FORECASTS BASED ON OFFICIAL BUSINESS INDICES IN JAPAN: A REAL-TIME ANALYSIS

Hidehiro Ikeno*

This paper analyses the performance of forecasts of real economic activity based on the Japanese official business indices in a real-time framework. As in other countries, preliminary numbers of the business indices are first announced, then they are revised. These revisions are often substantial, so that the business indices that are available on a real-time basis often differ from their revised form. This paper explores how this difference can affect forecasts. Results indicate that real-time forecasts based on the diffusion leading index are not significantly worse than forecasts based on the final revised form. Also, the real-time forecasts using the composite leading index are often comparable to those from final revised forms. However, real-time forecasts from the diffusion coincident index are significantly worse than forecasts that use the final revised form.

Key words and phrases: business index, probit model, real-time forecast

1. Introduction

This work examines real-time forecasts which are based on the official business indices in Japan. These official business indices, called “Keiki Doko Shisu” in Japanese, were formerly compiled by the Economic Planning Agency (EPA), and are currently compiled by the Economic and Social Research Institute (ESRI) in the Cabinet Office. Since promptness is regarded as important, the indices are first announced in a preliminary form, then they are revised; revisions are carried out repeatedly. Is there a significant difference between forecasts based on the revised and the real-time indices? This paper compares revised and real-time forecasts and evaluates how much they differ. However good the forecasts based on the final revised form are, if forecasts based on business indices that are available in real-time are unsatisfactory, then usefulness of these indices is limited from a practical point of view.

The same question over preliminary and revised numbers applies to other countries as well as Japan. Diebold and Rudebusch (1991a) show that real-time forecasts of production increase in the US based on leading business indices are hardly useful, though forecasts based on ex-post revised numbers perform satisfactorily. Diebold and Rudebusch (1991b) show similar results in regard to turning points of the business cycle in the US.

There are already many studies on the performance of the official business indices in Japan, but they use business indices that are revised many times and...
Performance of real-time forecasts is virtually unexplored in Japan. This paper examines the difference between real-time forecasts and forecasts based on ex-post analysis. The present work investigates forecasts of turning points in the business cycle and forecasts of the rate of change of the production index. The diffusion index is used for the former forecasts, and the composite index for the latter. In each case, first, it is confirmed that forecasts based on ex-post data perform satisfactorily. Then, it compares errors of ex-post forecasts and real-time forecasts and performs statistical tests to see whether the accuracies of those forecasts are the same. It is noted that there is a methodological difference between this paper and Diebold and Rudebusch (1991a). Their work examines usefulness of real-time forecasts by comparing them with forecasts that do not use the indices.

Below, Section 2 discusses revisions of the official business indices in Japan. Section 3 sets out the methodology used in this paper. Sections 4 and 5 respectively show and discuss empirical results concerning the diffusion indices and the composite index. Conclusions are drawn in Section 6.

2. Revisions of the Japanese business indices

Japan officially publishes two classes of business indices; diffusion indices (DIs) and composite indices (CIs). One could roughly mention that DIs represent something qualitative of the business cycle, and the CIs represent something quantitative. The DIs take account of the direction of change over the previous three months of the economic variables from which the indices are compiled. They represent the proportion of economic variables which move in the direction corresponding to the economic boom, and range between zero and one hundred by definition. They do not take account of the size of change. By contrast, the CIs take account of the size of changes in economic variables.

Each class consists of the leading, the coincident, and the lagged index. The first two indices are expected mainly to serve for analyzing and forecasting the future and current economic state. The last index is mainly used as an ex-post reference when turning points of the business cycle are discussed. Each index is compiled from around ten economic variables. In countries like the US and the UK, CIs are traditionally almost exclusively referred to, although the UK quite recently stopped publishing them. In Japan, however, the DIs are traditionally more widely quoted than the CIs. As in many countries, these indices in Japan are subject to substantial revisions, due mainly to two factors: correction of numbers that were provisionally announced, and methodological change in compilation.

Correction of provisionally announced numbers is carried out almost every month. Preliminary numbers of the indices are first announced for promptness, and these numbers are then revised over the following months. Revision typically takes place several times. Correction of preliminary numbers is virtually complete.

within several months.

Occasionally, the indices are revised due to a methodological change in compilation. Let us call it a “definitional change” in this paper. According to information publicly announced by the ESRI on its homepage, the Japanese DIs went through eight definitional changes during the period from 1960 to 2001. On average, therefore, a definitional change took place almost every five years. One reason is changes in economic environment in Japan. Some economic variables gradually become more important than others from the point of view of forecasting because of economic structural changes. Another reason is to improve the performance of business indices, as more is learned about which variables are good for prediction. When a definitional change is carried out, often indices are revised covering the preceding period for at least ten years back.

3. Methodology

The EPA used to publish monthly, and currently the ESRI publishes monthly, a booklet with the title “Business Indicators” (“Keiki Doko Shisu” in Japanese), which contains time series of official business indices and information on turning points of the business cycle. Indices given in Business Indicators include the diffusion leading index (DLI), the diffusion coincident index (DCI), and the composite leading index (CLI).

