"Turning point Prediction with the Composite Leading Index: An Ex Ante Analysis," in *Leading Economic Indicators: New Approaches and Forecasting Records* (Kajal Lahiri and Geoffrey H. Moore, eds.), 231-256, 1991. Cambridge: Cambridge University Press.

(21)

CHAPTER 14

Turning point prediction with the composite leading index: An ex ante analysis

Francis X. Diebold and Glenn D. Rudebusch

On the day of its release, the preliminary estimate of the Department of Commerce composite index of leading indicators (CLI) is widely reported in the popular and financial press. Although declines in the composite leading index are often regarded as a potential signal of the onset of a recession, evaluations of the ability of the CLI to predict turning points have been limited in most previous studies by the use of final, revised CLI data. However, the composite leading index is extensively revised after each preliminary estimate; not only are revisions made as more complete historical data become available for the components, but ex post, the statistical weights are updated and components are added or eliminated to improve leading performance. Forecasts constructed with an ex post, recomputed CLI may differ from real-time forecasts based on the contemporaneous, original construction CLI. In this chapter, we perform a completely ex ante, or real-time, evaluation of the ability of the CLI to predict turning points by using the original preliminary estimates and revisions as they became available in real time.

In section 14.1, we describe revisions in the CLI and our procedure for generating ex ante turning point probability forecasts from the CLI. The methodology is the Bayesian procedure described in Diebold and Rudebusch (1989a), adapted to a newly constructed ex ante dataset. This new dataset, which has over 70,000 elements, contains every preliminary, provisionally revised, and final estimate of the CLI since the inception of the index in 1968. This allows us to reproduce the precise

We would like to thank Bill Nelson, Suzanne Nace, and Gerhard Fries for research assistance, and Barry Beckman, Frank de Leeuw, Jim Hamilton, Kajal Lahiri, Johannes Ledolter, Geoffrey Moore, Jim Stock, Mark Watson, and Victor Zarnowitz for useful comments at various stages of this research program. The views expressed in this paper are not necessarily those of the Board of Governors of the Federal Reserve System or its staff.

information content in the CLI available to forecasters at any point in time. Our implementation also incorporates results on the nature of duration dependence in U.S. expansions and contractions. While these results, which are examined in an appendix, are of independent interest to students of the business cycle, they also provide requisite inputs for the turning point probability forecasts.

In section 14.2, we evaluate the ex ante forecasts in terms of Brier's (1950) quadratic probability score (QPS), the probability-forecast analog of mean squared error. We also examine an informative factorization of the joint density of forecasts and realizations. The performance of the ex ante Bayesian probability forecast is compared with that of a range of alternatives, including a naive "no change" forecast, the optimal constant probability forecast, and the well-known rule of three consecutive declines.

In section 14.3, to facilitate interpretation of the results, we describe the stochastic properties of the preliminary CLI release and subsequent revisions, both within the across definitional regimes. Particular attention is paid to the information content of the preliminary estimate relative to the final revised value. A characterization of the statistical properties of the revisions is given relative to the polar cases of efficient forecast error and classical measurement error.

14.1 Ex ante CLI probability forecasts

While the information content of preliminary estimates is a consideration in any real-time forecasting 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. Toward the end of each month, the Bureau of Economic Analysis (BEA) produces a preliminary estimate of the previous month's composite leading index on the basis of incomplete and preliminary source data, and it may also revise the index for any or all of the preceding eleven months. Thus, each initial estimate is subject to up to eleven revisions within the first year. These *statistical* revisions in the CLI occur because of statistical revisions in the component indicators (due to larger and/or more representative samples as time passes, etc.) and also because of late-arriving data that are included, for example, in the first revision but not in the preliminary estimate.

However, the currently available CLI data are not only of a revised statistical form, but the components have also been reweighted and reselected ex post to improve the performance of the index over the sample.

These *definitional* revisions in the composite leading index have several different forms:

1. Compositional changes due to changes in data availability, data timing, or cyclical lead performance
2. Changes in weights assigned to component indicators due to statistical updating as more data become available
3. Definitional changes in component indicators, which may be due to changes in component definitions or coverage, and so on.

A substantial number, about one every two years, of these definitional revisions have occurred since the first presentation of the index of leading indicators in the November 1968 *Business Conditions Digest (BCD)*. Compositional changes in the CLI occurred in August 1969, April 1975, February 1979, January 1982, January 1983, and January 1987. For example, a major revision occurred in January of 1983 when the BEA updated statistical factors, incorporated historical revisions in the component 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 ex post leading performance.

Given these extensive revisions, it is of interest to recreate a real-time forecasting environment for predictive evaluation. For forecasting cyclical turning points, a leading index is only as good as the rule used to interpret its movements and map these movements into turning point predictions.[1] The classic example of a turning point filter associated with the CLI is the "three consecutive declines" rule for signaling a downturn (e.g., Vaccara and Zarnowitz, 1977), but many other methods have been proposed (e.g., Hymans, 1973; Wecker, 1979; Zarnowitz and Moore, 1982). More recently, a class of sequential-analytic event-oriented leading indicator prediction rules has gained popularity. The approach originates in Neftci's (1982) ingenious application of Shiryayev's (1978) results on optimal detection of changes in the probability generating process. Neftci uses this technique in a business cycle context to forecast turning points, that is, the dates of transition between "expansion" and "contraction" regimes. This approach, which we denote as the sequential probability recursion (SPR), has been refined recently by Diebold and Rudebusch (1989a) and Hamilton (1989). Evaluation of real-time turning point forecasts produced via the SPR methodology, as well as various other simpler methodologies, is the subject of this chapter.

[1] For an ex ante analysis that considers the standard problem of forecasting the level of an economic series, such as aggregate output, see Diebold and Rudebusch (1989b).

Assume that the behavior of the economy differs during expansions and contractions. Given this nonlinearity, it is advantageous to forecast both the expected future value of an economic variable and the form of its future probability structure as delineated by turning points (see Neftci, 1982; Diebold and Rudebusch, 1989a). To formalize this forecasting procedure, let Y_t be a coincident time-series that moves with general economic activity and switches probability distribution at turning points, and let X_t be a leading time-series with turning points (i.e., changes in distribution) that occur before the turning points in the coincident series. Let Z be an integer-valued random variable that represents the time index date of the first period after the turning point in X_t. For example, in the prediction of a downturn:

$$\begin{aligned} X_t &\sim F_t^u(X_t) \qquad 1 \leq t < Z \\ X_t &\sim F_t^d(X_t) \qquad Z \leq t \end{aligned} \tag{1}$$

where F_t^u and F_t^d are the respective upturn and downturn distributions. Time-sequential observations on the leading indicator are received, so at time t, there are $(t + 1)$ observations denoted $\bar{x}_t = (x_0, x_1, \ldots, x_t)$. At time t, we calculate a probability for the event $Z \leq t$, that is, that by time t a turning point in X has occurred.

