Forecasting Output With the Composite Leading Index: A Real-Time Analysis

FRANCIS X. DIEBOLD and GLENN D. RUDEBUSCH*

We examine the ability of the composite index of leading economic indicators to predict future movements in aggregate economic activity. Previous examinations of predictive performance have evaluated either the in-sample residual errors from a forecasting equation fitted to the entire sample of data or the out-of-sample forecast errors from an equation fitted to a subsample of the data. Unlike previous evaluations, we perform a real-time analysis, which uses the provisional and partially revised data for the leading index that were actually available historically, along with recursive out-of-sample forecasts. We find a substantial deterioration of forecasting performance in the real-time framework.

KEY WORDS: Causality; Evaluation; Prediction.

The Government's best-known tool for economic forecasting—the index of leading indicators—will undergo a major revision by the end of the year. . . .

New York Times (May 29, 1987)

1. INTRODUCTION

The prediction of aggregate economic activity with the composite leading index (CLI) has enjoyed widespread popularity for several decades. Some empirical justification for this popularity has been provided by a number of recent formal evaluations of the forecasting ability of the CLI, which have found that the CLI *does* contain useful information for the prediction of aggregate economic activity. For example, Auerbach (1982) and Koch and Rasche (1988), in analyses of bivariate causality, found strong evidence that the CLI is useful in linear prediction of industrial production even after conditioning on lagged values of production and even in out-of-sample forecasts. In a multivariate context, Braun and Zarnowitz (1989) provided confirmatory evidence showing that inclusion of the CLI in commonly estimated vector autoregressive representations leads to a reduction of in-sample residual variance.

As in most examinations of predictive performance, the aforementioned studies evaluate either or both of two types of forecast errors: (1) the residual errors from a model fit to the entire sample of data, and (2) the out-of-sample forecast errors from a model fit to just a portion of the data set. Many have noted that these two types of errors may have very different characteristics. In particular, the out-of-sample forecast errors for a given model are usually much larger than its in-sample residual errors [for example, see Makridakis and Winkler (1989)]. This difference can often be traced to the overfitting of the (misspecified) model, which reduces the in-sample errors but does not improve forecast performance [e.g., Hendry (1980)]. In our analysis, we highlight a little-recognized limitation of forecast evaluations that examine either in-sample residuals or out-of-sample forecast errors, namely, that they are conducted with final revised values of the data. In contrast, actual real-time forecasting must rely on preliminary and partially revised data. In this article we introduce a third category of forecast errors that directly confronts this issue. We evaluate the *real-time* predictive performance of the CLI by not only examining out-of-sample forecasts but also by using the preliminary and partially revised CLI data that would have been available at the historical date of each forecast. Essentially, we are able to reproduce the CLI information set of a real-time forecaster.

The use of final revised CLI data in previous analyses of the predictive ability of the CLI would not be a critical flaw if the CLI were subject to only small revisions. Unfortunately, the CLI is extensively revised from its preliminary estimate to its final value. Not only are statistical revisions incorporated as more complete historical data become available for the components, but components are often added and eliminated, ex post, to improve the leading index's performance retrospectively. The existence of such revisions, especially the definitional ones, suggests the possibility that the good performance of the CLI in previous forecasting exercises may be spurious, in the sense that the CLI data actually available in real time were substantially less helpful in forecasting changes in real activity than the evaluations that use final CLI data suggest.

In Section 2 we discuss the nature of statistical and definitional revisions in the CLI and provide details of the CLI data matrix that was constructed for our analysis. Section 3 describes our prediction methodology, which proceeds in a number of stages as the forecaster's information set is progressively restricted from an ex post to a real-time, or ex ante, analysis. The empirical results are contained in Section 4, and Section 5 discusses the robustness of those results. Concluding remarks are offered in Section 6.

2. REVISIONS IN THE LEADING INDEX

Although the information content of preliminary and partially revised data is a consideration in any real-time fore-

* Francis X. Diebold is J. M. Cohen Assistant Professor, Department of Economics, University of Pennsylvania, Philadelphia, PA 19104-6297. Glenn D. Rudebusch is Economist, Division of Research and Statistics, Federal Reserve Board, Washington, DC 20551. Diebold gratefully acknowledges financial support from the National Science Foundation (Grant SES 89-2715), the University of Pennsylvania Research Foundation (Grant 3-71441), and the Institute for Empirical Macroeconomics. The authors thank the editor, associate editor, and two anonymous referees, as well as Alan Auerbach, Neil Ericsson, Lawrence Klein, Doug McManus, Andy Rose, Mark Rush, Herman Stekler, Peter Tinsley, and Ken West for useful comments. Bill Nelson and Roberto Sella provided outstanding research assistance. The views expressed here are those of the authors and are not necessarily shared by the Board of Governors of the Federal Reserve System or its staff.

Journal of the American Statistical Association
September 1991, Vol. 86, No. 415, Applications and Case Studies

casting situation, it is especially important when evaluating the performance of the composite index of leading indicators. The CLI is extensively revised from its preliminary estimate to its final form, undergoing both statistical and definitional revisions. This section describes the nature of these revisions. For descriptions of the components and the methodology used to construct the CLI, see Auerbach (1982) and Hertzberg and Beckman (1989).