Denote the series of indices contained in the period \(s\) issue of the Business Indicators as “indices of vintage \(s\)”. The series of indices of vintage \(s\) contain indices of each period prior to \(s\) and of period \(s\). Early entries of them have, probably, been already revised many times, and later entries are revised at most a few times or not at all.

Let \(DLI_s(t)\), \(DCI_s(t)\), and \(CLI_s(t)\) denote the DLI, the DCI, and the CLI of vintage \(s\) corresponding to period \(t\). Data used in this paper have been collected from monthly issues of the Business Indicators, from the January 1988 issue through the December 2001 issue. The first entry of the index series of all vintages is of January 1972. In the present notation, the start of \(s\) is in January 1988, and the end is in December 2001. The start of \(t\) is in January 1972 for all \(s\), and the end depends on \(s\); the end of the series of vintage \(s\) falls in period \(s\).

In fact, the monthly issue containing the series of the indices of vintage \(s\) is not published in period \(s\). Typically it is published two months after period \(s\). For example, the first preliminary numbers of the indices of December 2001 were published in February 2002. This is because the compilation and publication of even preliminary numbers takes about two months.

This paper also uses final revised numbers. The final numbers of the indices in this paper have been collected from the June 2002 issue of Business Indicators. These are denoted by \(DLI_{FINAL}(t)\), \(DCI_{FINAL}(t)\), and \(CLI_{FINAL}(t)\). Even these numbers are logically open to revision, but they will be considered as if they were the real final numbers in this paper.

Forecasting based on the composite coincident index (CCI) is not explored in this paper. The CCI is expected to contain some information on the current
state of the economy. On the other hand, compilation of the CCI is partly based on the industrial production index, which is used as a targeted production index in this paper. Therefore, a study of forecasting of the industrial production index by the CCI is not worthwhile.

3.1. Sizes of revisions

Table 1 shows the effects of revision. The table gives descriptive statistics of preliminary numbers and revised numbers of the DLI and the DCI for the period from January 1988 through December 2001. It also shows descriptive statistics of the rate of change of the CLI over past six months in a percentage form. The CLI is likely to be non-stationary as is shown later, so that they are transformed into the rate of change which is also shown later to be stationary.

Table 1. Descriptive statistics.

<table>
<thead>
<tr>
<th>Sample period</th>
<th>Full sample period A</th>
<th>Sample period B</th>
<th>Sample period A</th>
<th>Sample period B</th>
<th>Sample period B</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLI (A)</td>
<td>48.33</td>
<td>47.50†</td>
<td>†</td>
<td>50.19†</td>
<td>†</td>
</tr>
<tr>
<td>(B)</td>
<td>48.14</td>
<td>48.78</td>
<td>49.27</td>
<td>50.82</td>
<td>48.11</td>
</tr>
<tr>
<td>(C)</td>
<td>16.98</td>
<td>12.98</td>
<td>16.19</td>
<td>13.52</td>
<td>19.75</td>
</tr>
<tr>
<td>DCI (A)</td>
<td>50.14</td>
<td>52.02†</td>
<td>†</td>
<td>48.19†</td>
<td>†</td>
</tr>
<tr>
<td>(B)</td>
<td>51.68</td>
<td>52.93</td>
<td>53.21</td>
<td>49.97</td>
<td>50.06</td>
</tr>
<tr>
<td>(C)</td>
<td>19.14</td>
<td>14.78</td>
<td>17.81</td>
<td>17.00</td>
<td>16.39</td>
</tr>
<tr>
<td>Change (A) rate of</td>
<td>−0.27</td>
<td>−0.01†</td>
<td>†</td>
<td>−0.49†</td>
<td>†</td>
</tr>
<tr>
<td>(B)</td>
<td>−0.20</td>
<td>0.26</td>
<td>−0.01</td>
<td>−0.61</td>
<td>−0.96</td>
</tr>
<tr>
<td>CLI (C)</td>
<td>2.31</td>
<td>1.31</td>
<td>1.70</td>
<td>1.41</td>
<td>2.98</td>
</tr>
</tbody>
</table>

Notes: (A) the mean of preliminary numbers, (B) the mean of revised numbers of the base vintage, and (C) the root mean of squared values of revisions. † denotes that they are equal to the corresponding figures with ‡ by definition.

The preliminary numbers used in this table are those published for the first time in the Business Indicators, i.e., those based on numbers of the newest vintage of each period. The notation expresses them, respectively, as follows: \( DLI_t(t), DCI_t(t), \) and \( \{(CLI_t(t) - CLI_t(t - 6))/CLI_t(t - 6)\} \times 100, \) where \( t = \) January 1988, \( \ldots, \) December 2001. The revised numbers are indices of the base vintage, which means that they were last revised in the issue of the period of the base vintage. The table shows their means. Revisions are defined as the preliminary numbers minus the corresponding numbers of the base vintage. The table shows root means of their squared values.

