Making sense of economic indicators: a consumer's guide to indicators of real economic activity

Francesca Eugeni, Charles Evans, and Steven Strongin

Economic data are used primarily in two ways. Academic economists typically use data to build models of the economy in order to understand how the economy works. Business analysts, on the other hand, use economic data to forecast future economic activity. These two activities and groups of people are not truly distinct groups, nevertheless the two activities do involve some substantive differences. The problem facing the business analyst, and to a large extent the policymaker or businessman who has to make decisions based on the economic outlook, is how each piece of new information should be assessed. Does it portend higher growth or lower, a recession or a boom, slow growth or stasis? Such assessments are crucial to running a successful business and to the proper ongoing evaluation of economic policy. Yet economic analysis rarely focuses on precisely these questions. In the current article, we develop an organized structure for evaluating economic indicators and apply that structure to a wide variety of financial indicators and a selected group of real indicators as well.

This process is fundamentally more eclectic than the usual econometric analysis which looks for or constructs a "best" indicator, where "best" typically refers to winning some narrowly defined contest of general purpose forecasting ability measured over some preselected time span.[1] Unfortunately, experience tells us that such a search is likely to end in failure. Economic history is full of examples of indicators, such as stock prices and various monetary aggregates, which work for a short period of time after their discovery and then fail dramatically just as they become widely used. There are many reasons for this, but one stands out. As the following analysis will show, indicators do well at different things and at different times. Without an understanding of the limitations this implies, these "best" indicators are often stretched well beyond their capabilities. What the business analyst really needs to know is the type of information that an indicator possesses and the types of purposes to which it can reasonably be put.

Indicators, like people, perform better or worse depending on the context in which they operate. Efficient usage requires matching indicators both with appropriate questions and with other complementary indicators. For instance, some indicators, such as the Purchasing Managers' Index of the National Association of Purchasing Management (NAPM), do well at predicting short run changes in activity, but do not do very well at pinning down the level of activity over longer time spans. Other indicators, such as the growth in real M2, forecast short run phenomena poorly, but do better at predicting average activity over a longer time

Francesca Eugeni is an associate economist, Charles Evans is a senior economist, and Steven Strongin is vice president and economic advisor at the Federal Reserve Bank of Chicago.

The authors would like to thank the participants of the Special Meeting on Operating Procedures held at the Federal Reserve Bank of St. Louis, June 18 and 19, 1992, and the participants of the Federal Reserve Bank of Chicago's Macro Workshop held June 9, 1992, for their comments.

span such as a year. Also, while some indicators are very close substitutes, such as the twenty or so interest rates sometimes used in econometric studies, each providing little additional information beyond the first, other indicators possess substantial independent information, thus providing important confirming or contradicting information. The analyst needs to know how to match questions with indicators depending on current needs. A swiss army knife is a fine general purpose tool, but it is hardly a substitute for a well equipped workshop. It is not enough just to produce a "best" model; rather, it is important to understand what type of information is contained in a given indicator so that its message can be properly evaluated and also to determine how much weight to give that message given what else is also known.

This article develops and implements a set of procedures for evaluating indicators of economic activity that closely match the actual use of such indicators by policymakers and businessmen alike. We see that process as primarily involving the reassessment of short to medium term economic activity based on an indicator by indicator analysis, with the primary decision matrix being whether to revise the assessment of activity up or down. We do not address related issues of assessing long run growth, inflation, interest rates, or the value of the dollar. Evaluating indicators in this context has four primary parts: ranking candidate indicators; characterizing the nature of the information in those indicators; assessing their usefulness in practice; and determining what relative weight should be given to each indicator. The idea is to develop the information that an analyst needs in order to interpret information as it comes in and to choose which indicators to watch depending on the questions being asked.

All of our analyses will be carried out on a bivariate (two variable) basis. Multivariate regression models allow indicators to play off against one another making it impossible to determine exactly what information is in each indicator. This in no way reduces the generality of the methods developed in this study, in that the forecast of a given multivariate model can be treated as a single indicator, just like any other. In fact, the National Bureau of Economic Research (NBER) Experimental Leading Index examined in Section 4 is just such an indicator.