Toward the end of each month during our sample, the Bureau of Economic Analysis (BEA) issued a *Business Conditions Digest* (BCD), which contained a preliminary estimate of the previous month's value of the CLI on the basis of incomplete and preliminary source data. For example, the October 1988 issue of the BCD contained the initial estimate of the September 1988 value of the leading index. Each month, the BCD also provided revisions in the index for any or all of the preceding 11 months. These *statistical* revisions to the CLI occurred because of statistical revisions in the component indicators (reflecting the collection of larger, more representative source data samples as time passes) and because some components were released too late to be included in the preliminary estimate and could only be included in the first or second revision of the index.

Statistical revisions, of course, plague the real-time interpretation of most economic time series; in the case of the CLI, however, the final data have not only undergone these statistical revisions but, in addition, the components that make up the index have been reselected and reweighted ex post. These *definitional* revisions in the CLI involve a retrospective reevaluation of the CLI's performance over the enlarged historical sample and are attempts to improve the lead-time performance of the index during the historical sample. Major changes in the definition of the CLI are fairly frequent, occurring in August 1969, April 1975, February 1979, January 1982, January 1983, January 1987, and January 1989—about once every two years since the introduction of the CLI in the November 1968 issue of the BCD.

As a typical example, in the 1983 revision, the BEA recalculated component weights and standardization factors (on the basis of additional data) and replaced two of the components (crude materials price inflation and the change in liquid assets) with series that were broadly similar but produced a more consistent cyclical lead performance retrospectively. [A complete chronology and description of all compositional changes is provided in Diebold and Rudebusch (1988).] Clearly, such ex post definitional changes make inference about the real-time forecasting ability of the CLI very difficult.

In other work (Diebold and Rudebusch 1988), we have examined the magnitude of both statistical and definitional revisions to the CLI. From December 1969 through January 1986, the standard deviation of the revision from the preliminary estimate of the percent change in the CLI to the final estimate (as of December 1987) is .86%. This is large relative to the standard deviation of the percent change in the final estimate of the CLI over that same period, which is 1.11%; the signal-to-noise ratio is about 1.3. Within definitional regimes, when definitional revisions are not a factor, the standard deviations of revisions are smaller but still substantial. Such large revisions imply that the ex ante, real-time forecasting performance of the CLI may be much worse than the ex post evaluations conducted with final, revised data suggest.

In this article, we examine precisely how much of a deterioration in the forecasting performance of the CLI occurs from using real-time CLI data. For this purpose we have collected all preliminary and partially revised values of the CLI as published sequentially in issues of the BCD from November 1968 through December 1988. The data matrix containing all of these values is illustrated schematically in Table 1. Each row of the matrix provides the value of the CLI for a particular month, ranging from 1948:1 to 1988:11 (491 months). The entries of a given row vary across the columns as the value of the CLI for that month is revised. Each column of the matrix provides the data given in one

Table 1. Schematic of CLI Data Matrix

Data for this month	*Data released in this month (BCD issue date)*						
	j = 1968:11	j = 1968:12	. .	$j = t$	. .	j = 1988:11	j = 1988:12
i = 1948:1	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
i = 1948:2	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
i = 1948:3	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
.	.	.	. .	.	. .	.	.
.	.	.	. .	.	. .	.	.
i = 1968:10	$\mathrm{CLI}(i)_j^p$	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
i = 1968:11	0	$\mathrm{CLI}(i)_j^p$	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
i = 1968:12	0	0	. .	$\mathrm{CLI}(i)_j$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
.	.	.	. .	.	. .	.	.
$i = t - 1$	0	0	. .	$\mathrm{CLI}(i)_j^p$	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
$i = t$	0	0	. .	0	. .	$\mathrm{CLI}(i)_j$	$\mathrm{CLI}(i)_j$
.	.	.	. .	.	. .	.	.
i = 1988:10	0	0	. .	0	. .	$\mathrm{CLI}(i)_j^p$	$\mathrm{CLI}(i)_j$
i = 1988:11	0	0	. .	0	. .	0	$\mathrm{CLI}(i)_j^p$

NOTE: The CLI data matrix is 491 × 242. The entry $\mathrm{CLI}(i)_j$ denotes the value of the leading index in the ith month that was available and current as of the jth month; a zero entry indicates that no CLI data had been released for that month. A superscript p denotes the preliminary (initial) estimate for a given month. Moving across the columns from left to right yields varying nonzero entries (from revisions) and a progressively larger sample as history unfolds.

issue of the BCD; the leftmost column is from the November 1968 issue, which first introduced the CLI and presented historical CLI data from January 1948 through October 1968, the next column is from December 1968, and so forth, across the 242 issues. A zero value in the i, j entry indicates that data for the ith month (the row) were not yet released in the jth month (the column). Moving through real time across the columns, each additional month adds one more nonzero observation to the column (denoted by superscript p for preliminary) and changes any or all of the previous, nonzero entries. Every statistical and definitional revision in the CLI is thus represented. Statistical revisions change only the last few nonzero entries in the preceding column, whereas definitional revisions usually will change all of them. This data matrix, which contains about 90,000 nontrivial entries, allows us to reproduce the precise information content in the CLI available to forecasters at any point in time.

