What Have We Learned from a Survey of Japanese Professional Forecasters?

Taking Stock of Four Years of ESP Forecast Experience

Takao Komine [*]
Kanemi Ban [†]
Masaaki Kawagoe [‡]
Hiroshi Yoshida [§]

February 27, 2009

1 Introduction

A monthly survey of Japanese professional forecasters, the ESP Forecast ('ESPF' hereafter) [1], was launched in 2004. Although similar ad hoc surveys have been conducted in the past, the ESPF may have been the first regular publication to cover forecasts produced by business and academic economists. This initiative was roundly welcomed, and the survey has been established as a valuable information source – the Bank of Japan often presents the results in a figure in its Monthly Report of Recent Economic and Financial Developments, for example.

This paper will provide a bird's-eye view of a three-year research project funded by the Economic and Social Research Institute ('ESRI' hereafter) of the Cabinet Office. The ESRI encourages academics to study ESPF database, hoping to obtain some policy implications.

The structure of the paper is as follows. The following section will examine the details and characteristics of the ESPF. The section 3 will summarise major results of applying an annual performance evaluation scheme to the consensus forecasts. Section 4 will turn to testing the rational expectations hypothesis. Remaining promising research topics are discussed in section 5. Section 6 will conclude the paper.

[*]Hosei University
[†]Osaka University
[‡]Cabinet Office, Japanese Government
[§]Economic Planning Association

[1] 'ESP' is acronym of a public relations magazine of the Cabinet Office, 'Economy, Society, and Policy', and does not stand for extrasensory perception.

2 What is the ESPF?

2.1 Brief History

Before the start of the ESPF, there had been no reliable 'anchor' in macroeconomic forecasts. Media provided some information, although they usually cover a small number of economists – about a dozen at most – and is also conducted only on an ad hoc basis. The government's official outlook might play some role, but it seems to fail to meet business needs because of its 'official' nature. For example, it is usually available once a year in late December for the purpose of formulating an annual budget for the next fiscal year (April to March), unless a supplementary budget is planned. It might not be neutral due to intentional optimism in geared towards containing political pressures to increase public expenditures in a recession. In this situation, the launch of the ESPF could be a valuable step to fill the gap.

The ESPF was formally launched in May 2004 after a trial in April. The Economic Planning Association distributes questionnaires to participants around the 25th of each month and publishes the result around the 10th of the following month. The participants are requested to provide their *personal* [2] forecasts of 16 variables for the current and next fiscal year (March to April) – seven and five of which are GDP and its components, and financial variables, respectively – in addition to three macro variables for quarters during the forecast period [3] – 10 quarters at longest. They are requested to answers to some judgment questions. The number of participants was 38 at the start and, following a few dropouts and new entries every year, it is still the same as of January 2009.

2.2 Comparison and Extension

Compared with other surveys of forecasters (Table 1), it is rather obvious that the ESPF and the Blue Chip are similar. In fact, the ESPF was designed to follow a success of Blue Chip Economic Indicators. This is reflected in frequency of monthly publication, choice of forecasted indicators, especially annual forecast, and forecast period of two years. A difference is that variables forecasted quarterly are much fewer in the ESPF than in the Blue Chip. This is because of concerns at the time of the launch of the Survey that heavy burden might discourage forecasters to participate in it, which limits quarterly indicators to essential ones. Table 1 also shows the ESPF is a medium-sized survey in terms of the number of participants and variables forecasted, i.e. smaller than the Blue Chip, the Survey of Professional Forecasters, and NABE Outlook Survey, but larger than the BOE survey in its Inflation Report.

(Table 1 Comparison of the ESPF with Other Forecast Surveys)

[2] Some submit their forecasts on an institutional basis for the reason that the forecasts are the outcome of joint work of forecasting teams. This paper does not treat personal and institutional responses differently.

[3] For example, in January 2009, forecasters are requested to provide quarterly figures starting from 2008Q4 to 2011Q1, i.e. quarterly developments up to the end of FY 2001. The last quarter of the forecast period is fixed until December, 2009. Therefore, the forecast period is 10 quarters in January, gradually shortening to 6 quarters in December.

