How Did Leading Indicator Forecasts Perform During the 2001 Recession?

James H. Stock and Mark W. Watson

T he recession that began in March 2001 differed in many ways from other recessions of the past three decades. The twin recessions of the early 1980s occurred when the Federal Reserve Board, under Chairman Paul Volcker, acted decisively to halt the steady rise of inflation during the 1970s, despite the substantial employment and output cost to the economy. Although monetary tightening had reduced the growth rate of real activity in 1989, the proximate cause of the recession of 1990 was a sharp fall in consumption, a response by consumers to the uncertainty raised by Iraq's invasion of Kuwait and the associated spike in oil prices (Blanchard 1993). In contrast, the recession of 2001 started neither in the shopping mall nor in the corridors of the Federal Reserve Bank, but in the boardrooms of corporate America as businesses sharply cut back on expenditures—most notably, investment associated with information technology—in turn leading to declines in manufacturing output and in the overall stock market.

Because it differed so from its recent predecessors, the recession of 2001 provides a particularly interesting case in which to examine the forecasting performance of various leading economic indicators. In this article, we take a look at how a wide range of leading economic indicators performed during this episode. Did these leading economic indicators predict a slowdown of growth? Was that slowdown large enough to suggest that a recession was

imminent? Were the leading indicators that were useful in earlier recessions also useful in this recession? Why or why not?

We begin our analysis by examining the predictions of professional forecasters—specifically, the forecasters in the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia—during this episode. As we show in Section 2, these forecasters were taken by surprise: even as late as the fourth quarter of 2000, when industrial production was already declining, the median SPF forecast was predicting strong economic growth throughout 2001.

Against this sobering backdrop, Section 3 turns to the performance of individual leading indicators before and during the 2001 recession. Generally speaking, we find that the performance of specific indicators was different during this recession. Some indicators, in particular the so-called term spread (the difference between long-term and short-term interest rates on government debt) and stock returns, provided some warning of a slowdown in economic growth, although the predicted growth was still positive and these indicators fell short of providing a signal of an upcoming recession. Other, previously reliable leading indicators, such as housing starts and orders for capital goods, provided little or no indication of the slowdown.

In practice, individual leading indicators are not used in isolation; as Mitchell and Burns (1938) emphasized when they developed the system of leading economic indicators, their signals should be interpreted collectively. Accordingly, Section 4 looks at the performance of pooled forecasts based on the individual leading indicator forecasts from Section 3 and finds some encouraging results. Section 5 concludes.

1.   FORECASTING THE 2001 RECESSION: HOW DID THE PROS DO?

This section begins with a brief quantitative review of the 2001 recession. We then turn to professional forecasts during this episode, as measured in real time by the Philadelphia Fed's quarterly Survey of Professional Forecasters.

A Brief Reprise of the 2001 Recession

Figure 1 presents monthly values of the four coincident indicators that constitute the Conference Board's Index of Coincident Indicators: employment in nonagricultural businesses, industrial production, real personal income less transfers, and real manufacturing and trade sales.[1] These four series are also the primary series that the NBER Business Cycle Dating Committee uses to

---

[1] For additional information on the Conference Board's coincident and leading indexes, see www.tcb-indicators.org.

Figure 1 Coincident Indicators
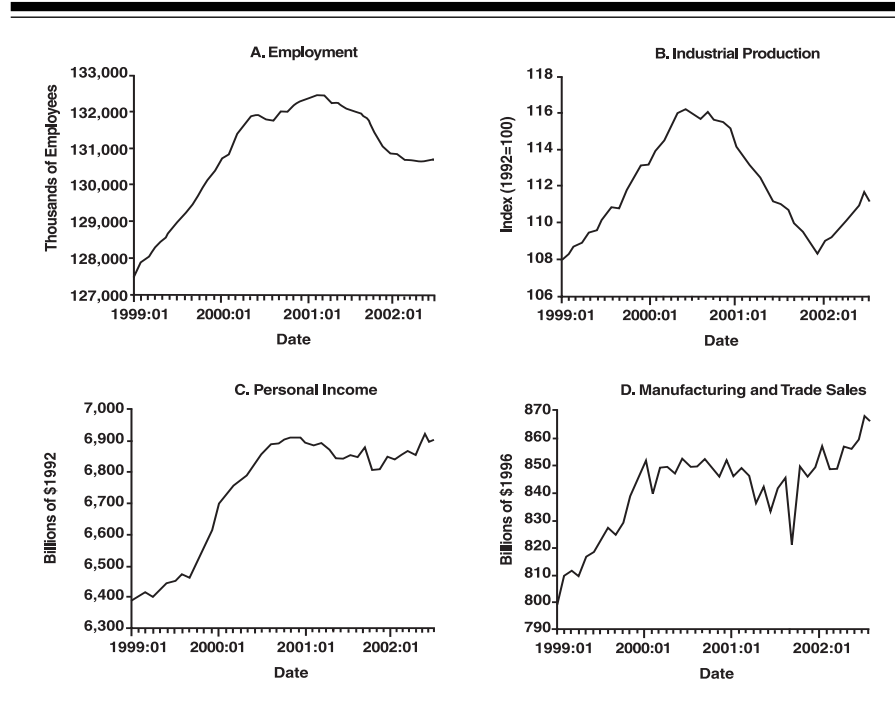
establish its business cycle chronology (Hall 2002). The percentage growth rates of these series, expressed at an annual rate, are plotted in Figure 2. In addition, Figure 2 presents the percentage growth of real GDP (at an annual rate); because GDP is measured quarterly and the time scale of Figure 2 is monthly, in Figure 2 the same growth rate of real GDP is attributed to each month in the quarter, accounting for the "steps" in this plot.

Figures 1 and 2 reveal that the economic slowdown began with a decline in industrial production, which peaked in June 2000. Manufacturing and trade sales fell during the first quarter of 2001, but employment did not peak until March 2001, the official NBER cyclical peak. Real personal income reached a cyclical peak in November 2000 and declined by 1.5 percent over the next twelve months. This relatively small decline in personal income reflected the unusual fact that productivity growth remained strong through this recession. Based on the most recently available data, real GDP fell during the first three quarters of 2001, with a substantial decline of 1.6 percent (at an annual rate) in the second quarter.