Table 1 also indicates the statistics for some sub-sample periods. Definitional changes of the indices were carried out twice during the period January 1988 through December 2001. The first change was in April 1996, and the second

\[ \text{This means that } CLI_t(t) \text{ in the expression involves newly published numbers, but } CLI_t(t - 6) \text{ may have been revised during the preceding five months.} \]
was in November 2001. Statistics from the sample period January 1988 through March 1996, i.e., Sample period A, with the base vintage of March 1996 involves effects that are caused exclusively by correction of preliminary numbers. Comparison between these statistics and those from the same sample period but with the base vintage of April 1996 shows the effects of the definitional change carried out in April 1996. A similar argument applies to statistics from the sample period April 1996 through October 2001, i.e., Sample period B, with the base vintages of October and November 2001.

Table 1 reveals the following points. First, the preliminary and the revised numbers are almost the same on average with DLI, DCI, and CLI. The revisions are not significantly different from zero when they are tested under the assumption of normality. Second, both revision factors, i.e., correction of preliminary numbers and definitional changes, are significant. The only exception is the DCI and the definitional change in November 2001. In fact, the list of economic variables from which the DCI is compiled changed only marginally in November 2001 while the counterpart of the DLI changed substantially. Third, aside from effects of definitional changes, effects of correction of preliminary numbers on the DCI increase over time, i.e., from Sample period A to Sample period B, whereas effects on the DLI and the CLI remained almost the same.

3.2. Forecasts based on the diffusion indices

The EPA formerly announced, and the ESRI currently announces, dates of the official turning points of the business cycle. The dates are decided not only based on the business indices but also based on other items of information on the Japanese economy. One problem is that recognition of turning points takes a long time. Usually, a turning point is announced around one to two years later. Information on official turning points of the business cycle that is recorded in the period s issue of Business Indicators will be called in this paper “information on turning points of vintage s”. Also, information on turning points in the June 2002 issue is taken to be the final revised information. Occasionally the date of a turning point that has been already announced is changed. However, such changes are small and are not a serious problem, compared to the delay of recognition.

Since delays in recognizing turning points are troublesome from a practical point of view, prompt forecasting of the current economic state by the coincident index is both interesting and desirable. A forecast by the coincident index in a preliminary form would certainly be useful from a practical point of view if it performs reasonably well.

Delays in recognition can affect forecasts of the business cycle on the real-time basis. Compared with forecasts based on ex-post knowledge of the business cycle, forecasts based on real-time information are handicapped, because some information on the business cycle is not available to the latter. In comparing forecasts based on the final revised data and real-time forecasts, two factors

---

3 Table 1 does not show standard errors of the revisions. However, these are almost the same as the root means of the squared values since the averages of the revisions are close to zero.
should be considered. As well as revision of the business indices, the delay of recognition of turning points of the business cycle should be taken into account when considering true real-time forecasts.

This paper explores the performance of the diffusion indices as a forecaster of the probability of downturn of the economy, using a probit model. Many studies have used a probit model to forecast the business cycle. Hirata and Ueda (1998), Komaki (2001), and Ikeno (2003) all use a probit model to forecast the Japanese business cycle. Estrella and Mishkin (1997, 1998) and Filardo (1999) use a probit model with the US data.

With the diffusion indices, four classes of forecasts are investigated. These are forecasts of two kinds of ex-post analyses, i.e., in-sample and out-of-sample analyses, and two kinds of real-time analyses.

For in-sample analysis, a forecast is based on all information in the final revised numbers and the final revised information on turning points. All entries of each index and information on turning points in the final revised form are used.

An in-sample forecast is carried out as follows: parameters are first estimated by maximizing the likelihood based on each of the following equations with all entries in the final revised form:

\[
P(R_{\text{FINAL}}(t) = 1) = F(\alpha + \beta DLI_{\text{FINAL}}(t - k))
\]

\[
P(R_{\text{FINAL}}(t) = 1) = F(\alpha + \beta DCI_{\text{FINAL}}(t))
\]

where \(R_{\text{FINAL}}(t)\) is a dummy variable for a recession, which takes the value unity when the economy is in a recession in period \(t\) and zero otherwise, based on the final revised information of turning points; \(F(\cdot)\) is the cumulative normal distribution function, and \(\alpha\) and \(\beta\) are parameters. The forecast period \(k\) should take a positive value. Fitted values based on \(DLI_{\text{FINAL}}(t - k)\) and \(DCI_{\text{FINAL}}(t)\) in (3.1) and (3.2) respectively are taken as a forecast of the probability that the economy is in recession in period \(t\).

Out-of-sample analysis uses only entries prior to the period when a forecast is supposed to be formed; entries in the final revised form are used. These forecasts are still far from real-time forecasts, but some studies suggest that they are expected as approximation of real-time forecasts in the literature.\(^4\)

The parameters in an out-of-sample forecast are estimated using the same equations as in the in-sample forecast, (3.1) and (3.2). However, estimation is carried out with entries up to period \(t - 2k\) and up to period \(t\) of each index, respectively, and turning point dates up to period \(t - k\) and up to period \(t\), respectively. Then, the fitted values \(DLI_{\text{FINAL}}(t - k)\) and \(DCI_{\text{FINAL}}(t)\) are regarded as a forecast. Estimation is carried out recursively, such that the parameters are re-estimated every time a new forecast is calculated.

