Forecast accuracy of the Japanese government: Its year-ahead GDP forecast is too optimistic

Ashiya, Masahiro

Forecast Accuracy of the Japanese Government: Its Year-Ahead GDP Forecast Is Too Optimistic *

Masahiro ASHIYA [+]

April 2005

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

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

Forecast Accuracy of the Japanese Government: Its Year-Ahead GDP Forecast Is Too Optimistic

This paper evaluates accuracy and efficiency of the real GDP forecasts made by the Japanese government over the past 22 years. The government's 16-months-ahead forecast has upward bias of 0.7 percentage points, and is significantly inferior to the mean forecast of private institutions or a vector autoregression forecast that uses real-time data only. Moreover, the government forecast is inferior to these benchmark forecasts not only in the 'recessionary' 1990s but also in the 'prosperous' 1980s. This result casts serious doubt on the assertion that the government intentionally produced inaccurate forecasts in time of recession.

1

1. Introduction

The Japanese economy finally hit the bottom in 2003, and the government predicts that it is now on the road to recovery. People doubt the strength of the economy, however, since they feel past forecasts of the government were too optimistic and thus the prediction above is also unreliable. Is this criticism legitimate? This is the first issue we address in this paper using the forecast record over the past 22 years. Our data covers both growth years (the 1980s) and recessionary periods (the 1990s). This unique feature of the data allows us to investigate the effect of business cycle on the accuracy of the government forecasts.

The Japanese government releases a forecast of real GDP growth rate for the current fiscal year and for the next fiscal year in every late December. (Section 2 explains the data in detail.) Figure 1 shows the records of these forecasts and the actual growth rates. We can find at a glance that the current-year forecasts ($f_{t,t}$) are relatively accurate but the year-ahead forecasts ($f_{t-1,t}$) in the 1990s are significantly higher than the actual growth rates (g_t). Indeed, Section 3 finds that the year-ahead forecasts are on average 0.7 percentage points higher than the actual growth rates, and those in the 1990s are on average 1.4 percentage points higher than the realizations. This upward bias is found to be statistically significant. Hence, the year-ahead forecast violates the weakest form of the rational expectations hypothesis.

Given this miserable record of the government forecast, the second question is whether we could make better forecasts using the same information the government possessed at that time. Section 4 is the first attempt in the literature to answer this question. We consider two alternative forecasts: a vector autoregression (VAR) forecast and a vector error correction model (VECM) forecast, both of which utilize only the real-time data available when the government released its forecasts. The encompassing test[1] reveals that both the current-year and the year-ahead forecasts of the government fail to exploit useful information contained in these VAR and VECM forecasts.[2] Namely, both of the government forecasts could be improved by simply taking account of past statistical relationships between basic economic variables. The mean squared forecast error of the government's year-ahead forecasts is significantly larger than that of the year-ahead VAR/VECM forecasts. Furthermore, given the VAR forecast or the VECM

forecast, the year-ahead forecast of the government provides no additional information useful in forecasting the growth rate. Therefore, during the past two decades, it was better to rely on a mechanical forecast like VAR rather than to employ the government forecast.

The analysis above, however, may be mere hindsight. To eliminate this possibility, we investigate what other forecasters did at the time the government released its forecasts. Section 4 examines the mean forecast of private institutions that release forecasts in every early December as another benchmark. These institutions could use no better information than the Japanese government could, but their mean forecasts are significantly more accurate than the corresponding government forecasts. (Ashiya (2005) finds that almost all of the year-ahead forecasts made by these individual institutions are significantly better than the government forecast. See note 15.) Moreover, both the current-year and the year-ahead forecasts of the government contain little additional information given the mean forecasts of private institutions. Therefore the government forecasts are inferior to the commercial forecasts.

Some might argue that the Japanese government in the 1990s knew the true economic conditions but she attempted to stimulate consumer spending by optimistic forecasts. Section 4 deals with this issue by dividing the data into two sub-samples, the 1980s and the 1990s. If the above argument were correct, the government forecasts in the 1980s must be accurate because the government would not have an incentive to release biased forecasts during the prosperous decade. However, Section 4-3 finds that the year-ahead forecasts of the government are inferior to the VAR forecasts and the commercial forecast in both the 1980s and the 1990s. Namely, the government forecasts are inaccurate regardless of the business conditions. This result casts serious doubt on the assertion that the government knew the true economic conditions in the 1990s.

As for the United States, Romer and Romer (2000) investigate inflation and growth rate forecasts of Federal Reserve. They find that Federal Reserve possesses valuable information not contained in the commercial forecasts. Our finding is in a striking contrast to theirs.

The paper is organized as follows. Section 2 explains the data, and Section 3 reports the results of rationality tests. Section 4 assesses relative performance of the government forecasts. Section 5 concludes the paper.

3

2. Data

The Japanese government releases a forecast of the Japanese real GDP growth rate for the ongoing fiscal year and for the next fiscal year in every late December. We name the former $f_{t,t}$ and the latter $f_{t,t+1}$: $f_{t,t}$ is a four-months-ahead forecast for fiscal year t, and $f_{t,t+1}$ is a 16-months-ahead forecast for fiscal year $t+1$. For example, the government released a forecast for fiscal year 2002 (from April 2002 to March 2003) and for fiscal year 2003 (from April 2003 to March 2004) in December 2002. We call the former $f_{2002,2002}$ and the latter $f_{2002,2003}$.

To avoid the effect of the second Oil Shock, we use the forecasts released from December 1980 on. That is, we use $f_{t,t}$ for the fiscal years 1980 through 2002 and $f_{t,t+1}$ for the fiscal years 1981 through 2002. Figure 1 presents the raw data.

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.[3] Figure 1 indicates that the 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.

3. Tests for rationality

A rational forecast should not be biased, and should use available information efficiently. This section analyzes unbiasedness and efficiency of the government forecasts based on the properties of the forecast error.[4] It will find that the four-months-ahead forecast ($f_{t,t}$) is unbiased and efficient, but that the 16-months-ahead forecast ($f_{t,t+1}$) is not.

3-1. Tests for unbiasedness

A forecast is unbiased if its average deviation from the outcome is zero. Holden and Peel (1990) suggest a simple regression of the forecast error on a constant term as

4

the test for unbiasedness. Define $FE_{t,t} \equiv f_{t,t} - g_t$ and $FE_{t,t+1} \equiv f_{t,t+1} - g_{t+1}$. We estimate

$$FE_{t,t} = \alpha + u_t \tag{1}$$

and $FE_{t,t+1} = \alpha + u_t$. $\tag{2}$

The null hypothesis is $\alpha = 0$. We also check serial correlation in u_t, since forecasts can be improved if there is serial correlation in the error term.

As for regression (2), we must take the information set of the government into account. Since the government does not know the actual growth rate of the ongoing fiscal year (g_t) when it releases $f_{t,t+1}$, $FE_{t,t+1}$ might exhibit a first-order moving average error. In that case, OLS estimate of standard error might be biased downward, although its estimate of α is consistent.

Table 1 shows the results. $Q(r)$ is the Ljung-Box Q-statistic, which tests serial correlation in u_t up to r-th order. The significance levels at which the null hypothesis (no bias or no serial correlation) might be rejected are given in parentheses. The first row shows that $f_{t,t}$ is not biased and that there is no auto-correlation in u_t. On the other hand, the second row shows that $f_{t,t+1}$ is remarkably optimistic: it is on average 0.6864 percentage points higher than the actual growth rate. Although there is no auto-correlation in u_t, the null of no bias is rejected at the 0.10 significance. Since unbiasedness is a necessary condition for efficiency, $f_{t,t+1}$ is considered to be inefficient and hence irrational.