The economy gained substantial strength in the final quarter of 2001 and throughout 2002, and all the monthly indicators were growing by December 2001. Thus, based on the currently available evidence, the recession appears

Figure 2 Coincident Indicators (Growth Rates, PAAR)
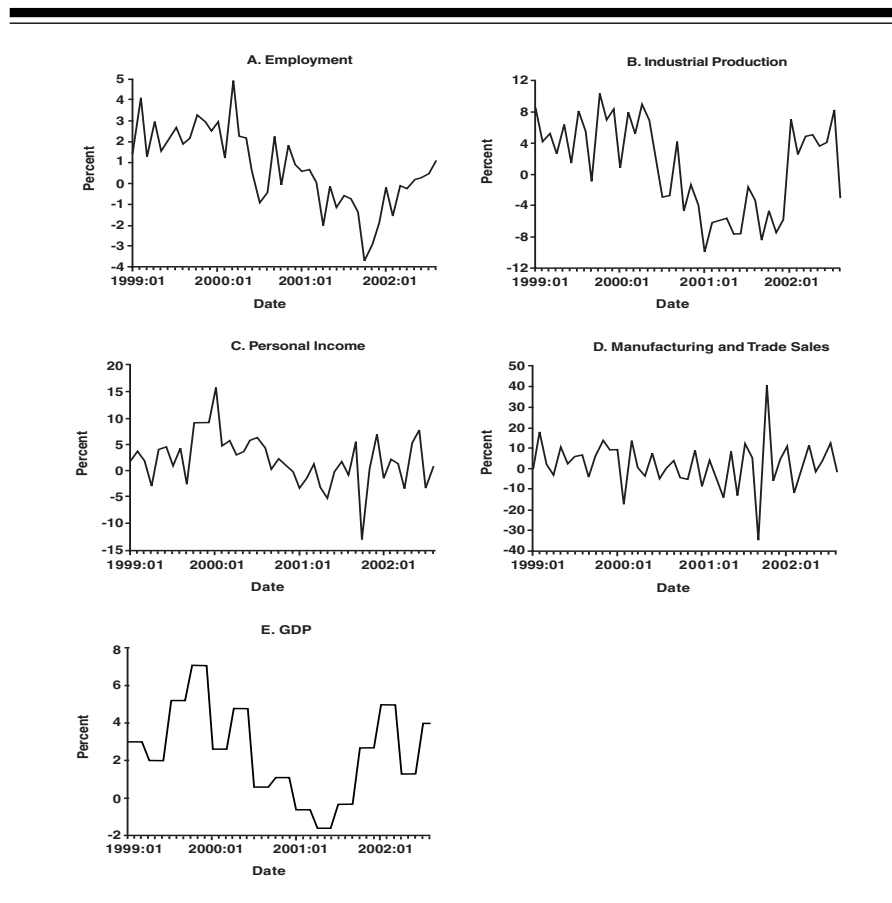
to have ended in the fourth quarter of 2001. When this article went into production, however, the NBER had yet to announce a cyclical trough, that is, a formal end to the recession.

Professional Forecasts During 2000 and 2001

In the second month of every quarter, the Research Department of the Federal Reserve Bank of Philadelphia surveys a large number of professional forecasters—in the first quarter of 2000, thirty-six forecasters or forecasting groups participated—and asks them a variety of questions concerning their short-term forecasts for the U.S. economy. Here, we focus on two sets of forecasts: the forecast of the growth rate of real GDP, by quarter, and the

Table 1 Median Forecasts of the Percentage Growth in Quarterly GDP from the Survey of Professional Forecasters

Target Date		Forecasts Made In							
		2000				2001			
Quarter	Actual Growth	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
2000Q4	1.1	2.9	3.1	**3.2**	3.2				
2001Q1	−0.6	2.8	2.6	3.0	**3.3**	0.8			
2001Q2	−1.6		2.9	2.7	3.2	**2.2**	1.2		
2001Q3	−0.3			3.2	3.3	3.3	**2.0**	1.2	
2001Q4	2.7				3.2	3.7	2.6	**2.8**	−1.9
2002Q1	5.0					3.7	3.1	2.7	**0.1**
2002Q2	1.3						3.6	3.0	2.4
2002Q3	4.0							3.9	3.6

Notes: Entries are quarterly percentage growth rates of real GDP, at an annual rate. One-quarter-ahead forecasts appear in bold. Actual GDP growth is from the 28 February 2003 GDP release by the Bureau of Economic Analysis. Forecasts are the median forecast from the Philadelphia Federal Reserve Bank's Survey of Professional Forecasters (various issues; see www.phil.frb.org/econ/spf).

probability that the forecasters assign to the event that GDP growth will be negative in an upcoming quarter.

The median growth forecasts—that is, the median of the SPF panel of forecasts of real GDP growth for a given quarter—are summarized in Table 1 for late 2000Q4 through 2002Q3. The first two columns of Table 1 report the quarter being forecast and its actual growth rate of real GDP, based on the most recently available data as of this writing. The remaining columns report the median SPF growth forecasts; the column date is the quarter in which the forecast is made for the quarter of the relevant row. For example, as of 2000Q1, the SPF forecast for 2000Q4 GDP growth was 2.9 percent at an annual rate (this is the upper-left forecast entry in Table 1). Over the course of 2000, as the fourth quarter approached, the SPF forecast of 2000Q4 growth rose slightly; as of 2000Q3, the forecast was 3.2 percent. Because the Bureau of Economic Analysis does not release GDP estimates until the quarter is over, forecasters do not know GDP growth for the current quarter, and in the 2000Q4 survey the average SPF forecast of 2000Q4 real GDP growth was 3.2 percent. As it happened, the actual growth rate of real GDP during that quarter was substantially less than forecasted, only 1.1 percent based on the most recently available data.

