2/16	1/16
Accum	1/16	3/16	6/16	10/16	13/16	15/16	16/16

Note. The table indicates the weight of each quarter in the yearly calculation, together with the proportion of the full year value (accumulated) known at each point in time. See also equation (A4).

Table A1 reveals that those forecasts based on information from the period lasting until the end of quarter 1 of the year prior to that referred to by the forecast do not include any information pertaining to outcome. One quarter later, which is to say in quarter 2 of the previous year, 1/16 of the outcome is known. By the point in time at which growth up to and

[24] Yearly growth in GDP is defined as the total of quarterly levels of the year T divided by the total of the quarterly levels of T-1.

including quarter 1 of the forecast year is known, 10/16 of the yearly outcome is known to the forecaster. Note that these calculations are stylised and do not consider current revisions of quarterly outcome.[25] We have, of course, observed that the closer to the date of full year outcome that the forecast is made, the more of the approaching outcome is known. The phenomenon in which an increasingly minor portion of outcome needs to be forecast as the horizon decreases is here designated the *outcome effect.*

In addition to outcome effect, there exists further reason as to why forecast error ought to decrease the closer to outcome date the forecast is made. Assume that variable y of the quarterly change develops as follows

$$\Delta y_t = \mu_y + \alpha \Delta y_{t-1} + z_t + \varepsilon_t. \tag{A5}$$

Equation (A5) indicates that the outcome of quarterly growth depends on quarterly growth in the previous period and other information. This other information is symbolised by the variable z_t and is assumed to be unknown in period t. The variable ε_t is randomly distributed and represents disturbances to variable y. Furthermore, we assume that the other information will develop according to

$$z_t = \mu_z + \beta z_{t-1} + \eta_t. \tag{A6}$$

If we replace z_t in Equation (A5) with Equation (A6), we obtain the following relationship

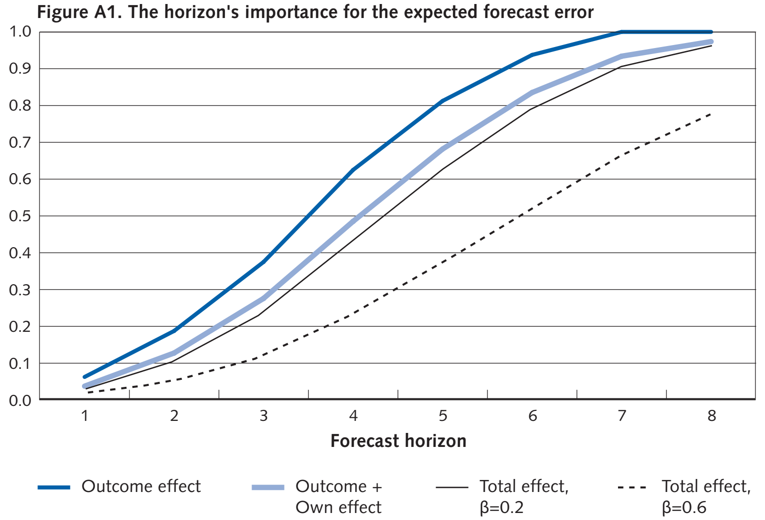
$$\begin{aligned} \Delta y_t &= \mu_y + \mu_z + \alpha \Delta y_{t-1} + \beta z_{t-1} + \varepsilon_t + \eta_t \\ &= \mu + \alpha \Delta y_{t-1} + \beta z_{t-1} + \xi_t \end{aligned} \tag{A7}$$

Equation (A7) provides a simplified description of the information held by the forecaster on each forecasting occasion and the manner in which this information is utilised. As a new outcome for Δy becomes available, this information is utilised to make forecasts. Parameter α (persistence of Δy) determines how far into the future the new outcome will remain important. This effect is known here as the *own effect*. Similar reasoning can be applied to the other information, z: when new outcome for z is registered, this can be utilised for forecasting purposes, with the period of time for which z will remain usable depending upon the parameter β. This effect is known here as the *information effect*.

[25] CPI is not normally revised, while GDP observations are revised back in time when a new outcome is published. Any seasonal adjustments made also lead to revision of historical GDP observations.

Given these three effects, it can be calculated how the anticipated forecast errors will develop as the time remaining until outcome is known decreases. We here demonstrate a calculation in which α is 0.4 and β is 0.2 and 0.6, respectively.[26] The forecast error arising before any information on outcome is known, i.e. quarter 1 of the year before the outcome year, has been normed as 1. The expected forecast error (disregarding ε_t and η_t) is presented in Figure A1 as a function of the horizon. The observed forecast error will deviate from the expected forecast error due to the disturbances ε_t and η_t, but, on average, the forecast error complies with that presented in Figure A1.

Figure A1. The horizon's importance for the expected forecast error



Note. The black line indicates the extent to which the expected forecast error depends on the outcome effect, the blue line indicates the extent to which the forecast error depends on the outcome and own effects and the two black lines (dashed and dotted) indicate the extent to which the forecast error depends on all effects, where β is 0.2 and 0.6, respectively.

Figure A1 demonstrates that the function form of the expected forecast error depends on the outcome effect and the parameters in equation (A7), which are unknown. In addition, possible revisions of data also contribute to the circumstance that the functional form, in practice, is unknown. In this study, we have decided to proceed from a flexible cubic function as an approximation of the manner in which the expected absolute forecast error declines as the horizon decreases. However, in the estimation, we test the significance of the horizon, removing from the specification those parts of the trend not significantly impacting the forecast error.

[26] The value of α is assessed with the aid of GDP data from 1980 with a dummy variable for the years 1991–93. In contrast, the values of β have been arbitrarily selected. The selected coefficients mean that the expected forecast error coincides with the expected *absolute* forecast error.