The ESPF has been expanded with addition of special questions: a question of possible timing and direction of changes in the policy target interest rate was introduced in October, 2006 [4], which was followed in January, 2008 by a question asking the next turning point of business cycle. These additions were motivated by a desire to address key policy questions. These special questions are extensively utilized in the NABE Outlook as well as the Blue Chip. The latest innovation is to set up questions about subjective distribution of GDP growth rate and CPI inflation, as is observed in the Survey of Professional Forecasters.

What is still missing in the ESPF concerns long-term projections, which are regular questions in the SPF. In particular, taking account of implications for monetary policy, introducing a question about long-term inflation expectation may be worthy of serious consideration. According to Croushore (1993, p.8), obtaining information about long-term inflation expectations is a major motive for FRB Philadelphia to take over the SPF from ASA and NBER in 1990. Although the Bank of Japan has just started to extract such information from its survey named 'the Opinion Survey on the General Public's Views and Behaviour', asking such a question may still be valuable.

3 Annual Performance Evaluation

3.1 Evaluation Scheme

The Economic Planning Association evaluates performance of individual forecasters and publicizes names of the best five forecasters every year, thereby encouraging the participants to improve their performance. The evaluation is based on RMSE (Root Mean Squared Error) of five variables [5] : two for fiscal year, real and nominal GDP growth rates, and three for quarterly developments, real GDP growth rate (s.a.a.r.), CPI y-o-y inflation rate and unemployment rate (s.a.).

First of all, participants' information sets need to be clear. They can observe March outcomes, say, released late April before submitting their May forecasts, the due date of which is usually around 5th. The time lag is longer for the quarterly estimate for GDP statistics: in the previous example, the last observation is the second estimate of Q4 figures, which is available in mid-March. The Q1 figures are released around 10th of May.

Now our evaluation scheme is shown as follows. Let f^k_{ijt} be a forecast placed by a forecaster i in at time $t-j$ for an economic indicator k of time t, and y^k_t its outcome. Figure 1 to 3 provide overall pictures of forecasts and actual values for FY and quarterly GDP growth rate and quarterly CPI inflation rate.

(Figure 1 Forecasts and Actual Values of Fiscal Year Growth Rate)
(Figure 2 Forecasts and Actual Values of Quarterly Growth Rate)
(Figure 3 Forecasts and Actual Values of Quarterly CPI inflation Rate)

[4] Note the zero-interest rate policy was lifted in mid-July, 2006.

[5] Note four variables except for FY nominal GDP growth rate are used in FY 2004 and 2005 evaluation, as shown in Table 2.

The forecast error is represented by:

$$e^k_{ijt} \equiv y^k_t - f^k_{ijt}. \tag{1}$$

The forecasts errors are multiplied by weights, w^k_{jt}, to keep uncertainty faced by forecasters constant across time. For example, forecasting FY 2007 growth rate should be easy in May 2008, because actual values are available up to 2007Q4 and only 2008Q1 has to be forecasted, but difficult in January 2007, because all the four quarterly figures have to be forecasted. The weights are analytically calculated using estimates of dynamic regression models [6].

The formulae for annual and quarterly figures are shown as follows,

$$RMSE^k_{it} = \left(\frac{1}{17}\sum_j (w^k_{jt} e^k_{ijt})^2\right)^{1/2}, \quad k = RGDPY, NGDPY. \tag{2}$$

$$RMSE^k_{it} = \left[\frac{1}{4}\sum_{Q\in t}\left(\frac{1}{6}\sum_j (w^k_{jQ} e^k_{ijQ})^2\right)\right]^{1/2}, \quad k = GDPQ, CPIQ, UNQ. \tag{3}$$

In order to rank overall performance of a forecaster, it is necessary to aggregate the five RMSEs obtained from the formulae. An RMSE for the fiscal year is calculated based on 17 responses from January to May next year. As for quarterly projections, the last six responses are used while more observations are available. However, note the quarterly growth rate of real GDP is more volatile than CPI inflation. This is why we calculate a mean of T-scores across the five variables for each forecaster to obtain an indicator for total evaluation as below,

$$TTL_{it} = \frac{1}{5}\sum_k \left(50 + 10 \times \frac{RMSE^k_{it} - Mean(RMSE^k_t)}{Std(RMSE^k_t)}\right). \tag{4}$$

Here smaller scores indicate better performances.