An examination of the one-quarter-ahead forecasts (for example, the 2000Q3 forecast of 2000Q4 growth) and the current-quarter forecasts (the 2000Q4 forecast of 2000Q4 growth) reveals that the SPF forecasters failed

Table 2 Probabilities of a Quarterly Decline in Real GDP from the Survey of Professional Forecasters

| Target Date | | Forecasts Made In | | | | | | | |
| | | 2000 | | | | 2001 | | | |
Quarter	Actual Growth	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
2000Q4	1.1	13%	9%	**7%**	4%				
2001Q1	−0.6	17	15	13	**11**	37%			
2001Q2	−1.6		18	16	17	**32**	32%		
2001Q3	−0.3			17	19	23	**29**	35%	
2001Q4	2.7				19	18	23	**26**	82%
2002Q1	5.0					13	18	20	**49**
2002Q2	1.3						13	16	27
2002Q3	4.0							15	18

Notes: Forecast entries are the probability that real GDP growth will be negative, averaged across SPF forecasters. The forecasted probability that growth will be negative in the quarter after the forecast is made (that is, the one-quarter-ahead forecast) appears in bold. See the notes to Table 1.

to predict the sharp declines in real GDP, even as they were occurring. The SPF one-quarter-ahead forecast of 2001Q1 growth was 3.3 percent, whereas GDP actually fell by 0.6 percent; the one-quarter-ahead forecast of 2001Q2 growth was 2.2 percent, but GDP fell by 1.6 percent; and the one-quarter-ahead forecast of 2001Q3 growth was 2.0 percent, while GDP fell by 0.3 percent. Throughout this episode, this average forecast was substantially too optimistic about near-term economic growth. Only in the fourth quarter of 2001 did the forecasters begin to forecast ongoing weakness—in part in reaction to the events of September 11—but, as it happened, in that quarter GDP was already recovering.

The SPF forecasters are also asked the probability that real GDP will fall, by quarter, and Table 2 reports the average of these probabilities across the SPF forecasters. In the fourth quarter of 2000, the forecasters saw only an 11 percent chance that GDP growth in the first quarter of 2001 would be negative, consistent with their optimistic growth forecast of 3.3 percent for that quarter; in fact, GDP growth was negative, falling by 0.6 percent. Throughout the first three quarters of 2001, the current-quarter predicted probabilities of negative growth hovered around one-third, even though growth was in fact negative in each of those quarters. When, in the fourth quarter of 2001, the SPF forecasters finally were sure that growth would be negative—the SPF probability of negative same-quarter growth was 82 percent—the economy in fact grew by a strong 2.7 percent. Evidently, this recession was a challenging time for professional forecasters.

Table 3 Relative MSFEs of Individual Indicator Forecasts of U.S. Output Growth, 1999Q1–2002Q3

Predictor	Transformation	GDP		IP	
		$h = 2$	$h = 4$	$h = 2$	$h = 4$
		Root Mean Squared Forecast Error			
Univariate autoregression		2.06	2.03	4.34	4.92
Predictor		*MSFE Rel. to Univariate AR Model*			
Random walk	level	1.26	1.11	1.56	1.17
Interest Rates					
Federal funds	Δ	1.01	0.71	0.97	0.78
90-day T-bill	Δ	1.01	0.76	1.02	0.88
1-year T-bond	Δ	1.17	0.96	1.22	1.06
5-year T-bond	Δ	1.37	1.24	1.38	1.23
10-year T-bond	Δ	1.36	1.26	1.21	1.23
Spreads					
Term spread (10 year–federal funds)*	level	0.86	0.65	0.77	0.72
Term spread (10 year–90-day T-bill)	level	0.87	0.62	0.70	0.62
Paper-bill spread (commercial paper–T-bill)	level	1.31	1.17	1.96	1.43
Junk bond spread (high yield–AAA corporate)	level	0.76	0.65	0.67	0.58
Other Financial Variables					
Exchange rate	$\Delta \ln$	0.85	0.87	0.85	0.80
Stock prices*	$\Delta \ln$	0.83	0.93	0.64	0.71
Output					
Real GDP	$\Delta \ln$			0.92	0.96
IP–total	$\Delta \ln$	0.98	1.01		
IP–products	$\Delta \ln$	1.03	0.99	1.03	0.96
IP–business equipment	$\Delta \ln$	1.00	1.01	1.05	1.06
IP–intermediate products	$\Delta \ln$	0.89	0.90	0.89	0.88
IP–materials	$\Delta \ln$	0.97	1.01	1.04	0.98
Capacity utilization rate	level	0.91	1.01	0.85	1.03
Labor Market					
Employment	$\Delta \ln$	0.96	1.00	0.96	0.99
Unemployment rate	Δ	1.24	1.08	1.31	1.09
Average weekly hours in manufacturing*	level	0.87	0.75	0.72	0.87
New claims for unemployment insurance*	$\Delta \ln$	0.75	0.84	0.74	0.81

Continued on next page

2. FORECASTS BASED ON INDIVIDUAL LEADING INDICATORS

Perhaps one reason for these difficulties was that the 2001 recession differed from its recent predecessors. If so, this difference would also be reflected

**Table 3 Relative MSFEs of Individual Indicator Forecasts of U.S.
            Output Growth, 1999Q1–2002Q3**

Other Leading Indicators					
Housing starts (building permits)*	$\Delta \ln$	1.30	1.07	1.52	1.14
Vendor performance*	level	1.02	0.97	1.19	0.97
Orders–consumer goods and materials*	$\Delta \ln$	0.77	0.83	0.81	0.83
Orders–nondefense capital goods*	$\Delta \ln$	1.02	1.03	0.92	1.09
Consumer expectations (Michigan)*	level	1.96	2.14	1.33	1.49
Prices and Wages					
GDP deflator	$\Delta^2 \ln$	1.00	0.94	0.94	0.84
PCE deflator	$\Delta^2 \ln$	1.01	1.05	0.99	0.99
PPI	$\Delta^2 \ln$	1.01	1.02	0.96	0.99
Earnings	$\Delta^2 \ln$	1.00	1.01	0.89	0.98
Real oil price	$\Delta^2 \ln$	1.13	1.18	1.07	1.11
Real commodity price	$\Delta^2 \ln$	1.04	1.00	1.12	1.09
Money					
Real M0	$\Delta \ln$	2.13	2.84	1.41	1.73
Real M1	$\Delta \ln$	1.09	1.07	1.57	1.12
Real M2*	$\Delta \ln$	2.06	1.82	2.13	1.94
Real M3	$\Delta \ln$	1.81	2.23	2.05	2.15

Notes: The entry in the first line is the root MSFE of the AR forecast, in percentage growth rates at an annual rate. The remaining entries are the MSFE of the forecast based on the individual indicator, relative to the MSFE of the benchmark AR forecast. The first forecast is made using data through 1999Q1; the final forecast period ends at 2000Q3. The second column provides the transformation applied to the leading indicator to make the forecast, for example, for the federal funds rate forecasts, X_t in (1) is the first difference of the federal funds rate.