The probability of $Z \leq t$ after observing the data $\bar{x}_t$ at time t can be decomposed by Bayes's formula:

$$P(Z \leq t \mid \bar{x}_t) = \frac{P(\bar{x}_t \mid Z \leq t)P(Z \leq t)}{P(\bar{x}_t)} \tag{2}$$

Define $\Pi_t = P(Z \leq t \mid \bar{x}_t)$ as the posterior probability of a turning point given the data available. Then, as shown in Diebold and Rudebusch (1989a), a very convenient recursive formula for the posterior probability of a downturn is available:

$$\Pi_t = \frac{[\Pi_{t-1} + \Gamma_t^u(1 - \Pi_{t-1})]f_t^d(\bar{x}_t)}{\{[\Pi_{t-1} + \Gamma_t^u(1 - \Pi_{t-1})]f_t^d(\bar{x}_t) + (1 - \Pi_{t-1})f_t^u(\bar{x}_t)(1 - \Gamma_t^u)\}} \tag{3}$$

where $\Gamma_t^u = P(Z = t \mid z \geq t)$, the probability of a peak in period t given that one has not already occurred, and f_t^u and f_t^d are the probability densities of the latest (tth) observation if it came from, respectively, an upswing or downswing regime (in X_t). (To use this formula in the prediction of troughs, exchange f_t^u with f_t^d and use the transition probability Γ_t^d, the probability of a trough in t given a continuing contraction.) With this formula, the probability Π_t can be calculated sequentially by using the previous probability Π_{t-1}, the "prior" (independent of $\bar{x}_t$) turning point probability that $Z = t$ (i.e., Γ_t^u or Γ_t^u), and the likelihoods of the most recent observation x_t based on the distribution of X_t in upswings and downswings. Given Π_t, a probability forecast about the value of Z, the forecaster maps this into the occurrence of a turning point in Y_t. In practice, the probability of a turning point in X_t is related to the probability of an imminent turning point in Y_t over a fixed horizon decided upon by the investigator.

To apply the above sequential probability recursion, we must first estimate the densities f_t^d and f_t^u, as well as the turning point transition probabilities Γ_t^u and Γ_t^d, and we must specify an initial condition Π_0. The specification of these elements has been explored to some degree in Diebold and Rudebusch (1989a), and we adopt their final specification with one crucial modification: we consider an ex ante forecasting exercise with rolling creation of the CLI upswing and downswing densities based only on observations that would have been available in historical time.

The sequential probability formula requires the probability densities of the leading series conditional on an expansion regime and conditional on a contraction regime. The leading series is assumed to have two stochastic generating structures, expansion and contraction, and this division of the leading series into regimes depends upon the underlying classification of economic activity. We have followed the *Business Conditions Digest* (see chart A in various issues) in denoting peaks and troughs of the CLI that correspond to the NBER business cycle. The procedure used to construct f_t^u and f_t^d involved fitting a normal density function to previous observations in each regime. In particular, if $\bar{x}_t = (x_0, x_1, \ldots, x_t)$ is the vector of sequential observations on the leading indicator observed up to time t, let $\bar{x}_t^u$ be the vector of those observations from the upswing regime and $\bar{x}_t^d$ be those observations from the downswing regime. Then f_t^u is a normal density with mean and variance equal to the sample mean and variance of $\bar{x}_t^u$, and f_t^d is a normal density with mean and variance equal to the sample mean and variance of the elements of $\bar{x}_t^d$. The composite leading index was first reported in the *BCD* in 1968 and was reported ex post back to 1948. Our scoring sample runs from December 1968 to December 1986, and for each month a new set of densities is computed based on previous data back to 1948. A twelve-month data lag is also built in, so that the last twelve observations are not used in constructing the densities. This is to allow a real-time forecaster sufficient time to recognize regime changes and classify observations.[2]

[2] For general references to the use of preliminary data in forecasting, see Howrey (1978). Three exceptions to the use of final, revised data in CLI evaluation are Stekler and Schepsman (1973) and Zarnowitz and Moore (1982), who find that the use of preliminary data increases false signals, and Hymans (1973), who finds little difference.

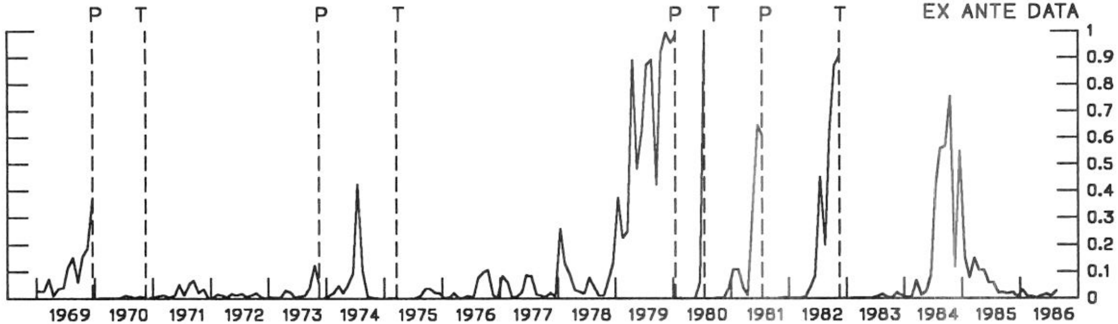

Figure 14.1. Ex ante CLI recession probabilities.

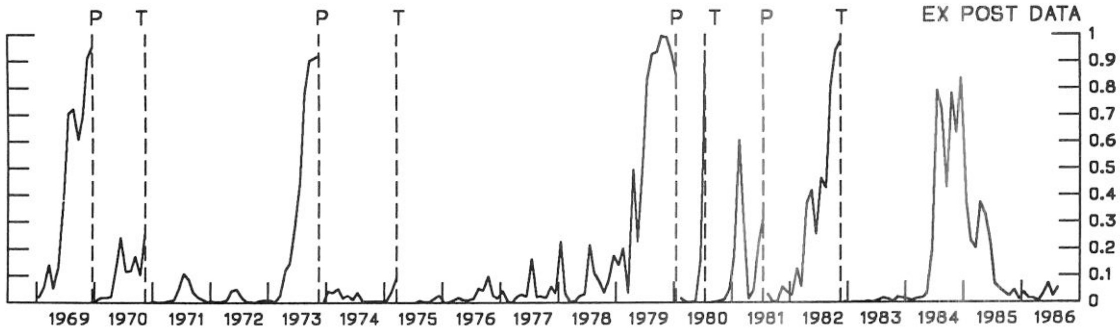

Figure 14.2. Ex post CLI recession probabilities.

The appendix provides evidence that, for the postwar period, the probability of a peak or a trough does not change significantly as the current regime progresses [also see the more sophisticated analysis in Diebold and Rudebusch (1990)]. For example, a long expansion is no more likely to end than a short one. Thus, we limit ourselves to time-invariant specifications of the transition conditional probabilities, that is, $\Gamma_t^u = \Gamma^u$ and $\Gamma_t^d = \Gamma^d$.

The final element in the recursive probability formula is last period's posterior probability of a turning point. There are two corrections made to this probability in practice. First, at the start of a new regime, a start-up probability of zero is used as the previous period's probability. Also, as is clear from the formula, if the posterior probability reaches one at any point, it will force all remaining probability forecasts to be one in the regime. Thus, we put an upper bound of 0.95 on the previous posterior probability as it enters the recursive probability formula.

Examples of turning point probability forecasts based on ex ante and ex post CLI data are given in Figures 14.1 and 14.2. (The forecasts shown use constant prior transition probabilities that are optimal in an average accuracy sense, to be defined rigorously in the next section, at a forecast horizon of seven months for expansions and three months for contractions.) The ex post probability forecasts perform quite well. Using an arbitrary critical probability of 0.9 [as advocated in Neftci (1982)] to signal turning points, the ex post forecasts would have signaled in advance three of the four peaks (missing the very sudden 1981 peak) and two of the four troughs with no false alarms. Using the real-time data, only one of the peaks is predicted and two of the troughs, again with no false alarms. While these results are indicative, they depend upon the critical probability value (.9) chosen. In the next section, we consider a more rigorous evaluation procedure that makes use of the information contained in the entire range of probability forecasts.