Some forecasters do not necessarily submit their projections for all variables each month. If such missing values are often observed in a variable [7], it is not evaluated. The overall evaluation for an individual forecaster requires that all five variables be evaluated. Note that if the number of missing values are rather small, evaluation is conducted, and a penalty is noted in the calculation, i.e. one and a half times the average of weighted MSE, $(w^k_{jt} e^k_{ijt})^2$, of submitted forecasts.

3.2 Evaluating the Consensus Forecast

Application of Eq.(2)-(4) to forecast error of the consensus forecast, e^k_{cjt}, shows it is a good forecast, ranked 5th to 10th place out of around 30 forecasters evaluated every

[6] Weights on unemployment rate come from estimates of AR model, while those for GDP growth and CPI inflation from bivariate dynamic models with Index of Industrial Production and CPI for Tokyo area as an additional independent variable, respectively. See Kawagoe (2008) for the details.

[7] The meaning of the tern 'often' is determined on an ad hoc basis each year, paying attention to the balance between the number of forecasters included in the evaluation and quality of the evaluation. For example, in the case of the evaluation conducted last September, those who failed to submit their forecasts for real GDP growth rate 3 out of 17 times are judged to be ineligible for the evaluation.

year (Table 2). Kawagoe (2007b) confirmed that this result was robust with regard to changes in weights on forecast errors as well as to variations in ways of aggregating RMSEs of forecasted variables. Relatively good performance of the consensus forecast is a well established finding in the US. But this performance evaluation exercise confirmed this finding probably for the first time in Japan.

(Table 2 Performance of the Consensus Forecast)

The merit of the consensus is usually explained by appealing to risk diversification, using an analogy common in portfolio investment. There are no perfect forecasts anywhere: each forecast misses something. In this situation, a 'portfolio of forecasts' does make sense in terms of improving performance. It is not clear why all the weights on forecasts should be equal, as is the case in the consensus forecast. However, a forecast portfolio with optimal weights derived analytically, usually fails to outperform the consensus, which was noted in surveys by Clemen (1989), Diebold and Lopez (1996), and Timmermann (2006). This point has not been examined in Japan, as far as we know. Another to be considered is to incorporate covariance matrix of disturbances of variables forecasted into aggregation of RMSEs because these variables do not move around independently, but are connected through some rules, such as identities, Phillips curve and Okun's law.

4 Testing Rational Expectations Hypothesis (REH)

Now let's turn to test the REH, $E(e^k_{ijt}|\Omega_{ij}) = 0$, where Ω_{ij} is the information set held by forecaster i at time $t-j$. Using the ESPF data, Ban (2009) conducted two kinds of tests in quarterly GDP growth rate by the following equation,

$$y_t = \beta_{i0} + \beta_{i1} f_{ijt} + \beta_{i2} Z_{ij} + v_{ijt}, \tag{5}$$

unbiasedness with the null of $(\beta_{i0}, \beta_{i1}) = (0,1)$, given $\beta_{i2} = 0$; and efficiency with the null of $(\beta_{i0}, \beta_{i1}, \beta_{i2}) = (0,1,0)$ for any variable $Z_{ij} \in \Omega_{ij}$. Note these hypotheses should be tested in individual regressions unless micro-homogeneity is satisfied.

Here we define Z as real GDP growth rates in the previous four quarters at the time of forecasting. Table 3 shows two-thirds and about 90 per cent out of 33 individual regressions are rejected in unbiasedness and efficiency tests, respectively. However, interestingly, as the horizon shrinks, the number of rejections decreases: if these hypotheses are tested in each of zero to two months forecast horizons, only a few are rejected. The failure of efficiency hypothesis may imply forecasters use the models different from traditional time series models.