*Included in the Conference Board's Index of Leading Indicators.

in the performance of leading indicators over this episode. In this section, we examine the performance of forecasts based on individual leading indicators during the 2001 recession. We begin by discussing the methods used to construct these forecasts, then turn to graphical and quantitative analyses of the forecasts.

Construction of Leading Indicator Forecasts

The leading indicator forecasts were computed by regressing future output growth over two or four quarters against current and past values of output growth and the candidate leading indicator. Specifically, let $Y_t = \Delta \ln Q_t$, where Q_t is the level of output (either the level of real GDP or the Index of Industrial Production), and let X_t be a candidate predictor (e.g., the term

spread). Let Y_{t+h}^h denote output growth over the next h quarters, expressed at an annual rate; that is, let $Y_{t+h}^h = (400/h)\ln(Q_{t+h}/Q_t)$. The forecasts of Y_{t+h}^h are made using the h-step-ahead regression model,

$$Y_{t+h}^h = \alpha + \sum_{i=0}^{p-1} \beta_i X_{t-i} + \sum_{i=0}^{q-1} \gamma_i Y_{t-i} + u_{t+h}^h, \qquad (1)$$

where u_{t+h}^h is an error term and α, $\beta_0, \ldots, \beta_{p-1}$, $\gamma_0, \ldots, \gamma_{q-1}$ are unknown regression coefficients. Forecasts are computed for two- and four-quarter horizons ($h = 2$ and $h = 4$).

To simulate real-time forecasting, the coefficients of equation (1) were estimated using only data prior to the forecast date. For example, for a forecast made using data through the fourth quarter of 2000, we estimate (1) using only data available through the fourth quarter of 2000. Moreover, the number of lags of X and Y included in (1), that is, p and q, were also estimated using only data available through the date of the forecast; specifically, p and q were selected using the Akaike Information Criterion (AIC), with $1 \le p \le 4$ and $0 \le q \le 4$.[2] Restricting the estimation to data available through the forecast date—in this example, 2000Q4—prevents the forecasts from being misleadingly accurate by using future data and also helps to identify shifts in the forecasting relation during the period that matters for forecasting, the end of the sample. This approach, in which all estimation and model selection is done using only data prior to the forecast date, is commonly called "pseudo out-of-sample forecasting"; for an introduction to pseudo out-of-sample forecasting methods and examples, see Stock and Watson (2003b, Section 12.7).

As a benchmark, we computed a multistep autoregressive (AR) forecast, in which (1) is estimated with no X_t predictor and the lag length is chosen using the AIC ($0 \le q \le 4$). As an additional benchmark, we computed a recursive random walk forecast, in which $\hat{Y}_{t+h|t}^h = h\hat{\mu}_t$, where $\hat{\mu}_t$ is the sample average of Y_s, $s = 1, \ldots, t$. Like the leading indicator forecasts, these benchmark forecasts were computed following the pseudo out-of-sample methodology.[3]

---

[2] The AIC is $\text{AIC}(p, q) = \ln(SSR_{p,q}/T) + 2(p + q + 1)/T$, where $SSR_{p,q}$ is the sum of squared residuals from the estimation of (1) with lag lengths p and q, and T is the number of observations. The lag lengths p and q are chosen to minimize $\text{AIC}(p, q)$ by trading off better fit (the first term) against a penalty for including more lags (the second term). For further explanation and a worked example, see Stock and Watson (2003b, Section 12.5).

[3] One way that this methodology does *not* simulate real-time forecasting is that we use the most recently available data to make the forecasts, rather than the data that were actually available in real time. For many of the leading indicators, such as interest rates and consumer expectations, the data are not revised, so this is not an issue. For others, such as GDP, revisions can be large, and because our simulated real-time forecasts use GDP growth as a predictor in equation (1), their performance in this exercise could appear better than it might have in real time, when preliminary values of GDP would be used.

A Look at Twelve Leading Indicators

We begin the empirical analysis by looking at the historical paths of twelve commonly used monthly leading indicators. After describing the twelve indicators, we see how they fared during the 2001 recession.
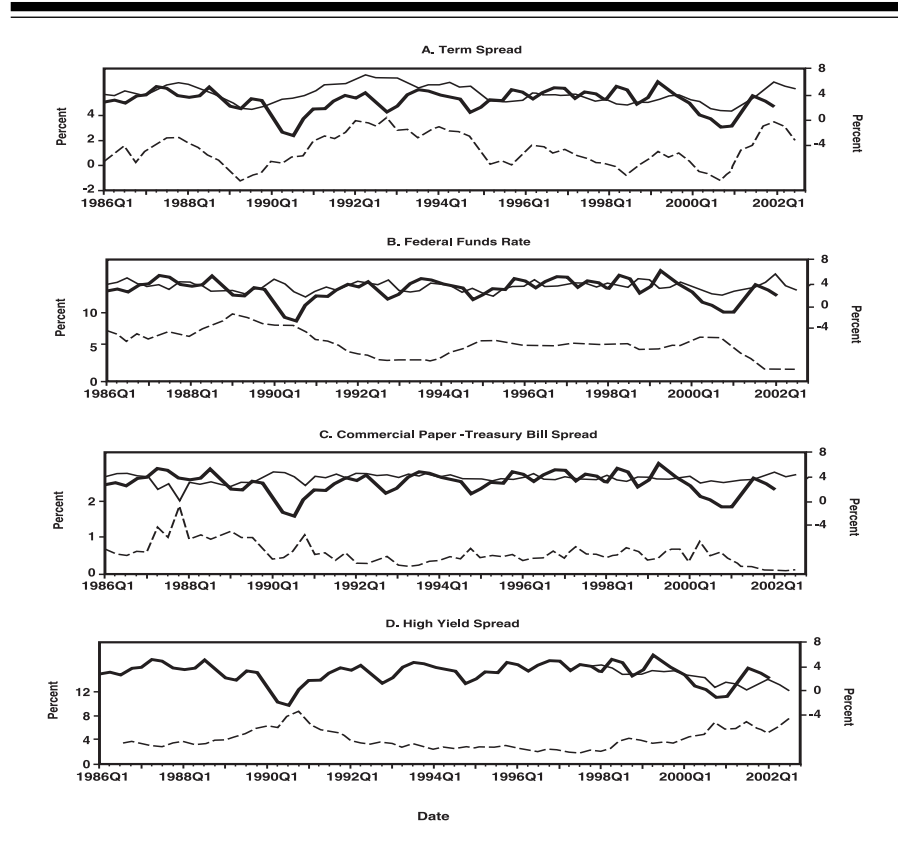
The Twelve Leading Indicators