14.2 Evaluation of the probability forecasts

Accuracy refers to the closeness, on average, of predicted probabilities and observed relative frequencies. Consider a time-series of T probability forecasts $\{\Pi_t\}_{t=1}^T$, where Π_t is the time-t probability forecast of a turning point over horizon H. Let $\{R_t\}_{t=1}^T$ be the corresponding time-series of realizations; R_t equals one if a turning point occurs within the horizon (i.e., between times t and $t + H$) and equals zero otherwise. The quadratic probability score (Brier, 1950) is given by:

$$\text{QPS} = 1/T \sum_{t=1}^{T} 2(\Pi_t - R_t)^2 \tag{4}$$

The QPS ranges from 0 to 2, with a score of 0 corresponding to perfect accuracy. The QPS is the unique strictly proper scoring rule that is a function of the divergence between predictions and realizations; extended discussion and motivation, as well as consideration of alternative loss functions and evaluation measures, may be found in Diebold and Rudebusch (1989a).

The quadratic probability scores for a variety of probability forecasting methods are presented in Table 14.1. The forecasts are scored separately in the prediction of peaks and troughs, and scoring horizons range from one to thirteen months. Three different applications of the SPR are scored. Two are produced with the final, revised CLI data as of January 1987: SPR^a uses upswing and downswing CLI densities formed with the complete sample of data, while SPR^b rolls through the final data sample and creates densities only with data temporally prior to the forecast. The third SPR forecast, SPR^c, is truly ex ante and is formed with precisely the information set that would have been available to a real-time forecaster. At each horizon, we present the QPS of these SPR forecasts optimized with respect to the constant transition probabilities. Thus, the forecasts are completely ex ante, conditional upon Γ^u and Γ^d.

Other non-leading-indicator turning point probability forecasts are also scored. The forecasting methods include a no-change, NAIVE forecast, which amounts to a constant zero probability forecast, $\Pi_t = 0$, of a downturn or upturn. This is the probability forecast analog of a random walk (in this case, QPS $= 2\overline{R}$). More generally, one can search in the zero–one interval for the number that is the most accurate as a probability prediction of turning points. Such optimal, CONSTANT probability forecasts are of the form $\Pi_t = \kappa^u$ during expansions and $\Pi_t = \kappa^d$ during contractions, where the constants are chosen to minimize QPS. In the fifth row of Table 14.1, for example, at a forecast horizon of five months, a 12 percent probability forecast of a downturn ($K^u = .12$, given in parentheses below the score) is the most accurate constant probability forecast. Finally, two variants on the "three consecutive declines" theme for the prediction of downturns were evaluated for expansions. A recession signaling rule of three consecutive declines (3CD) was applied that translates three declines in the CLI into successive probability forecasts of 1.0, .8, .6, .4, .2, and 0.0 (unless, of course, three more consecutive declines occur, at which time the probability forecast returns to 1.0). This was applied to both the final data ($3CD^b$) and the real-time data ($3CD^c$). We attach no particular importance to this "rule-of-three," but rather take it to be indicative of various rules of thumb that have appeared in the literature. No similar rule of thumb for the

Table 14.1. *QPS as a function of horizon for various forecasting methods*

	Forecast horizon (in months)						
Method	1	3	5	7	9	11	13
Prediction of peaks							
SPR[a]	.05	.09	.12	.14	.22	.29	.36
(Γ^u)	(.00001)	(.003)	(.007)	(.02)	(.03)	(.04)	(.11)
SPR[b]	.05	.08	.12	.14	.21	.29	.36
(Γ^u)	(.00001)	(.002)	(.006)	(.02)	(.03)	(.04)	(.10)
SPR[c]	.04	.11	.19	.25	.31	.37	.42
(Γ^u)	(.00002)	(.0005)	(.003)	(.01)	(.03)	(.04)	(.05)
NAIVE	.05	.15	.25	.35	.45	.56	.66
CONSTANT	.05	.14	.22	.29	.35	.40	.44
(κ^u)	(.02)	(.07)	(.12)	(.18)	(.23)	(.28)	(.33)
3CD[b]	.14	.13	.08	.11	.21	.32	.41
3CD[c]	.11	.17	.24	.29	.39	.47	.55
Prediction of troughs							
SPR[a]	.10	.30	.45	.49	.56		
(Γ^d)	(.005)	(.05)	(.18)	(.29)	(.42)		
SPR[b]	.10	.30	.46	.48	.52		
(Γ^d)	(.005)	(.05)	(.21)	(.33)	(.51)		
SPR[c]	.10	.35	.60	.71	.57		
(Γ^d)	(.0001)	(.001)	(.005)	(.34)	(.61)		
NAIVE	.16	.49	.82	1.10	1.35		
CONSTANT	.15	.37	.48	.50	.44		
(κ^d)	(.08)	(.25)	(.41)	(.55)	(.67)		

The scoring sample is Dec. 1968–Dec. 1986. For each CONSTANT and SPR score, the associated constant prior transition probability is given beneath in parentheses. Superscripts on the forecasting methodologies refer to: (a) Based on the final revised CLI data as of January 1987, with SPR densities formed from final revised data. (b) Based on the final revised CLI data as of January 1987, with rolling SPR densities. (c) Based on ex ante real-time CLI data, with rolling SPR densities.

prediction of troughs appears in the literature; we therefore construct and score this forecast only for expansions. The linear decay that we adopt prevents abrupt dropoffs of Π_t from 1.0 to 0.0 and improves the performance of the "raw" rule of three at most horizons.

The results in Table 14.1 indicate that there is clearly information in the final revised CLI data for the prediction of both peaks and troughs. Both the simple 3CD rule-of-thumb and the more rigorous SPR substantially outperform, in an average accuracy sense, the naive and constant probability forecasts at a variety of horizons. The use of rolling densities formed from the ex post data in the construction of SPR forecasts (SPR^b) does not change this result.

The situation shifts dramatically, however, when the CLI data contemporaneous to the forecast are used in forming forecasts (SPR^c and $3CD^c$). With preliminary data, the simple rule-of-thumb $3CD^c$ never outperforms the constant probability forecast. The SPR^c does improve upon the constant probability forecast, though the enhancement at most horizons is not as great as for the ex post forecast SPR^b. Furthermore, during downswings, SPR^c performance is worse than CONSTANT at the longer forecasting horizons.

The deterioration of the SPR forecasts from ex post to ex ante can be decomposed into (a) that due to different ex ante and ex post densities f^u and f^d characterizing upswings and downswings, and (b) that due to different preliminary and revised CLI values. Comparing the SPR^a, SPR^b, and SPR^c rows of Table 14.1, we conclude that use of ex ante CLI data, as opposed to real-time density estimates, is responsible for most of the forecast divergence.

14.3 Characterization of revisions in the CLI

It was noted earlier that differences in ex ante and ex post turning point forecasting performance can be traced to one or both of the following: use of real-time CLI data and use of real-time estimated densities in the SPR. We saw that the first of these, not the second, was responsible for most of the difference; as such, we now study the properties of both intra- and inter-definitional revisions in the CLI.

