



Forecast accuracy and product differentiation of Japanese Institutional Forecasters

Ashiya, Masahiro

(Citation)

International Journal of Forecasting, 22(2):395-401

(Issue Date)

2006-04

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90000268>



Forecast Accuracy and Product Differentiation of Japanese Institutional Forecasters

Masahiro ASHIYA ⁺

July 2005

JEL Classification Codes: E37; C53; E17.

Keywords: Economic forecasts; Forecast evaluation; Forecast accuracy.

⁺ Faculty of Economics, Kobe University, 2-1 Rokko-dai, Nada, Kobe, 657-8501, Japan;

E-mail: ashiya@econ.kobe-u.ac.jp

Forecast Accuracy and Product Differentiation of Japanese Institutional Forecasters

This paper investigates whether some forecasters consistently outperform others using real GDP forecast data of 53 Japanese institutions over the past 24 years. It finds that the accuracy rankings are not significantly different from those that might be expected when all institutions had equal forecasting ability. On the other hand, their rankings of the relative forecast levels are significantly different from a random one. These results suggest that the macroeconomic forecasting business is competitive and each institution chooses the degree of “product differentiation” of its forecast so that accuracy and publicity are optimally balanced.

JEL Classification Codes: E37; C53; E17.

Keywords: Economic forecasts; Forecast evaluation; Forecast accuracy.

1. Introduction

It is a regular year-end event in Japan that institutional forecasters release their economic forecasts for the ongoing fiscal year and for the next fiscal year. Some of these forecasts turn out to be accurate, and others not (Figure 1 shows their forecasts and the realizations). A natural question is whether there are institutions that consistently outperform others. This is the first issue we address using the track records of real GDP forecasts over the past 24 years.

Our study is unique in the following three aspects. Firstly, it is the first attempt to examine the relative accuracy among the Japanese forecasters. Secondly, the unbalanced panel data contains more than 50 forecasters. Thirdly, the data is long enough to cover both prosperous periods (the 1980s) and recessionary periods (the 1990s) of the Japanese economy.

We select the data of 53 institutions, which participated in ten or more surveys (See Section 2 for details). Forecast accuracy varies considerably among these institutions. Table 1 shows the descriptive statistics of the mean absolute forecast error (MAFE) of each institution for the current fiscal year and for the next fiscal year. As for the current-year forecast, the MAFE of the best institution is 0.373%, and that of the worst institution is 0.700%. The average MAFE is 0.529% and the standard deviation is 0.069. As for the year-ahead forecast, the best is 1.110% and the worst is 1.833%.

Are these differences in forecast accuracy large enough to suggest that there are differences in forecasting ability among these institutions? To answer this question, we must take into account that some years are more difficult to forecast than others. The variance of the absolute forecast errors tends to be larger in these difficult-to-forecast years. It follows that the level of MAFE is mainly determined by the performance in the difficult-to-forecast years. Therefore MAFE is not an appropriate measure of forecast accuracy.

We consider the accuracy ranking of the institutions instead. Following Kolb and Stekler (1996), we employ the non-parametric test of ranking developed by Skillings and Mack (1981), which is robust to changes in the variance of the forecast variables. The result in Section 3 shows that the accuracy rankings are not significantly different from those that might be expected when all institutions had equal forecasting ability. Namely we cannot reject the hypothesis that all forecasters are equal. Our result

strengthens those of Batchelor (1990), Batchelor and Dua (1990a, b), and Kolb and Stekler (1996), since our data cover 24 years while their data contain at most 11 years.

Next we investigate whether some institutions are consistently more optimistic (or more pessimistic) than others. When the institutions are ranked according to the relative levels of their forecasts, we find this ranking significantly different from a random one. Furthermore, the optimism/pessimism ranking of the year-ahead forecasts was relatively stable over the sample period that covers four business cycles (Batchelor and Dua (1990b) examine the Blue Chip forecasters for 11 years and find similar results).