Six of these indicators are based on interest rates or prices: a measure of the term spread (the ten-year Treasury bond rate minus the federal funds rate); the federal funds rate; the paper-bill spread (the three-month commercial paper rate minus the Treasury bill rate); the high-yield "junk" bond spread (the difference between the yield on high-yield securities[4] and the AAA corporate bond yield); the return on the S&P 500; and the real price of oil. Research in the late 1980s (Stock and Watson 1989; Harvey 1988, 1989; Estrella and Hardouvelis 1991) provided formal empirical evidence supporting the idea that an inverted yield curve signals a recession, and the term spread is now one of the seven indicators in the Conference Board's Index of Leading Indicators (ILI). The federal funds rate is included because it is the instrument of monetary policy. Public-private spreads also have been potent indicators in past recessions (Stock and Watson 1989; Friedman and Kuttner 1992); the second of these, the junk bond spread, was proposed by Gertler and Lown (2000) as an alternative to the paper-bill spread, which failed to move before the 1991 recession. Stock returns have been a key financial leading indicator since they were identified as such by Mitchell and Burns (1938), and the S&P 500 return is included in ILI.[5] Finally, fluctuations in oil prices are widely considered to be a potentially important source of external economic shocks and have been associated with past recessions (e.g., Hamilton 1983).

The next five indicators measure different aspects of the real economy. Three of these are in the ILI: new claims for unemployment insurance; housing starts (building permits); and the University of Michigan Index of Consumer Expectations. Because corporate investment played a central role in the 2001 recession, we also look at two broad monthly measures of business investment: industrial production of business equipment and new orders for capital goods. Finally, we consider a traditional leading indicator, the growth rate of real M2, which also enters the ILI.

---

[4] Merrill Lynch, U.S. High Yield Master II Index.

[5] For a review of the extensive literature over the past fifteen years on the historical and international performance of asset prices as leading indicators, see Stock and Watson (2001).

Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast


Graphical Analysis

Figure 3 plots the time path of these twelve leading indicators from 1986Q1 through 2002Q3, along with actual two-quarter real GDP growth and its forecast based on that indicator. For each series in Figure 3, the solid lines are the actual two-quarter GDP growth (thick line) and its indicator-based forecast (thin line); the dates correspond to the date of the forecast (so the value plotted for the first quarter of 2001 is the forecasted and actual growth of GDP over the second and third quarters, at an annual rate). The dashed line is the historical values of the indicator itself (the value of the indicator plotted in the first quarter of 2001 is its actual value at that date). The scale for the solid lines is given on the right axis and the scale for the dashed line is given on the left axis.

Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast

E. Stock Prices

F. Real Oil Price

G. Unemployment Insurance Claims

H. Building Permits

Date
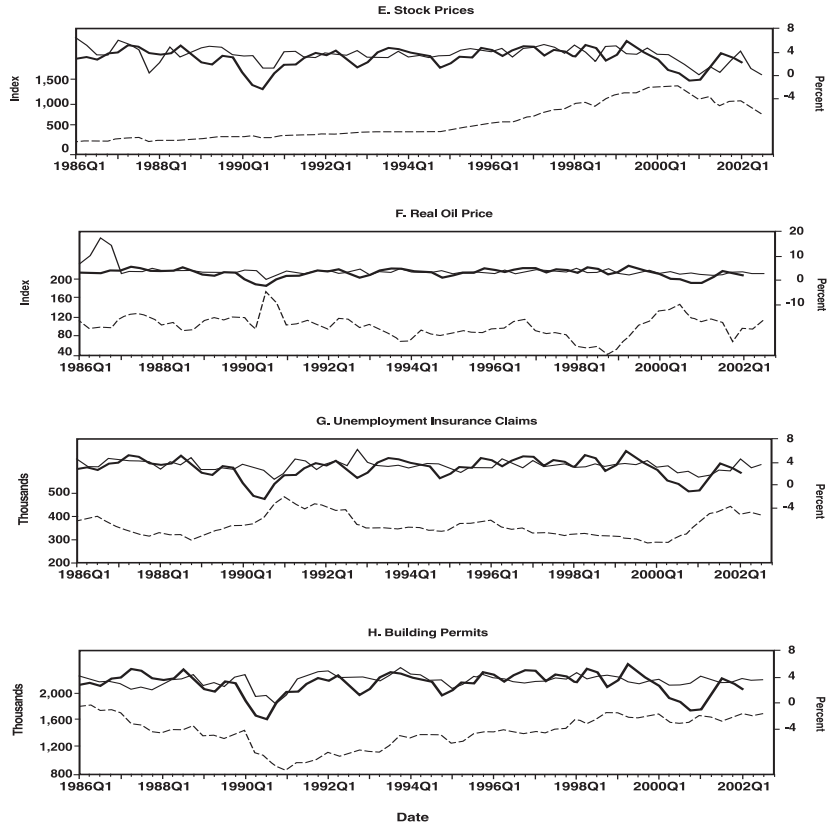
Inspection of Figure 3 reveals that some of these indicators moved in advance of the economic contraction, but others did not. The term spread provided a clear signal that the economy was slowing: the long government rate was less than the federal funds rate from June 2000 through March 2001. The decline in the stock market through the second half of 2000 also presaged further declines in the economy. New claims for unemployment insurance rose sharply over 2000, signaling a slowdown in economic activity. In contrast, other indicators, particularly series related to consumer spending, were strong throughout the first quarters of the recession. Housing starts fell sharply during the 1990 recession but remained strong through 2000. The consumer expectation series remained above 100 throughout 2000, reflecting overall positive consumer expectations. Although new capital goods orders dropped