Once the indicators are assessed and characterized, the last section of the article formally addresses the question of how to weight the information in one indicator relative to another. This is done through mixing models, which effectively produce a forecast based on the weighted average of the individual forecasts generated by the indicators. There are a number of advantages that are derived from using this mixing approach over the classical multivariate forecasting techniques. First, when one of the indicators begins to fail, which they do, you can reweight or at least temporarily just ignore that indicator. Second, by using only the primary information in each indicator, these models are less subject to the type of overfitting arising from interactions between indicators that plagues large econometric models. Third and most important, the mixing approach allows a much more precise assessment of exactly the type and value of information that is contained in each indicator and thus allows analysts to reoptimize their choice of indicators based on the type of question being asked.

Our investigation indicates that this type of analysis is crucial to the effective use of indicators. First, we find that a number of commonly used indicators, such as the monetary base and M1, actually contain negative information, in the sense that forecasts based purely on the past history of activity, ignoring these indicators, do better in practice than forecasts which include the information in these indicators. Second, we find that long term interest rate levels provide no additional information about future economic activity beyond that contained in short term interest rate levels, while the slope of the term structure contains substantial additional information. This would seem to indicate that a rise in long term interest rates is associated with an improvement in the near term outlook of the economy. It is interesting to note that this is contrary to popular wisdom, according to which a scenario with declining short term interest rates and increasing long term rates is viewed as negative. Third, we find that some indicators, such as the spread between the 3 month eurodollar rate and the 3 month Treasury bill rate, do a very good job of forecasting growth during expansions, but rarely signal recessions, while others, such as real M1 and the mix between bank and nonbank financing do better at forecasting during recessions, even though they are poor forecasters in general. Fourth, we find that composite indicators, such as the Department of Commerce Composite Index of Leading Indicators and the NBER Experimental Leading Index, are very good predictors of economic activity over a two quarter horizon, while real M2 and the slope of the term structure are more useful over a one year horizon.

This last finding illustrates a crucial point: the forecast horizon is fundamental to the choice of indicators. Short horizons favor interest rate risk spreads, such as the difference between the 6 month commercial paper rate and the 6 month Treasury bill rate (risk spreads are yield differences between private and public debt instruments with the same maturity), and activity based indicators, such as the Purchasing Managers' Index and the Sensitive Materials Price Index. Longer horizons, on the other hand, favor monetary indicators, such as real M2, and interest rate term spreads, such as the difference between the 12 month Treasury bill rate and the overnight federal funds rate (term spreads are yield differences between two public debt instruments with different maturities). This indicates that different types of information are important for forecasting growth at different forecast horizons.

Methodology

As noted above, the primary focus of this article is the examination of various data series as indicators of changes in real economic activity, which we measure as annualized quarterly log changes in real GDP, except in the sections of the article which focus on issues of timing, in which case the annualized monthly log changes in employment are used. Since the employment data series is available at the monthly frequency, it allows for more precise estimation of the pattern of impact over time.

Throughout the article the indicators are used to produce forecasts of economic activity. The specific functional form of the forecasting equation is always the same. One year of data for the indicator and one year of lagged economic activity are included in the regression. Thus, the exercise is strictly equivalent to a bivariate vector autoregression (VAR) with one year of lags: four quarters of lags for the real GDP models and twelve months of lags for the employment models. The models are estimated in log differences and rates of change are annualized. Interest rates, interest rate spreads, and some of the composite indicators are used in their level form. In many of the tables an additional forecast is provided with the label "NONE." In this case, the forecast is based solely on the past history of economic activity, that is, a pure autoregressive model with one year of lagged data. This pure autoregressive forecast is referred to as the no-indicator forecast. When the horizon of the forecast is varied, we simply change the dependent variable in the regression rather than dynamically iterate the one period ahead forecast. This optimizes the parameterization for the forecast horizon in question, rather than multiplicatively combining estimation errors forward. Symbolically the forecasting equation can be written:

$$(1)\quad Y_{t+k} - Y_t = A(L)\Delta Y_{t-1} + B(L)I_{t-1} + \omega_t;$$

where Y_t is the log of economic activity at time t, I_t is the indicator at time t, k is the number of periods in the forecast horizon, and $A(L)$ and $B(L)$ are polynomials in the lag operator L of order one year.