We first consider the nature of revisions across definitional and compositional changes. The size of revisions to the CLI provides an indication of the information content of the preliminary estimates. Over the entire sample from December 1968 to January 1987, the standard deviation of the revision from the preliminary estimate of the CLI percentage change to the final estimate as given in January 1987 is .86 percentage points. Thus, for example, if the preliminary increase is 1.0 percent, one can only be 80 percent confident that the final estimate will be greater than −.10 and less than 2.10 percent (assuming normality). Within the most recent subsample of January 1983 to February 1986 (this allows for a final, eleventh revision through January of 1987), where definitional revisions are not a factor, the standard deviation from the preliminary estimate to eleventh revision is .49, and the corresponding 80 percent confidence interval is ±.63 percent.

We now examine statistical revisions *within* two recent definitional regimes, in particular, the periods February 1979 to December 1981 and January 1983 to January 1987. These represent timely and comparatively long regimes, and they provide an interesting contrast in terms of aggregate economic activity. For each date in each sample, we have twelve estimates available, which we denote Y1, Y2, . . . , Y12, where Y1 is the preliminary number and Y12 is the final revised number. We therefore have eleven non-overlapping revisions for each calendar date, defined by Y2Y1 = Y2 − Y1, . . . , Y12Y11 = Y12 − Y11.

It may be useful to classify the stochastic properties of revisions relative to the polar cases of classical measurement error and efficient forecast error, as in Mankiw, Runkle, and Shapiro (1984) and Mankiw and Shapiro (1986). The intuition behind the dichotomy is simple: If a provisional estimate differs from the revised value by only measurement error, then the revision is uncorrelated with the revised value but correlated with the provisional information set. On the other hand, if a provisional estimate represents an efficient forecast (i.e., rational, or minimum mean squared error conditional on available information), then the revision is correlated with the revised value but uncorrelated with the provisional information set. By determining where the CLI revisions lie within this spectrum, we can gain insight into the potential for achieving improvement in the preliminary numbers. If the intra-definitional-regime revisions behave as efficient forecast errors, then they are optimal estimates of the final, revised numbers. To the extent that the final numbers produce the better forecasts, then, efficient forecast error revisions are desirable.

We consider first the January 1983–January 1987 sample. Descriptive statistics, for varying degrees of revision collapse, are shown in Table 14.2. Note that the standard deviations of Y1, . . . , Y12 are all in the neighborhood of .86 percent, whereas the standard deviations of the revisions begin around .5 (for the earliest revisions) and eventually decrease to around .1 (for the last revisions). Thus, the standard deviation of the revisions (particularly the early revisions) is quite large rela-

Table 14.2. *Revisions in the composite leading index, 1983–7*

Variable	*N*	Mean	SD	*T* ratio
Y1	49	0.52	0.86	4.26
Y2	48	0.60	0.92	4.48
Y3	47	0.55	0.83	4.57
Y4	46	0.54	0.85	4.34
Y5	45	0.54	0.85	4.27
Y6	44	0.55	0.86	4.23
Y7	43	0.57	0.88	4.26
Y8	42	0.57	0.88	4.16
Y9	41	0.60	0.87	4.41
Y10	40	0.60	0.89	4.28
Y11	39	0.58	0.88	4.12
Y12	38	0.56	0.88	3.92
Y3Y1	47	0.03	0.45	0.49
Y5Y3	45	0.00	0.16	0.00
Y7Y5	43	0.01	0.16	0.29
Y9Y7	41	0.01	0.09	0.70
Y12Y9	38	−0.03	0.13	−1.46
Y5Y1	45	0.04	0.48	0.53
Y9Y5	41	0.02	0.18	0.86
Y12Y9	38	−0.03	0.13	−1.46
Y6Y1	44	0.04	0.47	0.58
Y12Y6	38	−0.01	0.16	−0.21
Y12Y1	38	0.07	0.49	0.82

Note: Y*m*Y*n* denotes the revision from the *n*th estimate to the *m*th estimate of the percent change in the CLI.

tive to the standard deviation of the percent-change CLI estimates. This implies that all of the CLI growth rate estimates, and particularly that of Y1, have large associated confidence intervals. The *t*-tests detect no bias in any of the revisions.

If revisions are efficient forecast errors, then the variances of Y1 through Y12 should be monotonically increasing, because an efficient forecast is necessarily smoother than the series being forecast. Conversely, if revisions are measurement errors, then the variances of Y1, . . . , Y12 should be decreasing. The data do not distinguish these two

Table 14.3. *Revisions and revised values: correlations and P-values, 1983–7*

	Y1	Y5	Y9	Y12
Y5Y1	−0.23	0.33	0.28	0.28
	0.12	0.03	0.08	0.09
Y9Y5	0.00	−0.11	0.09	0.07
	0.98	0.48	0.56	0.69
Y12Y9	−0.06	−0.09	−0.15	−0.01
	0.72	0.60	0.35	0.97

Note: Y*m*Y*n* denotes the revision from the *n*th estimate to the *m*th estimate of the percent change in the CLI.

cases, as the estimated standard deviations of Y1, . . . , Y12 display little variation.

Correlations between levels and three broad revisions are given in Table 14.3. Under the null of efficient forecast errors, the above-diagonal entries should be significant, while the below-diagonal entries should be insignificant. The table appears roughly consistent with the rational forecast error scenario; in particular, the entries of the first above-diagonal row of the table are significant at the 10 percent level and large in absolute value; for a more detailed analysis, see Diebold and Rudebusch (1988). The other above-diagonal entries are insignificant, perhaps because revisions after the fourth estimate contain little information, and the correlations cannot be estimated with precision.

The results for the earlier sample (1979–81) are quite different. There is a dropoff in variance as we move from Y1 to Y2 (Table 14.4) that is not consistent with forecast efficiency, and the correlations reported in Table 14.5 indicate a measurement error component, as evidenced by the lack of significant above-diagonal correlations as well as a highly significant below-diagonal correlation.

We interpret these results as indicating that the definitional change implemented in January 1983 significantly enhanced the statistical properties of the CLI revisions. One obvious source of measurement error in the preliminary estimate is that it is based on incomplete data, for not all component indicators are included in the preliminary (and sometimes even the second and third) releases. To the extent that better forecasts for the missing component indicators can be found, an element

Table 14.4. *Revisions in the composite leading index, 1979–81*

Variable	*N*	Mean	SD	*T* ratio
Y1	35	−.32	1.77	−1.07
Y2	34	−.24	1.55	−.88
Y3	33	−.26	1.57	−.96
Y4	32	−.23	1.56	−.84
Y5	31	−.15	1.57	−.53
Y6	30	−.17	1.60	−.58
Y7	29	−.20	1.62	−.68
Y8	28	−.20	1.61	−.64
Y9	27	−.14	1.64	−.44
Y10	26	−.14	1.61	−.45
Y11	25	−.16	1.61	−.48
Y12	24	−.21	1.64	−.63
Y3Y1	33	.08	.59	.79
Y5Y3	31	.01	.24	.15
Y7Y5	29	−.07	.13	−2.68
Y9Y7	27	−.01	.14	−.41
Y12Y9	24	.01	.21	.19
Y5Y1	31	.07	.57	.73
Y9Y5	27	−.08	.19	−2.03
Y12Y9	24	.01	.21	.19
Y6Y1	30	.04	.57	.42
Y12Y6	24	−.03	.27	−.61
Y12Y1	24	−.02	.74	−.14

Note: Y*m*Y*n* denotes the revision from the *n*th estimate to the *m*th estimate of the percent change in the CLI.

of measurement error is immediately introduced into the revisions. In the 1979–81 sample, two components, net business formation and the change in inventories, were not available for any of the preliminary numbers, and inventory change was also omitted from twenty-six of thirty-five first revisions and from one second revision. For the more recent sample from 1983–7, only the preliminary numbers suffer from omitted components.[3]

[3] After the most recent compositional redefinition of the CLI (see Hertzberg and Beckman, 1989), only components that will be available for the preliminary estimate were included in the newly reconstructed CLI.

