Are 16-month-ahead forecasts useful? A directional analysis of Japanese GDP forecasts

Ashiya, Masahiro

(Citation)
Journal of Forecasting,25(3):201-207

(Issue Date)
2006-02-27

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000267

Are 16-Months-Ahead Forecasts Useful?: A Directional Analysis of Japanese GDP Forecasts

Masahiro ASHIYA

Past literature casts doubt on the ability of long-term macroeconomic forecasts to predict the direction of change. We reexamine this issue using the Japanese GDP forecast data of 37 institutions, and find that their 16-months-ahead forecasts contain valuable information on whether the growth rate accelerates or not.

JEL Classification Codes: E37; C53; E17.

Keywords: Economic forecasts; Directional analysis; Forecast evaluation; Rational expectations hypothesis.

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

1. Introduction

Past literature on forecast rationality has found considerable evidence against the directional accuracy of long-term macroeconomic forecasts.[1] Leitch and Tanner (1995) focus on the direction of change of the growth rate (i.e., $\operatorname{sgn}\Delta g_t$). They investigate the forecasts of 42 professional institutions, and find that the average accuracy of their three-quarters-ahead forecasts is only 55%. Öller and Barot (2000) analyze the forecasts of $\operatorname{sgn}\Delta g_t$ for 13 European countries made by the OECD and 13 national institutes, and show only one-half of their one-year-ahead forecasts are better than a naïve model that always predicts the same sign. Ash, Smyth, and Heravi (1998) consider the forecasts of the sign of the growth rate (i.e., $\operatorname{sgn} g_t$) for the G7 economies made by the OECD, and show six in seven one-year-ahead forecasts and all seven 18-months-ahead forecasts have no predictive power. Artis (1996) evaluates the 1973-94 forecasts of $\operatorname{sgn}\Delta g_t$ for the G7 economies made by the IMF, and finds that four in seven one-year-ahead forecasts fail to maintain directional accuracy. The only positive evidence is Pons (2000) that evaluates the 1973-95 forecasts of $\operatorname{sgn}\Delta g_t$ for the G7 economies made by the OECD and the IMF.[2] It finds the one-year-ahead forecasts have predictive power except those for Japan and Italy.[3]

We reexamine this issue using the Japanese GDP forecast data of 37 institutions, and find that they contain useful information on the acceleration/deceleration of the growth rate. Section 2 explains the methodology. Section 3 explains the data, and Section 4 reports the results. Discussion is in Section 5, and Section 6 concludes the paper.

2. Methodology

Most of the studies that evaluate economic predictions focus on a quantitative nature of the forecast error. Although these studies find significant bias,[4] this bias may be consistent with the rational expectations hypothesis. Lai (1990) summarizes the following four reasons. (a) The growth process may change over time and the forecasters may gradually learn it (Lewis, 1989). (b) The possibility of a major policy change may affect the forecasts even if the change does not occur during the sample period (Krasker, 1980). (c) The loss function for forecast errors may not be quadratic

(Zellner, 1986), and (d) the forecasters may not reveal their information. Furthermore, Ashiya (2005b) shows that forecasters compromise accuracy to gain publicity for their firms. He finds that, the more forecasters' wages depend on publicity, the more extreme and the less accurate their forecasts are. Hence we consider a less stringent criterion than the unbiasedness, and examine the extent the forecasts predict the direction of change correctly. Forecasts have positive value if they predict the acceleration/deceleration better than the random forecasts.

Suppose forecaster *i* in year *t* releases two forecasts, a short-term growth forecast for year t, $f^i_{t,t}$, and a long-term forecast for year $t+1$, $f^i_{t,t+1}$. Let g_t be the actual growth rate in year t, and let $\varepsilon^i_{t,t}$ ($\varepsilon^i_{t,t+1}$) be the forecast error of $f^i_{t,t}$ ($f^i_{t,t+1}$). Then, by definition, $f^i_{t,t} = g_t + \varepsilon^i_{t,t}$ and $f^i_{t,t+1} = g_{t+1} + \varepsilon^i_{t,t+1}$.