3. METHODOLOGY

We shall consider the usefulness of the CLI for predicting industrial production (IP) in a linear forecasting framework. In particular, we examine the marginal predictive information content of the CLI beyond that embodied in lags of industrial production. There are five basic forecasting scenarios that we consider, each differing by the inclusion or exclusion of CLI lags and by the degree to which they mimic an actual real-time forecasting environment. The first two scenarios involve in-sample residuals and constitute the usual ex post Granger causality test. The second two are partially ex ante exercises that use recursive estimates of forecasting models, that is, the sample is enlarged by one observation at a time and the model is reestimated with each new observation and used to produce out-of-sample forecasts. Finally, the last scenario produces real-time forecasts with the CLI by using recursive estimates and by using the preliminary and partially revised estimates of the CLI that were historically available.

The first two forecasting scenarios are based on the entire sample of final revised data. The first of these estimates the regression,

$$\mathrm{IP}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \, \mathrm{IP}_{t-i-k+1} + \varepsilon_t, \qquad (1)$$

where IP is the log of the level of industrial production. The forecasting equation for this scenario will be referred to as the FF1 regression (for *F*ull sample, *F*inal data, and lags of *1* variable on the right side, namely, IP). The number of months ahead the prediction is being made (k) is varied over 1, 4, 8, and 12. The lag length of the estimated regression (p) is also varied over 1, 4, 8, and 12. (The sample of IP data used on the left side of FF1, and FF2 later, always ranges from 1950:1 to 1988:12. Of course, depending on p and k, the ranges of the right side IP data and, in FF2, of the CLI data vary.)

The second ex post regression includes on the right side lagged values of the final revised CLI. (In terms of the earlier-discussed CLI data matrix, the CLI data used are from the rightmost column.) Specifically, we estimate (from 1950:1 to 1988:12) the regression:

$$\mathrm{IP}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \, \mathrm{IP}_{t-i-k+1} + \sum_{i=1}^{p} \gamma_i \, \mathrm{CLI}_{t-i-k+1} + \varepsilon_t, \qquad (2)$$

where CLI is the log of the level of the composite leading index, and again we let $k = 1, 4, 8, 12$ and $p = 1, 4, 8, 12$. We denote this as the FF2 regression (*F*ull sample, *F*inal data, and lags of *2* variables on the right side, namely, IP and CLI).

From the FF1 and FF2 regressions, we save the 231 residuals from 1969:10 to 1988:12. (This sample of residuals is used for comparison with the later recursively estimated regressions. Given November 1968 as the first historical release date of the CLI, the first possible 12-step-ahead, real-time forecast was for the 1969:10 value of IP.) The causality test of Granger (1969), which was employed in this context by Auerbach (1982), provides a formal comparison of the sums of squares of these residuals. The FF1 regression provides the predictive content of only own-lags of IP, and the FF2 regression gives predictions enhanced with the marginal contribution from the CLI. In terms of subsequent results, it will prove useful to think of the fitted values from the FF1 and FF2 regressions as "forecasts"; clearly, however, these "forecasts" are completely ex post, as the estimation uses the *full* sample of *final revised* data. As noted in the introduction, the "forecast errors" in these two ex post scenarios are merely in-sample regression residuals. These errors are used to construct an associated mean squared prediction error (MSPE) and mean absolute prediction error (MAPE) from each regression. The MSPE is simply a normalized restricted sum of squares (in the FF1 case) and a normalized unrestricted sum of squares (in the FF2 case).

In the second pair of forecasting scenarios, the ex post nature of FF1 and FF2 is partially relaxed via recursive estimation of the equations to mimic real-time forecasting (which is affected by subsample sampling variability in coefficient estimates); however, the final revised CLI data are still used. In this out-of-sample forecasting exercise, Equations (1) and (2) are denoted by RF1 and RF2 (for *R*ecursive, *F*inal data regressions). An example will make the procedure clear. Consider the estimation of RF2, and suppose that $p = k = 4$. Then, we first estimate, over the period 1950:1 through 1969:6,

$$\mathrm{IP}_t = \beta_0 + \sum_{i=1}^{4} \beta_i \, \mathrm{IP}_{t-i-4+1} + \sum_{i=1}^{4} \gamma_i \, \mathrm{CLI}_{t-i-4+1} + \varepsilon_t, \qquad (3)$$

where the CLI is the final revised series (1988:12 definition). The four-step-ahead forecast of $\mathrm{IP}_{1969:10}$ (which is made with the final data through 1969:6) is then constructed from the estimated coefficients as

$$\hat{\mathrm{IP}}_{1969:10} = \hat{\beta}_0 + \sum_{i=1}^{4} \hat{\beta}_i \, \mathrm{IP}_{1969:10-i-4+1} + \sum_{i=1}^{4} \hat{\gamma}_i \, \mathrm{CLI}_{1969:10-i-4+1}. \qquad (4)$$

The regression is then reestimated with one more observation using data from 1950:1 through 1969:7, the forecast for 1969:11 is constructed, and so forth. We progress recursively through the sample in this fashion—continually reestimating and reforecasting—until the forecast for 1988:12 is constructed, at which point the sample is exhausted. For RF1 and RF2 and for each combination of k and p, we obtain a 231-element vector of out-of-sample forecasts. An MSPE and an MAPE are calculated for each of the associated vectors of prediction errors.