The indicators are split into four groups, which we call families. Each family is meant to represent a natural division of indicators into groups which are likely to share similar characteristics. The first family we examine is interest rate levels, the second is money based measures, the third is interest rate spreads, and the fourth is composite indicators, such as the Department of Commerce Composite Index of Leading Indicators and the Standard and Poor's 500 Stock Index. The fourth group also contains those series which do not fit neatly into the overall classification scheme.

The idea is to first examine the indicators within a family, characterize the information, and find out which indicators within each family produce the best forecasts and contain the most independent information. Then we take these "best" indicators from all four families and examine what is to be gained by mixing the information from different families. This serves a number of purposes. First, breaking the large list of potential indicators into smaller groups makes each examination more manageable. Second, using natural groupings allows us to look at questions such as what is the best interest rate or the best money measure in a natural way. Third, one key issue for indicators is the degree to which they actually contain independent information. Focusing on groups which are already thought to have similar information provides a natural way to learn if these preconceptions are accurate or if some of these groups contain more than one type of information. Lastly, by first selecting the best indicators at the family level and then mixing between families, we can produce a mixed forecast which, as noted above, closely approximates the way indicators are used in practice.

Each family of indicators is subjected to the same analysis. First, each family of indicators is described. Then each of the indicators is subjected to four evaluations: classical goodness-of-fit rankings; indicators' performance in practice; characterization of fit; and encompassing tests. The results of our evaluations are summarized in tables numbered as follows: the first digit in the table's number refers to the family of indicators (for example, interest rate levels constitute our first family), while the second digit refers to the type of statistics discussed (for example, multiperiod forecast results are summarized in the second table of each family). For example, Table 1.2 is the second table in our first family of indicators.

The first part of our analysis focuses on classical goodness-of-fit statistics, which are based on simple full sample regressions estimated on data from January 1962 to December 1991. The results are presented in Table _.1[2] of each family analysis section. In this table we report the correlation coefficients produced by the regression, and we rank the indicators in each family according to their R^2s. The idea is that the best indicators are the ones that produce the best fit as measured by the R^2 of the regression. This closely approximates the standard notions of evaluating indicators of economic activity. It is also closely linked to the notion of Granger causality, which statistically measures whether or not the indicator actually helps forecast economic activity. The probability value for this test is also included in the table. Low probability values, especially below .05, are normally thought to indicate that a variable is valuable in generating forecasts.

The second evaluation switches the focus to how well the indicators are likely to work in practice. To this end, goodness-of-fit is reinterpreted in a way closer to the way forecasts are actually used. First, Table _.2 shows goodness-of-fit rankings recalculated for a series of forecast horizons using standard regression analysis to provide a bench mark for evaluating out-of-sample forecasts. The one quarter horizon used in Table _.1 is first presented and then a two quarter forecast horizon evaluation and a four quarter forecast horizon evaluation.[3] Table _.3 in each section then repeats this analysis using forecasting equations which do not contain any prior information. Specifically, the forecasting equations are estimated using Kalman filtering techniques which recursively compute minimum mean squared errors using only data available prior to the forecasting period. This analysis provides a more accurate assessment of how an indicator is likely to perform in practice, since this is the regression an analyst would have actually estimated just prior to making the forecast, rather than the regression the analyst would generate today using all of the data since the forecast period. These forecasts are then ranked by the root mean squared error (RMSE) (the average size of the error) of the forecasts from July 1973 onward. To see how the indicators perform under different circumstances, we look at Kalman forecasts in recessions and expansions, and re-rank the indicators according to their RMSEs, as shown in Table _.4.