Of the two real-time forecasts, the one called a strong real-time forecast in this paper is calculated as follows: for a forecasted probability that the economy

\[^4\] For example, in regard to the out-of-sample forecasts, Estrella and Mishkin (1998) state that “this type of procedure leads to a fairer and more realistic test of the predictive abilities of the various models than the in-sample results”.


is in recession in period $t$, a probit model is estimated with the series of each index of vintage $t - k$ with the DLI and those of vintage $t$ with the DCI, together with turning points up to the last one in the information on turning points of vintage $t - k$ and those of vintage $t$, respectively. This is truly real-time in the sense that setting up the forecasts does not require any knowledge beyond either vintage $t - k$ or vintage $t$. Even information of vintage $t - k$ does not have a record of the business cycle up to period $t - k$, because recognition of a turning point often takes a long time after the time it occurs. Hence, some of the later entries of the business index of vintage $t - k$ are not used for estimation, because there is no counterpart information on the business cycle. A similar argument applies to information of vintage $t$. The fitted values based on $DLI_{t-k}(t-k)$ and $DCI_t(t)$ are then taken as the forecast.

The other real-time forecast, called a weak real-time forecast in this paper, is based on real-time series of the business indices but on the final revised information on turning points. It is hypothetical and not truly real-time. Estimation of this forecast nevertheless exposes the effects of revision of the indices while holding other factors constant through comparison with the ex-post forecast.

If real-time forecasts perform as well as in-sample or out-of-sample forecasts, one should not hesitate even when using preliminary numbers of the index as long as results from in-sample or out-of-sample forecasts are satisfactory. In that case, the practical usefulness of the index as a forecaster of present and subsequent economic activity is confirmed. If not, then one should be cautious when using preliminary numbers to forecast economic activity. Most studies that empirically examine business indices employ ex-post analysis, i.e., either in-sample or out-of-sample analysis. Unless these are compared with real-time forecasts, one cannot infer real-time usefulness from results in those preceding studies.

Why is a probit model used, rather than a Neftci model or a Markov-switching model? Although the Neftci model used to be widely used and a Markov-switching model is increasingly used in the literature, a probit model has the following advantages:\footnote{A Markov-switching model with a business index is used by Hamilton and Perez-Quiros (1996), Layton (1996), and Layton and Katsuura (2001a, b).}

1. A Neftci model requires a prior probability calculation, which often allows some arbitrariness. A probit model does not require a prior probability.

2. A Markov-switching model does not use information contained in the official dates of turning points of the business cycle. A probit model can exploit such information.

3. A Markov-switching model involves an extremely heavy calculational burden. Although it is an established method, it is often claimed that the numerical results of estimation depend critically on the initial values used.\footnote{This problem is recognized even with a popular econometric program such as RATS.} A probit model suffers no such drawbacks.
3.3. Forecasts based on the composite index

This paper uses the rate of change of the CLI over a preceding period to forecast the rate of change of the industrial production index (IPI) over a subsequent period, because the levels of both variables are likely to follow a non-stationary process. Dickey-Fuller tests and KPSS tests are executed to study stochastic processes of the following variables: levels, rates of change over the previous three months, and the rates of change over the previous six months of the CLI and the IPI. For the levels, a model with a trend and a constant is assumed as the null hypothesis, and for the rates of change, a model with a constant is assumed. All variables are based on the final revised numbers.

Table 2 shows results with the sample period January 1974 through December 2001. The number of lags is set to four in all cases. Most results do not change substantially with more lags. Two kinds of statistics are shown from the Dickey-Fuller tests, i.e., $\tau$ and $T(\rho - 1)$. The null hypothesis of unit-root is strongly rejected for all rates of change. The null hypothesis is not rejected at all for the level of the IPI, but results for the level of the CLI are mixed. On the other hand, with the KPSS tests, the null hypothesis of stationarity is strongly for both levels, but is not rejected at all for their rates of change. It therefore appears that the levels of the CLI and the IPI are non-stationary but all rates of change stationary. Hence, to analyze the CLI and the IPI, rates of change are used in the following investigations.

Table 2. Stationarity of the CLI and the IPI: sample period January 1974 to December 2001.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Level of CLI</th>
<th>3-m change rate of CLI</th>
<th>6-m change rate of CLI</th>
<th>Level of IPI</th>
<th>3-m change rate of IPI</th>
<th>6-m change rate of IPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-F: $\tau$</td>
<td>$-3.116$</td>
<td>$-5.878^*$</td>
<td>$-6.130^*$</td>
<td>$-0.829$</td>
<td>$-4.015^*$</td>
<td>$-4.283^*$</td>
</tr>
<tr>
<td>D-F: $T(\rho - 1)$</td>
<td>$-23.928^*$</td>
<td>$-67.000^*$</td>
<td>$-63.664^*$</td>
<td>$-4.170$</td>
<td>$-39.482^*$</td>
<td>$-48.518^*$</td>
</tr>
<tr>
<td>KPSS: $\eta_\tau$</td>
<td>$0.748^*$</td>
<td>$0.104$</td>
<td>$0.111$</td>
<td>$1.142^*$</td>
<td>$0.410$</td>
<td>$0.435$</td>
</tr>
</tbody>
</table>

'D-F' means Dickey-Fuller. * significant at the one-percent level, and ** significant at the five-percent level.