A possible explanation of these results is that institutional forecasters pursue both accuracy and publicity (Batchelor & Dua, 1990b; Laster et al., 1999), and that the macroeconomic forecasting industry is competitive. Forecasters have incentives to be extreme in order to generate publicity, but inaccurate forecasters would be driven out of business by the competitive pressure. Thus they would be constantly extreme if and only if being extreme does not worsen forecast accuracy significantly. Section 5 tests this hypothesis, and finds positive evidence. As for the current-year forecasts, the extreme forecasters are inaccurate and the optimism/pessimism ranking is not stable. As for the year-ahead forecasts, the extremists are as accurate as the moderates and the optimism/pessimism ranking is stable over the sample period.

The paper is organized as follows. Section 2 explains the data. Section 3 evaluates the variance of forecast accuracy, and Section 4 evaluates the variance of forecast level. Discussions are in Section 5, and Section 6 concludes the paper.

2. Data

Toyo Keizai Inc. has published the forecasts of about 70 Japanese institutions in the February or March issue of “Monthly Statistics (Tokei Geppo)” since the 1970s (Ashiya (2003, 2005, forthcoming) also uses this data). In every December, institution i releases forecasts of the Japanese real GDP growth rate for the ongoing fiscal year and for the next fiscal year. We call the former $f_{t,t}^i$ and the latter $f_{t,t+1}^i$. For example, the February 2004 issue contains forecasts for fiscal year 2003 (from April 2003 to March 2004) and for fiscal year 2004 (from April 2004 to March 2005). We treat the former as $f_{2003,2003}^i$ and the latter as $f_{2003,2004}^i$.

Since the participation rate was very low throughout the 1970s (on average 13.8 institutions per year), we use the forecasts published from February 1981 on. That is, we use $f_{t,t}^i$ for the fiscal years 1980 through 2003 and $f_{t,t+1}^i$ for the fiscal years 1981 through 2003. We exclude institutions that participated in less than 10 surveys, leaving 53 institutions. The average number of observations per institution is 18.42 for current-year forecasts ($f_{t,t}^i$) and 18.21 for year-ahead forecasts ($f_{t,t+1}^i$).

As for the actual growth rate g_t , Keane and Runkle (1990) argue that the revised data introduces a systematic bias because the extent of revision is unpredictable for the forecasters (See also Stark and Croushore, 2002). For this reason, we use the initial announcement of the Japanese government usually released in June (We obtain the same results by using the revised data of g_t released in June of year $t + 2$).

Figure 1 shows the forecast distributions and the actual growth rates. The vertical line shows the support of the forecast distribution of the 53 institutions. The closed diamond shows the mean forecast. The open square shows the actual growth rate.

3. Forecast accuracy

To test whether all institutional forecasters had equal forecasting ability, we consider the accuracy *ranking* of the institutions. Skillings and Mack (1981) generalize Friedman's (1937) distribution-free test and develop the following non-parametric test applicable to unbalanced panels.

Suppose the panel data consists of N institutions and M periods. Let $N_t (\leq N)$ be the number of institutions that release forecasts in year t . Let $r_t^i \in \{1, \dots, N_t\}$ be the rank of the absolute forecast error of institution i in year t . If ties occur, we use average ranks. If institution i does not participate in year t , we assume $r_t^i = 0.5(1 + N_t)$. We define the adjusted rank of institution i in year t , A_t^i , as

$$A_t^i \equiv \left(\frac{12}{1 + N_t} \right)^{0.5} \left(r_t^i - \frac{1 + N_t}{2} \right). \quad (1)$$

The first term of A_t^i compensates for the difference in observations. The second term measures relative performance. A negative (positive) A_t^i indicates that the rank of

institution i in year t is above (below) the median. $A^i \equiv \sum_{t=1}^M A_t^i$ denotes the sum of the adjusted ranks. If A^i is close to zero, forecast accuracy of institution i is on average similar to other institutions. If A^i is significantly smaller (larger) than zero, the forecast accuracy of institution i is on average better (worse) than that of other institutions. Let $A \equiv (A^1, \dots, A^N)'$.