To investigate whether $f_{t,t+1}$ is biased throughout the sample period, we divide the sample in the middle point. Namely, we break the full sample at fiscal year 1992, and rerun (1) and (2). Table 2 shows that $f_{t,t}$ is not biased in either sub-sample (the first and the second rows) and that $f_{t,t+1}$ is not biased in the first sub-sample, i.e., from $t = 1980$ to 1990 (the third row). The fourth row shows, however, that $f_{t,t+1}$ is on average 1.436 percentage points higher than the actual growth rate in the second sub-sample ($t = 1991$ to 2001). The null of no bias is rejected at the 0.05 significance. This result indicates that the Japanese government was optimistic about the future

5

exactly while the Japanese economy was stagnant.

3-2. Tests for efficiency

A forecast is efficient if its forecast error is not related to information available when the forecast was made. A test of weak efficiency is represented by the realization-forecast equations

$$g_t = \alpha + \beta \cdot f_{t,t} + u_t \tag{3}$$

and $\quad g_{t+1} = \alpha + \beta \cdot f_{t,t+1} + u_t . \tag{4}$

The null hypothesis is $(\alpha, \beta) = (0,1)$. If β is significantly different from one, there is correlation between $f_{t,t}$ ($f_{t,t+1}$) and $FE_{t,t}$ ($FE_{t,t+1}$) and hence the forecast could be improved. If g_t, $f_{t,t}$, and $f_{t,t+1}$ are non-stationary and follow unit root process, however, these regressions would yield downwardly biased estimates for β. Therefore we first examine stationarity of the data.

Table 3 shows the results of the augmented Dickey-Fuller (ADF) test for the unit roots. The test regressions are alternatively specified with an intercept term and both intercept term and deterministic trend. The favored specification is determined using the Schwarz BIC. Similarly, lag order k is selected by means of the BIC. The significance levels for the null of non-stationarity are given in parentheses. Since we cannot reject the null hypothesis of unit root for g_t, $f_{t,t}$, and $f_{t,t+1}$, we follow the methodology of Engle and Yoo (1991) and Aggarwal et al. (1995, 2000), and estimate the three-step error correction model below. Take the case of $f_{t,t}$ for example.

Step 1. Estimate the regression (3), and calculate $\hat{u}_t = g_t - \hat{\alpha} - \hat{\beta} \cdot f_{t,t}$.

Step 2. Estimate

$$\Delta g_t = -\gamma_0 \cdot \hat{u}_{t-1} + \gamma_1 \cdot \Delta f_{t,t} + \gamma_2 \cdot \Delta g_{t-1} + \gamma_3 \cdot \Delta f_{t-1,t-1} + \varepsilon_t$$

where $\Delta g_t \equiv g_t - g_{t-1}$ and $\Delta f_{t,t} \equiv f_{t,t} - f_{t-1,t-1}$

and calculate $\hat{\varepsilon}_t$.

Step 3. Estimate

$$\hat{\varepsilon}_t = \delta_0 + \delta_1 \left(\hat{\gamma}_0 \cdot f_{t-1,t-1} \right) + \mu_t .$$

The corrected estimate of β in regression (3), β^*, is given as

$$\beta^* = \hat{\beta} + \hat{\delta}_1.$$

The standard error of β^* equals that of $\hat{\delta}_1$. The corrected estimate of β in regression (4) is calculated by the same way.

Table 4 reports the results of regression (3) and (4). Standard errors of estimated coefficients are in parentheses. For either regression, the null of $\alpha = 0$, $\beta = 1$, and $(\alpha, \beta) = (0,1)$ are not rejected at the 0.10 significance level. However, these results are not meaningful because the forecasts and the realizations are non-stationary and thus estimated β is downwardly biased.

Table 5 shows the results of the ADF tests for cointegration, which examine whether the residuals of (3) and (4) are stationary. Lag order k is selected by means of the BIC, and the significance levels for the null are given in parentheses. The result shows that the null hypothesis of no-cointegration is rejected for either regression. Namely, both $(g_t, f_{t,t})$ and $(g_{t+1}, f_{t,t+1})$ are found to be cointegrated.

Table 6 presents the corrected estimate of β, β^*, in regression (3) and (4) based on the three-step error correction model explained above. Standard errors are in parentheses. Rationality requires that the cointegrating vector β^* must be unity. The null of $\beta^* = 1$ is not rejected for regression (3), but is rejected at the 0.05 significance level for regression (4).

To confirm the above results, we modify (3) and (4) and test the following regressions:

$$g_t - g_{t-1} = \alpha + \beta\left(f_{t,t} - g_{t-1}\right) + u_t \tag{3'}$$

$$\text{and} \quad g_{t+1} - g_t = \alpha + \beta\left(f_{t,t+1} - g_t\right) + u_t. \tag{4'}$$

The null hypothesis is $(\alpha, \beta) = (0,1)$. We also estimate the modified regression Jansen and Kishan (1996) propose:

$$g_{t+1} - f_{t,t+1} = \alpha + \gamma_1\left(f_{t,t+1} - f_{t,t}\right) + \gamma_2\left(f_{t,t} - g_{t-1}\right) + u_t. \tag{4''}$$

The null hypothesis of (4'') is $\alpha = \gamma_1 = \gamma_2 = 0$. Table 7 shows the results of (3'), (4'), and (4''). Again we find significant inefficiency in $f_{t,t+1}$: the null of $(\alpha, \beta) = (0,1)$ is

7

rejected at the 0.05 significance level for regression (4').

To sum up, while there is no evidence of bias and inefficiency in the four-months-ahead forecast ($f_{t,t}$), severe upward bias and inefficiency are detected in the 16-months-ahead forecast ($f_{t,t+1}$). Therefore the 16-months-ahead forecast is inconsistent with rationality.

4. Relative accuracy

As a further test of efficiency, this section measures accuracy of the government forecasts relative to other forecasts that are based only on information available prior to the release of the government forecasts. Subsection 4-1 introduces five benchmark forecasts. Subsection 4-2 evaluates relative accuracy of the four-months-ahead forecast ($f_{t,t}$), and Subsection 4-3 evaluates the 16-months-ahead forecast ($f_{t,t+1}$). Subsection 4-4 relaxes the assumption of the quadratic loss function and considers the "sign test", which is valid for non-quadratic, asymmetric, or discontinuous loss functions. This section will find that the government forecasts are inferior to the mean forecasts of private institutions, and that they contain little additional information given VAR forecasts or the mean forecasts of private institutions.

4-1. Benchmark forecasts

We consider the following five forecasts as a benchmark. The first two are naïve forecasts.

(a) Same-change forecast ($f_{t,t}^{Same}$ and $f_{t,t+1}^{Same}$):

$$f_{t,t}^{Same} \equiv f_{t,t+1}^{Same} \equiv g_{t-1}$$

(Remember that the government does not know g_t when it releases $f_{t,t+1}$). This "naïve" forecast might be optimal if g_t would follow a random walk process.

(b) Trend forecast ($f_{t,t}^{Trend}$ and $f_{t,t+1}^{Trend}$):

$$f_{t,t}^{Trend} \equiv \frac{1}{10} \sum_{i=1}^{10} g_{t-i} \text{ , and}$$

$$f_{t,t+1}^{Trend} \equiv \frac{1}{10} \sum_{i=1}^{9} g_{t-i} + \frac{1}{10} f_{t,t}^{Trend}$$

8

$$= \frac{11}{100} \sum\nolimits_{i=1}^{9} g_{t-i} + \frac{g_{t-10}}{100}.$$

These forecasts assume instant mean reversion.