(Table 3 Tests of Individual Forecasts)

Confirming micro-homogeneity across 33 forecasters, $\beta_{i0} = \beta_0$ and $\beta_{i1} = \beta_1$, Ban (2009) went on to test the hypotheses in pooled data. Otherwise, consistency could not be ensured, except for a trivia case, $(1/N)\sum_i \beta_{i1} = \beta_1$. The hypotheses are rejected in all the forecast horizons in the pooled data (Table 3). The different results for the short

horizons might be affected by inconsistent pooled coefficients due to constant cross-sectional realization (Zarnowitz, 1985). Note Zarnowitz (1985)'s argument holds, even if forecasters are homogeneous, as long as all the variables are stationary (Bonham and Cohen, 2001).

Use of the consensus requires care. Even if micro-homogeneity holds, there is still a risk of private information bias (Bonham and Cohen, 2001). Fingleski and Watchel (1983) showed the unbiasedness test by applying OLS to Eq.(5) with the consensus as the regressor, $f_{ijt} = f_{cjt}$, may produce inconsistent estimator because the presence of private information may cause correlation between the regressor, f_{cjt}, and the disturbance, v_{cjt}. Hence, Ban (2009) tested unbiasedness of the consensus by IV with real GDP growth rate in the previous eight quarters at the time of forecasting as instrument variables. As Table 4 shows, the null is rejected in only one case with the longest forecast horizon. This may illustrate what a misleading result use of the consensus in the unbiasedness test leads to. Nevertheless, we think this result suggests the consensus is still quite useful for forecasting purposes, even though the care is needed for estimation. As for efficiency, the null is rejected in all cases except for those with the three shortest forecast horizons, which is consistent with results from individual regressions.

(Table 4 Tests of the Consensus Forecast)

5 Promising Research

5.1 Real-time Data Analysis

We have assumed there is *an* actual value. But this is not true. In Japan, the first estimate of quarterly GDP is released a month and a half after the end of the quarter, as explained above. The estimate is subsequently revised with the second estimate a month later and incoming annual statistics in December the following year. Furthermore, re-calculation of seasonal adjustments conducted with every release of new estimates also usually brings about non-negligible changes.

The analyses in the previous sections implicitly assume forecasters try to find the value of the first estimate. This assumption seems plausible as a first approximation to forecasters' behaviour, judging from our experience and anecdotal evidence. However, the revisions to the first estimates are too significant to be ignored.

According to Kawagoe (2007a), MAE of revision from the first to the most recent estimates is larger than 2 percentage points, almost comparable to mean growth rate, 2 and a half per cent (s.a.a.r.) (Table 5). Note that this size of revision is about median for OECD countries, but, *alas*, it is the largest among the G7 countries [8].

(Table 5 Revisions to first quarterly estimates of real GDP growth rate)

The large revisions posed a difficulty in evaluating individual performance last year. A series of strong figures of first quarterly estimates, followed by large downward revisions, ranked optimists high for quarterly, but low for fiscal year forecasts. Are they good forecasters?

[8] See Tosetto (2006) for the details.

As a first step to understand natures of revisions, Kawagoe (2007a) examined whether they are noise or news, following Mankiw and Shapiro (1986) [9]. A modified Eq.(5) is used,

$$R_t^{i1} = \beta_0 + \beta_1 y_t^{1E} + \beta_2 Z_t + v_t, \tag{6}$$

where R_t^{i1} is a revision from the first estimate, y_t^{1E} to i th estimate, y_t^{iE}. Here two revisions are focused: 1st to 2nd, R_t^{21} and 1st to final (or the most recent), R_t^{F1}. If the null of $\beta_1 = 0$ can be rejected with the alternative of $\beta < 0$, the noise view holds: the revisions are predictable and should be eliminated beforehand, i.e. they are 'noise'. Otherwise, they are not predictable and therefore contain new information, i.e. the news view holds. Also note unbiasedness and efficiency tests are applied in a similar manner to Section 4.