Next we consider the covariance matrix, V , of the random vector A . Define m_{ij} as the number of years containing forecasts from institutions i and j . Then the elements of V , σ_{ij} , are defined as

$$\sigma_{ij} = \begin{cases} -m_{ij} & \text{if } i \neq j \\ \sum_{k \neq i} m_{ik} & \text{if } i = j \end{cases}$$

Let V_{11} be the upper left $N-1$ by $N-1$ submatrix of V , and let V_{11}^{-1} be the inverse of V_{11} . Define $\hat{A} \equiv (A^1, \dots, A^{N-1})'$. Skillings and Mack (1981) show that, under the null hypothesis that there is no difference in forecasting ability, the statistic

$$S \equiv \hat{A}' V_{11}^{-1} \hat{A} \quad (2)$$

has an asymptotic chi-squared distribution with $N-1$ degrees of freedom. A significantly large S indicates that institutions were not equal in their forecasting abilities.

The first row of Table 2 presents the values of the S -statistic calculated from the absolute forecast errors of 53 institutions over 24 years. We obtain $S = 48.28$ for the current-year forecasts and $S = 40.87$ for the year-ahead forecasts. Since neither statistic is significant at the 0.10 level (their P -values are larger than 0.60), we find no significant difference in the forecasters' abilities.

To confirm the above result, we estimate the following fixed-effects model used by O'Brien (1990). Let $AFE_{t,t}^i \equiv |f_{t,t}^i - g_t|$ be the absolute forecast error of the current-year forecast made by institution i in year t . $dum^i(j)$ ($j = 1, \dots, 52$) denotes the individual dummy and

$$dum^i(j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$dum_s(s)$ ($s = 1980, \dots, 2002$) denotes the year dummy and

$$dum_t(s) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}.$$

The regression we consider is

$$AFE_{t,t}^i = \alpha + \sum_{j=1}^{52} \beta_j \cdot dum^i(j) + \sum_{s=1980}^{2002} \gamma_s \cdot dum_t(s) + u_{t,t}^i. \quad (3)$$

If β_j is significantly smaller (larger) than zero, the absolute forecast error of institution j is smaller (larger) than that of other institutions conditional on year effects. The null hypothesis is $\beta_1 = \dots = \beta_{52} = 0$, i.e., institutions are homogeneous in average forecast accuracy.

The second row of Table 2 presents the result of the F -test on the coefficients of the individual dummies of equation (3). It shows that the coefficient of the individual effect is not significant at the 0.10 level. Hence, there is no evidence that institutions differ systematically in forecast accuracy.

4. Forecast level

This section examines whether some institutions consistently release optimistic (or pessimistic) forecasts. To address this question, the observed distribution of the level of their forecasts is compared with the distribution expected if their relative forecast levels each year were purely random.

First, we employ the ranking-based test of Skillings and Mack (1981), explained in Section 3. The first row of Table 3 shows the values of the S -statistic, which was defined in equation (2). We obtain $S = 81.39$ for the current-year forecasts and $S = 196.14$ for the year-ahead forecasts, both of which are significant at the 0.01 level. It indicates that some forecasters were relatively optimistic while others were relatively pessimistic during the sample period.

We obtain the same result when we consider the following fixed-effects model:

$$f_{t,t}^i = \alpha + \sum_{j=1}^{52} \beta_j \cdot dum^i(j) + \sum_{s=1980}^{2002} \gamma_s \cdot dum_t(s) + u_{t,t}^i. \quad (4)$$

The null hypothesis is $\beta_1 = \dots = \beta_{52} = 0$, i.e., institutions are homogeneous in their average forecast level conditional on year effects. The second row of Table 3 shows the results of the F -test. The null hypothesis is clearly rejected for both the current-year and the year-ahead forecasts.

Next we analyze whether this optimism/pessimism relationship is stable over the sample period. Since our data covers both prosperous periods (the 1980s) and recessionary periods (the 1990s), we divide the data into two sub-samples, 1980-1991 and 1992-2003. Let $A_t^i(f_{t,t}^i)$ be the adjusted rank of forecast level of $f_{t,t}^i$ (defined in equation (1)). Then we calculate the Spearman Rank Correlation coefficient between (a) the average of $A_t^i(f_{t,t}^i)$ for each institution in the first half and (b) the average of $A_t^i(f_{t,t}^i)$ for each institution in the second half. We find that the rank correlation is -0.014 for the current-year forecasts and 0.293 for the year-ahead forecasts. Namely the optimism/pessimism relationship is found to be stable for the year-ahead forecasts only.