Figure 3 Twelve Leading Indicators from 1986 to 2002, Two-Quarter Growth in Real GDP, and Its Leading-Indicator-Based Forecast
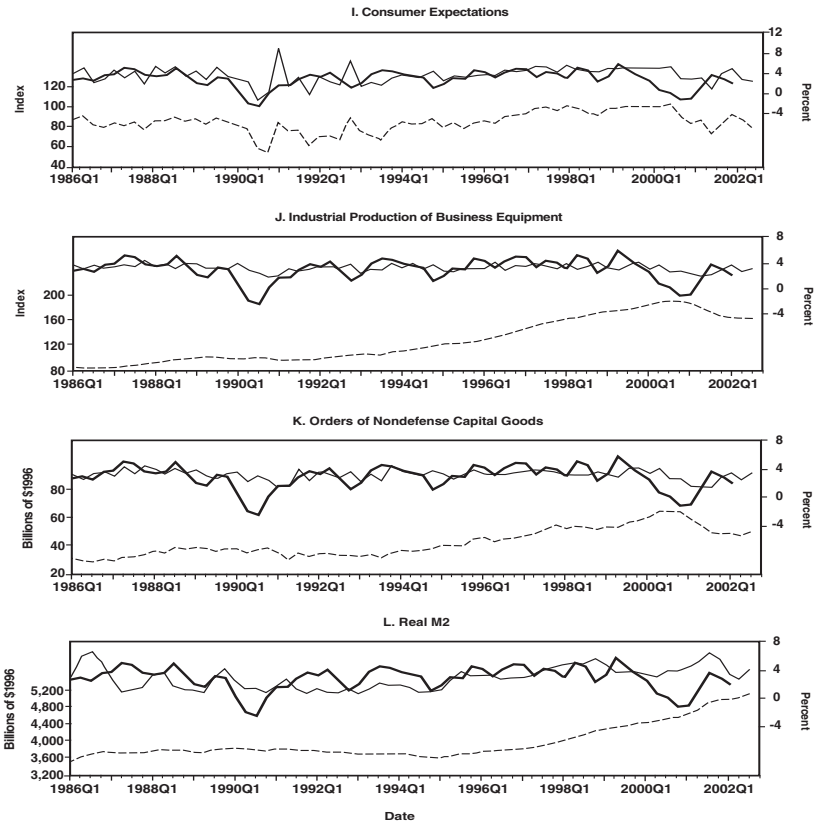
Notes: The solid lines are actual two-quarter GDP growth (thick line) and its indicator-based forecast (thin line), aligned so that the plotted date is the date of the forecast. The dashed line is the historical values of the indicator itself. The scale for the solid lines is given on the right axis and the scale for the dashed line is given on the left axis.

off sharply, that decline was contemporaneous with the decline in GDP, and in this sense new capital goods orders did not forecast the onset of the recession. The paper-bill spread provided no signal of the recession: although it moved up briefly in October 1998, October 1999, and June 2000, the spread was small and declining from August 2000 through the end of 2001, and the forecast of output growth based on the paper-bill spread remained steady and strong. In contrast, the junk bond spread rose sharply in 1998, leveled off, then rose

again in 2000. The junk bond spread correctly predicted a substantial slowing
in the growth rate of output during 2001; however, it incorrectly predicted a
slowdown during 1998. Finally, real M2 performed particularly poorly; the
strong growth of the money supply before and during this recession led to
M2-based output forecasts that were far too optimistic.

Quantitative Analysis of Forecast Errors

The graphical analysis shows that many of these indicators produced overly
optimistic forecasts, which in turn led to large forecast errors. However, some
indicators performed better than others. To assess forecast performance more
precisely, we examine the mean squared forecast error over this episode of
the different indicators relative to a benchmark autoregressive forecast. The
mean squared forecast error is the most common way, but not the only way, to
quantify forecasting performance, and we conclude this section with a brief
discussion of the results if other approaches are used instead.

Relative Mean Squared Forecast Error

The relative mean squared forecast error (MSFE) compares the performance
of a candidate forecast (forecast i) to a benchmark forecast; both forecasts
are computed using the pseudo out-of-sample methodology. Specifically, let
$\hat{Y}_{i,t+h|t}^{h}$ denote the pseudo out-of-sample forecast of Y_{t+h}^{h}, computed using data
through time t, based on the ith individual indicator. Let $\hat{Y}_{0,t+h|t}^{h}$ denote the
corresponding benchmark forecast made using the autoregression. Then the
relative MSFE of the candidate forecast, relative to the benchmark forecast, is

$$\text{relative MSFE} = \frac{\sum_{t=T_1}^{T_2-h}(Y_{t+h}^{h} - \hat{Y}_{i,t+h|t}^{h})^2}{\sum_{t=T_1}^{T_2-h}(Y_{t+h}^{h} - \hat{Y}_{0,t+h|t}^{h})^2}, \tag{2}$$

where T_1 and $T_2 - h$ are, respectively, the first and last dates over which the
pseudo out-of-sample forecast is computed. For this analysis, we set T_1 to
1999Q1 and T_2 to 2002Q3. If the relative MSFE of the candidate forecast is
less than one, then the forecast based on that leading indicator outperformed
the AR benchmark in the period just before and during the 2001 recession.

In principle, it would be desirable to report a standard error for the relative
MSFE in addition to the relative MSFE itself. If the benchmark model is not
nested in (that is, is not a special case of) the candidate model, then the
standard error can be computed using the methods in West (1996). Clark and
McCracken (2001) show how to test the hypothesis that the candidate model
provides no improvement in the more complicated case that the candidate
model nests the benchmark model. Unfortunately, neither situation applies

Table 4 Relative MSFEs of Combination Forecasts, 1999Q1–2002Q3

Combination Forecast Method	GDP		IP	
	$h = 2$	$h = 4$	$h = 2$	$h = 4$
Based on All Indicators				
Mean	0.95	0.94	0.95	0.95
Median	0.96	0.95	0.97	0.95
Inverse MSFE weights	0.97	0.98	0.95	0.96
Excluding Money				
Mean	0.94	0.91	0.91	0.92
Median	0.96	0.94	0.92	0.94
Inverse MSFE weights	0.96	0.95	0.93	0.94

Notes: Entries are the relative MSFEs of combination forecasts constructed using the full set of leading indicator forecasts in Table 3 (first three rows) and using the subset that excludes monetary aggregates (final three rows).

here because the lag length is chosen every quarter using the AIC; in some quarters the candidate model nests the benchmark, but in other quarters it does not. Because methods for this mixed case have yet to be worked out, the empirical results below report relative MSFEs but not standard errors.