Table 14.5. *Revisions and revised values: correlations and P-values, 1979–81*

	Y1	Y5	Y9	Y12
Y5Y1	−0.54	−0.25	−0.25	−0.23
	0.00	0.17	0.20	0.28
Y9Y5	−0.12	−0.13	−0.01	−0.03
	0.55	0.53	0.97	0.88
Y12Y9	−0.39	−0.37	−0.37	−0.26
	0.06	0.08	0.07	0.22

Note: YmYn denotes the revision from the nth estimate to the mth estimate of the percent change in the CLI.

14.4. Summary and conclusions

We have used a Bayesian algorithm to produce ex ante probability forecasts of peaks and troughs from the CLI. Most notably, the forecasts were constructed using the original preliminary data and revisions as they became available in real time. The forecasts were evaluated, and compared with ex post forecasts and forecasts generated by alternative methods, using proper probability forecast scoring rules. Finally, in order to better understand the differences between ex ante and ex post forecast performance, we characterized the properties of CLI revisions. Our main findings include the following:

1. A deterioration in turning point forecasting performance occurs when ex ante data are used, regardless of the forecasting method adopted. In the prediction of peaks, the real-time SPR maintains a small margin of superiority over its competitors. The deterioration in ex ante SPR forecast performance is relatively more severe for the prediction of troughs, leading to mixed results for comparative predictive ability, depending on the forecast horizon. The real-time SPR appears to maintain slight superiority at short horizons, but fares slightly worse than less sophisticated methods at longer horizons. This may be due simply to the short lengths of most contractions, so that good forecasting at long horizons is trivially simple (but not useful) merely by setting Π to a large enough value.

2. Deterioration in SPR forecast performance is due mostly to the move to ex ante data, as opposed to the use of rolling probability densities in the SPR. Examination reveals that the size and volatility of CLI revisions, both within and across definitional regimes, are high relative

to the magnitude of the revised percentage change in the CLI. Moreover, the CLI revisions appear to contain a measurement error component, which may be partially explained by the missing indicators in the preliminary CLI estimate. The measurement error component does not appear to be too severe in practice, however, and may be becoming less pronounced over time, due to beneficial definitional revisions.

3. There is no indication that turning point probabilities increase with the age of an expansion or contraction, in the period since World War II. Overall, postwar expansions and contractions show only weak, if any, duration dependence. This means that the transition probabilities used in the SPR, Γ_t^u and Γ_d^t, may be taken as approximately constant.

Appendix: Duration dependence in U.S. business cycles

Key elements of the SPR procedure for forecasting peaks and troughs are the probabilities of a turning point conditional only upon the expansion or contraction length-to-date. These probabilities, denoted by Γ_t^u during upswings and Γ_t^d during downswings, are the prior probabilities for the Bayesian recursion. Figure 14A.1 shows two examples of the possible relationship between the probability that an ongoing expansion will reach a peak and the age of that expansion.[4] The linear upward sloping hazard function (solid line, λ_2), which corresponds to a process with positive duration dependence, indicates that as an expansion progresses, the probability of a peak increases. The horizontal hazard (dashed line, λ_1), on the other hand, for which the transition probability is constant, corresponds to an absence of duration dependence. The resulting distributions of lengths of expansions and contractions are illustrated in Figure 14A.2. The duration distribution associated with the constant hazard is exponential. In discrete time, the distribution is geometric, with the probability of a regime of duration τ given by:

$$P(\text{duration} = \tau) = (1 - p)^{\tau-1}p \qquad (0 < p < 1) \qquad \text{(A1)}$$

where the probability p of a turning point is a constant. This is shown as the monotonically declining dashed line in Figure 14A.2. The duration distribution corresponding to the increasing hazard, on the other hand, is non-geometric; its explicit shape will depend on the explicit nature of the hazard. In general, however, its probability mass will be more concentrated than that of the geometric, an implication of the

[4] This same duration analysis applies to the probability of a trough and the age of the preceding contraction. However, the slope and position of the lines will differ across expansions and contractions to reflect different average regime lengths.

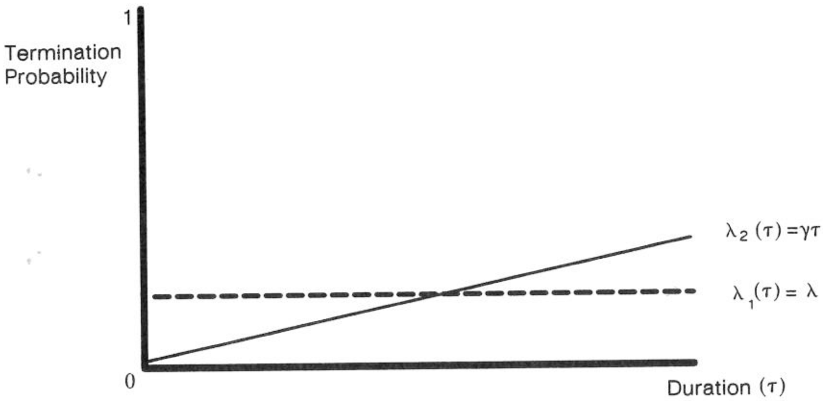

Figure 14.A1. Increasing and constant hazard functions.

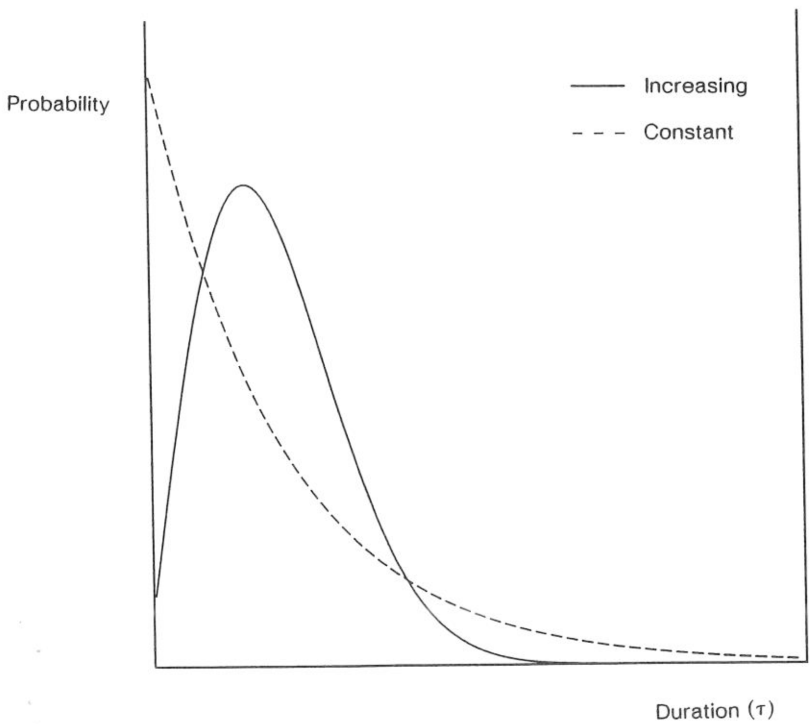

Figure 14.A2. Duration distributions associated with increasing and constant hazards.

turning point probability rising with duration length.[5] Consider, for example, the case of $p_\tau = \tau/100$; then the resulting density is

$$P(\text{duration} = \tau) = (1 - p_\tau)^{\tau-1}p_\tau \tag{A2}$$

which was used to generate the solid curve in Figure 14A.2. The hazard probability begins at .01 and rises by steps of .01 per period, leading to regime durations clustered around an intrinsic period.