We also compare each institution's relative forecast level for the current-year forecast with that for the year-ahead forecast. Then the correlation between $A_t^i(f_{t,t}^i)$ and $A_t^i(f_{t,t+1}^i)$ is 0.605, which indicates that an institution that is optimistic (pessimistic) for the current year is also optimistic (pessimistic) for the next year.

5. Discussions

This section discusses the economic rationale behind the following results we have obtained:

- (a) no significant difference in the forecasters' abilities,
- (b) systematic differences in their forecast levels,
- (c) unstable optimism/pessimism relationships for the current-year forecasts, and
- (d) stable optimism/pessimism relationships for the year-ahead forecasts.

One possible explanation of these results is that the year-ahead forecasts are fairly arbitrary, and hence the forecasters stick to their favorite positions (I thank an anonymous referee for this suggestion). Real decreases in the forecast error start with the forecasts made early in the target year. As the forecast error decreases, the optimism/pessimism bias becomes insignificant. This might be the reason why stable optimism/pessimism relationships disappear from the current-year forecasts.

Another explanation is that, as Batchelor and Dua (1990b) and Laster et al. (1999) suggest, institutional forecasters pursue two objectives: accuracy and publicity. Inaccurate forecasters will be driven out of business by the competitive pressure. On the other hand, an extreme forecast attracts public attention when it turns out to be true,

while a conventional forecast never comes to public notice. Moreover, an extreme forecast will pick up the occasional extreme events (boom and recession), and this may be valued by users just as much as a low average error (I thank an anonymous referee for this suggestion). Therefore, forecasters would be constantly extreme (i.e. optimistic or pessimistic) if and only if forecast accuracy is not significantly impaired.

We test the latter hypothesis by investigating the relationship between accuracy and extremeness of forecasts. Define $A_t^i(AFE_{t,t}^i)$ as the adjusted rank of the absolute forecast error for $f_{t,t}^i$: a negative (positive) $A_t^i(AFE_{t,t}^i)$ indicates that the forecast accuracy of $f_{t,t}^i$ is better (worse) than that of other institutions. Next define $|A_t^i(f_{t,t}^i)|$ as the absolute value of the adjusted rank of the forecast level for $f_{t,t}^i$. $|A_t^i(f_{t,t}^i)|$ indicates the degree of extremeness of $f_{t,t}^i$: the larger $|A_t^i(f_{t,t}^i)|$ is, the more extreme $f_{t,t}^i$ is. Let $\overline{A_t^i(AFE_{t,t}^i)}$ and $\overline{|A_t^i(f_{t,t}^i)|}$ be the average of $A_t^i(AFE_{t,t}^i)$ and $|A_t^i(f_{t,t}^i)|$ for each institution.

We calculate the Spearman Rank Correlation coefficient between $\overline{A_t^i(AFE_{t,t}^i)}$ and $\overline{|A_t^i(f_{t,t}^i)|}$. Since the optimism/pessimism relationship is unstable over the sample period for the current-year forecasts, extreme forecasts would be sufficiently inaccurate for the current-year forecasts. That is, we should find *positive* correlation between $\overline{A_t^i(AFE_{t,t}^i)}$ and $\overline{|A_t^i(f_{t,t}^i)|}$. On the other hand, since the optimism/pessimism relationship is stable for the year-ahead forecasts, we should find *insignificant* correlation between $\overline{A_t^i(AFE_{t,t+1}^i)}$ and $\overline{|A_t^i(f_{t,t+1}^i)|}$.

The rank correlation between $\overline{A_t^i(AFE_{t,t}^i)}$ and $\overline{|A_t^i(f_{t,t}^i)|}$ turns out to be 0.455, and the rank correlation between $\overline{A_t^i(AFE_{t,t+1}^i)}$ and $\overline{|A_t^i(f_{t,t+1}^i)|}$ turns out to be -0.043. These figures are consistent with the “accuracy and publicity hypothesis”.

