Testing homogeneity of Japanese CPI forecasters

Ashiya, Masahiro

(Citation)
Journal of Forecasting,29(5):435-441

(Issue Date)
2010-08

(Resource Type)
journal article

(Version)
Accepted Manuscript

Testing homogeneity of Japanese CPI forecasters*

Masahiro ASHIYA [+]

February 2009

JEL Classification Codes: E37; C53; E17.

Keywords: Macroeconomic Forecast; Forecast evaluation; Analysis of variance.

* I gratefully acknowledge financial supports from Grant-in-Aid for Encouragement of Young Scientists from the Japanese Ministry of Education, Science and Culture.

[+] Faculty of Economics, Kobe University, 2-1 Rokko-dai, Nada, Kobe, 657-8501, Japan;
E-mail: ashiya@econ.kobe-u.ac.jp

Testing homogeneity of Japanese CPI forecasters

February 2009

This paper investigates whether some forecasters consistently outperform others using Japanese CPI forecast data of 42 forecasters over the past 18 quarters. It finds that the accuracy rankings of zero, one, two, and five-month forecasts are significantly different from those that might be expected when all forecasters had equal forecasting ability. Moreover, their rankings of the relative forecast levels are also significantly different from a random one.

JEL Classification Codes: E37; C53; E17.

Keywords: Macroeconomic Forecast; Forecast evaluation; Analysis of variance.

1. Introduction

The world economy has gone through tumultuous changes over the past few years; the rise and fall of the commodity prices, and the boom and bust of the financial markets. These unprecedented shocks have made the task of macroeconomic forecasters formidable. As a result, many of them have failed to foresee the volatile fluctuation of output and prices. This experience leads us to the following question: Were all forecasters equally successful (or unsuccessful) in this period? Was there any significant difference in their forecast accuracy? We answer this question using the monthly data of 42 Japanese consumer price index (CPI) forecasters from April 2004 through August 2008.

Section 2 explains the data. Section 3 evaluates the homogeneity of forecast accuracy. It finds that forecasting ability is unequal among the forecasters. More precisely, the accuracy rankings of zero, one, two, and five-month forecasts are significantly different from those that might be expected when all forecasters had equal forecasting ability. This result is contrary to the findings of Batchelor (1990), Batchelor and Dua (1990a, b), Kolb and Stekler (1996), and Ashiya (2006).

Section 4 investigates the biases of the relative forecast level of the individual forecasters. It finds that the rankings of the relative forecast levels are significantly different from a random one for three, four, and five-month forecasts. Namely, forecasters differ systematically in their forecast levels. This result is consistent with the findings of Batchelor and Dua (1990b) and Ashiya (2006). Section 5 concludes the paper.

2. Data

The Economic Planning Association has conducted a monthly survey of professional forecasters, "ESP Forecast Survey," since April 2004. We use the forecast data of the consumer price index (CPI) through August 2008. We select the data of 42 forecasters (out of 44 forecasters), who participated in 18 surveys or more (the excluded forecasters participated in five surveys).

Let CPI_t be the CPI of month t. Then the rate of change over the year, p_t, is computed by the following equation:

$$p_t \equiv \frac{CPI_t - CPI_{t-12}}{CPI_{t-12}} \times 100 .$$

The quarterly average change over the year is calculated as the simple arithmetic mean of p_t. More specifically, the quarterly average change over the year from month $t-2$ to month t,

q_t, is defined as

$$q_t \equiv (p_{t-2} + p_{t-1} + p_t)/3 .$$

Let $f_{t-k,t}^{i}$ be the k-month-ahead forecast of forecaster i with respect to q_t, which is released in month $t-k$. The forecast error is defined as $FE_{t-k,t}^{i} \equiv f_{t-k,t}^{i} - q_t$. Its absolute value is $AFE_{t-k,t}^{i} \equiv \left| f_{t-k,t}^{i} - q_t \right|$.