Let us define

$$\Delta fg^i_{t,t} \equiv f^i_{t,t} - g_{t-1}$$

and $\Delta g_t \equiv g_t - g_{t-1}$.

$\Delta fg^i_{t,t}$ is positive if and only if forecaster *i* in year *t* predicts that the growth rate will accelerate in year t. Δg_t is positive if and only if the actual growth rate accelerated in year t. We compare the sign of $\Delta fg^i_{t,t}$ (i.e. $\operatorname{sgn}\Delta fg^i_{t,t}$) with the sign of Δg_t (i.e. $\operatorname{sgn}\Delta g_t$) to evaluate the usefulness of the short-term forecasts.

Similarly, define

$$\Delta fg^i_{t,t+1} \equiv f^i_{t,t+1} - g_t$$

and $\Delta ff^i_{t,t+1} \equiv f^i_{t,t+1} - f^i_{t,t}$.

Positive $\Delta fg^i_{t,t+1}$ and positive $\Delta ff^i_{t,t+1}$ both indicate that forecaster *i* in year *t* predicts the growth rate will accelerate in year $t+1$. Forecasters in year *t*, however, do not know the value of g_t when they release $f^i_{t,t+1}$. Furthermore, $\operatorname{sgn}\Delta ff^i_{t,t+1}$ is more accurate predictor of $\operatorname{sgn}\Delta g_{t+1}$ than $\operatorname{sgn}\Delta fg^i_{t,t+1}$ is if $\varepsilon^i_{t,t}$ and $\varepsilon^i_{t,t+1}$ are highly correlated. Therefore we analyze both $\operatorname{sgn}\Delta fg^i_{t,t+1}$ and $\operatorname{sgn}\Delta ff^i_{t,t+1}$.

To test the forecasting ability in the direction of change, we can use the Fisher's (1922) exact test based on contingency tables (See Henriksson and Merton, 1981).

Consider the case of the short-term forecasts for example. The null hypothesis is that $\operatorname{sgn}\Delta fg_{t,t}^{i}$ and $\operatorname{sgn}\Delta g_t$ are independent. Let n_{00} be the number of forecasts in which $\Delta fg_{t,t}^{i} > 0$ and $\Delta g_t > 0$, n_{10} be the number of forecasts in which $\Delta fg_{t,t}^{i} \le 0$ and $\Delta g_t > 0$, n_{01} be the number of forecasts in which $\Delta fg_{t,t}^{i} > 0$ and $\Delta g_t \le 0$, n_{11} be the number of forecasts in which $\Delta fg_{t,t}^{i} \le 0$ and $\Delta g_t \le 0$, and $n \equiv n_{00} + n_{01} + n_{10} + n_{11}$ be the total number of forecasts. Then the probability that this outcome came form a population that satisfies the null hypothesis is

$$P_i \equiv \sum_{x=n_{11}}^{n^*} \binom{n_{10}+n_{11}}{x}\binom{n_{00}+n_{01}}{n_{01}+n_{11}-x} \Big/ \binom{n}{n_{01}+n_{11}}$$

where $n^* \equiv \min\{n_{10}+n_{11}, n_{01}+n_{11}\}$.

We can reject the null hypothesis when P_i is sufficiently small. We can test the null hypothesis that $\operatorname{sgn}\Delta fg_{t,t+1}^{i}$ ($\operatorname{sgn}\Delta ff_{t,t+1}^{i}$) and $\operatorname{sgn}\Delta g_{t+1}$ are independent by the similar methods.

3. 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 1970s.[5] 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, February 2003 issue contains forecasts for fiscal year 2002 (from April 2002 to March 2003) and for fiscal year 2003 (from April 2003 to March 2004) released in December 2002. We treat the former as $f_{2002,2002}^{i}$ and the latter as $f_{2002,2003}^{i}$. Therefore $f_{t,t}^{i}$ is a four-months-ahead forecast for year t, and $f_{t,t+1}^{i}$ is a 16-months-ahead forecast for year $t+1$.