We examine directly whether a histogram of historical duration lengths conforms to a geometric distribution, as it must under the non-periodic null hypothesis of constant turning point probabilities. Similar tests have been applied by McCulloch (1975), Savin (1977), and de Leeuw (1987). These studies are limited by the choice of a particular set of assumptions used in the construction of the histograms for the tests, which may account for the somewhat conflicting results obtained. We provide a sensitivity analysis exploring the whole range of possible assumptions.

We proceed as follows. For any given vector of expansion or contraction durations x, the data are first transformed by a minimum duration factor t_0 as $x^* = x - (t_0 - 1)$, which shifts the origin to reflect the length of the shortest possible regime. Minimum allowable expansion and contraction durations arise from definitional aspects of the NBER reference cycle dating procedure. Moore and Zarnowitz (1986), for example, indicate that expansions and contractions of less than six months would be very unlikely to qualify. (Note that under the geometric null, the unconditional distribution of τ is the same as the distribution of τ conditional on τ taking on a value greater than or equal to t_0.) Given the number of histogram bins K to be used, the bin width W is defined by $(x^*_{\max} - x^*_{\min})/K$, where $x^*_{\max}$ and $x^*_{\min}$ are the largest and smallest elements of the observed duration sequence x^*. The element, x^*_i, is grouped in bin n if $(x^*_{\min} + (n - 1)W) \leq x^*_i < (x^*_{\min} + nW)$. The histogram bin heights are computed by dividing the number of bin members by $N \cdot W$, that is, the duration sample size multiplied by the bin width.

We also compute exact finite sample confidence intervals under the geometric null. The maximum likelihood estimate of the hazard parameter of the best-fitting geometric distribution is $\hat{\lambda} = 1/\bar{x}^*$, where $\bar{x}^* = \Sigma_{i=1}^{N} x^*_i/N$ is the sample mean. A sample of N pseudorandom deviates is drawn from this geometric distribution, and the histogram with cell boundaries identical to the original is computed. (Generated deviates falling below x^*_{min} or above x^*_{max} are classified as members of bin 1 or bin

[5] This insight provides a link between the concepts of duration dependence and periodicity. See Diebold and Rudebusch (1990) for detailed discussion.

K, respectively.) This procedure is replicated 5,000 times. This allows construction of confidence intervals around individual bin heights. The goodness-of-fit test statistic also can be calculated as

$$S = \sum_{i=1}^{K} [(O_i - E_i)^2/E_i] \tag{A3}$$

where O_i is the observed number of elements of bin i and E_i is the expected number of elements of bin i under the geometric null (the average across simulations). Using the 5,000 simulated samples as observations allows construction of the exact distribution of the test statistic, which for our small sample sizes typically deviates from its asymptotic χ^2 distribution.

The lengths of expansions and contractions (in months) are derived from the business cycle turning dates as designated by the National Bureau of Economic Research. The entire sample of thirty-one expansions and thirty contractions, every business cycle since 1854, is given in Table 14A.1. Nine different subsamples are considered, including pre– and post–World War II expansions and contractions, as well as peacetime expansions. We generally favor the entire expansion and contraction samples since, as pointed out by Romer (1986), the evidence of structural shift between the pre- and postwar economies is not completely convincing. The choice of a proper sample depends upon which cycles are considered part of the intrinsic structure of the economy and which are attributed to special non-cyclical events (e.g., wars). We also consider the sensitivity of our results to the number of histogram cells (K), two through five. Statistical theory provides some guide in the construction of a histogram as to the correct number of cells to be distinguished. Terrell and Scott (1985) show that the minimum number of cells required for an optimal histogram is[6]

$$K^* = \{(2N)^{1/3}\} \tag{A4}$$

where the special brackets indicate rounding up to the nearest integer.[7] Histograms formed with this optimal minimum cell number have been shown to perform very well in practice. Finally, we consider a variety of

[6] The optimality is in terms of minimum deviation [in the Kullback-Liebler (1951) sense] of the estimated histogram cell heights from the true, but unknown, values of the probability distribution.

[7] The choice of cell number is important; too coarse a partition yields an uninformative distribution estimate, while too fine a partition yields a very jagged (and hence equally uninformative) estimate.

Table 14.A1. *NBER business cycle reference dates and durations*

Trough	Peak	Contraction	Expansion
December 1854	June 1857	NA	30
December 1858	October 1860	18	22
June 1861	April 1865	8	46
December 1867	June 1869	32	18
December 1870	October 1873	18	34
March 1879	March 1882	65	36
May 1885	March 1887	38	22
April 1888	July 1890	13	27
May 1891	January 1893	10	20
June 1894	December 1895	17	18
June 1897	June 1899	18	24
December 1900	September 1902	18	21
August 1904	May 1907	23	33
June 1908	January 1910	13	19
January 1912	January 1913	24	12
December 1914	August 1918	23	44
March 1919	January 1920	7	10
July 1921	May 1923	18	22
July 1924	October 1926	14	27
November 1927	August 1929	13	21
March 1933	May 1937	43	50
June 1938	February 1945	13	80
October 1945	November 1948	8	37
October 1949	July 1953	11	45
May 1954	August 1957	10	39
April 1958	April 1960	8	24
February 1961	December 1969	10	106
November 1970	November 1973	11	36
March 1975	January 1980	16	58
July 1980	July 1981	6	12
November 1982	?	16	72[a]

[a]The 72-month duration of the expansion beginning in November of 1982 is intended as a conservative estimate, implying that it ended in November 1988.
Note: Wartime expansions are underlined.