The second group is time-series forecasts. Since Section 3 has found that the growth rate might be nonstationary, a proper model specification is either a vector error correction model (VECM) or a vector autoregression model (VAR) with the lag-specification of Toda and Yamamoto (1995). However, (1) these methods are so sophisticated that the government in the early 1990s would not have known them, and (2) our motive here is to check whether we could make better forecasts using the same information the government possessed at that time. For these two reasons, we consider the VAR in differences and the VECM, both of which are based only on the real-time data available when the government released its forecasts.[5] Note that the government could compute these VAR forecasts by itself before it released own forecast.

(c) VAR forecast ($f_{t,t}^{VAR}$ and $f_{t,t+1}^{VAR}$).

We construct the VAR in differences with four lags and a constant term consists of four variables: real GNP (Y), GNP deflator (P), real private final-consumption-expenditure (C), and the call money rate (R). The model is specified as follows:

$$\Delta X_t = \alpha^X + \sum\nolimits_{i=1}^{4} \beta_i^X \Delta Y_{t-i} + \sum\nolimits_{i=1}^{4} \gamma_i^X \Delta P_{t-i} + \sum\nolimits_{i=1}^{4} \delta_i^X \Delta C_{t-i} + \sum\nolimits_{i=1}^{4} \omega_i^X \Delta R_{t-i} + \varepsilon_t^X$$

where $X \in \{Y, P, C, R\}$.

We focus on pseudo out-of-sample forecasts in order to simulate how the VAR forecasts would have been computed in real time. More precisely, to calculate the VAR forecasts of year t, we first estimate the model using real-time quarterly data from 1953 through the second quarter of fiscal year t (i.e. July-September of year t).[6] Then the estimated coefficients are used to generate a two-quarter-ahead forecast, $f_{t,t}^{VAR}$, and a six-quarter-ahead forecast, $f_{t,t+1}^{VAR}$.

(d) VECM forecasts ($f_{t,t}^{VECM}$ and $f_{t,t+1}^{VECM}$)

We use cointegrating vectors estimated in the Johansen (1991) test procedure to construct a vector error correction model (VECM). The Johansen test suggests one cointegrating vector in the data used for forecasts of year 1994-1999 and 2002 and two cointegrating vectors in other data. Hence, the model is specified as follows:

9

$$\Delta X_t = \alpha^X + \mu_1^X EC(1)_{t-1} + \mu_2^X EC(2)_{t-1} + \sum_{i=1}^{4} \beta_i^X \Delta Y_{t-i} + \sum_{i=1}^{4} \gamma_i^X \Delta P_{t-i} + \sum_{i=1}^{4} \delta_i^X \Delta C_{t-i} + \sum_{i=1}^{4} \omega_i^X \Delta R_{t-i} + \varepsilon_t^X$$

where $X \in \{Y, P, C, R\}$ and $EC(n)$ is n-th cointegrating relation error correction term (Only $EC(1)$ is included for the estimations of year 1994-1999 and 2002). The VECM forecasts are generated by the same way as the VAR forecasts.

The last benchmark forecast is the consensus forecast of private sector. We use the data from "Monthly Statistics (Tokei Geppo)".[7]

(e) Consensus (mean) forecast of private institutions ($f_{t,t}^{Mean}$ and $f_{t,t+1}^{Mean}$)

Toyo Keizai Inc. has published the forecasts of about 70 Japanese private institutions in the February or March issue of "Monthly Statistics (Tokei Geppo)" since the 1970s. In every early December, each institution makes forecasts of the Japanese real GDP growth rate for the ongoing fiscal year and for the next fiscal year. We call their mean forecasts for the ongoing/next fiscal year $f_{t,t}^{Mean} / f_{t,t+1}^{Mean}$.[8]

4-2. Relative accuracy of $f_{t,t}$

This subsection evaluates relative accuracy of the government's four-months-ahead forecast ($f_{t,t}$). Table 8 presents the root mean square error (RMSE) and the mean absolute error (MAE) for each forecast. $f_{t,t}$ is superior to the same-change forecast ($f_{t,t}^{Same}$) and the trend forecast ($f_{t,t}^{Trend}$). However, $f_{t,t}$ is no better than $f_{t,t}^{VAR}$ or $f_{t,t}^{VECM}$, and is less accurate than $f_{t,t}^{Mean}$.

Next we test whether there exists statistically significant difference in forecast accuracy among these six models. Ericsson (1992, p.483) shows that "forecast encompassing" proposed by Chong and Hendry (1986) is the sufficient condition of the RMSE dominance. The encompassing test evaluates two forecasting systems, f^1 and f^2, by the following regression:

$$g_t = \delta f_{t,t}^1 + (1-\delta) f_{t,t}^2 + u_t. \tag{5}$$

If δ is positive and significantly different from zero, it indicates that f^1 provides some information concerning g_t, which information is excluded from f^2. In this case f^1 is said to "encompass" f^2. Similarly, if $1-\delta$ is positive and significantly

different from zero, then f^2 is said to encompass f^1.

Ericsson (p.485, l.1) argues, however, that "the forecast-encompassing test may have no power when the dependent variable is $I(1)$." Since Section 3 has shown that g_t may contain a unit root, we follow Ericsson (1992) and Jansen and Kishan (1996) and estimate the transformed encompassing test

$$g_t - f_{t,t}^2 = \delta\left(f_{t,t}^1 - f_{t,t}^2\right) + u_t. \tag{6}$$

If δ is significantly larger than zero, then f^1 encompasses f^2. If δ is significantly smaller than unity, then f^2 encompasses f^1. If $\delta = 0$ ($\delta = 1$), then f^1 (f^2) provides no additional information useful in forecasting g_t given f^2 (f^1).[9]

Table 9 presents the results of (6). Standard errors are in parentheses. The significance levels for the null of $\delta = 0$ are given in "P-value ($\delta = 0$)", and the significance levels for the null of $\delta = 1$ are given in "P-value ($\delta = 1$)".[10]

The first and second rows of Table 9 show that $f_{t,t}$ encompasses both $f_{t,t}^{Same}$ and $f_{t,t}^{Trend}$ at the 0.01 significance. Furthermore, $\delta > 1$ in both cases. Therefore the government forecast efficiently utilizes all information in these naïve forecasts, and its RMSE is significantly smaller than that of naïve forecasts. On the other hand, the third row shows that $f_{t,t}$ and $f_{t,t}^{VAR}$ do not encompass each other. This indicates that $f_{t,t}$ and $f_{t,t}^{VAR}$ contain independent information and that $f_{t,t}$ could be improved by using information in past economic statistics. This result is remarkable because the government has better information than the VAR model: it has observed the economic trend in October and November of year t, while the VAR model has not.

What is more important is that both $f_{t,t}^{VECM}$ and $f_{t,t}^{Mean}$ encompass $f_{t,t}$ at the 0.10 significance (the fourth and the fifth rows). It indicates that the RMSE of the government forecast ($f_{t,t}$) is significantly larger than that of $f_{t,t}^{VECM}$ (the VECM model) or $f_{t,t}^{Mean}$ (the mean forecast of private institutions). Moreover, given $f_{t,t}^{VECM}$ or $f_{t,t}^{Mean}$, $f_{t,t}$ provides little additional information useful in forecasting g_t.[11]

We also conduct the generalized encompassing test

11

$$g_t = \alpha + \delta_1 \cdot f_{t,t}^1 + \delta_2 \cdot f_{t,t}^2 + u_t . \tag{5'}$$

Regression (5') relaxes the restrictions of $\alpha = 0$ and $\delta_1 + \delta_2 = 1$ in (5) and (6).[12] If δ_i ($i = 1, 2$) is significantly larger than zero, then f^i contains useful information not in the constant term and f^j ($j \neq i$). If $\delta_i = 0$, f^i provides no additional information useful in forecasting g_t given f^j.