We use the forecasts for the fiscal years 1981 to 2002 in order to avoid the effect of the second Oil Shock. We exclude institutions whose participation rates are less than 70% (i.e. those who participate in less than 15 surveys), leaving 37 institutions. The total number of forecast sets, $\left(f_{t,t}^{i}, f_{t,t+1}^{i}\right)$, is 728: the average number of observations per institution is 19.68.

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 Japanese government usually released in June. Japanese economy experienced four business cycles in our sample period: the peaks were 1984, 1990, 1996, and 2000, and the troughs were 1981, 1986, 1993, 1998, and 2001.[6]

4. Results

First we check the proportion of correct forecasts for the individual institutions. Define $P\left(\Delta fg_{t,t}^{i}\right)$ as the proportion of times that forecaster i correctly predicts $\operatorname{sgn}\Delta g_t$ by $\operatorname{sgn}\Delta fg_{t,t}^{i}$. Similarly, define $P\left(\Delta fg_{t,t+1}^{i}\right)$ ($P\left(\Delta ff_{t,t+1}^{i}\right)$) as the proportion of times that forecaster i correctly predicts $\operatorname{sgn}\Delta g_{t+1}$ by $\operatorname{sgn}\Delta fg_{t,t+1}^{i}$ ($\operatorname{sgn}\Delta ff_{t,t+1}^{i}$). Table 1 shows the summary statistics.

The short-term forecasts turn out to be very accurate. The average of $P\left(\Delta fg_{t,t}^{i}\right)$ is 0.933, and the standard deviation is 0.055. The best institutions predict $\operatorname{sgn}\Delta g_t$ perfectly. As for the long-term forecasts, $P\left(\Delta fg_{t,t+1}^{i}\right)$ and $P\left(\Delta ff_{t,t+1}^{i}\right)$ yield significantly different results: the average of $P\left(\Delta fg_{t,t+1}^{i}\right)$ is only 0.599, while the average of $P\left(\Delta ff_{t,t+1}^{i}\right)$ is 0.754.[7] $P\left(\Delta ff_{t,t+1}^{i}\right)$ of the worst institution (0.650) is much higher than $P\left(\Delta fg_{t,t+1}^{i}\right)$ of average institutions.

Table 2 shows the results of the non-parametric analysis (explained in Section 2) for the individual institutions. The first column indicates the number of institutions for which the null hypothesis of no correlation between forecasts and realizations is rejected at the 0.10 significance level. The second and the third column indicate the number of institutions for which the null is rejected at the 0.05 and 0.01 significance level. The first row shows that the null hypothesis that $\operatorname{sgn}\Delta fg_{t,t}^{i}$ and $\operatorname{sgn}\Delta g_t$ are independent is rejected for all 37 institutions at the 0.05 significance level. It follows that $\operatorname{sgn}\Delta fg_{t,t}^{i}$ is useful in predicting $\operatorname{sgn}\Delta g_t$. In contrast, the second row shows that $\operatorname{sgn}\Delta fg_{t,t+1}^{i}$ has no power in predicting $\operatorname{sgn}\Delta g_{t+1}$. The null hypothesis that $\operatorname{sgn}\Delta fg_{t,t+1}^{i}$

and $\operatorname{sgn}\Delta g_{t+1}$ are independent is rejected for only three institutions.

The third row of Table 2 compares $\operatorname{sgn}\Delta ff^{i}_{t,t+1}$ with $\operatorname{sgn}\Delta g_{t+1}$, and finds a striking result. Among 37 institutions, the null hypothesis that $\operatorname{sgn}\Delta ff^{i}_{t,t+1}$ and $\operatorname{sgn}\Delta g_{t+1}$ are independent is rejected for 33 institutions at the 0.10 significance level. (The null is rejected for 22 institutions at the 0.05 significance level.) It indicates that about 90 percent of the long-term forecasts for the next year have predictive power when combined with the short-term forecasts for the ongoing year.