Empirical Results

The relative MSFEs for thirty-seven leading indicators (including the twelve in Figure 3) are presented in the final four columns of Table 3 for two- and four-quarter-ahead forecasts of GDP growth and IP growth; the indicator and its transformation appear in the first two columns.

The mixed forecasting picture observed in Figure 3 is reflected in the MSFEs in Table 3. The relative MSFEs show that some predictors—the term spread, short-term interest rates, the junk bond spread, stock prices, and new claims for unemployment insurance—produced substantial improvements over the benchmark AR forecast. For example, the mean squared forecast error of the four-quarter-ahead forecast of GDP based on either measure of the term spread was one-third less than the AR benchmark. The two-quarter-ahead forecast of real GDP growth based on unemployment insurance claims had an MSFE 75 percent of the AR benchmark, another striking success.

In contrast, forecasts based on consumer expectations, housing starts, long-term interest rates, oil prices, or the growth of monetary aggregates all performed worse—in some cases, much worse—than the benchmark autoregression. Overall, the results from Table 3 reinforce the graphical analysis based on Figure 3 and provide an impression of inconsistency across indicators and, for a given indicator, inconsistency over time (e.g., the differing behavior of housing starts during the 1990 and 2001 recessions). This instability of

forecasts based on individual leading indicators is consistent with other re-
cent econometric evidence on the instability of forecast relations in the United
States and other developed economies; see, for example, the review of fore-
casts with asset prices in Stock and Watson (2001).

Results for Other Loss Functions

The mean squared forecast error is based on the most commonly used forecast
loss function, quadratic loss. Quadratic loss implies a particular concern
about large mistakes (a forecast error twice as large is treated as four times
as "costly"). Although the theoretical literature abounds with other forecast
loss functions, after quadratic loss the next most frequently used loss function
in practice is mean absolute error loss, which in turn leads to considering the
relative mean absolute forecast error (MAFE). The MAFE is defined in the
same way as the MSFE in equation (2), except that the terms in the summation
appear in absolute values rather than squared. The MAFE imposes less of a
penalty for large forecast errors than does the MSFE.

    We recomputed the results in Table 3 using the relative MAFE instead
of the relative MSFE (to save space, the results are not tabulated here). The
qualitative conclusions based on the relative MAFE are similar to those based
on the relative MSFE. In particular, the predictors that improved substantially
upon the AR as measured by the MSFE, such as the term spread and new
claims for unemployment insurance, also did so as measured by the MAFE;
similarly, those that fared substantially worse than the AR under the relative
MSFE, such as consumer expectations and housing starts, also did so using
the MAFE.

    This analysis has focused on forecasts of growth rates. A different tack
would be to consider forecasts of whether the economy will be in a recession,
that is, predicted probabilities that the economy will be in a recession in the
near future. This focus on recessions and expansions can be interpreted as
adopting a different loss function, one in which the most important thing is to
forecast the decree of the NBER Business Cycle Dating Committee. Because
this episode has had only one turning point so far, the peak of March 2001, we
think that more information about leading indicator forecasts during this period
can be gleaned by studying quarterly growth rate forecasts than by focusing on
binary recession event forecasts. Still, an analysis of recession event forecasts
is complementary to our analysis, and recently Filardo (2002) looked at several
probabilistic recession forecasting models. One of his findings is that the
results of these models depend on whether final revisions or real-time data
are used (the forecasts based on finally revised data are better). He also finds
that a probit model based on the term spread, the paper-bill spread, and stock
returns provided advance warning of the 2001 recession, a result consistent
with the relatively good performance of the term spread and stock returns in
Table 3.

3. COMBINATION FORECASTS

The SPF forecasts examined in Tables 1 and 2 are the average of the forecasts by the individual survey respondents. Such pooling of forecasts aggregates the different information and models used by participating forecasters, and studies show that pooled, or combination, forecasts regularly improve upon the constituent individual forecasts (see Clemen 1989; Diebold and Lopez 1996; and Newbold and Harvey 2002). Indeed, in their original work on leading indicators, Mitchell and Burns (1938) emphasized the importance of looking at many indicators, because each provides a different perspective on current and future economic activity.

In this section, we pursue this line of reasoning and examine the performance during the 2001 recession of combination forecasts that pool the forecasts based on the individual leading indicators examined in Section 3. The literature on forecast combination has proposed many statistical methods for combining forecasts; two important early contributions to this literature are Bates and Granger (1969) and Granger and Ramanathan (1984). Here we consider three simple methods for combining forecasts: the mean, the median, and an MSFE-weighted average based on recent performance.

The mean combination forecast is the sample average of the forecasts in the panel. The median modifies this by computing the median of the panel of forecasts instead of the mean, which has the potential advantage of reducing the influence of "crazy" forecasts, or outliers. This is the method that was used to produce the SPF combination forecasts in Table 1. The MSFE-weighted average forecast gives more weight to those forecasts that have been performing well in the recent past. Here we implement this combination forecast by computing the forecast error for each of the constituent forecasts over the period from 1982Q1 through the date that the forecast is made (thereby following the pseudo out-of-sample methodology), then estimating the current mean squared forecast error as the discounted sum of past squared forecast errors, with a quarterly discount factor of 0.95. The weight received by any individual forecast in the weighted average is inversely proportional to its discounted mean squared forecast error, so the leading indicators that have been performing best most recently receive the greatest weight.

The results are summarized in Table 4. The combination forecasts provide consistent modest improvements over the AR benchmark. During this episode, the simple mean performed better than either the median or inverse MSFE-weighted combination forecasts.