The estimation results are summarized in Table 10. Standard errors are in parentheses. The null of the F-test is $(\alpha, \delta_1 + \delta_2) = (0,1)$. Since this joint null hypothesis is not rejected at the 0.10 significance for any regression, the restrictions imposed on (6) are found to be not binding. Therefore it is natural that we find the same result in Table 10 as in Table 9. The coefficients of the government forecasts, δ_1, are significant and larger than unity in the first and second rows. It indicates that $f_{t,t}$ contains all information in these naïve forecasts. On the other hand, δ_2 is much larger than δ_1 in the third, fourth, and fifth rows: $f_{t,t}$ neglects information contained in $f_{t,t}^{VAR}$, $f_{t,t}^{VECM}$, or $f_{t,t}^{Mean}$.

We further investigate the relative accuracy of $f_{t,t}$ by dividing the full sample in the middle point, i.e. $t = 1992$. Table 11 presents the results of regression (6). The first, second, and the third rows show that $f_{t,t}$ is encompassed by $f_{t,t}^{VAR}$, $f_{t,t}^{VECM}$, and $f_{t,t}^{Mean}$ in the first sub-sample ($t = 1980$ to 1991). On the other hand, the fourth and the fifth rows show that $f_{t,t}$ encompasses $f_{t,t}^{VAR}$ and $f_{t,t}^{VECM}$ in the second sub-sample ($t = 1992$ to 2002).

4-3. Relative accuracy of $f_{t,t+1}$

This subsection evaluates relative accuracy of the government's 16-months-ahead forecast ($f_{t,t+1}$). Table 12 presents the root mean square error (RMSE) and the mean absolute error (MAE) for each forecast. $f_{t,t+1}$ is superior to the same-change forecasts and the trend forecasts, but is inferior to other three forecasts.[13]

As for the encompassing test, Table 13 shows the results of regression (6):

$$g_{t+1} - f_{t,t+1}^2 = \delta\left(f_{t,t+1}^1 - f_{t,t+1}^2\right) + u_t.$$

Standard errors are in parentheses. The significance levels for the null of $\delta = 0$ are given in "P-value ($\delta = 0$)", and the significance levels for the null of $\delta = 1$ are given in "P-value ($\delta = 1$)".[14]

   Table 13 reveals three important points. The first point to notice is that the government forecast is significantly superior to the naïve forecasts: the first and second rows show that $f_{t,t+1}$ encompasses both $f_{t,t+1}^{Same}$ and $f_{t,t+1}^{Trend}$ at the 0.01 significance. The second point is that both $f_{t,t+1}^{VAR}$ and $f_{t,t+1}^{VECM}$ encompass $f_{t,t+1}$. It indicates that vector autoregression forecasts are significantly more accurate than $f_{t,t+1}$, though the government must have had informational advantage over the VAR models. Moreover, since δ in the third and fourth rows are nearly equal to zero, $f_{t,t+1}$ provides no additional information useful in forecasting g_{t+1}, given $f_{t,t+1}^{VAR}$ or $f_{t,t+1}^{VECM}$. The third point is that $f_{t,t+1}^{Mean}$ also encompasses $f_{t,t+1}$ and $\delta < 0$ in this regression. It shows that the mean forecast of private institutions is significantly more accurate than the government forecast. Furthermore, it indicates that the government forecast provides no additional information given the mean forecast of private institutions.[15]

   Table 14 shows the results of the generalized encompassing test (5'):

$$g_{t+1} = \alpha + \delta_1 \cdot f_{t,t+1}^1 + \delta_2 \cdot f_{t,t+1}^2 + u_t.$$

Standard errors are in parentheses. The null of the F-test is $(\alpha, \delta_1 + \delta_2) = (0, 1)$, which is not rejected at the 0.10 significance for any regression. The result in Table 14 is the same as that in Table 13. The coefficients of the government forecasts, δ_1, in the first and second rows are significant and larger than unity, which indicate $f_{t,t+1}$ contains all information in these naïve forecasts. The third, fourth, and fifth rows show that the coefficients of $f_{t,t+1}^{VAR}$, $f_{t,t+1}^{VECM}$ and $f_{t,t+1}^{Mean}$ are all significant and larger than unity. It proves that $f_{t,t+1}$ neglects information contained in $f_{t,t+1}^{VAR}$, $f_{t,t+1}^{VECM}$, or $f_{t,t+1}^{Mean}$. Moreover, the coefficients of $f_{t,t+1}$ in these rows are not significantly different from zero. It indicates that $f_{t,t+1}$ contains no additional information given $f_{t,t+1}^{VAR}$, $f_{t,t+1}^{VECM}$, or

13

$f_{t,t+1}^{Mean}$.

Observing this result, some might worry that it ignores the possibility that the Japanese government uses an asymmetric loss function. The Japanese economy was in deep trouble throughout the 1990s. If the government was attempting to assure consumers and encourage consumer spending during this recession, then it may be optimal to produce a rosy picture of the truth. In other words, the government may be very averse to producing a forecast below the true growth rate, since it might discourage consumer confidence. During the boom, however, the government need not stimulate consumer spending and hence need not release a biased forecast. Since the Japanese economy had enjoyed the longest boom in the 1980s, the government forecast must be accurate in the 1980s if an asymmetric loss function is the cause of the government's poor forecast record.

Table 15 shows the estimation result of regression (6) using the data of the 1980s and the 1990s separately. It shows that $f_{t,t+1}$ is encompassed by $f_{t,t+1}^{VAR}$, $f_{t,t+1}^{VECM}$, and $f_{t,t+1}^{Mean}$ in both sub-samples. Namely, it confirms that the year-ahead forecast of the Japanese government is inaccurate no matter what phase the business cycle is. This result casts doubt on the assertion that the government adopts an asymmetric loss function.

4-4. Generalized loss function

Diebold and Mariano (1995, p253) argue that "realistic economic loss functions frequently do not conform to stylized textbook favorites like mean squared prediction error." Therefore this subsection relaxes the assumption that the loss associated with a particular forecast error is proportional to its square. More specifically, we evaluate relative accuracy of the government forecasts using the "sign test" explained below.

Let T be the sample size, and let n be the number of years the government "wins", i.e., its absolute forecast error is smaller than that of the benchmark forecast. Then

$$S \equiv \frac{n - 0.5T}{\sqrt{0.25T}}$$

is asymptotically distributed according to $N(0,1)$. If S is significantly larger (smaller) than zero, then the government forecast is superior (inferior) to the benchmark forecast.

14

Note that this test is valid for non-quadratic, asymmetric, or discontinuous loss functions.

Table 16 shows the result, which is consistent with the previous subsections. The government forecast is significantly better than the naïve forecasts (except $f_{t,t+1}^{Trend}$). However, its year-ahead forecast is significantly inferior to the VECM forecast and the mean forecast of private institutions.

5. Conclusions

This paper evaluates accuracy and efficiency of the real GDP forecasts made by the Japanese government over the past 22 years. The government 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.