We analyze zero through five-month-ahead forecasts in this paper. Zero-month forecasts and three-month forecasts are released in March, June, September, and December. One-month forecasts and four-month forecasts are released in February, May, August, and November. Two-month forecasts and five-month forecasts are released in January, April, July, and October. The sample period of each forecast series is as follows: from the first quarter of 2004 through the second quarter of 2008 (18 quarters) for zero-month forecast; from the second quarter of 2004 through the second quarter of 2008 (17 quarters) for one-month forecast; from the second quarter of 2004 through the third quarter of 2008 (18 quarters) for two-month forecast and three-month forecast; from the third quarter of 2004 through the third quarter of 2008 (17 quarters) for four-month forecast and five-month forecast.

Table 1 presents the values of several traditional measures of forecast accuracy for the individual forecasters. The first row in Table 1 shows the summary statistics (average, standard deviation, minimum, and maximum) of the mean absolute error (MAE). As for zero-month forecast, the average of the MAE among forecasters is 0.072 percentage points, and the MAE of the best forecaster is zero. The second row shows the summary statistics of the root mean square error (RMSE).

The third row of Table 1 shows the summary statistics of modified Theil's U, constructed as the ratio of the RMSE of each forecaster to the RMSE of the "same-as-the-last-month" forecast. More specifically, define $T_{t-k,t}^{i}$ as the set of quarters in which forecaster i released $f_{t-k,t}^{i}$. Let $U_{t-k,t}^{i}$ be the Theil's U of forecaster i for $f_{t-k,t}^{i}$. Then $U_{t-k,t}^{i}$ is defined as

$$U_{t-k,t}^{i} = \sqrt{\sum_{t \in T_{t-k,t}^{i}} \left(f_{t-k,t}^{i} - q_t \right)^2} \Big/ \sqrt{\sum_{t \in T_{t-k,t}^{i}} \left(p_{t-k-1} - q_t \right)^2} \quad \text{for} \quad k = 0,1,\cdots,5 .$$

If $U_{t-k,t}^{i} > 1$, then forecaster i is inferior to the "same-as-the-last-month" forecast. Table 1 shows that $U_{t-k,t}^{i}$ is on average 2.269 for zero-month forecast and 1.104 for one-month forecast.

These descriptive statistics seem to indicate that there are some differences in forecasting ability among the forecasters. However, they are not appropriate measures of forecast accuracy for the following reason. To evaluate the ability of the forecasters, we must take into account that some periods are more difficult to forecast than others. The variance of the forecast errors tends to be larger in these difficult-to-forecast periods. It follows that the level of the MAE (or the RMSE) is mainly determined by the performance in the difficult-to-forecast periods. The next section employs a ranking-based test to deal with this problem.

3. Tests for homogeneity of forecast accuracy

This section evaluates whether all forecasters had equal forecasting ability. Following Kolb and Stekler (1996), we first 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. Then we consider the alternative method of O'Brien (1990). Both methods produce the same result: forecasters are not equal in their forecasting abilities.

3.1 The methodology of the ranking test

To test whether all forecasters had equal forecasting ability, we consider the accuracy *ranking* of the forecasters. 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* forecasters and *M* quarters. Let N_t ($\le N$) be the number of forecasters that release forecasts in the *t*-th quarter ($t \in \{1, \cdots, M\}$). Let $r_t^i \in \{1, \cdots, N_t\}$ be the rank of the absolute forecast error of forecaster *i* in the *t*-th quarter. If ties occur, we use average ranks. If forecaster *i* does not participate in the *t*-th quarter, we assume $r_t^i = 0.5(1 + N_t)$.

We define the adjusted rank of forecaster *i* in the *t*-th quarter, A_t^i, as

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

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 forecaster *i* in the *t*-th quarter is above (below) the median.

The sum of the adjusted ranks,

$$A^i \equiv \sum_{t=1}^{M} A_t^i \tag{2}$$

indicates the relative performance over the sample period. If A^i is close to zero, forecast accuracy of forecaster *i* is on average similar to others. If A^i is significantly smaller (larger) than zero, the forecast accuracy of forecaster *i* is on average better (worse) than that of others.