In the recursive regression RF2 each "forecast" is completely out-of-sample with regard to parameter estimation but, as should be clear from the earlier discussion of definitional revisions in the CLI, these "forecasts" may incorporate subtle ex post information through retrospective reconstruction of the CLI by the BEA. To eliminate this ex post information, our final scenario constructs real-time ex ante forecasts with the CLI by recursively estimating (2), but at each month in which the k-step-ahead forecast is made using only the preliminary and partially revised CLI data that were actually available in that month. Thus, as a new regression is estimated for each month, the CLI data vector used for the regression changes. For each new month the CLI data vector will have one more element for the latest preliminary number, but the 11 previous elements, in the case of statistical revisions, or perhaps all previous elements, in the case of definitional revisions, may have changed. We shall denote this real-time ex ante forecasting regression by RP2 (*R*ecursive, *P*reliminary CLI data forecasts). The MSPE and MAPE of this regression can be compared with RF2, determining the forecast deterioration due to provisional data, and with RF1, providing a real-time ex ante causality test analog. [The CLI data vector used in RP2 corresponds to a particular column of the real-time data matrix, moving from the leftmost column to the rightmost column as time passes. In terms of Eq. (3) (with $p = k = 4$), the four-step-ahead forecast for 1969:10 is made with the CLI data matrix column that contains the preliminary number for 1969:6. If we took into account the one-month lags in releasing the CLI and IP data, the real-time forecast dates would be advanced one month, but the data used and the forecasts obtained would not change.]

In the next section, we analyze the forecast errors that result from application of the aforementioned procedures. We also have explored, as shall be described in Section 5, several variations and extensions of these methods to assess the robustness of our results.

4. EMPIRICAL ANALYSIS

The methodology outlined previously produces 80 vectors of forecast errors, corresponding to all combinations of steps ahead forecasted (k = 1, 4, 8, and 12) and number of lags included (p = 1, 4, 8, and 12) for each regression strategy FF1, FF2, RF1, RF2, and RP2. Each vector of errors has 231 elements, corresponding to forecasts for 1969:10 through 1988:12. For each vector the MSPE and the MAPE were calculated; these are presented in Tables 2 and 3, respectively.

Let us first consider the MSPE results in Table 2. There are four separate comparisons that are of particular interest. The first is the ex post Granger causality test, which compares the MSPE for the in-sample residuals from the FF1 regression [denoted by MSPE (FF1)] with the MSPE for the in-sample residuals from the FF2 regression [denoted by MSPE (FF2)]. For all p and k combinations, a consistent pattern emerges:

$$\text{MSPE(FF2)} < \text{MSPE(FF1)}. \qquad \text{(Result A)}$$

This confirms the bivariate causality results of Auerbach (1981); that is, the CLI has marginal predictive content, ex post.

The second comparison of interest extends the Granger causality methods to out-of-sample forecasts from regressions estimated recursively with the final data. In this case,

$$\text{MSPE(RF2)} < \text{MSPE(RF1)}. \qquad \text{(Result B)}$$

Thus this out-of-sample analog to the Granger test also indicates that the CLI has marginal predictive power beyond own-lags of IP.

Table 2. Mean Squared Prediction Error, Industrial Production

Steps ahead (k)	Lags included (p)	Forecast scenario				
		FF1	FF2	RF1	RF2	RP2
1	1	.87	.63	.88	.70	1.08
1	4	.64	.50	.65	.54	.65
1	8	.64	.47	.66	.54	.66
1	12	.64	.47	.67	.56	.66
4	1	7.66	4.68	8.11	5.89	11.85
4	4	6.29	3.47	6.74	4.42	7.64
4	8	6.29	3.16	6.80	4.09	7.21
4	12	6.47	3.24	7.24	4.55	7.89
8	1	20.55	12.92	22.87	16.90	34.15
8	4	18.69	9.21	20.96	12.22	25.54
8	8	19.10	8.99	21.89	12.82	26.08
8	12	19.28	9.21	22.80	14.50	28.31
12	1	34.77	24.57	40.13	31.26	56.09
12	4	33.38	18.63	39.41	24.61	48.69
12	8	33.83	18.51	41.12	27.02	50.52
12	12	33.14	18.06	41.56	28.59	52.36

NOTE: Entries are MSPE × 10,000.

Table 3. Mean Absolute Prediction Error, Industrial Production

Steps ahead (k)	Lags included (p)	Forecast scenario				
		FF1	FF2	RF1	RF2	RP2
1	1	.68	.61	.69	.66	.78
1	4	.58	.53	.58	.57	.62
1	8	.58	.52	.59	.56	.61
1	12	.59	.62	.60	.58	.62
4	1	2.00	1.68	2.13	1.92	2.51
4	4	1.80	1.47	1.91	1.66	2.09
4	8	1.80	1.39	1.93	1.56	1.99
4	12	1.79	1.40	1.93	1.63	2.06
8	1	3.51	2.88	3.80	3.23	4.14
8	4	3.30	2.36	3.60	2.71	3.52
8	8	3.32	2.30	3.67	2.69	3.51
8	12	3.34	2.34	3.73	2.86	3.72
12	1	4.78	4.04	5.20	4.58	5.39
12	4	4.67	3.40	5.15	3.93	5.02
12	8	4.68	3.48	5.23	4.12	5.11
12	12	4.65	3.44	5.25	4.18	5.24

NOTE: Entries are MAPE × 100.