Section 3 has found that the four-months-ahead forecast and the actual growth rate are cointegrated with factor one. However, the 16-months-ahead forecast is on average 0.7 percentage points higher than the actual growth rate, and the cointegrating vector is significantly different from unity. The bias is much larger in the 1990s, during which the Japanese economy was in turmoil. These results demonstrate that the four-months-ahead forecast supports the rational expectations hypothesis, but that the 16-months-ahead forecast does not.

Section 4 estimated the encompassing test, and has revealed five important points. First, the government forecasts efficiently utilize all information in the naïve forecasts. Second, the 16-months-ahead forecast of the government is significantly inferior to the corresponding VAR forecasts, independent of the assumption of the quadratic loss function. Third, the government forecast fails to exploit useful information contained in the VAR forecasts. What is worse, its 16-months-ahead forecast contains no additional information given the VAR forecasts. Fourth, the mean forecasts of Japanese private institutions encompass the corresponding government forecasts. In other words, both the four-months-ahead forecast and the 16-months-ahead forecast of the Japanese government are significantly less accurate than the corresponding mean forecasts of private institutions. Finally, when the data of the 1980s and the 1990s are analyzed separately, the 16-months-ahead forecast of the government in either sub-sample is

15

inferior to the VAR forecasts and the mean forecast of private institutions. It indicates that the Japanese government produces inferior forecasts regardless of the phases of the business cycle.

The Japanese government could take various fiscal and monetary measures to accomplish the predicted growth rate. Moreover, it had a strong incentive to do so. The official GDP forecast is consented by the cabinet and is reported to the National Diet. The cabinet faces severe criticism and suffers political damage if it fails to achieve the predicted growth rate. Taking account of these circumstances, the messy forecast record of the government astonishes us. The government might have been carrying out inadequate economic policies based on an illusory picture of the Japanese economy.

Notes

1. See Cooper and Nelson (1975), Chong and Hendry (1986), Mizon and Richard (1986), Ericsson (1992), and Jansen and Kishan (1996).

2. The conclusion that the government forecast is less accurate than the VAR/VECM forecasts does not change even if we employ non-quadratic or asymmetric loss functions. See Section 4-4.

3. We obtain the same results by using the revised data of g_t released in June of year $t+2$.

4. See Batchelor and Dua (1991), Artis (1996), Jansen and Kishan (1996), Pons (2000), Joutz and Stekler (2000), Öller and Barot (2000), Fildes and Stekler (2002), and references therein for the related literature.

5. We obtain the same result when we (inadequately) use the forecasts of the VAR in levels.

6. The government has released the realizations of the second quarter by every early December. Hence, it must have possessed these data when it released $f_{t,t}$ and $f_{t,t+1}$.

7. Ashiya (2003) also uses the data from "Monthly Statistics (Tokei Geppo)".

8. Ashiya (2005) investigates individual forecasts of 38 private institutions that participated in 15 or more surveys of "Monthly Statistics" since December 1980.

9. When $\delta < 0$ ($\delta > 1$), f^1 (f^2) is so inaccurate that the minimum squared error composite requires "short selling" of it (Cooper and Nelson, 1975, p.8).

10. Since $\delta > 1$ in the first and second rows, "P-value ($\delta = 1$)" in these rows are meaningless for the encompassing test.

11. Ashiya (2005) finds that, among 38 forecast series of individual private institutions, 13 series encompass the government forecast and 19 series are encompassed by the government forecast.

12. Cooper and Nelson (1975), Fair and Shiller (1990), Joutz and Stekler (2000), Romer and Romer (2000), and Batchelor (2001) estimate (5') for the test of encompassing.

13. As for the US economy, Zarnowitz and Braun (1993) investigate NBER-ASA survey GDP forecasts from 1968 to 1990. They find that two-quarter-ahead Bayesian VAR forecast is inferior to 75% of individual forecasts, but that five-quarter-ahead BVAR forecast is superior to 75% of individual forecasts. On the other hand, Anderson et al.

(2002) compare various year-ahead forecasts of real GDP and CPI for 1990-1998, and find that, based on the RMSE, VECM forecast is inferior to federal policymakers or other institutional forecasters.

14. Since $\delta > 1$ in the first and second rows, "*P*-value ($\delta = 1$)" in these rows are meaningless for the encompassing test. Similarly, since $\delta < 0$ in the fifth row, "*P*-value ($\delta = 0$)" in this row is meaningless for the encompassing test.

15. Ashiya (2005) finds that 35 out of 38 forecast series of individual private institutions encompass the government forecast, and no series is encompassed by the government forecast.

Acknowledgements

References

Aggarwal, R., Mohanty, S., Song, F., 1995. "Are Survey Forecasts of Macroeconomic Variables Rational?" Journal of Business, 68, 99-119.

Aggarwal, R., Mohanty, S., 2000. "Rationality of Japanese Macroeconomic Survey Forecasts: Empirical Evidence and Comparisons with the US." Japan and the World Economy, 12, 21-31.

Anderson, Richard G., Hoffman, Dennis L., Rasche, Robert H., 2002, "A Vector Error-Correction Forecasting Model of the US economy." Journal of Macroeconomics, 24, 569-598.

Artis, Michael J., 1996. "How Accurate are the IMF's Short-term Forecasts? Another Examination of the World Economic Outlook." International Monetary Fund Working Paper, WP/96/89.

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, Masahiro, 2005. "Twenty-two Years of Japanese Institutional Forecasts." Applied Financial Economics Letters, 1, 79-84.

Batchelor, Roy, 2001. "How Useful Are the Forecasts of Intergovernmental Agencies? The IMF and OECD versus the Consensus." Applied Economics, 33, 225-235.

Batchelor, Roy, Dua, Pami, 1991. "Blue Chip Rationality Tests." Journal of Money, Credit, and Banking, 23(4), 692-705.

Chong, Yock Y., Hendry, David F., 1986, "Econometric Evaluation of Linear Macro-Economic Models." Review of Economic Studies, 53, 671-690.

Cooper, J. Phillip, Nelson, Charles R., 1975, "The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictors." Journal of Money, Credit, and Banking, 7(1), 1-32.

Diebold, Francis X., Mariano, Roberto S., 1995, "Comparing Predictive Accuracy." Journal of Business and Economic Statistics, 13, 253-263.

Engle, R.F., Yoo, S., 1991, "Cointegrated Economic Time Series: An Overview with New Results." in Engle, R.F. and Granger, C.W.J. eds. *Long-Run Economic Relationships*, Oxford University Press, Oxford.

Ericsson, Neil R., 1992, "Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration." Journal of Policy Modeling, 14, 465-495.

Fair, Ray C., Shiller, Robert J., 1990,"Comparing Information in Forecasts from Econometric Models." American Economic Review, 80, 375-389.

Fildes, Robert, Stekler, Herman, 2002, "The State of Macroeconomic Forecasting." Journal of Macroeconomics, 24, 435-468.

Holden, K., Peel, D.A., 1990, "On Testing for Unbiasedness and Efficiency of Forecasts." The Manchester School, 58, 120-127.

Jansen, Dennis W., Kishan, Ruby Pandey, 1996, "An Evaluation of Federal Reserve Forecasting." Journal of Macroeconomics, 18, 89-109.

Johansen, Soren, 1991, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." Econometrica, 59, 1551-1580.

Joutz, Fred, Stekler, H.O., 2000, "An Evaluation of the Predictions of the Federal Reserve." International Journal of Forecasting, 16, 17-38.

Keane, M.P., Runkle, D.E., 1990. "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." American Economic Review, 80, 714-735.

Mizon, Grayham E., Richard, Jean-Francois, 1986, "The Encompassing Principle and Its Application to Testing Non-Nested Hypotheses." Econometrica, 54, 657-678.