Let $A \equiv \left(A^1, \cdots, A^N\right)'$. If A is significantly different from the zero vector, then it indicates that the forecasters vary in forecasting ability. To test this hypothesis, we consider the covariance matrix, *V*, of the random vector A. Define m_{ij} as the number of quarters containing forecasts from forecasters *i* and *j*. Then the elements of *V*, σ_{ij}, are defined as

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

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 $\hat{A} \equiv \left(A^1, \cdots, A^{N-1}\right)'$. Skillings and Mack (1981) show that under the null hypothesis there is no difference in forecasting ability, the statistic

$$S \equiv \hat{A}' V_{11}^{-1} \hat{A} \tag{3}$$

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

3.2 The results of the ranking test

Table 2 presents the values of the *S*-statistic calculated from the absolute forecast errors of 42 forecasters. The first row is the result of zero-month forecast. We obtain $S = 113.98$, which is significant at the 0.01 level (the *P*-value is in the second column). The second row shows that $S = 55.64$ for one-month forecast, which is significant at the 0.10 level. The third row shows that $S = 59.75$ for two-month forecast, which is significant at the 0.05 level. The sixth row shows that $S = 54.05$ for five-month forecast, which is significant at the 0.10 level.

These results demonstrate that the accuracy rankings of zero, one, two, and five-month forecasts are significantly different from those that might be expected when all forecasters had equal forecasting ability. Namely, we find that forecasting ability is unequal among the forecasters. This conclusion is in stark contrast to the results of Batchelor (1990), Batchelor and Dua (1990a, b), Kolb and Stekler (1996), and Ashiya (2006).

3.3 The regression test for homogeneity

To confirm the above result, we estimate the following fixed-effects model used by O'Brien (1990) and Ashiya (2006). Let $dum^i(j)$ ($j = 1, \cdots, 41$) be the individual dummy:

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

Let $dum_t(s)$ ($s = 1, \cdots, 17$) be the quarter dummy:

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

The regression we consider is

$$AFE^i_{t-k,t} = \alpha + \sum_{j=1}^{41} \beta_j \cdot dum^i(j) + \sum_{s=1}^{17} \gamma_s \cdot dum_t(s) + u^i_{t,t}. \quad (4)$$

If β_j is significantly smaller (larger) than zero, the absolute forecast error of forecaster *j* is smaller (larger) than that of others conditional on the quarter-specific effects. The null hypothesis is $\beta_1 = \cdots = \beta_{41} = 0$, i.e., forecasters are homogeneous in average forecast accuracy.

Table 3 presents the result of the *F*-test on the coefficients of the individual dummies of equation (4). It shows that the coefficient of the individual effect is significant at the 0.05 level for every forecast span (i.e. zero through five-month forecasts). Hence we have clear evidence that forecasters differ systematically in forecast accuracy.

4. The test for homogeneity of the forecast level

This section examines whether some forecasters consistently release extremely large (or extremely small) 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 quarter were purely random.

First, we employ the ranking-based test of Skillings and Mack (1981). Table 4 shows the values of the *S*-statistic calculated from equation (3). We find that the rankings of the relative forecast levels are significantly different from a random one for three, four, and five-month forecasts.

Next we consider the fixed-effects model of equation (4), substituting $f^i_{t-k,t}$ for $AFE^i_{t-k,t}$. Table 5 shows the results of the *F*-test on the coefficients of the individual dummies. We find that the coefficient of the individual effect is significant at the 0.05 level for two, three, four, and five-month forecasts. These results indicate that some forecasters tend to produce relatively large forecasts, while others tend to produce relatively small forecasts during the sample period.

5. Conclusions

This paper has used the monthly survey of 42 Japanese CPI forecasters from April 2004 to August 2008, and has tested the hypothesis that all forecasters are equal in forecasting ability. This hypothesis was rejected by the ranking test for zero, one, two, and five-month forecasts. Furthermore, the hypothesis was rejected by the panel regression for zero through five-month forecasts. This result presents a striking contrast to the past literature.