Next, Figure _.1 in each section graphs the cumulative residuals for the Kalman forecasts. These charts allow us to determine if these forecasts tend to perform badly during recessions or if there was some particular point in the past where they did especially well or poorly. It also tells us if the forecasts have tended to miss in some systematic fashion over time. The residuals are measured as the actual growth in economic activity minus the forecasted growth. Therefore, although a flat cumulated residuals' slope indicates good overall performance, a path consistently close to the zero horizontal line would be ideal. On the other hand, a downward trend in the cumulative residuals would indicate a period of overpredicting growth in activity, while an upward trend would indicate a period of underforecasting.

The third evaluation seeks to characterize the type of information in the indicator. Typically the question can be thought of as follows: if the indicator goes up today how does that change my expectations about economic activity in the future? This is analyzed by calculating the dynamic response path of employment for each of the indicator forecasting equations, which shows how a one standard deviation[4] increase in the indicator changes expectations about the future growth rate of employment for each month for the next 36 months.[5] This allows us to characterize the information in the indicator based on how fast economic activity responds, how much it responds and how long the change in activity lasts. Figure _.2 in each family section graphs the dynamic response path for selected indicators in the family, as well as the two standard deviation bands on the estimates of the dynamic response paths to show the amount of uncertainty about the response path.

The fourth evaluation switches the focus to independence of information. As noted earlier, one of the most important factors to understand about indicators is whether or not they contain independent information relative to some other indicator. This allows the analyst to assess whether a new piece of information actually contains any additional information or whether it is simply the same information with a different label. This is evaluated through a set of techniques called encompassing tests. In the context of this paper, indicator A is said to encompass indicator B if, given the forecast implicitly based on A, there is no additional information in indicator B. Indicator A is said to dominate indicator B if A encompasses B and B does not encompass A. The simplest way to test this is to run a regression with economic activity as the dependent variable and the forecast of activity based on indicator A and the forecast of activity based on indicator B as the independent variables. Symbolically this can be written:

(2) $\Delta GDP_t = \phi\, for(A)_t + (1-\phi) for(B)_t + \varepsilon$;

where $for(A)_t$ and $for(B)_t$ are the forecasts of GDP_t based on indicators A and B respectively and ϕ is the relative weight an ordinary least squares (OLS) regression assigns to $for(A)_t$ and $for(B)_t$. If ϕ is significantly different from 0 then we can reject that $for(A)$ is encompassed by $for(B)$. Likewise if $1 - \phi$ is significantly different from 0 then we can reject that $for(B)$ is encompassed by $for(A)$. If neither is encompassed then both indicators contain independent information and a better forecast can be obtained by mixing both sets of information with the relative weights given by ϕ. If only one is encompassed, then it is said to be dominated and only the other is necessary to produce an efficient forecast. If both are encompassed then either indicator alone can produce an efficient forecast. This occurs when there is a very high degree of collinearity and the standard error of the parameter estimate is large. In this case the indicator which has the best historical track record would likely be the superior choice. The generalization to longer horizons is straightforward, though the calculations of the standard errors are more complicated since the errors are no longer independent.

Table _.5 in each family section contains the encompassing tests. The table is read as follows. The indicators are listed both along the top and along the side of the matrix. The numbers in the table refer to the test that the indicator listed along the side is encompassed by the indicator along the top. The statistics reported are the significance levels for the test that the indicator along the top does in fact contain all the information in the indicator along the side. Values below .05 indicate substantial independent information possessed by the indicator listed along the side. For the sake of readability, such values are replaced by a dash in the table. In general, the lower the number, the more likely it is that the indicator listed along the side possesses independent information and the higher the number, the more likely it is that the indicator listed along the top encompasses the indicator along the side.