It is also interesting to note the deterioration in performance of the out-of-sample forecasts compared with the in-sample residuals. For all p and k, MSPE(FF1) < MSPE(RF1) and MSPE(FF2) < MSPE(RF2).

Finally, we consider real-time forecast comparisons that involve the preliminary CLI data. Our third result compares the recursive forecasting regression using the final CLI data with the recursive forecasting regression using the preliminary CLI data and finds that

$$\text{MSPE(RF2)} < \text{MSPE(RP2)}. \qquad \text{(Result C)}$$

This indicates that the final revised CLI data are more helpful in forecasting IP than the preliminary data; thus statistical and definitional revisions clearly improve the CLI's predictive power. Our final comparison considers whether the preliminary CLI data have any marginal predictive power over one-lags of IP. Here, we summarize our findings as

$$\text{MSPE(RP2)} \approx \text{MSPE(RF1)} \qquad \text{(Result D)}$$

to indicate that, in general, the MSPE of these scenarios are quite close and there is no clear ranking across all combinations of p and k for these two MSPE's. Thus inclusion of the CLI in real time does not improve and, often, *worsens* predictive performance relative to simply forecasting with an IP autoregression. None of these results change when MSPE is replaced by MAPE (Table 3).

The MSPE and MAPE results are illustrated in Figure 1 for the case of four-step-ahead forecasts ($k = 4$) using eight lags ($p = 8$). The sawtooth pattern suggested by Results A–D is clearly evident. MSPE falls from FF1 to FF2 with inclusion of final CLI data. It rises again to an even higher level in the recursive RF1 but falls with the addition of revised CLI data in RF2. Finally, however, the MSPE rises when the preliminary CLI data are included in RP2.

Although the foregoing rankings are indicative, it is important to examine formally the statistical significance of the comparisons. A test for significance of the MSPE difference between two unbiased forecasts is readily available. Let e_1 denote a given forecast error, and let e_2 denote a second comparison forecast error. Then the difference in the MSPE's, which reduces to the difference in variances, by unbiasedness, is simply the covariance of $(e_1 + e_2)$ and $(e_1 - e_2)$. That is,

$$E[(e_1 + e_2)(e_1 - e_2)] = \sigma_1^2 - \sigma_2^2. \qquad (5)$$

Thus testing for significance of $\sigma_1^2 - \sigma_2^2$ is equivalent to testing significance of $\text{cov}[(e_1 + e_2), (e_1 - e_2)]$. [See Granger and Newbold (1986), who noted that this test is uniformly most powerful invariant.]

We estimate this covariance and test its significance using methods robust to serial correlation of unknown form; such an approach is particularly important for $k > 1$, because multiple-step-ahead forecast errors are generally serially correlated. Define $x = e_1 + e_2$ and $y = e_1 - e_2$. We proceed by noting that the covariance (5) is just the cross-covariance of x and y at a displacement of 0, that is, $\gamma_{xy}(0)$, which can be consistently estimated in the usual manner. It is well known that under standard assumptions [see Priestley (1980, pp. 692–693)] the variance of the sample covariance is

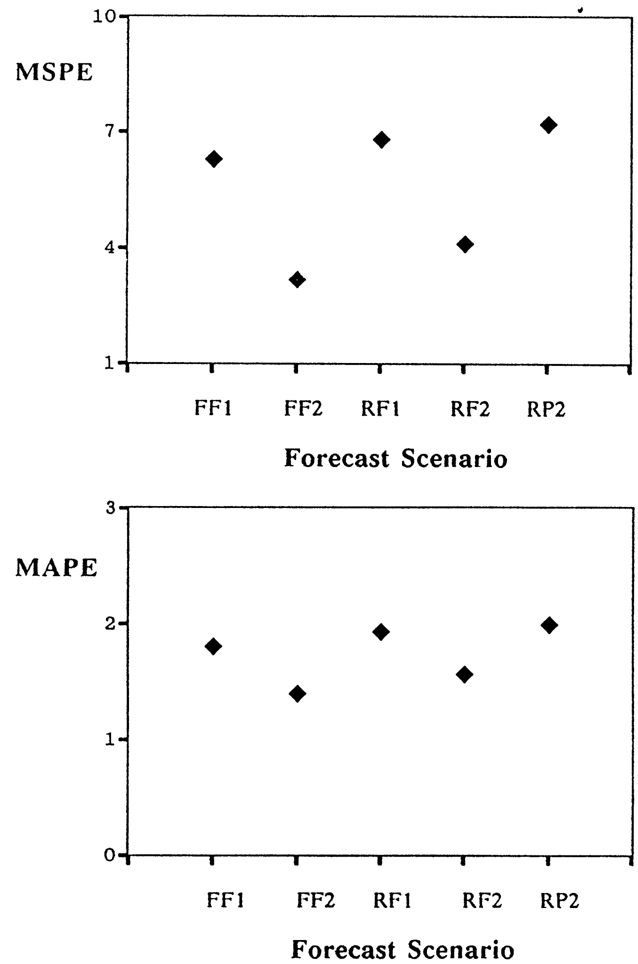

Figure 1. Performance of IP Forecasts: Four-Step-Ahead Forecast With Eight Lags (k = 4, p = 8).