Table 6 shows the noise view holds in most cases. In addition, both unbiasedness and efficiency hypothesis are rejected. This result is likely to pose an interesting question of how forecasters interpret revisions and utilize them when updating their forecasts: do they really believe in the incoming data and change their projections, or do they discount the changes and stick to their previous projections? Clearly, further studies are awaited.

(Table 6 Noise or News)

5.2 Subjective Distribution

Kawagoe (2007b) analyzed distribution of individual forecasts and their corresponding actual values every month during FY 2004 and 2005 and found the actual real GDP growth rates tend to be 'outliers' from forecast distribution: all the forecasters tend to make same (rather significant) mistake. Remember the micro-homogeneity hypothesis holds across individual forecasters.

The result has motivated us to undertake two projects: to show uncertainty around the consensus forecast, and to encourage non-specialists to join forecasting, thereby making information sets used for forecasting more heterogeneous. This subsection deals with the first, and the next does the second.

It is well known that dispersion measures, such as standard deviation, of individual forecasts do not represent their overall uncertainty (Zarnowitz and Lambros, 1987). Kawagoe (2007b) showed actual values far from individual forecasts could be well justified once their overall uncertainty is considered.

Let $k(y_t - f_{it}; h)$ be uncertainty forecaster i attach to his point forecast, f_{it} [10]. k stand for the stochastic density function called the kernel function, which assumes normal distribution with standard deviation of h, which represents uncertainty attached to point forecast and also controls smoothness of the overall distribution. The Kernel estimator of realized value is distribution over n forecasters as follows,

$$\hat{g}(y_t) = \frac{1}{n}\sum_{i=1}^{n} k(y_t - f_{it}; h). \tag{7}$$

[9] See also Mankiw, Runkle and Shapiro (1984) and Faust, Rogers and Wright (2005).

[10] Here subscript j for forecasting date is omitted for simplicity.

A non-parametric method sets h equal to $h^* = 0.24$ %, which minimizes *MISE* (Mean Integrated Squared Error), $E\{\int[\hat{g}(y) - g(y)]^2 dy\}$, and is given below [11],

$$h^* = \left(\frac{4}{3n}\right)^{1/5} \bar{\sigma} \tag{8}$$

$$\bar{\sigma} \equiv median(\mid f_{it} - median(f_{it}) \mid)/0.6745. \tag{9}$$

An alternative way is to plug an equation standard error of an AR model of y_t adjusted for forecast horizon, $h^{**} = 1.28$ %, into h, in Eq.(7).

The two methods result in distributions shown by solid and dotted lines, respectively, in Figure 4. The figure suggests the realised value seems unexpectedly low in terms of the solid line distribution, but may be a well expected event in terms of the dotted one. The large difference between the two distributions can explain why so many real values seem outliers far from clustering individual forecasts. The point is that the uncertainty attached to point forecasts are large, compared to distance among individual forecasts.

(Figure 4 Uncertainty around Individual Forecasts and Their Actual Values)

The result implies it is necessary to show explicitly uncertainty around the consensus: otherwise, users may underestimate its uncertainty. Studying FRB Philadelphia's SPF and BOE's external forecaster survey in the Inflation Report, the ESPF Steering Committee, which includes two of the authors as its members, decided to ask forecasters to submit their subjective distribution of real GDP growth rate and CPI inflation rate . The results have been available since June 2008.

5.3 Non-specialist Forecasts

A simple way to overcome forecast homogeneity is to ask the same question to other people, while, in practice, it may be difficult to do so to people on the street! If arguments of 'Wisdom of Crowds' (Surowiecki, 2004) holds, a well designed survey of general public could beat experts. But each survey has its own defects. As for inflation expectations, a casual look at Figure 5, which plots inflation expectations in the next 12 months calculated from the Monthly Consumer Confidence Survey, and their corresponding outcomes, may conclude they are biased, especially in a deflation period, and heavily affected by recent developments they face.