Transformation of the CLI into the rate of change is carried out at each vintage, and the CLI in the final revised form is also transformed into the rate of change. The classification of a vintage and the final revised form then remains valid even after the transformation. Three classes of forecasts are explored: in-sample forecasts, out-of-sample forecasts, and weak real-time forecasts. The basic idea of each class corresponds to its counterpart in the diffusion indices.

It is often stressed that the business cycle involves the movement of many economic variables as well as the IPI. However, the IPI is one of very few production indices for which data are available on a monthly basis in Japan. Use of the GDP in place of the IPI would greatly decrease the number of observations in this paper.

For the in-sample forecast, the following two equations are estimated by the

---

7 For the KPSS test, see Kwiatowski et al. (1992).
OLS, with all entries in the final revised form:

\[
IPI(t) = \gamma + \sum_{h=0}^{a} \lambda_h Z_{FINAL}(t-h) + \sum_{h=0}^{a} \phi_l IPI(t-k-h) \tag{3.3}
\]

\[
IPI(t) = \gamma + \sum_{h=0}^{a} \lambda_h Z_{FINAL}(t-h) \tag{3.4}
\]

where \(IPI(t)\) is the rate of change of the IPI from period \(t\) to period \(t+k\) in percentage form, and, \(Z_{FINAL}(t) \equiv \left\{ (CLI_{FINAL}(t) - CLI_{FINAL}(t-k)) / CLI_{FINAL}(t-k) \right\} \times 100.\)

The fitted value is taken as the forecast of \(IPI(t)\). Equations (3.3) and (3.4) are both studied to see whether results depend on the inclusion of lagged terms in the rate of change of the IPI.

For the out-of-sample forecast of \(IPI(t)\), the same equations as above are used but entries only up to \(t-k\) are used in estimation of the parameters.

The equations used for the weak real-time forecast are as follows:

\[
IPI(t) = \gamma + \sum_{h=0}^{a} \lambda_h Z_{t}(t-h) + \sum_{h=0}^{a} \phi_l IPI(t-k-h) \tag{3.5}
\]

\[
IPI(t) = \gamma + \sum_{h=0}^{a} \lambda_h Z_{t}(t-h) \tag{3.6}
\]

where \(Z_{t}(u) \equiv \left\{ (CLI_{t}(u) - CLI_{t}(u-k)) / CLI_{t}(u-k) \right\} \times 100.\)

The parameters are estimated with entries up to period \(t-k\) in the series of vintage \(t\). The fitted value based on the series of vintage \(t\) is taken as the forecast.

The IPI is an economic variable that is subject to revision after announcement of preliminary numbers. The IPI used in this paper, which is denoted by \(IPI(t)\), is the revised number. In this paper, they are taken as the final revised numbers, which were last revised in January 2003.\(^8\) Although the IPI in Japan is also subject to substantial revision due to methodological changes in compilation, revision due to correction of preliminarily announced numbers is relatively small.

There is no counterpart to the strong real-time forecasts of the DIs. The evaluation of forecasts in this paper is based on the IPI in the final revised form. In the case of the DIs, changes to the published turning point dates of the business cycle are minor, but the delay is a serious problem. The main point of the strong real-time forecasts is to examine the use of the preliminary DI numbers and the effect of the delay. Therefore, their evaluation based on the final revised data is not seriously affected by the changes to the official turning point dates. In the case of the CI, the delay in announcement of the IPI is not a serious problem; the IPI is more quickly published than the business indices. If the strong real-time

\(^8\) These numbers for the industrial production index were downloaded from the homepage of the Ministry of Economy, Trade and Industry in February 2003.
forecasts are based on preliminary numbers of the CLI and preliminary numbers of the IPI, the difference between the weak real-time and the strong real-time forecasts would therefore be due almost wholly to revision of the IPI, which is outside the scope of this paper.

4. Empirical results: the DLI and the DCI

In this section, significance of the indices is first confirmed in the framework of ex-post analysis, i.e., in the framework of in-sample forecasts. Then, the four classes of forecasts are compared.

Table 3 shows estimates using the probit model from in-sample analysis. The start of the sample period is in January 1974 and the end is in October 2000, which is the last turning point in the final revised information. For the DLI, a forecast period, denoted by \( k \), takes values three and six. This range of forecast periods is in accordance with the popularly expected use of the DLI. For the DCI, there is no such forecast period because of its expected function of coincidence. For the DLI, standard errors are corrected for autocorrelations in errors arising from overlapping of forecast periods, using the method of Estrella and Rodrigues (1998). The results indicate that the DLI and the DCI are significant at the conventional level in all cases. They confirm that the DLI and the DCI contain useful information for forecasting the business cycle in the framework of ex-post analysis.

Table 3. Estimation of the probit models: in-sample analysis (3.1) and (3.2).