One remaining problem is that it is difficult ex ante to identify the institution whose $\operatorname{sgn}\Delta ff^{i}_{t,t+1}$ is accurate. Fortunately, the analysis below demonstrates that all we have to do is to follow the majority. Let $\overline{ff}_{t,t+1}$ be the number of institutions of positive $\Delta ff^{i}_{t,t+1}$ minus the number of institutions of negative $\Delta ff^{i}_{t,t+1}$. Then Table 3 shows that the proportion of times that the sign of $\overline{ff}_{t,t+1}$ coincides with $\operatorname{sgn}\Delta g_{t+1}$ is 0.773. The null hypothesis that the sign of $\overline{ff}_{t,t+1}$ and $\operatorname{sgn}\Delta g_{t+1}$ are independent is rejected at the 0.05 significance level. Incidentally, the second row of Table 3 shows that $\overline{fg}_{t,t+1}$ (defined in the same way as $\overline{ff}_{t,t+1}$) has no predictive power of $\operatorname{sgn}\Delta g_{t+1}$.

5. Discussions

The last section has shown that (a) $\operatorname{sgn}\Delta fg^{i}_{t,t}$ is highly correlated with $\operatorname{sgn}\Delta g_{t}$, (b) $\operatorname{sgn}\Delta fg^{i}_{t,t+1}$ and $\operatorname{sgn}\Delta g_{t+1}$ are independent, and (c) $\operatorname{sgn}\Delta ff^{i}_{t,t+1}$ is highly correlated with $\operatorname{sgn}\Delta g_{t+1}$. Since

$$\operatorname{sgn}\Delta fg^{i}_{t,t} \equiv \operatorname{sgn}\left[\Delta g_{t} + \varepsilon^{i}_{t,t}\right],$$

$$\operatorname{sgn}\Delta fg^{i}_{t,t+1} \equiv \operatorname{sgn}\left[\Delta g_{t+1} + \varepsilon^{i}_{t,t+1}\right], \text{ and}$$

$$\operatorname{sgn}\Delta ff^{i}_{t,t+1} \equiv \operatorname{sgn}\left[\Delta g_{t+1} + \varepsilon^{i}_{t,t+1} - \varepsilon^{i}_{t,t}\right],$$

it follows that (a') the absolute value of $\varepsilon^{i}_{t,t}$ is small, (b') $\varepsilon^{i}_{t,t+1}$ tends to be negative (positive) when $\Delta g_{t+1} \equiv g_{t+1} - g_{t}$ is positive (negative), and (c') there is a positive correlation between $\varepsilon^{i}_{t,t}$ and $\varepsilon^{i}_{t,t+1}$.

These systematic forecast errors are consistent with both a behavioral explanation and a Bayesian learning model Lewis (1989) suggests. There might be a common disturbance term in $\varepsilon^i_{t,t}$ and $\varepsilon^i_{t,t+1}$ such as optimism/pessimism,[8] or the forecaster might gradually learn about a shift in the process of fundamentals.

6. Conclusions

This paper evaluates the directional accuracy of the long-term GDP forecasts using the data of 37 Japanese institutions. Each institution releases a four-months-ahead forecast for the ongoing fiscal year and a 16-months-ahead forecast for the next fiscal year in every December. It is found that the sign of the difference between these two forecasts is impressively useful to predict the acceleration/deceleration of the growth rate of the next fiscal year, although the 16-months-ahead forecast alone is not. This result presents a sharp contrast to the argument of past literature that forecasts with one year or longer horizon are no better than random forecasts in predicting the direction of change.

The methodology developed in this paper is useful for evaluating other forecasts. Ashiya (2003b) applies it to the real GDP forecasts of G7 countries made by the IMF, and finds that combining current-year forecasts and year-ahead forecasts improves the directional accuracy significantly.