(Figure 5 Expected CPI Inflation rate: the Monthly Consumer Confidence Survey)

Iiduka and Kawagoe (2009) tried to collect information from economists, but not specialists of forecasting. They distributed a simplified version of ESPF's questionnaire to members of the Japan Association of Business Cycle Studies (JABCS hereafter), and

[11] $\bar{\sigma}$ calculated by Eq.(9) is used rather than standard form of σ in Eq.(8) so that sample outliers do not affect the result. See Silverman (1986), Wand and Jones (1995), and Bowman and Azzalini (1997) for the details of the method.

compared the responses with the ESPF. While two-thirds of them turned out to have some forecast experiences [12], we will call them 'non-specialists'.

Non-specialists' forecasts are usually different from specialists'(Table 6): cumulative distributions are significantly different between the two, except for 2 cases, according to Kolmogorov-Smirnov test. Interestingly, the specialists can not always win: JABCS beats ESPF no less than 5 times out of 17 matches.

(Table 7 Comparison between specialists' and non-specialists' forecasts)

ESPF is better at (quarterly) CPI inflation forecast than JABCS, probably because quarterly CPI figures are directly inferred from incoming monthly data [13], to which specialists are naturally likely to devote more resources than non-specialists. Advantages for specialists are less obvious in forecasting the real GDP growth rate because there are no direct links between monthly indicators and the quarterly GDP estimate, and much uncertainty remains. Thus, some information held by non-specialist, but not by specialists, may improve the performance of the consensus when there is a great deal of uncertainty, if it is properly combined with the information of specialists.

This is just a single. Our colleagues are now undertaking another, setting up a game called 'Challenge economists!' on a website [14], with the co-operation of Prediction Inc., in which participants bet by token and increase their wealth when beating the ESPF consensus. Although there were only a few participants at the start, there are now more than 60. Some preliminary result will be available in the near future.

5.4 Recent Developments since the Lehman Shock

An advantage of the ESPF is its monthly release, especially in terms of uncertainty of the consensus. This is very valuable in the current rapidly changing situation following the Lehman shock in September, 2008.

The consensus forecast of 2008Q4 growth rate (s.a.a.r.) was 1.1 per cent in August just before the shock, but then revised downwards to a negative rate, -0.6 per cent, in November, and further to -5.1 and -10.6 per cent in January and February, 2009, respectively. Then actual figure turned out to be -12.7 per cent.

The deterioration of the prospects of forecasters is extraordinary, but point forecasts could be well documented in other sources. Our clear strength is shown in monthly developments of mean probabilities of FY growth rate (Figure 6). Given rapid deterioration in economic conditions, they provide the government as well as private sector with valuable information of possible size of uncertainty. Individual behaviours are also interesting, but such dynamics hidden in the mean probabilities remain to be explored.

(Figure 6 Changes in subjective distributions)

[12] A tenth of them are engaged in forecasting now, but excluding them does not affect major results here.

[13] In addition, CPI in Tokyo area is released two weeks earlier than the national CPI, and, therefore, it plays a role as a leading indicator.

[14] Please visit http://prediction.jp/espf, which is, unfortunately, in Japanese only.

6 Conclusion

It has been about five years since the launch of the ESPF. Now the consensus of the ESPF plays a role of anchoring economic prospects. This paper has reviewed the major results of the three year research programme funded by the ESRI. The consensus forecast performs well, compared to individual forecasters, which is consistent with US experience. The consensus forecast of real GDP growth rate is found to be unbiased in shorter horizon than 12 months, and may be useful to forecasting. However, as is well known, the results have misleading implications for individual forecasts: their unbiasedness and efficiency are rejected in most of the cases, except for zero to two month horizon.