Öller , Lars-Erik, Barot, Bharat, 2000. "The Accuracy of European Growth and Inflation Forecasts." International Journal of Forecasting, 16, 293-315.

Pons, Jordi, 2000, "The Accuracy of IMF and OECD Forecasts for G7 Countries." Journal of Forecasting, 19, 53-63.

Romer, Christina D., Romer, David H., 2000, "Federal Reserve Information and the Behavior of Interest Rates." American Economic Review, 90, 429-457.

Stark, Tom, Croushore, Dean, 2002, "Forecasting with a Real-time Data Set for Macroeconomists." Journal of Macroeconomics, 24, 507-531.

Toda, Y. Hiro, Yamamoto, Taku, 1995, "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes." Journal of Econometrics, 66, 225-250.

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

Zarnowitz, Victor, Braun, Phillip, 1993. "Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance." in Stock, James H., Watson, Mark W. eds. *Business cycles, indicators, and forecasting. NBER Studies in Business Cycles, vol. 28.* Chicago and London: University of Chicago Press, 11-84.

Figure 1: Forecasts and realizations
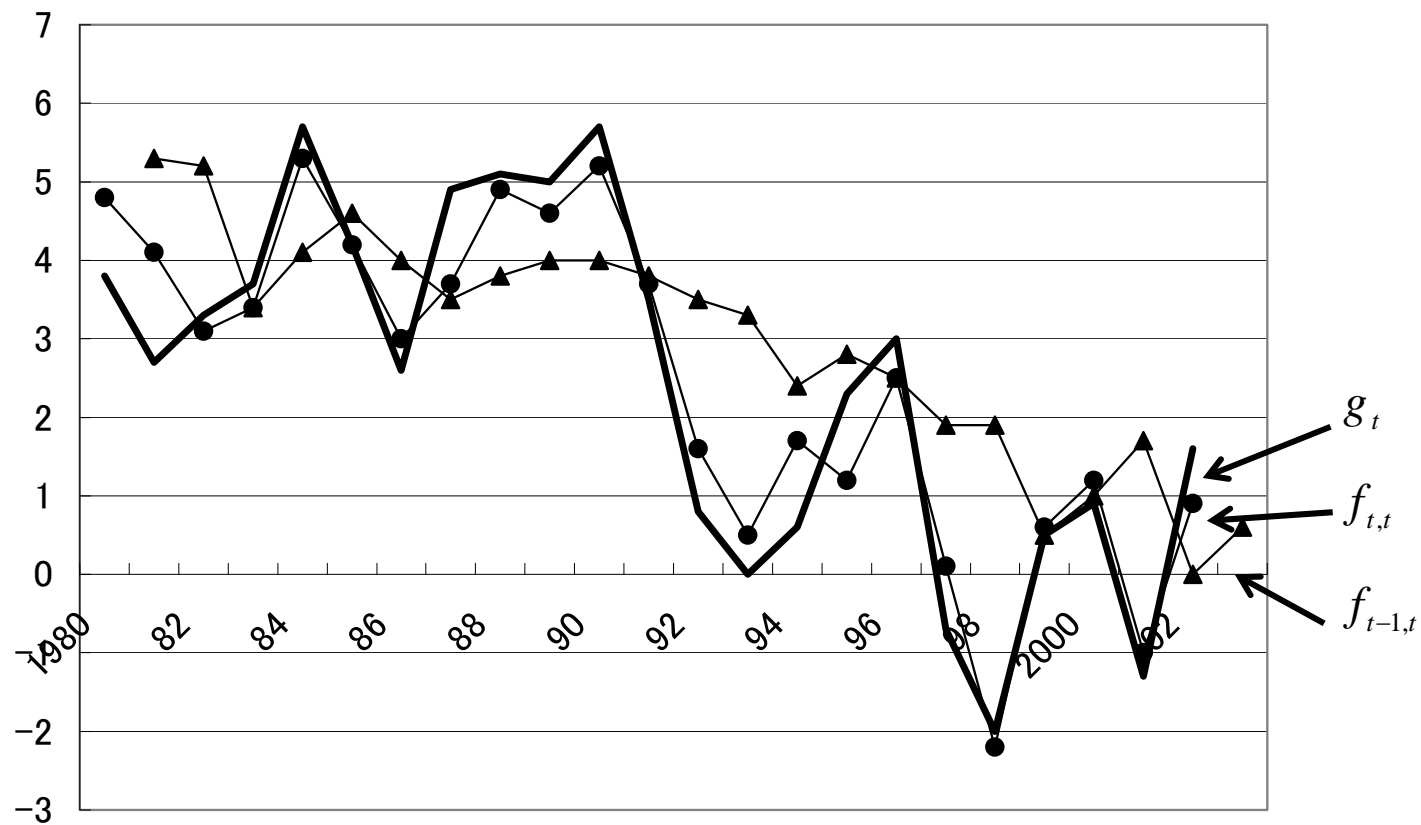
g_t : realization, $f_{t,t}$: current-year forecast, $f_{t-1,t}$: year-ahead forecast.

22

Table 1: Tests for unbiasedness

Model: (1) $FE_{t,t} = \alpha + u_t$

  (2) $FE_{t,t+1} = \alpha + u_t$

Model	α	$Q(1)$	$Q(2)$	$Q(3)$	$Q(4)$
(1)	0.0522 (0.717)	0.237 (0.627)	1.001 (0.606)	1.041 (0.791)	1.867 (0.760)
(2)	0.6864 (0.088)	1.695 (0.193)	2.378 (0.305)	2.386 (0.496)	2.387 (0.665)

Notes

$FE_{t,t} \equiv f_{t,t} - g_t$ and $FE_{t,t+1} \equiv f_{t,t+1} - g_{t+1}$.

$Q(r)$ is the Ljung-Box Q-statistic, which tests serial correlation in u_t up to r-th order.

The significance levels for the null are given in parentheses.

Table 2: Tests for unbiasedness (sub-sample)

Model: (1) $FE_{t,t} = \alpha + u_t$

(2) $FE_{t,t+1} = \alpha + u_t$

Model	t	α	$Q(1)$	$Q(2)$	$Q(3)$
(1)	1980-1991	-0.017 (0.935)	0.793 (0.373)	0.853 (0.653)	0.864 (0.834)
(1)	1992-2002	0.127 (0.554)	0.201 (0.654)	1.315 (0.518)	1.323 (0.724)
(2)	1980-1990	-0.064 (0.891)	1.313 (0.252)	2.244 (0.326)	2.342 (0.505)
(2)	1991-2001	1.436 (0.025)	0.148 (0.700)	1.940 (0.379)	2.067 (0.559)

Notes

$FE_{t,t} \equiv f_{t,t} - g_t$ and $FE_{t,t+1} \equiv f_{t,t+1} - g_{t+1}$.

$Q(r)$ is the Ljung-Box Q-statistic, which tests serial correlation in u_t up to r-th order.

The significance levels for the null are given in parentheses.

Table 3: Unit root tests

	ADF (*P*-value)	trend [a, b]	k [a, c]
g_t	-2.9743 (0.168)	1	0
$f_{t,t}$	-3.0446 (0.150)	1	0
$f_{t,t+1}$	-1.1423 (0.846)	1	2
Δg_t	-4.9158 (0.001)	0	1
$\Delta f_{t,t}$	-5.4507 (0.000)	0	1
$\Delta f_{t,t+1}$	-7.3280 (0.000)	1	1

Notes

The significance levels for the null of non-stationarity are given in parentheses.

a: The favored specification is determined using the Schwarz BIC.

b: Those with trend=1 include both the intercept term and the deterministic trend. Those with trend=0 include the intercept term only.

c: Lag order.