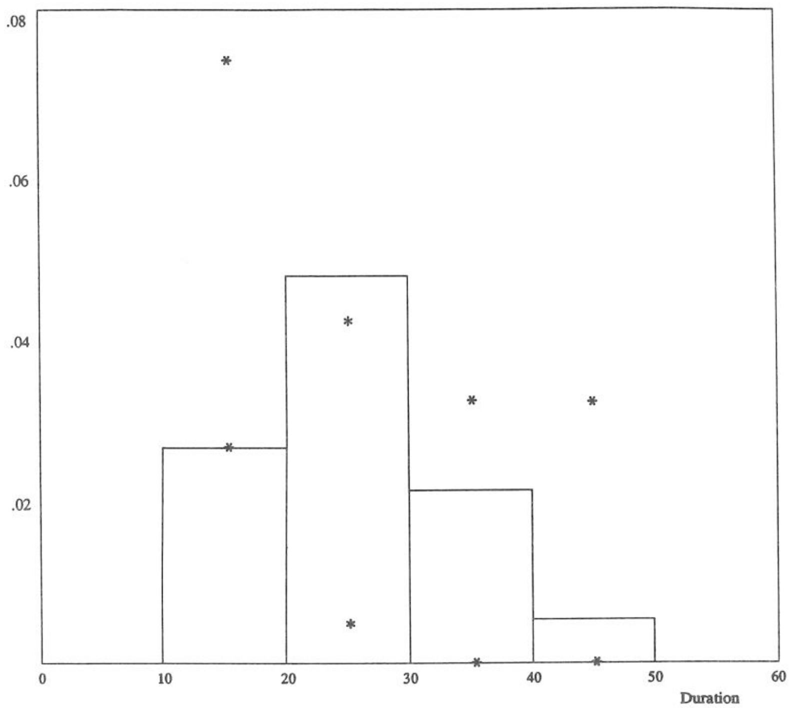

Figure 14.A3. Pre–World War II, peacetime expansions (E6).

minimum duration values. For each subsample, t_0 values up to the shortest expansion or contraction duration in that subsample are considered; in this way, we ensure that all values of x^* remain positive.

Probability-values (p-values) for the goodness-of-fit test statistic based on nonparametric distribution estimates are shown in Table 14A.3, and selected corresponding histograms are shown in Figures 14A.3, 14A.4, and 14A.5. The p-values represent the likelihood of obtaining the value of the test statistic actually obtained, under the geometric null of no duration dependence; large p-values therefore imply that the transition probabilities Γ^u and Γ^d should be invariant to the age of the ongoing regime. The range of samples investigated, denoted E1 through E6 and C1 through C3 for expansions and contractions, respectively, is identified in Table 14A.2.

In Table 14A.3 our choice for a single preferred p-value for each sample is underlined, though our conclusions based on these preferred prob-

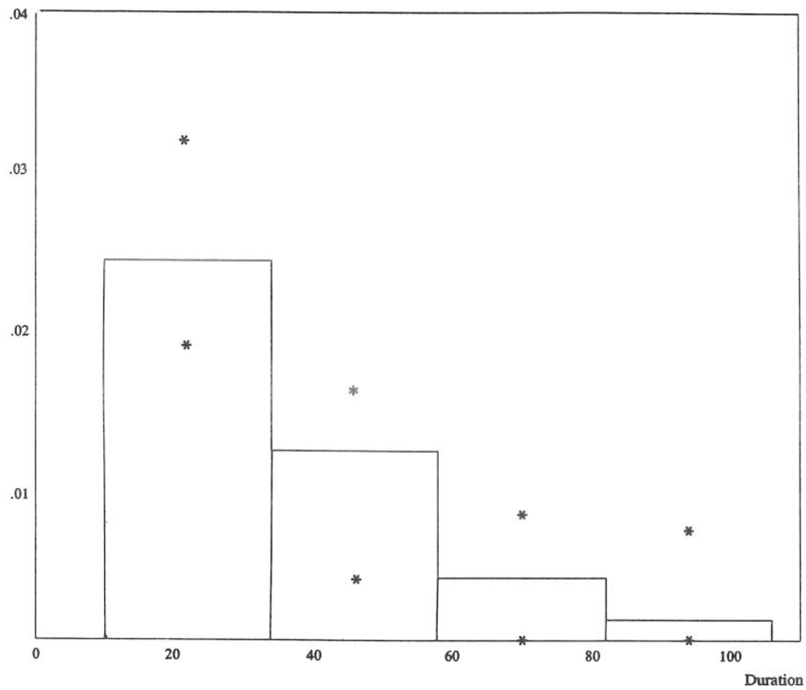

Figure 14.A4. Expansions, complete sample (E1).

abilities will always be tempered by their sensitivity to the number of histogram cells (K) used to characterize the distribution and the minimum duration values (t_0).[8] A reasonable choice for t_0 is the actual shortest observed duration, which is six months for contractions, the length of the 1980 contraction, and ten months for expansions. Our preferred cell number is the Terrell-Scott optimal bin number. Setting $K = K^*$ for our samples implies that observations should be grouped into four cells for all samples except the postwar ones, where three cells should be used.

Of the underlined p-values for the nine samples investigated in Table 14A.3, only one indicates significant duration dependence at the 5 percent level. This is sample E6, the set of all prewar, peacetime expansions. However, for the sample of all prewar expansions, duration dependence

[8] Previous researchers, such as McCulloch, Savin, or de Leeuw, have essentially focused on only a few of the entries in Table 14A.3, without an examination of the sensitivity of the results to their assumptions.

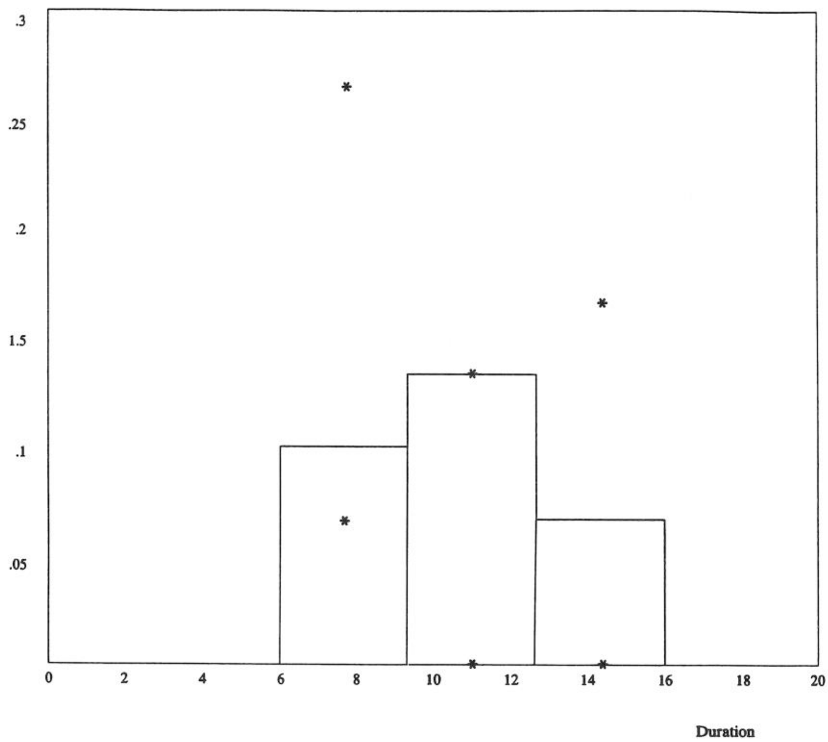

Figure 14.A5. Post–World War II contractions (C2).