Notes

1. There is mixed evidence for the directional accuracy of the short-term forecasts. See Schnader and Stekler (1990), Stekler (1994), Ash et al. (1998), and Joutz and Stekler (2000).
2. Artis and Pons employ different vintages of realization data. Artis chooses the data published in October of year $t+1$ as the actual growth rate of year t, whereas Pons chooses the latest available data.
3. As for other economic variables, Lai (1990) evaluates survey exchange rate forecasts, and finds that three-months-ahead forecasts have little forecasting ability of the direction of change. Öller and Barot (2000) and Ash et al. (2002) analyze inflation forecasts and find that they are not useful guidance on the direction of change. Pons (2001) investigates the price forecasts made by the IMF for the G7 countries, and finds that only two out of seven one-year-ahead forecasts are valuable in predicting the direction of price increases. Greer (2003) examines the directional accuracy of one-year-ahead forecasts for long-term interest rates, and finds only eleven forecasters out of eighty-six are better than random forecasts at the 0.10 significance level.
4. Ashiya (2002) finds that neither the rational expectations hypothesis nor reputation models with rational and strategic forecasters is consistent with the forecast revisions of Japanese institutions.
5. Ashiya (2003a, 2005a) also uses the data from "Monthly Statistics (Tokei Geppo)".
6. The initial announcements of the actual growth rates for fiscal years 1980 to 2002 were 3.8%, 2.7%, 3.3%, 3.7%, 5.7%, 4.2%, 2.6%, 4.9%, 5.1%, 5.0%, 5.7%, 3.5%, 0.8%, 0.0%, 0.6%, 2.3%, 3.0%, -0.7%, -2.0%, 0.5%, 0.9%, -1.3%, and 1.6%.
7. The accuracy rate of a naïve model that predicts the same direction as last year turns out to be 0.500.
8. Ashiya (2002) finds that the Japanese institutional forecasters as a whole are pessimistic (optimistic) when they obtain positive (negative) information.

References

Artis MJ. 1996. How accurate are the IMF's short-term forecasts? Another examination of the World Economic Outlook. IMF Working Paper, WP/96/89.

Ash JCK, Smyth DJ, Heravi SM. 1998. Are OECD forecasts rational and useful?: a directional analysis. *International Journal of Forecasting* **14**: 381-391.

Ash JCK, Easaw JZ, Heravi SM, Smyth DJ. 2002. Are Hodrick-Prescott 'Forecasts' rational? *Empirical Economics* **27**: 631-643.

Ashiya M. 2002. Accuracy and rationality of Japanese institutional forecasters. *Japan and the World Economy* **14**: 203-213.

Ashiya M. 2003a. Testing the rationality of Japanese GDP forecasts: the sign of forecast revision matters. *Journal of Economic Behavior and Organization* **50**: 263-269.

Ashiya M. 2003b. The directional accuracy of 15-months-ahead forecasts made by the IMF. *Applied Economics Letters* **10**: 331-333.

Ashiya M. 2005a. Twenty-two years of Japanese institutional forecasts. *Applied Financial Economics Letters*, 1, 79-84.

Ashiya M. 2005b. Strategic bias and professional affiliations of macroeconomic forecasters. Discussion Paper Series No. 0429, Graduate School of Economics, Kobe University.

Fisher RA. 1922. On the interpretation of χ^2 from contingency tables, and the calculation of *P*. *Journal of the Royal Statistical Society* **85**: 87-94

Greer M. 2003. Directional accuracy tests of long-term interest rate forecasts. *International Journal of Forecasting* **19**: 291-298.

Henriksson RD, Merton RC. 1981. On market timing and investment performance. Two. Statistical procedures for evaluating forecasting skills. *Journal of Business* **54**: 513-533.

Joutz F, Stekler HO. 2000. An evaluation of the predictions of the Federal Reserve. *International Journal of Forecasting* **16**: 17-38.

Keane MP, Runkle DE. 1990. Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review* **80**: 714-735.

Krasker WS. 1980. The 'Peso Problem' in testing the efficiency of forward exchange markets. *Journal of Monetary Economics* **6**: 269-276.

Lai KS. 1990. An evaluation of survey exchange rate forecasts. *Economics Letters* **32**: 61-65.

Leitch G, Tanner JE. 1995. Professional economic forecasts: are they worth their costs? *Journal of Forecasting* **14**: 143-157.