Table 4: Tests for efficiency

Model: (3) $g_t = \alpha + \beta \cdot f_{t,t} + u_t$

(4) $g_{t+1} = \alpha + \beta \cdot f_{t,t+1} + u_t$

Eq.	α (S.E.)	β (S.E.)	$F(2, Obs-2)$ [a]	Obs.
(3)	-0.137 (0.228)	1.034 (0.071)	0.183	23
(4)	-0.668 (0.945)	0.994 (0.281)	1.527	22

Notes

Standard errors of estimated coefficients are in parentheses.

a: F-value for the joint restriction of $(\alpha, \beta) = (0,1)$. The 10% critical value of $F(2,20)$ is 2.589.

Table 5: Cointegration tests

	ADF (*P*-value)	k [a, b]
u_t of (3)	-4.606 (0.008)	1
u_t of (4)	-3.297 (0.093)	0

Notes

The significance levels for the null of no-cointegration are given in parentheses.

a: The favored specification is determined using the Schwarz BIC.

b: Lag order.

Table 6: Estimation of cointegrating vectors

	β^* (S.E.)
(3)	1.014 (0.056)
(4)	1.506 (0.217)**

Notes

β^* is the corrected estimate of β in regression (3) and (4). Standard errors are in parentheses.

**: Significantly different from unity at the 0.05 level.

Table 7: Further tests for efficiency

Model: (3') $g_t - g_{t-1} = \alpha + \beta\left(f_{t,t} - g_{t-1}\right) + u_t$

(4') $g_{t+1} - g_t = \alpha + \beta\left(f_{t,t+1} - g_t\right) + u_t$

(4") $g_{t+1} - f_{t,t+1} = \alpha + \gamma_1\left(f_{t,t+1} - f_{t,t}\right) + \gamma_2\left(f_{t,t} - g_{t-1}\right) + u_t$

Eq.	α (S.E.)	β (S.E.)	γ_1 (S.E.)	γ_2 (S.E.)	F-stat.	Obs.
(3')	-0.028 (0.136)	1.167 (0.095)			1.606 [b]	23
(4')	-0.392 (0.389)	0.498 (0.254)*[a]			3.770** [b]	22
(4")	-0.480 (0.423)		-0.235 (0.396)	-0.354 (0.292)	1.931 [c]	22

Notes

Standard errors of estimated coefficients are in parentheses.
a: The null is $\beta = 1$.
b: F-value for the joint restriction of $(\alpha, \beta) = (0,1)$.
c: F-value for the joint restriction of $\alpha = \gamma_1 = \gamma_2 = 0$.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 8: Prediction performance of $f_{t,t}$

	RMSE	MAE
$f_{t,t}$	0.668	0.548
$f_{t,t}^{Same}$	1.784	1.500
$f_{t,t}^{Trend}$	2.176	1.763
$f_{t,t}^{VAR}$	0.658	0.577
$f_{t,t}^{VECM}$	0.645	0.569
$f_{t,t}^{Mean}$	0.633	0.506

Notes

$f_{t,t}^{Same} \equiv g_{t-1}$ and $f_{t,t}^{Trend} \equiv \frac{1}{10}\sum_{i=1}^{10} g_{t-i}$. $f_{t,t}^{VAR}$ is prediction of the VAR with four lags consists of real GNP, GNP deflator, real consumption, and the call money rate. $f_{t,t}^{VECM}$ is prediction of the VECM with four lags. $f_{t,t}^{Mean}$ is the consensus (mean) forecast of private institutions.

Table 9: Encompassing test for $f_{t,t}$

Model: (6) $g_t - f_{t,t}^2 = \delta\left(f_{t,t}^1 - f_{t,t}^2\right) + u_t$

$\left(f^1, f^2\right)$	δ (S.E.)	P-value ($\delta = 0$)	P-value ($\delta = 1$)
$\left(f_{t,t}, f_{t,t}^{Same}\right)$	1.169 (0.093)	0.000	0.082 [a]
$\left(f_{t,t}, f_{t,t}^{Trend}\right)$	1.123 (0.072)	0.000	0.101 [a]
$\left(f_{t,t}, f_{t,t}^{VAR}\right)$	0.459 (0.343)	0.195	0.128
$\left(f_{t,t}, f_{t,t}^{VECM}\right)$	0.429 (0.285)	0.146	0.057
$\left(f_{t,t}, f_{t,t}^{Mean}\right)$	0.321 (0.371)	0.396	0.081

Notes
Standard errors are in parentheses.

$f_{t,t}^{Same} \equiv g_{t-1}$ and $f_{t,t}^{Trend} \equiv \dfrac{1}{10}\sum_{i=1}^{10} g_{t-i}$. $f_{t,t}^{VAR}$ is prediction of the VAR with four lags consists of real GNP, GNP deflator, real consumption, and the call money rate. $f_{t,t}^{VECM}$ is prediction of the VECM with four lags. $f_{t,t}^{Mean}$ is the consensus (mean) forecast of private institutions.

a: This value is meaningless for the encompassing test because $\delta > 1$.

1

Table 10: Generalized encompassing test for $f_{t,t}$

Model: (5') $g_t = \alpha + \delta_1 \cdot f_{t,t}^1 + \delta_2 \cdot f_{t,t}^2 + u_t$

$\left(f^1, f^2\right)$	α (S.E.)	δ_1 (S.E.)	δ_2 (S.E.)	$F(2, \text{Obs} - 3)$ [a]	Obs.
$\left(f_{t,t}, f_{t,t}^{Same}\right)$	-0.076 (0.222)	1.182 (0.112)***	-0.163 (0.098)	0.059	23
$\left(f_{t,t}, f_{t,t}^{Trend}\right)$	0.604 (0.394)	1.170 (0.090)***	-0.300 (0.135)**	1.189	23
$\left(f_{t,t}, f_{t,t}^{VAR}\right)$	-0.227 (0.223)	0.424 (0.351)	0.646 (0.365)*	0.569	23
$\left(f_{t,t}, f_{t,t}^{VECM}\right)$	-0.281 (0.217)	0.389 (0.289)	0.686 (0.300)**	0.851	23
$\left(f_{t,t}, f_{t,t}^{Mean}\right)$	-0.042 (0.226)	0.343 (0.418)	0.681 (0.406)	0.069	23

Notes
Standard errors are in parentheses.
a: The null of the F-test is $(\alpha, \delta_1 + \delta_2) = (0,1)$.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 11: Encompassing test for $f_{t,t}$ (sub-sample)

Model: (6) $g_t - f_{t,t}^2 = \delta\left(f_{t,t}^1 - f_{t,t}^2\right) + u_t$

$\left(f^1, f^2\right)$	δ (S.E.)	P-value ($\delta = 0$)	P-value ($\delta = 1$)
$t = 1980 - 1991$			
$\left(f_{t,t}, f_{t.t}^{VAR}\right)$	-0.107 (0.442)	0.813 [a]	0.029
$\left(f_{t,t}, f_{t.t}^{VECM}\right)$	-0.275 (0.352)	0.451 [a]	0.004
$\left(f_{t,t}, f_{t.t}^{Mean}\right)$	-0.114 (0.498)	0.823 [a]	0.047
$t = 1992 - 2002$			
$\left(f_{t,t}, f_{t.t}^{VAR}\right)$	0.911 (0.498)	0.097	0.861
$\left(f_{t,t}, f_{t.t}^{VECM}\right)$	0.838 (0.397)	0.061	0.692
$\left(f_{t,t}, f_{t.t}^{Mean}\right)$	0.688 (0.546)	0.236	0.581

Notes
Standard errors are in parentheses.

a: This value is meaningless for the encompassing test because $\delta < 0$.