The way to interpret Table _.5 is that a side indicator whose row is blank contains information that is independent of every other indicator in the family. A top indicator whose column is full of high numbers is said to encompass the indicators on the side. An indicator that did both would be said to dominate the family. In general, we search for the set of indicators in each family which contains all the information in the family using as few indicators as possible. This will mean that the best variable from the previous tests will be included together with additional indicators which contain independent information, that is, the indicators that add the most. Formally, this means that we include all indicators that are not encompassed by any other indicators in the family plus whatever additional indicators are necessary to fully encompass or cover all of the other indicators in the family. This is analogous to finding a set of minimally sufficient statistics.

The indicators that make it through this process will then be tested in the mixing model section of the article in between-family encompassing tests, which examine whether or not there is independent information between families. Then a set of "best" indicators will be selected in order to develop mixing models of indicators which contain independent information for each of the forecasting horizons. These models will contain estimates of the appropriate relative weights that should be applied to the individual indicator-based forecasts. Completing the circle of policy forecasts, the mixing models will be time varying to see if there is any gain from adjusting the weight applied to these individual forecasts based on recent performance.

1. Interest rate levels

As shown in Table 1.1, we selected the following levels of interest rates for investigation: the federal funds rate (FF); the 3 and 6 month Treasury bill rates (TB03 and TB06); the 1, 3, 5, and 10 year Treasury constant maturity bond rates (CM01, CM03, CM05, and CM10); the 3 month eurodollar rate (EURO3); the 6 month commercial paper rate (CP6); and the BAA corporate bond rate (BAA). Goodness-of-fit tests show that all of these interest rates are negatively correlated with real GDP, which indicates that an increase in interest rates this period is associated with a decline in real output.

The eurodollar rate, the commercial paper rate, and the federal funds rate have the three largest absolute correlation coefficients with real GDP and produce the best fit to the model as measured by their individual R^2s, ranking first, second, and third, respectively. The strength of such relationships is not surprising given the role that these instruments play in money markets. For example, because the federal funds rate is a key instrument of monetary policy and a bench mark for other money market interest rates, fluctuations in the rate are strongly associated with future movements in real economic activity.

The predictive power of our interest rate family is then tested at different forecast horizons using standard regression analysis over the full sample period. The in-sample results of Table 1.2 show that while EURO3 loses some of its strength as the forecast horizon increases, as shown by the recalculated rankings, the fit of both CP6 and FF improves at longer forecast horizons, with FF having the strongest predictive power at the four quarter forecast horizon.

TABLE 1.1

Classical goodness-of-fit statistics

Indicator	R^2	Correlation with real GDP	P-value	Rank
FF	0.338	–0.353	0.0000	3
TB03	0.293	–0.299	0.0001	6
TB06	0.304	–0.295	0.0000	5
CM01	0.309	–0.282	0.0000	4
CM03	0.279	–0.257	0.0002	7
CM05	0.268	–0.251	0.0003	8
CM10	0.253	–0.237	0.0009	10
EURO3	0.354	–0.352	0.0000	1
CP6	0.348	–0.342	0.0000	2
BAA	0.258	–0.269	0.0007	9

NOTE: Sample period is January 1962 - December 1991, quarterly data.

To determine how interest rates would actually perform as indicators of economic activity, we use Kalman filtering techniques to produce out-of-sample forecasts using only data available prior to the forecasting period. When we rank the resulting RMSEs in Table 1.3 it becomes clear, once again, that the overall performance of short term interest rates improves when we expand the forecast horizon. FF continues to perform best at the one year forecast horizon, while maintaining a standing similar to the in-sample results at

TABLE 1.2

Multiperiod forecasts, in-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	R^2	Rank	R^2	Rank	R^2	Rank
FF	0.338	3	0.463	3	0.530	1
TB03	0.293	6	0.402	5	0.496	3
TB06	0.304	5	0.406	4	0.487	5
CM01	0.309	4	0.397	6	0.443	6
CM03	0.279	7	0.350	7	0.377	7
CM05	0.268	8	0.332	8	0.346	8
CM10	0.253	10	0.296	10	0.307	10
EURO3	0.354	1	0.471	2	0.490	4
CP6	0.348	2	0.475	1	0.516	2
BAA	0.258	9	0.329	9	0.315	9
NONE	0.118	11	0.123	11	0.076	11

NOTE: Sample period is January 1962 - December 1991, quarterly data.