Lewis KK. 1989. Changing beliefs and systematic rational forecast errors with evidence from foreign exchange. *American Economic Review* **79**: 621-636.

Öller L-E, Barot B. 2000. The accuracy of European growth and inflation forecasts. *International Journal of Forecasting* **16**: 293-315.

Pons J. 2000. The accuracy of IMF and OECD forecasts for G7 countries. *Journal of Forecasting* **19**: 53-63.

Pons J. 2001. The rationality of price forecasts: a directional analysis. *Applied Financial Economics* **11**: 287-290.

Schnader MH, Stekler HO. 1990. Evaluating predictions of change. *Journal of Business* **63**: 99-107.

Stark T, Croushore D. 2002. Forecasting with a real-time data set for Macroeconomists. *Journal of Macroeconomics* **24**: 507-531.

Stekler HO. 1994. Are economic forecasts valuable? *Journal of Forecasting* **13**: 495-505.

Toyo Keizai Inc. *Monthly Statistics* (Tokei Geppo).

Zellner A. 1986. Biased predictors, rationality and the evaluation of forecasts. *Economics Letters* **21**: 45-48.

Table 1: The accuracy rates of the individual institutions

	Avg.	Std.	Min.	Max.
$P\left(\Delta fg_{t,t}^{i}\right)$	0.933	0.055	0.765	1.000
$P\left(\Delta fg_{t,t+1}^{i}\right)$	0.599	0.060	0.500	0.750
$P\left(\Delta ff_{t,t+1}^{i}\right)$	0.754	0.058	0.650	0.909

Notes:

$P\left(\Delta fg_{t,t}^{i}\right)$ is the proportion of times that forecaster i correctly predicts the sign of $\left(g_t - g_{t-1}\right)$ by the sign of $\left(f_{t,t}^{i} - g_{t-1}\right)$.

$P\left(\Delta fg_{t,t+1}^{i}\right)$ is the proportion of times that forecaster i correctly predicts the sign of $\left(g_{t+1} - g_t\right)$ by the sign of $\left(f_{t,t+1}^{i} - g_t\right)$.

$P\left(\Delta ff_{t,t+1}^{i}\right)$ is the proportion of times that forecaster i correctly predicts the sign of $\left(g_{t+1} - g_t\right)$ by the sign of $\left(f_{t,t+1}^{i} - f_{t,t}^{i}\right)$.

Table 2: Non-parametric tests of independence for 37 institutions

	$P_i < 0.10$	$P_i < 0.05$	$P_i < 0.01$
$\mathrm{sgn}\,\Delta fg^i_{t,t}$	37 [a]	37 [a]	36 [a]
$\mathrm{sgn}\,\Delta fg^i_{t,t+1}$	3 [b]	3 [b]	0 [b]
$\mathrm{sgn}\,\Delta ff^i_{t,t+1}$	33 [c]	22 [c]	4 [c]

Notes

a: The null hypothesis is that $\mathrm{sgn}\,\Delta fg^i_{t,t}$ and $\mathrm{sgn}\,\Delta g_t$ are independent.

b: The null hypothesis is that $\mathrm{sgn}\,\Delta fg^i_{t,t+1}$ and $\mathrm{sgn}\,\Delta g_{t+1}$ are independent.

c: The null hypothesis is that $\mathrm{sgn}\,\Delta ff^i_{t,t+1}$ and $\mathrm{sgn}\,\Delta g_{t+1}$ are independent.

Table 3: Non-parametric tests of the majority forecasts

	accuracy	P_i
$\overline{ff}_{t,t+1}$	0.773	0.015 [a]
$\overline{fg}_{t,t+1}$	0.591	0.335 [b]

Notes

a: The null hypothesis is that the sign of $\overline{ff}_{t,t+1}$ and $\operatorname{sgn}\Delta g_{t+1}$ are independent.

b: The null hypothesis is that the sign of $\overline{fg}_{t,t+1}$ and $\operatorname{sgn}\Delta g_{t+1}$ are independent.