6. Conclusions

This paper analyzes the real GDP forecast record of 53 Japanese institutions over 24

years. First, we evaluate the relative forecast accuracy among these institutions. When performance is measured by the mean absolute forecast error, there appears to be a significant difference in accuracy among them (Table 1). The difference vanishes, however, when either a ranking-based test is employed or year effects are explicitly considered (Table 2). Namely, we cannot reject the hypothesis that all institutions are equal in forecasting ability. We next evaluate the relative forecast level among these institutions. Our data provide clear evidence that institutions differ systematically in their forecast levels (Table 3).

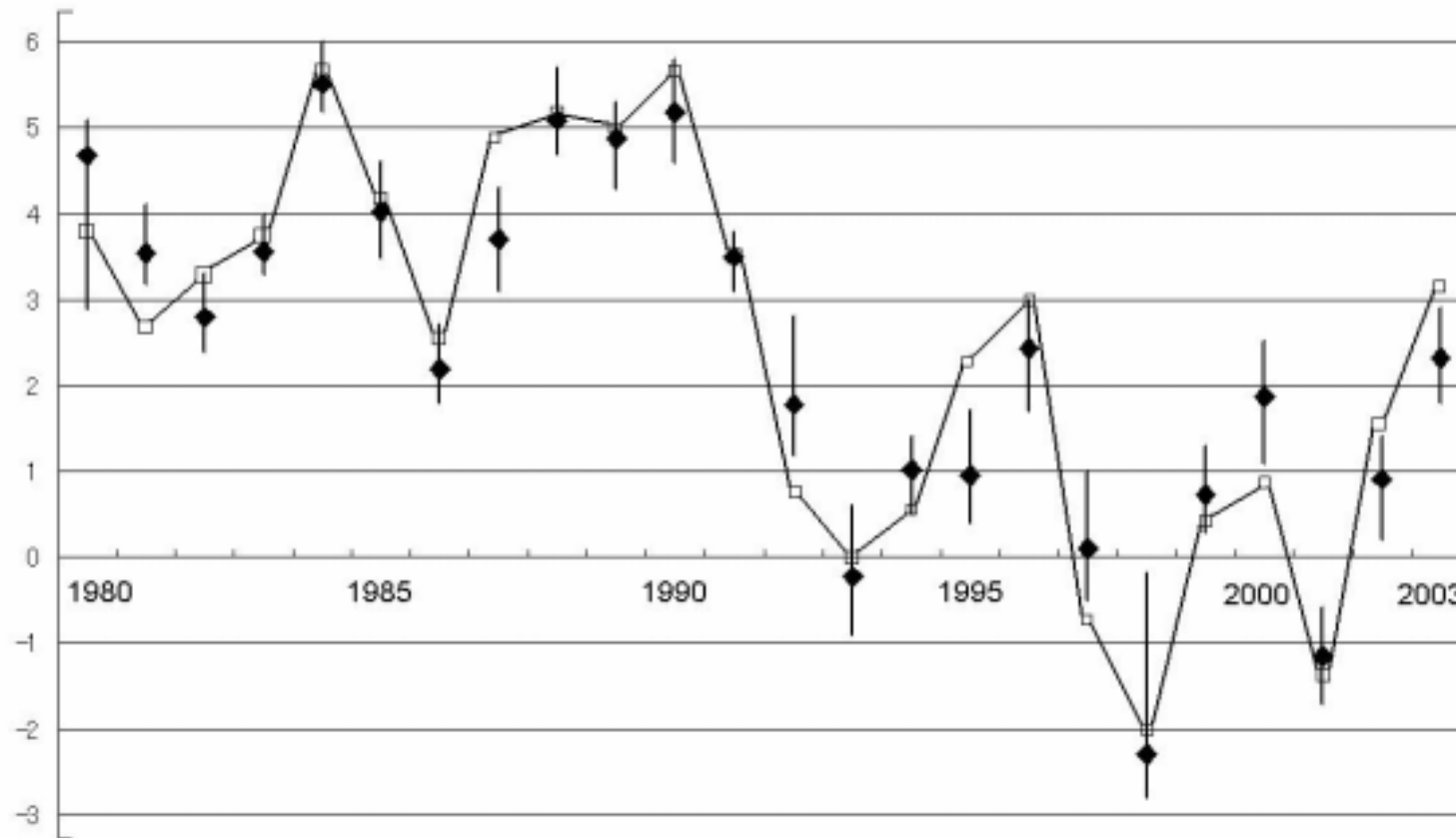
These findings are in accord with the notion that the forecasting industry is competitive and each institution seeks publicity. The competitive pressure eliminates inaccurate forecasters, and consequently all remaining institutions are equal in ability. These institutions try to draw public attention without compromising forecast accuracy by modestly differentiating their forecasts from others.