$$\text{var}(\hat{\gamma}_{xy}(0)) \approx \frac{1}{T} \sum_{\tau=-\infty}^{\infty} [\gamma_{xx}(\tau)\gamma_{yy}(\tau) + \gamma_{xy}(\tau)\gamma_{yx}(\tau)],$$

a consistent estimate of which requires consistent estimates of $\gamma_{xx}(\tau)$, $\gamma_{yy}(\tau)$, $\gamma_{xy}(\tau)$, and $\gamma_{yx}(\tau)$, which can be obtained in the usual manner. One remaining issue is the truncation of the doubly infinite sum. We explored a range of truncation points; the empirical results, which are contained in Table 4, use a truncation at lag 25, but the estimates were robust to other choices. The MSPE differences associated with Results A, B, and C are almost uniformly significant at all steps ahead and for all lags included. The MSPE differences associated with Result D, conversely, are insignificant. We conclude that, subject to the qualifications and caveats noted below, inclusion of the CLI in real-time forecasting equations is of little value.

5. VARIATIONS AND EXTENSIONS

The results reported above are robust to a number of variations. In particular, qualitatively similar results were obtained when first differences rather than levels of all variables were used, as shown in the MSPE comparisons in Table 5. Qualitatively similar results were also obtained,

Table 4. Tests for Significance of MSPE Differences

Steps ahead (k)	Lags included (p)	Result A, FF1–FF2	Result B, RF1–RF2	Result C, RF2–RP2	Result D, RF1–RP2
1	1	.021	.033	.018	.266
1	4	.000	.005	.040	.451
1	8	.000	.002	.025	.426
1	12	.000	.006	.037	.378
4	1	.035	.066	.020	.202
4	4	.004	.018	.021	.393
4	8	.001	.010	.016	.480
4	12	.001	.008	.009	.459
8	1	.057	.108	.019	.173
8	4	.014	.038	.008	.295
8	8	.013	.039	.009	.338
8	12	.012	.038	.007	.320
12	1	.085	.142	.027	.183
12	4	.043	.071	.014	.269
12	8	.048	.079	.020	.333
12	12	.044	.078	.017	.325

NOTE: Entries are marginal significance levels (p values) for the null hypothesis that the true MSPE difference is 0.

as shown in Table 6, after eliminating from the sample the pre-1960 data, which are of lower quality.

From the perspective of those who originally developed the leading indexes of economic activity, a potential objection to our results is that we evaluate the CLI using a questionable criterion, namely, the ability to forecast movements in IP. Instead, they might argue that the CLI is designed to signal broad changes in business conditions and, particularly, business recessions. One response to this would be to consider a more comprehensive indicator of aggregate economic activity. (IP includes roughly one-third of total output.) With this in mind, we also conducted the analysis of Section 4 after replacing IP with the composite index of coincident economic indicators, a much broader monthly measure of economic conditions. Little difference was found. As shown in Table 7, there is an obvious sawtooth pattern

Table 5. Mean Squared Prediction Error, Growth in Industrial Production

Steps ahead (k)	Lags included (p)	Forecast scenario				
		FF1	FF2	RF1	RF2	RP2
1	1	6.48	5.85	6.57	5.95	6.51
1	4	6.39	5.19	6.52	5.41	5.85
1	8	6.42	4.89	6.60	5.42	5.75
1	12	6.44	4.90	6.79	5.77	6.48
4	1	8.48	7.30	8.72	7.55	8.82
4	4	8.53	6.72	8.83	7.17	8.71
4	8	8.58	6.44	9.05	7.33	9.16
4	12	8.53	6.59	9.36	8.02	9.94
8	1	8.73	8.28	8.99	8.51	9.47
8	4	8.78	7.98	9.17	8.44	9.70
8	8	8.78	8.22	9.50	9.19	19.58
8	12	8.72	8.16	9.49	9.45	10.88
12	1	9.05	8.82	9.55	9.33	9.31
12	4	8.95	8.75	9.65	9.54	9.79
12	8	8.89	8.50	9.64	9.45	10.05
12	12	8.74	8.46	9.60	9.66	9.94

NOTE: Entries are MSPE × 100,000.

for the MSPE for forecasting the composite coincident index that accords with the earlier results.

More fundamentally, however, there is some evidence that those who originally constructed the leading index intended to use it for the prediction of business cycle turning points rather than for forecasting changes in the level of aggregate economic activity over the entire cycle. In this view attention is placed on the ability of the leading index to predict an economic event (a turning point) rather than its ability to forecast future values of economic time series. We have some sympathy for this view; indeed, in Diebold and Rudebusch (1989), we employed a turning point filter that translated movements in the CLI (using final revised data) into probabilities of an imminent recession. However, in Diebold and Rudebusch (1991), a companion article to this one, we showed that the ability of the CLI to forecast turning points on a real-time basis (using the data matrix of Table 1) is substantially worse than its ability on an ex post basis using final revised data. Thus we believe that our qualitative conclusions are robust to use of an evaluation criterion of turning point prediction. To provide further evidence more directly in line with the present investigation, we examined regressions where IP on the left side was replaced in each scenario by a dummy variable signaling expansion or recession. The relative ranking of prediction errors followed that in Section 4. [Further evidence is provided by three earlier studies—Stekler and Schepsman (1973), Hymans (1973), and Zarnowitz and Moore (1982)—that have employed preliminary CLI data along with simple turning point filters, for example, the use of three consecutive declines as a recessionary signal. These studies have also found that use of preliminary data reduces the usefulness of the CLI in turning point forecasting.]