Table 14.A2. *Business cycle subsamples investigated, with associated size*

Sample	Sample size
Expansions	
E1. Entire sample	31
E2. Entire sample, excluding wars	26
E3. Post–World War II	9
E4. Post–World War II, excluding wars	7
E5. Pre–World War II	21
E6. Pre–World War II, excluding wars	19
Contractions	
C1. Entire sample	30
C2. Pre–World War II	21
C3. Post–World War II	9

Table 14.A3. *Goodness-of-fit tests, expansion, and contraction samples (probability values under the geometric null)*

		Sample								
K	t_0	E1	E2	E3	E4	E5	E6	C1	C2	C3
2	4	.47	.17	.47	.71	.82	.80	.77	.49	.77
2	5	.47	.15	.47	.70	.82	.80	.76	.51	.76
2	6	.47	.14	.46	.70	.82	.79	.77	.51	.75
2	7	.62	.13	.47	.70	.82	.80	NA	NA	.77
2	8	.61	.21	.46	.71	.82	.80	NA	NA	NA
2	9	.62	.21	.72	.70	.82	.80	NA	NA	NA
2	10	.80	.21	.71	.69	.64	.79	NA	NA	NA
2	11	NA	NA	.72	.69	NA	NA	NA	NA	NA
2	12	NA	NA	.71	.69	NA	NA	NA	NA	NA
3	4	.78	.40	.60	.25	.14	.03	.48	.04	.63
3	5	.78	.48	.60	.25	.17	.06	.59	.06	.65
3	6	.85	.49	.64	.25	.26	.07	.63	_.20_	.64
3	7	.85	.58	.66	.33	.27	.09	NA	NA	.74
3	8	.91	.62	.66	.33	.35	.14	NA	NA	NA
3	9	.91	.65	.65	.38	.46	.14	NA	NA	NA
3	10	.96	.75	_.73_	_.37_	.50	.21	NA	NA	NA
3	11	NA	NA	.72	.37	NA	NA	NA	NA	NA
3	12	NA	NA	.74	.41	NA	NA	NA	NA	NA
4	4	.55	.15	.05	.23	.01	.00	.82	.60	.98
4	5	.61	.18	.05	.25	.01	.00	.84	.79	.98
4	6	.68	.22	.08	.28	.02	.01	_.77_	.91	_.96_
4	7	.72	.27	.09	.30	.03	.01	NA	NA	.96
4	8	.75	.32	.10	.33	.05	.02	NA	NA	NA
4	9	.79	.38	.11	.35	.09	.03	NA	NA	NA
4	10	_.85_	_.44_	.13	.40	_.13_	_.04_	NA	NA	NA
4	11	NA	NA	.13	.47	NA	NA	NA	NA	NA
4	12	NA	NA	.14	.45	NA	NA	NA	NA	NA
5	4	.45	.59	.34	.09	.00	.00	.95	.01	.93
5	5	.51	.67	.39	.10	.00	.00	.97	.04	.92
5	6	.58	.72	.43	.13	.00	.00	.97	.11	.92
5	7	.64	.79	.45	.15	.00	.00	NA	NA	.87
5	8	.69	.85	.51	.18	.01	.00	NA	NA	NA
5	9	.74	.90	.55	.18	.02	.01	NA	NA	NA
5	10	.80	.95	.59	.21	.04	.02	NA	NA	NA
5	11	NA	NA	.67	.23	NA	NA	NA	NA	NA
5	12	NA	NA	.70	.26	NA	NA	NA	NA	NA

Note: The definition of samples and sample key is given in Table 14.A2. Our preferred (K, t_0) combination for each sample is underlined. NA = not applicable.

is also significant at the 5 percent level when a slightly smaller t_0 value is used or when observations are placed into five cells. The nonparametric duration distribution estimate for sample E6 is shown in Figure 14A.3 (where $K = 4$ and $t_0 = 10$). For each histogram cell, the high and low points of the 95 percent confidence interval for that individual cell height under the geometric null are indicated by asterisks (*). The distribution of prewar, peacetime expansions shows a clear peak, representing a clustering of durations, unlike a steadily declining geometric distribution. In contrast, for the sample of all expansions, shown in Figure 14A.4, the cell heights are not significantly different from their values under the geometric null, as suggested by the associated p-value of .81; a similar distributional shape is found for the sample of all contractions (not shown). The p-value for postwar contractions is rather small, especially for slightly smaller t_0 values, although the null is not rejected at conventional significance levels. The nonparametric distribution estimate for this sample (with $K = 3$, $t_0 = 6$), given in Figure 14A.5 shows a small, insignificant peak, though our ability to discriminate between alternatives is limited by the small sample.

The sample period that is relevant for our forecasting evaluation is the postwar period, and there is little evidence of duration dependence in postwar expansions and contractions.[9] Obviously, our failure to reject the geometric null hypothesis does not imply its acceptance; nevertheless, if duration dependence is present, it would appear to be a very weak phenomenon.

[9] This result does not necessarily imply, however, that business fluctuations amount to "Monte Carlo cycles," as claimed by McCulloch. In particular, entire business cycles (peak-to-peak or trough-to-trough) may display strong duration dependence even though expansions and contractions do not. See Diebold and Rudebusch (1990) and Zarnowitz (1987).

REFERENCES

Brier, G. W. (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review, 75,* 1–3.

de Leeuw, F. (1987), "Do Expansions Have Memory?," Discussion Paper No. 16, U.S. Department of Commerce, Bureau of Economic Analysis.

Diebold, F. X. and G. D. Rudebusch (1988), "Stochastic Properties of Revisions in the Index of Leading Indicators," In *Papers and Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1987.* Washington, D.C.: American Statistical Association.

(1989a), "Scoring the Leading Indicators," *Journal of Business, 62,* 369–91.

(1989b), "Forecasting Output with the Composite Leading Index: An Ex Ante Analysis," FEDS Working Paper, Federal Reserve Board.

(1990), "A Nonparametric Investigation of Duration Dependence in the American Business Cycle," *Journal of Political Economy 98,* 596–616.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica, 57,* 357–84.

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

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

Hymans, S. H. (1973), "On the use of Leading Indicators to Predict Cyclical Turning Points," *Brookings Papers on Economic Activity, 2,* 339–84.

Kullback, S., and R. A. Liebler (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics, 22,* 79–86.

Mankiw, N. G., D. E. Runkle, and M. D. Shapiro (1984), "Are Preliminary Estimates of the Money Stock Rational Forecasts?," *Journal of Monetary Economics, 14,* 15–27.

Mankiw, N. G., and M. D. Shapiro (1986), "News or Noise? An Analysis of GNP Revisions," *Survey of Current Business,* May.

McCulloch, J. H. (1975), "The Monte Carlo-Cycle in Business Activity," *Economic Inquiry, 13,* 303–21.

Moore, G. H., and V. Zarnowitz (1986), "The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies," in R. J. Gordon, ed., *The American Business Cycle.* Chicago: University of Chicago Press.

Neftci, S. N. (1982), "Optimal Prediction of Cyclical Downturns," *Journal of Economic Dynamics and Control, 4,* 225–41.

Romer, C. D. (1986), "Is the Stabilization of the Postwar Economy a Figment of the Data?," *American Economic Review, 76,* 314–34.

Savin, N. E. (1977), "A Test of the Monte Carlo Hypothesis: Comment," *Economic Inquiry, 15,* 613–17.

Shiryayev, A. N. (1978), *Optimal Stopping Rules.* New York: Springer-Verlag.

Stekler, H. O., and M. Schepsman (1973), "Forecasting with an Index of Leading Series," *Journal of the American Statistical Association, 68,* 291–6.

Terrell, G. R., and D. W. Scott (1985), "Oversmoothed Nonparametric Density Estimates," *Journal of the American Statistical Association, 80,* 209–14.

Vaccara, B., and V. Zarnowitz (1977), "How Good are the Leading Indicators?," *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1976.* Washington, D.C.: American Statistical Association.

Wecker, W. E. (1979), "Predicting the Turning Points of a Time Series," *Journal of Business, 52,* 35–50.

Zarnowitz, V. (1987), "The Regularity of Business Cycles," National Bureau of Economic Research Working Paper No. 2381.

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