One qualification is that our result relies on the crucial assumption that forecasters aim to minimize their forecast errors. There are various reasons for rational forecasters to announce forecasts different from the conditional expected value. Ashiya (2009) finds that the Japanese GDP forecasters in industries that emphasize publicity tend to make less accurate but more extreme forecasts in order to gain publicity for their firms. Whether this “publicity effect” can explain our result is an important topic for future research.

References

Ashiya, M. (2006) “Forecast Accuracy and Product Differentiation of Japanese Institutional Forecasters.” International Journal of Forecasting, 22, 395-401.

Ashiya, M. (2009) “Strategic Bias and Professional Affiliations of Macroeconomic Forecasters.” Journal of Forecasting, 28, 120-130.

Batchelor. R.A. (1990) “All Forecasters Are Equal.” Journal of Business and Economic Statistics, 8, 143-144.

Batchelor, R.A., and Dua, P. (1990a) “Forecaster Ideology, Forecasting Technique, and the Accuracy of Economic Forecasts.” International Journal of Forecasting, 6, 3-10.

Batchelor, R.A., and 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.

Kolb, R.A., and Stekler, H.O. (1996) “How Well Do Analysts Forecast Interest Rates?” Journal of Forecasting, 15, 385-394.

O’Brien, P.C. (1990) “Forecast Accuracy of Individual Analysts in Nine Industries.” Journal of Accounting Research, 28(2), 286-304.

Skillings, J.H., and Mack, G.A. (1981) “On the Use of a Friedman-Type Statistic in Balanced and Unbalanced Block Designs.” Technometrics, 23, 171-177.

Table 1: The descriptive statistics

Zero-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.072	0.056	0.000	0.270
RMSE	0.108	0.071	0.000	0.383
$U^i_{t,t}$	2.269	1.522	0.000	8.573

One-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.107	0.054	0.033	0.325
RMSE	0.149	0.071	0.058	0.453
$U^i_{t-1,t}$	1.104	0.899	0.608	5.228

Two-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.184	0.072	0.067	0.471
RMSE	0.250	0.090	0.082	0.615
$U^i_{t-2,t}$	1.062	0.307	0.577	2.051

Three-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.225	0.061	0.100	0.341
RMSE	0.300	0.089	0.114	0.450
$U^i_{t-3,t}$	0.959	0.273	0.546	1.874

Four-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.268	0.075	0.083	0.380
RMSE	0.361	0.105	0.108	0.501
$U^i_{t-4,t}$	0.811	0.200	0.510	1.519

Five-month forecast	Average	Std. Dev.	Minimum	Maximum
MAE	0.362	0.122	0.133	0.817
RMSE	0.478	0.151	0.183	0.979
$U^i_{t-5,t}$	0.962	0.170	0.558	1.395

MAE: mean absolute error.
RMSE: root mean square error.
$U^i_{t-k,t} = \sqrt{\sum_{t \in T^i_{t-k,t}} \left(f^i_{t-k,t} - q_t\right)^2} \Big/ \sqrt{\sum_{t \in T^i_{t-k,t}} \left(p_{t-k-1} - q_t\right)^2}$ for $k = 0,1,\cdots,5$.

Table 2: The ranking test for homogeneity of the absolute forecast error

k	*S*	*P*-value
0	113.98	0.000 ***
1	55.64	0.063 *
2	59.75	0.029 **
3	47.62	0.221
4	43.00	0.386
5	54.05	0.083 *

***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 3: The regression test for homogeneity of the absolute forecast error

k	*F*-test	*P*-value
0	6.020	0.000 ***
1	3.326	0.000 ***
2	3.086	0.000 ***
3	1.946	0.001 ***
4	1.603	0.012 **
5	2.194	0.000 ***

***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 4: The ranking test for homogeneity of the forecast level

k	*S*	*P*-value
0	32.73	0.818
1	26.82	0.958
2	45.20	0.301
3	60.85	0.024 **
4	64.54	0.011 **
5	90.62	0.000 ***

***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 5: The regression test for homogeneity of the forecast level

k	*F*-test	*P*-value
0	0.886	0.675
1	0.877	0.690
2	1.360	0.071 *
3	1.352	0.075 *
4	1.697	0.005 ***
5	2.515	0.000 ***

***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.