One could also accord much more statistical expertise to the real-time forecaster than we have done; in particular, there are real-time prediction procedures that are more sophisticated in extracting information from preliminary data

Table 6. Mean Squared Prediction Error, Post-1960 Industrial Production

Steps ahead (k)	*Lags included (p)*	*Forecast scenario*				
		FF1	*FF2*	*RF1*	*RF2*	*RP2*
1	1	.87	.61	.89	.66	.88
1	4	.64	.49	.71	.59	.67
1	8	.64	.46	.74	.59	.69
1	12	.64	.44	.77	.61	.72
4	1	7.70	4.30	8.40	5.39	8.69
4	4	6.20	3.26	7.10	4.39	6.35
4	8	6.19	2.97	7.46	4.18	6.35
4	12	6.10	2.77	8.03	4.30	6.58
8	1	20.69	11.85	24.40	16.06	26.75
8	4	18.42	8.67	22.30	12.93	22.19
8	8	18.39	7.81	24.50	13.29	24.45
8	12	18.04	7.64	25.14	14.67	26.39
12	1	35.01	22.98	43.79	32.57	51.23
12	4	32.27	17.80	42.07	28.94	48.04
12	8	32.09	17.32	43.36	31.97	52.18
12	12	31.01	16.85	43.41	35.23	55.50

NOTE: Entries are MSPE × 10,000.

than the regression RP2. As discussed by Howrey (1978, 1984), one could model and forecast future data revisions in a real-time, recursive procedure and incorporate these expected revisions into forecasts (via the Kalman filter). Given the frequency of the unforecastable definitional CLI revisions (and the resulting change in the properties of revisions as components are replaced), however, this procedure would appear to be of little practical use for forecasting with the CLI.

Finally, we note that the IP data used throughout this article are always of final revised form (as of May 1, 1989). We took such an approach, in part, because our intent is to measure the differences in forecast performance when using the preliminary as opposed to revised CLI *while holding other factors constant*. The alternative, of course, would be to use the preliminary and partially revised IP data as

Table 7. Mean Squared Prediction Error, Composite Coincident Index

Steps ahead (k)	*Lags included (p)*	*Forecast scenario*				
		FF1	*FF2*	*RF1*	*RF2*	*RP2*
1	1	.66	.48	.66	.48	.67
1	4	.49	.41	.50	.42	.48
1	8	.49	.39	.51	.43	.49
1	12	.49	.37	.52	.43	.50
4	1	5.80	3.25	6.16	3.49	6.16
4	4	4.42	2.64	4.73	2.87	4.26
4	8	4.44	2.46	4.86	2.84	4.30
4	12	4.34	2.36	4.94	2.89	4.52
8	1	16.46	8.64	18.61	9.87	18.26
8	4	13.66	7.14	15.53	8.08	14.77
8	8	14.04	6.93	16.32	8.29	15.31
8	12	13.66	6.96	16.40	8.63	15.51
12	1	29.87	16.68	35.28	19.85	31.75
12	4	26.63	13.87	32.13	16.15	29.49
12	8	27.10	13.95	33.51	16.77	30.03
12	12	26.50	14.02	33.97	17.45	30.29

NOTE: Entries are MSPE × 10,000.

they were available contemporaneously. This choice emerges for both the left side and the right side of the forecasting equations.

First, consider the use of final IP on the left side. We use final rather than contemporaneous IP as the regressand because we are interested in evaluating the ability of the CLI to forecast truth, which is taken to be the final IP value. An alternative approach might examine the accuracy of predictions of the initial estimate (or kth revision, $k = 1, 2, 3, \ldots$) of each month's IP. It is conceivable that the CLI is better at predicting early releases of IP than at predicting final values of IP (in part, perhaps, because the former use labor input data to infer physical output data released subsequently). To reiterate, however, we are interested in prediction of the best, that is, final, estimate of movements in aggregate output.

Second, consider the use of final IP on the right side. The alternative would involve constructing a matrix of preliminary and partially revised IP data, similar to the CLI matrix illustrated in Table 1, and using each column sequentially in forecasting regressions. This would provide a completely real-time analysis. Clearly, the outcome of this exercise would depend on the nature and size of revisions in the IP index. First, we note that definitional revisions of IP, which would cause conceptual complications if important, do occur but are of negligible importance during our sample. The three relevant revisions, which occurred in 1971, 1975, and 1985, primarily affected the industry-level classification scheme with only very small effects on the monthly movements in the aggregate IP index. [See Hosley and Kennedy (1985) and Federal Reserve Board (1986) for details.] Statistical revisions were also very small relative to those in the CLI, with a signal-to-noise ratio for the initial estimate of IP of about 2.8, twice as large as for the CLI [see Kennedy (1990)]. It is conceivable that adding this small amount of noise to the right side IP lags in RF1 and RP2 could increase the forecast errors relatively more in RF1, thus overturning our Result D. In such a (truly) real-time scenario, the uncertainty about recent movements in IP would allow the CLI to aid in prediction. Although we recognize such a possibility, we would argue that it is in some sense inconsequential. Such a result would provide a justification for use of the CLI that is quite different from the one adduced in the earlier ex post studies described in the introduction. Namely, the CLI would have little intrinsic leading ability but could possibly supplant the inadequacies of other data series.