References

- Ashiya, M. (2003). Testing the rationality of Japanese GDP forecasts: the sign of forecast revision matters. *Journal of Economic Behavior and Organization*, 50, 263-269.
- Ashiya, M. (2005). Twenty-two years of Japanese institutional forecasts. *Applied Financial Economics Letters*, 1, 79-84.
- Ashiya, M. (forthcoming). Are 16-months-ahead forecasts useful?: a directional analysis of Japanese GDP forecasts. *Journal of Forecasting*.
- Batchelor, R. A. (1990). All forecasters are equal. *Journal of Business and Economic Statistics*, 8, 143-144.
- Batchelor, R. A., & Dua, P. (1990a). Forecaster ideology, forecasting technique, and the accuracy of economic forecasts. *International Journal of Forecasting*, 6, 3-10.
- Batchelor, R. A., & Dua, P. (1990b). Product differentiation in the economic forecasting industry. *International Journal of Forecasting*, 6, 311-316.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32, 675-701.
- Keane, M.P., & Runkle, D.E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, 80, 714-735.
- Kolb, R.A., & Stekler, H.O. (1996). How well do analysts forecast interest rates? *Journal of Forecasting*, 15, 385-394.
- Laster, D., Bennett, P., & Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *Quarterly Journal of Economics*, 114, 293-318.
- O'Brien, P. C. (1990). Forecast accuracy of individual analysts in nine industries. *Journal of Accounting Research*, 28, 286-304.
- Skillings, J. H., & Mack, G. A. (1981). On the use of a Friedman-type statistic in balanced and unbalanced block designs. *Technometrics*, 23, 171-177.
- Stark, T., & Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24, 507-531.