<table>
<thead>
<tr>
<th>Index</th>
<th>Forecast period ((k))</th>
<th>Sample period</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>Pseudo (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLI 3</td>
<td>1974:01–2000:10</td>
<td>1.897**</td>
<td>−0.043*</td>
<td>0.355</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.755)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLI 6</td>
<td>1974:01–2000:10</td>
<td>1.872**</td>
<td>−0.043**</td>
<td>0.357</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.963)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCI</td>
<td>not applicable</td>
<td>1974:01–2000:10</td>
<td>2.151*</td>
<td>−0.045*</td>
<td>0.623</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.226)</td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: * significant at the one-percent level, and ** significant at the five-percent level. Standard errors are in parentheses. For the pseudo \(R^2\), see Estrella (1998).

The performances of the forecasts are evaluated by the quadratic probability scores (QPSs). Tables 4 to 6 show QPSs from the four classes of forecasts. If the start of a sample period is in period \( t_0 \) and the end is in period \( t_0 + n - 1 \), the QPS is expressed as follows:

\[
QPS = \frac{1}{n} \sum_{t=t_0}^{t_0+n-1} \{ R_{\text{FINAL}}(t) - \text{(forecasted probability)} \}^2.
\]

The start of regressions for estimation of \( \alpha \) and \( \beta \) is in January 1974 for all classes of forecasts, and the end depends on the class of forecast. The start of the full-sample period for the QPSs is \( k \) periods after January 1988 for the DLI, and in January 1988 for the DCI. The end is in the period of the last turning
point in the final revised information, i.e., December 2000. In addition to the full-sample period, Sample periods A and B are chosen so that one may see results that are free from the two definitional changes in April 1996 and November 2001. Equation (4.1) indicates that the evaluation of performance is based on the record of recession using the final revised information on turning points.

Tables 4 and 5 show performances of forecasts by the DLI, with \( k = 3 \) and \( k = 6 \) respectively. They indicate that there is virtually no difference in performances among the four classes of forecasts with the full-sample period. Results from the sub-sample periods indicate that the relatively good performance of the two real-time forecasts does not change even when effects of the two definitional changes are excluded.

Table 6 shows the performances of forecasts by the DCI. The two classes of real-time forecast perform much worse than the other two classes with the full-sample period. Results from the sub-sample periods imply that their poorer performances do not greatly improve even when effects of the definitional changes are excluded. The QRSs of in-sample and out-of-sample forecasts are much lower than their counterparts in Tables 3 and 4 with the full-sample and the two sub-sample periods.

For formal comparison of performances of forecasts, Diebold-Mariano (DM) tests were executed. Diebold and Mariano (1995) present a statistical test aimed at comparing predictive accuracy. The statistic follows the standard normal
Table 7. Forecasting power of weak real-time forecasts: DM statistics, full-sample period.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>in-sample forecasts vs. weak real-time forecasts</th>
<th>out-of-sample forecasts vs. weak real-time forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLI with (k = 3)</td>
<td>0.262</td>
<td>−0.370</td>
</tr>
<tr>
<td>DLI with (k = 6)</td>
<td>1.020</td>
<td>0.070</td>
</tr>
<tr>
<td>DCI</td>
<td>4.590*</td>
<td>3.971*</td>
</tr>
</tbody>
</table>

Note: * significant at the one-percent level, and ** significant at the five-percent level.

Table 8. Forecasting power of strong real-time forecasts: DM statistics, full-sample period.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>in-sample forecasts vs. strong real-time forecasts</th>
<th>out-of-sample forecasts vs. strong real-time forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLI with (k = 3)</td>
<td>0.354</td>
<td>−0.293</td>
</tr>
<tr>
<td>DLI with (k = 6)</td>
<td>0.793</td>
<td>−0.145</td>
</tr>
<tr>
<td>DCI</td>
<td>5.192*</td>
<td>4.790*</td>
</tr>
</tbody>
</table>

Note: * significant at the one-percent level, and ** significant at the five-percent level.

distribution under the null hypothesis that two forecasts have the same degree of accuracy. Harvey et al. (1997) present a modification of the DM test statistic for a small sample.

Tables 7 and 8 show DM test statistics from contests of the four classes of forecast with the full-sample periods. The test is not executed with the sub-sample periods because of their relatively small sample sizes. The statistics are calculated with rectangular windows for the asymptotic variance, and the order of autocorrelations is assumed to be \(k - 1\) while the QPSs are used as loss functions. The DM statistics modified by the method of Harvey et al. (1997) are virtually the same as those without modification, so that they are not reported.\(^9\) With the DCI, no serial autocorrelation is assumed since forecast periods do not overlap with each other. In Table 7, the null hypotheses are that in-sample forecasts are as accurate as weak real-time forecasts, and that out-of-sample forecasts are as accurate as weak real-time forecasts. In Table 8, the null hypotheses are that in-sample forecasts are as accurate as strong real-time forecasts, and that out-of-sample forecasts are as accurate as strong real-time forecasts. As might be expected from a casual observation of the QPSs in Tables 4 and 5, the null hypotheses are not rejected at any reasonable level with the DLI in Table 7, nor in Table 8. On the other hand, the null hypotheses are rejected with the DCI in both Tables 7 and 8.