6. SUMMARY AND CONCLUDING REMARKS

We have analyzed the value of the CLI of economic indicators in linear prediction of IP. We found that the CLI performed admirably in an ex post evaluation, confirming the results of others. By constructing a series of scenarios that progressively approached a real-time analysis (i.e., by contracting the forecaster's CLI information set until it matched that available in real time), however, we were able to chart a severe deterioration in the CLI's predictive performance. In forecasting frameworks with squared error loss and absolute error loss, we found that inclusion of the CLI in real-time forecasting equations generally failed to improve forecast performance.

Do our cumulative results indicate that the CLI is of no value as a forecasting tool? Perhaps, but not necessarily. In addition to the reservations expressed in Section 5, there are two general qualifications to such a conclusion. First, our evaluation, like previous evaluations, involved straightforward application of the linear model. It is possible that seasoned and skilled users can use their expert knowledge to extract useful predictive information from the CLI via more sophisticated procedures. Second, it is possible that the "mistakes" in the selection of components made in the early history of the construction of the CLI were part of a process of learning about cyclical behavior and statistical indexes that will not be repeated. In this interpretation, future definitional revisions in the CLI would be minor, and the CLI might perform much better on an ex ante basis. [Here, perhaps, the research of Stock and Watson (1989) may prove fruitful.] On balance, however, our results should be interpreted as sounding a strong cautionary note on forecasting with the CLI as presently constructed by the BEA.

Finally, from a methodological point of view, it is worth noting that our general approach may have wide applicability for questions of model selection and inference. In particular, the examination of out-of-sample or real-time causality using tests for MSPE equality on *recursive* residuals from restricted and unrestricted models may provide a more powerful tool to discriminate between models than conventional causality tests.

[*Received June 1989. Revised January 1991.*]

REFERENCES

Auerbach, A. J. (1982), "The Index of Leading Indicators: 'Measurement Without Theory' Thirty-Five Years Later," *Review of Economics and Statistics*, 64, 589–595.

Braun, P., and Zarnowitz, V. (1989), "Major Macroeconomic Variables and Leading Indexes: Some Estimates of Their Interrelations, 1886–1982," NBER Working Paper 2812.

Diebold, F. X., and Rudebusch, G. D. (1988), "Stochastic Properties of Revisions in the Index of Leading Indicators," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 712–717.

——— (1989), "Scoring the Leading Indicators," *Journal of Business*, 64, 369–391.

——— (1991), "Turning Point Prediction With the Composite Leading Index: An Ex Ante Analysis," in *Leading Economic Indicators: New Approaches and Forecasting Records*, eds. K. Lahiri and G. H. Moore, Cambridge, U.K.: Cambridge University Press, pp. 231–256.

Federal Reserve Board (1986), *Industrial Production, 1986 Version*, Washington, DC: Author.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–428.

Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series* (2nd ed.), New York: Academic Press.

Hendry, D. F. (1980), "Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money," in *Economic Modelling*, ed. P. Ormerod, London: Heinemann Educational Books, pp. 217–242.

Hertzberg, M. P., and Beckman, B. A. (1989), "Business Cycle Indicators: Revised Composite Indexes," *Business Conditions Digest*, 29, 97–102.

Hosley, J. D., and Kennedy, J. E. (1985), "A Revision of the Index of Industrial Production," *Federal Reserve Bulletin*, 71, 487–501.

Howrey, E. P. (1978), "The Use of Preliminary Data in Econometric Forecasting," *Review of Economics and Statistics*, 60, 193–200.

——— (1984), "Data Revision, Reconstruction, and Prediction: An Application to Inventory Investment," *Review of Economics and Statistics*, 66, 386–393.

Hymans, S. H. (1973), "On the Use of Leading Indicators to Predict Cyclical Turning Points," *Brookings Papers on Economic Activity*, 2, 339–384.

Kennedy, J. E. (1990), "An Analysis of Revisions to the Industrial Production Index," Federal Reserve Board, Economic Activity Section Working Paper 109.

Koch, P. D., and Rasche, R. H. (1988), "An Examination of the Commerce Department Leading-Indicator Approach," *Journal of Business & Economic Statistics*, 6, 167–187.

Makridakis, S., and Winkler, R. L. (1989), "Sampling Distributions of Post-sample Forecasting Errors," *Applied Statistics*, 38, 331–342.

Priestley, M. B. (1980), *Spectral Analysis and Time Series*, New York: Academic Press.

Stekler, H. O., and Schepsman, M. (1973), "Forecasting With an Index of Leading Series," *Journal of the American Statistical Association*, 68, 291–296.

Stock, J. H., and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual, 1989*, eds. O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, pp. 351–394.

Zarnowitz, V., and Moore, G. (1982), "Sequential Signals of Recession and Recovery," *Journal of Business*, 55, 57–85.