Figure 1: Forecast distributions and the actual growth rates

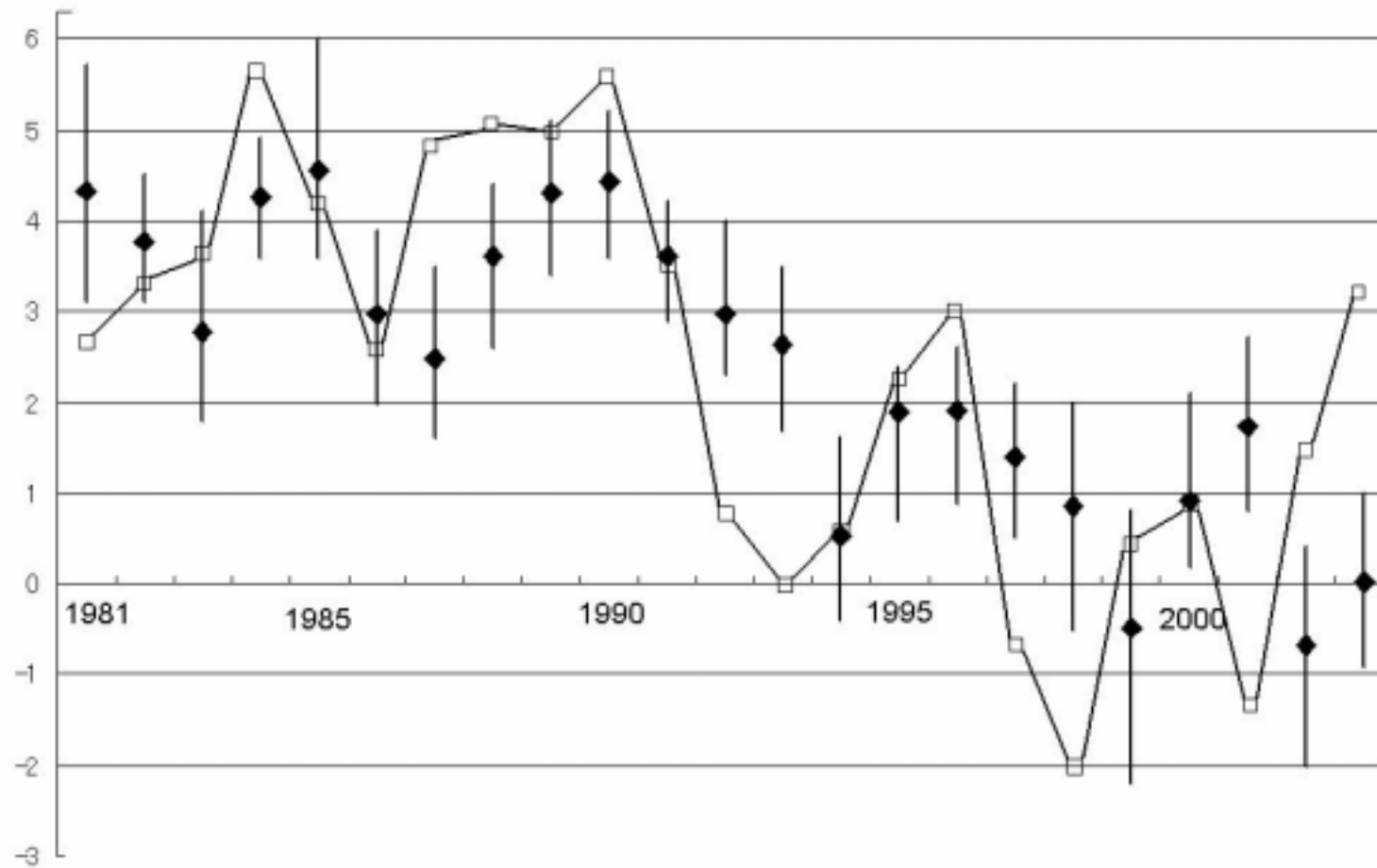
(a) Current-year forecasts



◆ (the closed diamond): the mean forecast, □ (the open square): the realization.

The vertical line shows the support of the forecast distribution of 53 forecasters.

(b) Year-ahead forecasts



◆ (the closed diamond): the mean forecast, □ (the open square): the realization.

The vertical line shows the support of the forecast distribution of 53 forecasters.

Table 1: Mean absolute forecast error of 53 institutions

	Current year	Year-ahead
Min.	0.373	1.110
Max.	0.700	1.833
Avg.	0.529	1.403
STD	0.069	0.144

Table 2: Variance tests for the absolute forecast error

	Current year	Year-ahead
<i>S</i> -statistic	48.28	40.87
(<i>P</i> -value)	(0.621)	(0.867)
<i>F</i> -test	1.163	0.813
(<i>P</i> -value)	(0.204)	(0.824)

Notes

The 95% (90%) critical value for the chi-squared distribution with 52 degrees of freedom is 69.83 (65.42).

Table 3: Variance tests for the forecast level

	Current year	Year-ahead
<i>S</i> -statistic	81.39	196.14
(<i>P</i> -value)	(0.006)	(0.000)
<i>F</i> -test	1.453	4.372
(<i>P</i> -value)	(0.022)	(0.000)

Notes

The 99% critical value for the chi-squared distribution with 52 degrees of freedom is 78.62.

Acknowledgements

The author is grateful for useful comments from the associate editor and two anonymous referees. The author is solely responsible for the contents.

Author's biography:

Masahiro ASHIYA is Associate Professor of Economics at Kobe University. He has published papers in various journals including Journal of Forecasting, Journal of Economic Behavior and Organization, International Journal of Industrial Organization, and Japan and the World Economy.