5. Empirical results: the CLI

Again, significance of the business index is first confirmed in the framework of ex-post analysis, and then the three classes of forecast are compared.

Table 9 shows estimates from in-sample analysis, based on Equations (3.3)\(^9\) They also recommend using a \(t\)-distribution instead of a normal distribution when one obtains a significance level. Even when this point is accounted for, the present results virtually do not change.
and (3.4), with $a = 1$ and the sample period from January 1974 to December 2001. Here $k$ is set to three and six, which are in accordance with the range that is popularly expected. Standard errors are corrected for serial autocorrelations in errors arising from overlapping of forecast periods, using the method of Newey and West (1987). All results indicate that at least one of the CLI rate of change terms is highly significant in each equation although not all of them are significant. The significance is confirmed of the rate of change of the CLI in forecasting the subsequent rate of change of the IPI in the framework of in-sample analysis. With different values of $a$, ranging from zero to three, the results are essentially unchanged while details are not reported to save space.


<table>
<thead>
<tr>
<th>Equation</th>
<th>$k$</th>
<th>$\gamma$</th>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>$\phi_0$</th>
<th>$\phi_1$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3.3)</td>
<td>3</td>
<td>0.261</td>
<td>0.271*</td>
<td>0.076</td>
<td>−0.103</td>
<td>0.319*</td>
<td>0.436</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.138)</td>
<td>(0.068)</td>
<td>(0.059)</td>
<td>(0.067)</td>
<td>(0.066)</td>
<td></td>
</tr>
<tr>
<td>(3.4)</td>
<td>3</td>
<td>0.349**</td>
<td>0.169*</td>
<td>0.222*</td>
<td>0.386</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.133)</td>
<td>(0.059)</td>
<td>(0.051)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.3)</td>
<td>6</td>
<td>0.731**</td>
<td>0.397*</td>
<td>−0.002</td>
<td>−0.177</td>
<td>0.198**</td>
<td>0.445</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.325)</td>
<td>(0.099)</td>
<td>(0.100)</td>
<td>(0.099)</td>
<td>(0.082)</td>
<td></td>
</tr>
<tr>
<td>(3.4)</td>
<td>6</td>
<td>0.759*</td>
<td>0.322*</td>
<td>0.067</td>
<td>0.440</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.290)</td>
<td>(0.093)</td>
<td>(0.098)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: * significant at the one-percent level, and ** significant at the five-percent level. Standard errors are in parentheses.

Tables 10 and 11 indicate results from forecasts made using the CLI with $a = 1$ and $a = 3$, with the full-sample periods and the sub-sample periods that do not include the periods of the definitional changes. The tables show performances of the three classes of forecast; the performances are evaluated by mean squared prediction errors (MSPEs). When the start of a sample period is in period $t_0$ and the end is in period $t_0 + n - 1$, the MSPE is expressed as follows:

$$MSPE = \frac{1}{n} \sum_{t=t_0}^{t_0+n-1} \{IPIR(t) - (\text{forecasted rate of change})\}^2.$$ 

Again, the start of regression for estimation of $\gamma$ and $\delta$ is in January 1974 for all classes of forecast, and the end depends on the class of forecast. The start of the sample period for the MSPEs is $k$ periods after January 1988, and the end is in December 2001 with the full-sample period.

There is no essential difference between the cases of $k = 3$ and $k = 6$. With the full-sample period, the in-sample forecast gives the best score, i.e., the smallest MSPE, and the weak real-time forecast gives the worst score, i.e., the largest MSPE. This result is intuitive. The differences in the scores are relatively small in many cases. The weak real-time forecasts also perform relatively well with the sub-sample periods. These results do not depend on whether lagged terms of the rate of change of the IPI are included.
Table 10. MSPE of forecasts by the CLI with $k = 3$.

<table>
<thead>
<tr>
<th>Period</th>
<th>Equations (3.3) and (3.5): $a = 1$</th>
<th>Equations (3.4) and (3.6): $a = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full-sample</td>
<td>Sample</td>
</tr>
<tr>
<td>Forecasts</td>
<td></td>
<td>period</td>
</tr>
</tbody>
</table>

Table 11. MSPE of forecasts by the CLI with $k = 6$.

<table>
<thead>
<tr>
<th>Period</th>
<th>Equations (3.3) and (3.5): $a = 3$</th>
<th>Equations (3.4) and (3.6): $a = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full-sample</td>
<td>Sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td>period</td>
</tr>
</tbody>
</table>