Because real money has been an unreliable leading indicator of output for many years in many developed economies (Stock and Watson 2001)—a characteristic that continued in the 2001 recession—it is also of interest to consider combination forecasts that exclude the monetary aggregates. Not surprisingly given the results in Table 3, the combination forecasts excluding money exhibit better performance than those that include the monetary aggregates.

Of course, the sample size is small and we should refrain from drawing strong conclusions from this one case study. Moreover, the improvements of the combination forecasts over the AR benchmark are less than the improvements shown by those individual indicators, such as new claims for unemployment insurance, that were, in retrospect, most successful during this episode. Still, the performance of the simple combination forecasts results is encouraging.

4.   DISCUSSION AND CONCLUSIONS

Leo Tolstoy opened *Anna Karenina* by asserting, "Happy families are all alike; every unhappy family is unhappy in its own way." So too, it seems, with recessions. While the decline of the stock market gave some advance warning of the 2001 recession, it was not otherwise a reliable indicator during the 1980s and 1990s. Building permits and consumer confidence, which declined sharply preceding and during the 1990 recession, maintained strength well into the 2001 recession. While the term spread indicated an economic slowdown in 2001, it did not give an early signal in the 1990 recession. The varying performance of these indicators reflects the differences in the shocks and economic conditions prior to the 1990 and 2001 recessions.

In retrospect, the performance of the various individual indicators is generally consistent with the view that this recession was a joint consequence of a sharp decline of the stock market (perhaps nudged by some monetary tightening) and an associated pronounced decline in business investment, especially in information technology. These shocks affected manufacturing and production but diffused only slowly to general employment, incomes, and consumption. But without knowing these shocks in advance, it is unclear how a forecaster would have decided in 1999 which of the many promising leading indicators would perform well over the next few years and which would not.

The failure of individual indicators to perform consistently from one recession to the next, while frustrating, should not be surprising. After all, the U.S. economy has undergone important changes during the past three decades, including an expansion of international trade, the development of financial markets and the concomitant relaxing of liquidity constraints facing consumers, and dramatic increases in the use of information technology in manufacturing and inventory management. Moreover, the conduct of monetary policy arguably has shifted from being reactionary, using recessions to quell inflation, to more proactive, with the Fed acting as if it is targeting inflation (see Goodfriend 2002). As we discuss elsewhere (Stock and Watson 2001, 2003a), these and other macroeconomic changes could change the relation between financial leading indicators and economic activity and, to varying degrees, could contribute to the reduction in volatility of GDP that the United States (and other countries) have enjoyed since the mid-1980s.

Our conclusion—that every decline in economic activity declines in its own way—is not new. Indeed, one of the reasons that Mitchell and Burns (1938) suggested looking at many indicators was that each measured a different feature of economic activity, which in turn can play different roles in different recessions. In light of the variable performance of individual indicators and the evident difficulty professional forecasters had during this episode, the results for the combination forecasts are encouraging and suggest that, taken together, leading economic indicators did provide some warning of the economic difficulties of 2001.

REFERENCES

Bates, J. M., and Clive W. J. Granger. 1969. "The Combination of Forecasts." *Operations Research Quarterly* 20: 451–68.

Blanchard, Olivier J. 1993. "Consumption and the Recession of 1990–1991." *American Economic Review Papers and Proceedings* 83(2): 270–75.

Clark, Todd E., and Michael W. McCracken. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 105 (November): 85–100.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5(4): 559–83.

Diebold, Francis X., and J. A. Lopez. 1996. "Forecast Evaluation and Combination." In *Handbook of Statistics*, edited by G. S. Maddal and C. R. Rao. Amsterdam: North-Holland.

Estrella, Arturo, and Gikas Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." *Journal of Finance* 46 (June): 555–76.

Filardo, Andrew J. 2002. "The 2001 U.S. Recession: What Did Recession Prediction Models Tell Us?" Manuscript, Bank for International Settlements.

Friedman, Benjamin M., and Kenneth N. Kuttner. 1992. "Money, Income, Prices and Interest Rates." *American Economic Review* 82 (June): 472–92.

Gertler, Mark, and Cara S. Lown. 2000. "The Information in the High Yield Bond Spread for the Business Cycle: Evidence and Some Implications." NBER Working Paper 7549 (February).

Goodfriend, Marvin. 2002. "The Phases of U.S. Monetary Policy: 1987 to 2001." Federal Reserve Bank of Richmond *Economic Quarterly* 88 (Fall): 1–17.

Granger, Clive W. J., and Ramu Ramanathan. 1984. "Improved Methods of Combining Forecasting." *Journal of Forecasting* 3: 197–204.

Hall, Robert E. 2002. "Dating Business Cycles: A Perspective." Manuscript, Stanford University.

Harvey, Campbell R. 1988. "The Real Term Structure and Consumption Growth." *Journal of Financial Economics* 22 (December): 305–33.

——————. 1989. "Forecasts of Economic Growth from the Bond and Stock Markets." *Financial Analysts Journal* 45 (September/October): 38–45.

Hamilton, James D. 1983. "Oil and the Macroeconomy Since World War II." *Journal of Political Economy* 91 (April): 228–48.

Merrill Lynch. www.mlindex.ml.com.

Mitchell, Wesley C., and Arthur F. Burns. 1938. *Statistical Indicators of Cyclical Revivals*, NBER Bulletin 69. Reprinted as Chapter 6 of *Business Cycle Indicators,* edited by G. H. Moore. Princeton, N.J.: Princeton University Press, 1961.

Newbold, Paul, and David I. Harvey. 2002. "Forecast Combination and Encompassing." In *A Companion to Economic Forecasting,* edited by M. P. Clements and D. F. Hendry. Oxford: Blackwell Press.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." In *NBER Macroeconomics Annual 1989*, edited by O. J. Blanchard and S. Fischer. Cambridge: MIT Press, 352–94.

——————. 2001. "Forecasting Output and Inflation: The Role of Asset Prices." NBER Working Paper 8180 (March).

——————. 2003a. "Has the Business Cycle Changed and Why?" In *NBER Macroeconomics Annual 2002*, edited by Mark Gertler and Kenneth Rogoff. Cambridge: MIT Press.

——————. 2003b. *Introduction to Econometrics*. Boston: Addison-Wesley.

West, Kenneth D. 1996. "Asymptotic Inference About Predictive Ability." *Econometrica* 64 (September): 1067–84.