There are still many areas for further research. Data revisions are too large to be ignored in the case of real GDP growth rate. It remains to be seen how they affect the updating of individual forecasts. Another interesting topic is forecast uncertainty. Dispersion of individual forecasts may lead to underestimation of uncertainty of the consensus forecast. This introduced a new question to ask forecasters for their subjective distribution of their point forecasts in June last year. This also motivated new research of how different forecast specialists and non-specialists are from each other in terms of economic predictions. Monthly update of subjective distribution of growth rate should be quite useful for analysis of how economic prospects are changed in the face of current turbulent situations. We hope the ESPF data will be widely used and shed a new light on expectations formation.

References

Ban, Kanemi (2009) "Testing the Rational Expectations Hypothesis Using a Survey of Professional Forecasters in Japan.".*mimeo.*

Bonham, Carl S. and Richard H. Cohen (2001) "To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data." *Journal of Business and Economic Statistics.* Vol. 19. pp. 278–290.

Bowman, Adrian W. and Adelchi Azzalini (1997) *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations.* Oxford, UK: Oxford University Press.

Clemen, Robret T. (1989) "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting.* Vol. 5. pp. 559–583.

Croushore, Dean (1993) "Introducing: The Survey of Professional Forecasters." *FRB Philadelphia Business Review.* pp. 3–13.Nov/Dec.

Diebold, Francis X. and Jose A. Lopez (1996) "Forecast Evaluation and Combination." In G. S. Maddala and C. R. Rao. eds. *Handbook of Statistics vol.14.* Amsterdam, Netherlands: North-Holland. pp. 241–268.

Faust, Jon, John H. Rogers, and Jonathan H. Wright (2005) "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit and Banking.* Vol. 37. pp. 403–419.

Fingleski, Stephen and Paul Watchel (1983) "Rational Expectations, Informational Efficiency, and Tests Using Survey Data: A Reply." *Review of Economics and Statistics.* Vol. 65. pp. 529–531.

Iiduka, Nobuo and Masaaki Kawagoe (2009) "How different Are Non-specialists' Forecasts from Specialists'?: Part I.".*mimeo.* (in Japanese).

Kawagoe, Masaaki (2007a) "How Is Economic Growth Rate Revised Afterwards?: Real-time Data Analysis of Real GDP Growth Rate.".*Quarterly Journal of National Accounts of Japan*, No.134. Economic and Social Research Institute, Cabinet Office, Japanese Government (in Japanese).

Kawagoe, Masaaki (2007b) "Is the Consensus Forecast Just the Average?: Reexamining an Evaluation of the ESP Forecast.".*ESRI Discussion Paper Series* No.180, Economic and Social Research Institute, Cabinet Office, Japanese Government (in Japanese).

Kawagoe, Masaaki (2008) "A Technical Note on Evaluation of Participants in FY 2007 ESP Forecast.".available on the Economic Planning Association HP (http://www.epa.or.jp/esp/fcst/fcst.html). (in Japanese).

Mankiw, N. Gregory and Matthew D. Shapiro (1986) "News or Noise: An Analysis of GNP Revisions." *Survey of Current Business.* Vol. May. pp. 20–25.

Mankiw, N. Gregory, David E. Runkle, and Matthew D. Shapiro (1984) "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics.* Vol. 14. pp. 15–27.

Silverman, Bernard W. (1986) *Density Estimation for Statistics and Data Analysis.* London, UK: Chapman and Hall.

Surowiecki, James (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than the Few.* London, UK: Abacus.

Timmermann, Allan (2006) "Forecast Combinations." In Graham Elliott, Clive W. J. Granger, and Allan Timmermann. eds. *Handbook of Economic Forecasting.* Amsterdam, Netherlands: North-Holland. pp. 135–196.

Tosetto, E. (2006) "Revisions in Quarterly GDP of OECD Countries: An Update.".presented at OECD National Account Working Party.

Wand, M. P. and M. C. Jones (1995) *Kernel Smoothing.* London, UK: Chapman and Hall.

Zarnowitz, Victor and Louis A. Lambros (1987) "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy.* Vol. 95. No. 3. pp. 591–621.

Zarnowitz, Victor (1985) "Rational Expectations and Macroeconomic Forecasts." *Journal of Business and Economic Statistics.* Vol. 3. .