Table 12 shows the DM statistics in the case of $k = 3$ with the full-sample period, to compare formally the performances of the forecasts. The method of calculation of the statistics is similar to that used for the probit models. The DM statistics are rather high compared with those from the DLI forecasts. With lagged terms of the rate of change of the IPI, the null hypothesis of the set of the in-sample and the weak real-time forecasts is rejected. However, the null hypothesis of the set of the out-of-sample and the weak real-time forecasts is not rejected in any of the cases. Consequently, even the weak real-time forecasts are often comparable to in-sample forecasts, and do not perform worse than out-
Table 12. Forecasting power of weak real-time forecasts: DM statistics, full-sample period, with $k = 3$.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>in-sample forecasts vs. weak real-time forecasts</th>
<th>out-of-sample forecasts vs. weak real-time forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equations (3.3) and (3.5): $a = 1$</td>
<td>1.786*</td>
<td>1.451</td>
</tr>
<tr>
<td>Equations (3.3) and (3.5): $a = 3$</td>
<td>1.993*</td>
<td>1.541</td>
</tr>
<tr>
<td>Equations (3.4) and (3.6): $a = 1$</td>
<td>1.599</td>
<td>1.361</td>
</tr>
<tr>
<td>Equations (3.4) and (3.6): $a = 3$</td>
<td>1.597</td>
<td>1.240</td>
</tr>
</tbody>
</table>

Note: * significant at the one-percent level, and ** significant at the five-percent level.

Table 13. Forecasting power of weak real-time forecasts: DM statistics, full-sample period, with $k = 6$.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>in-sample forecasts vs. weak real-time forecasts</th>
<th>out-of-sample forecasts vs. weak real-time forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equations (3.3) and (3.5): $a = 1$</td>
<td>1.707*</td>
<td>0.991</td>
</tr>
<tr>
<td>Equations (3.3) and (3.5): $a = 3$</td>
<td>1.853*</td>
<td>0.895</td>
</tr>
<tr>
<td>Equations (3.4) and (3.6): $a = 1$</td>
<td>1.567</td>
<td>1.074</td>
</tr>
<tr>
<td>Equations (3.4) and (3.6): $a = 3$</td>
<td>1.577</td>
<td>1.024</td>
</tr>
</tbody>
</table>

Note: * significant at the one-percent level, and ** significant at the five-percent level.

of-sample forecasts. It follows that weak real-time forecasts are comparable to forecasts based on ex-post data in many cases. Table 13 shows the DM statistics for $k = 6$ to compare formally the performances of the forecasts. The results are basically the same as those in the case of $k = 3$.

6. Conclusion

Empirical results in this paper indicate that making real-time forecasts based on the Japanese official business indices sometimes requires caution.

For the DLI, both real-time forecasts perform almost as well as the in-sample and out-of-sample forecasts. No significant difference is found in their performances. These results basically imply that one should not feel refrained too much from making real-time forecasts provided that forecasts from ex-post analysis are satisfactory. Revision of numbers and delay in recognition of turning points are found not to cause serious problems from the point of view of forecasting. This point confirms the practical usefulness of the DLI. In considering previous studies, the out-of-sample and the real-time forecasts give almost the same predictive accuracy.

In the case of the CLI, the weak real-time forecasts perform as well as the out-of-sample forecasts, and their performances are often comparable to those of in-sample forecasts. This confirms the practical usefulness of the CLI. Even preliminary numbers of the CLI are useful in forecasting the subsequent increasing rate of production provided that one is satisfied with forecasts from the ex-post analysis.

However, in the case of the DCI, the real-time forecasts perform much worse than the in-sample and out-of-sample forecasts. Statistically, these differences
in performances are significant. It is obviously desirable to discern the current state of the business cycle as rapidly as possible. But the results for the DCI show that caution is necessary in making real-time forecasts based on preliminary numbers of the DCI. Even if one is satisfied with performances of forecasts from ex-post analysis, real-time forecasts do not perform as well as expected. Any evaluation of forecasts based on final revised data of the DCI does not provide reliable information on performance of real-time forecasts.

The results from the sub-sample periods suggest that this problem with the DCI is not due simply to the definitional changes. Even when effects of the definitional changes are excluded, the real-time forecasts perform much worse than forecasts based on ex-post data.

What causes the real-time forecasts by the DCI to be so unreliable? Two causes can be identified that could give rise to the difference between the forecasts from ex-post analysis and the real-time forecasts. These are revision of the numbers, and delay in recognizing or change to the timing of turning points in the business cycle. The weak real-time and the strong real-time forecasts both fail to perform comparably with the in-sample and out-of-sample forecasts. This tells us that revision is the main cause of the poor performance of the real-time forecasts. Even without the delay in recognizing turning points, one would not be able to make a reliable forecast using only preliminary numbers of the DCI. The descriptive statistics and the results of forecasts with the sub-sample periods also suggest that correction of preliminarily announced numbers, rather than definitional changes, is mainly responsible for this problem with the DCI.

The QPSs in Tables 4 to 6 indicate that the in-sample forecasts using the DCI perform quite well when compared with those from the DLI. This implies that the DCI performs much better than the DLI in the framework of ex-post analysis. However, there is virtually no difference in forecast performance between the DLI and the DCI in the framework of real-time analysis.

The present results also indicate that caution is necessary with the use of preliminary numbers of the DCI when performance is sought comparable to that of forecasts using ex-post analysis. Compilation of preliminary numbers of the DCI should be improved and the promptness should be maintained. These problems remain even without definitional changes.

Acknowledgements

This paper was presented at the annual meeting of the Japan Statistical Society, held at Meijo University in September 2003. I am grateful to Professor Kazuo Mori and two referees of this journal for their comments. I am solely responsible for remaining errors in this paper.

References