Table 12: Prediction performance of $f_{t,t+1}$

	RMSE	MAE
$f_{t,t+1}$	1.886	1.541
$f_{t,t+1}^{Same}$	2.456	2.009
$f_{t,t+1}^{Trend}$	2.321	1.854
$f_{t,t+1}^{VAR}$	1.637	1.348
$f_{t,t+1}^{VECM}$	1.553	1.305
$f_{t,t+1}^{Mean}$	1.613	1.311

Notes

$f_{t,t+1}^{Same} \equiv g_{t-1}$ and $f_{t,t+1}^{Trend} \equiv \frac{11}{100}\sum_{i=1}^{9} g_{t-i} + \frac{g_{t-10}}{100}$. $f_{t,t+1}^{VAR}$ is prediction of the VAR with four lags consists of real GNP, GNP deflator, real consumption, and the call money rate. $f_{t,t+1}^{VECM}$ is prediction of the VECM with four lags. $f_{t,t+1}^{Mean}$ is the consensus (mean) forecast of private institutions.

1

Table 13: Encompassing test for $f_{t,t+1}$

Model: (6) $g_{t+1} - f_{t,t+1}^2 = \delta\left(f_{t,t+1}^1 - f_{t,t+1}^2\right) + u_t$

$\left(f^1, f^2\right)$	δ (S.E.)	P-value ($\delta = 0$)	P-value ($\delta = 1$)
$\left(f_{t,t+1}, f_{t,t+1}^{Same}\right)$	1.095 (0.285)	0.001	0.741 [a]
$\left(f_{t,t+1}, f_{t,t+1}^{Trend}\right)$	1.786 (0.456)	0.001	0.100 [a]
$\left(f_{t,t+1}, f_{t,t+1}^{VAR}\right)$	0.158 (0.314)	0.620	0.014
$\left(f_{t,t+1}, f_{t,t+1}^{VECM}\right)$	0.083 (0.289)	0.775	0.005
$\left(f_{t,t+1}, f_{t,t+1}^{Mean}\right)$	-0.248 (0.438)	0.577 [b]	0.010

Notes
Standard errors are in parentheses.

$f_{t,t+1}^{Same} \equiv g_{t-1}$ and $f_{t,t+1}^{Trend} \equiv \dfrac{11}{100}\sum_{i=1}^{9} g_{t-i} + \dfrac{g_{t-10}}{100}$. $f_{t,t+1}^{VAR}$ is prediction of the

VAR with four lags consists of real GNP, GNP deflator, real consumption, and the call money rate. $f_{t,t+1}^{VECM}$ is prediction of the VECM with four lags.

$f_{t,t+1}^{Mean}$ is the consensus (mean) forecast of private institutions.

a: This value is meaningless for the encompassing test because $\delta > 1$.

b: This value is meaningless for the encompassing test because $\delta < 0$.

2

Table 14: Generalized encompassing test for $f_{t,t+1}$

Model: (5') $g_{t+1} = \alpha + \delta_1 \cdot f^1_{t,t+1} + \delta_2 \cdot f^2_{t,t+1} + u_t$

$\left(f^1, f^2\right)$	α (S.E.)	δ_1 (S.E.)	δ_2 (S.E.)	$F(2, \text{Obs} - 3)$ [a]	Obs.
$\left(f_{t,t+1}, f^{Same}_{t,t+1}\right)$	-0.858 (1.000)	1.233 (0.458)**	-0.193 (0.289)	1.648	22
$\left(f_{t,t+1}, f^{Trend}_{t,t+1}\right)$	-0.010 (1.272)	1.437 (0.634)**	-0.556 (0.710)	0.407	22
$\left(f_{t,t+1}, f^{VAR}_{t,t+1}\right)$	-1.211 (0.879)	0.209 (0.413)	1.272 (0.530)**	1.128	22
$\left(f_{t,t+1}, f^{VECM}_{t,t+1}\right)$	-1.274 (0.825)	0.216 (0.353)	1.256 (0.420)***	1.383	22
$\left(f_{t,t+1}, f^{Mean}_{t,t+1}\right)$	0.204 (0.987)	-0.346 (0.727)	1.294 (0.654)*	0.022	22

Notes
Standard errors are in parentheses.
a: The null of the F-test is $(\alpha, \delta_1 + \delta_2) = (0,1)$.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Table 15: Encompassing test for $f_{t,t+1}$ (sub-sample)

Model: (6) $g_{t+1} - f_{t,t+1}^2 = \delta\left(f_{t,t+1}^1 - f_{t,t+1}^2\right) + u_t$

$\left(f^1, f^2\right)$	δ (S.E.)	P-value ($\delta = 0$)	P-value ($\delta = 1$)
$t = 1980 - 1990$			
$\left(f_{t,t+1}, f_{t.t+1}^{VAR}\right)$	0.474 (0.283)	0.125	0.093
$\left(f_{t,t+1}, f_{t.t+1}^{VECM}\right)$	0.353 (0.223)	0.144	0.016
$\left(f_{t,t+1}, f_{t.t+1}^{Mean}\right)$	-0.052 (0.529)	0.923 [a]	0.074
$t = 1991 - 2001$			
$\left(f_{t,t+1}, f_{t.t+1}^{VAR}\right)$	-0.786 (0.682)	0.276 [a]	0.026
$\left(f_{t,t+1}, f_{t.t+1}^{VECM}\right)$	-1.186 (0.756)	0.148 [a]	0.016
$\left(f_{t,t+1}, f_{t.t+1}^{Mean}\right)$	-0.388 (0.697)	0.590 [a]	0.075

Notes
Standard errors are in parentheses.

a: This value is meaningless for the encompassing test because $\delta < 0$.

Table 16: The sign test

$$S \equiv \frac{n - 0.5T}{\sqrt{0.25T}} \quad \sim \quad N(0,1)$$

$\left(f^1, f^2\right)$	$(n, T-n)$	S
$\left(f_{t,t}, f_{t,t}^{Same}\right)$	$(20,3)$	5.489***
$\left(f_{t,t}, f_{t,t}^{Trend}\right)$	$(20,3)$	5.489***
$\left(f_{t,t}, f_{t,t}^{VAR}\right)$	$(13,10)$	0.969
$\left(f_{t,t}, f_{t,t}^{VECM}\right)$	$(12,11)$	0.323
$\left(f_{t,t}, f_{t,t}^{Mean}\right)$	$(12,11)$	0.323
$\left(f_{t,t+1}, f_{t,t+1}^{Same}\right)$	$(15,7)$	2.612***
$\left(f_{t,t+1}, f_{t,t+1}^{Trend}\right)$	$(13,9)$	1.306
$\left(f_{t,t+1}, f_{t,t+1}^{VAR}\right)$	$(10,12)$	-0.653
$\left(f_{t,t+1}, f_{t,t+1}^{VECM}\right)$	$(8,14)$	-1.959*
$\left(f_{t,t+1}, f_{t,t+1}^{Mean}\right)$	$(7,15)$	-2.612***

Notes

T: The sample size.

n: The number of years the government "wins", i.e., its absolute forecast error is smaller than that of the benchmark forecast.

***: Significant at the 0.01 level.

**: Significant at the 0.05 level.

*: Significant at the 0.10 level.