TABLE 1.3

Kalman multiperiod forecasts, out-of-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
FF	3.793	2	2.859	3	2.160	1
TB03	3.969	9	3.075	6	2.260	5
TB06	3.862	4	3.000	5	2.251	4
CM01	3.826	3	2.996	4	2.356	6
CM03	3.876	5	3.094	7	2.483	7
CM05	3.936	7	3.144	8	2.552	8
CM10	3.949	8	3.249	10	2.683	9
EURO3	3.622	1	2.754	1	2.222	3
CP6	3.880	6	2.827	2	2.216	2
BAA	4.006	10	3.197	9	2.725	10
NONE	4.015	11	3.358	11	2.819	11

NOTE: Sample period is July 1973 - December 1991, quarterly data.

shorter horizons. CP6, on the other hand, experiences an out-of-sample deterioration at the one quarter horizon, but ranks second at both the two quarter and one year forecast horizons. In general, our results indicate that shorter maturity instruments, namely FF and EURO3, outperform longer maturity bonds, such as the 3, 5, and 10 year Treasury bonds.

Once the general strength of an indicator is established, it becomes important to determine how the indicator would perform under different economic circumstances, and Table 1.4 tells us how well or how poorly our interest rate family performs during recessions and expansions. The strength of FF deteriorates somewhat during both recessions and expansions, when compared to other interest rates. On the other hand, EURO3 continues to perform strongly especially during recessions, and CP6's ranking improves during expansionary periods. It is also interesting to note that our autoregressive indicator "NONE" ranks first in the Kalman forecasts during expansions. This result demonstrates that sometimes indicators can be misleading during expansionary periods.

The cumulated residuals from the Kalman forecasts in Figure 1.1 show that, overall, the indicators in our interest rate family consistently underforecasted real GDP between 1974 and 1982. The upward trend in the cumulated residuals during this period can be explained in part by an unprecedented increase in inflation, which caused interest rates to rise without the normally anticipated decline in output. On the other hand, between 1983 and 1989, FF, CP6, EURO3, and all of the Treasury bill rates performed well, as shown by the flattening of their cumulated residuals' slopes during this period. Between 1990 and 1991, however, the indicators' performance deteriorated again, as all of the interest rates missed the 1990-91 recession and consistently overforecasted real GDP.

Figure 1.2 shows the dynamic response of the forecasted growth rate of employment when FF increases. Because the response paths of our interest rate family are virtually identical across all indica-

TABLE 1.4

Kalman 1 quarter ahead forecasts in recessions and expansions

	Real GDP					
	Actual		Recession		Expansion	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
FF	3.793	2	3.753	4	3.801	4
TB03	3.969	9	4.108	8	3.941	9
TB06	3.862	4	3.780	6	3.878	6
CM01	3.826	3	3.663	2	3.857	5
CM03	3.876	5	3.722	3	3.905	7
CM05	3.936	7	3.814	7	3.959	10
CM10	3.949	8	3.766	5	3.983	11
EURO3	3.622	1	3.605	1	3.625	2
CP6	3.880	6	4.928	10	3.642	3
BAA	4.006	10	4.377	9	3.930	8
NONE	4.015	11	5.817	11	3.563	1

NOTE: Sample period is July 1973 - December 1991, quarterly data.

TABLE 1.5

Multiperiod encompassing tests

(Probability value for null hypothesis: X is encompassed by Y)

X \ Y	FF	TB03	TB06	CM01	CM03	CM05	CM10	EURO3	CP6	BAA	Maximum P-value
					Real GDP (1 quarter)						
FF	n.a.	—	—	—	—	—	—	0.932	0.856	—	0.932
TB03	0.482	n.a.	0.796	0.830	0.061	—	—	0.391	0.262	—	0.830
TB06	0.945	0.204	n.a.	0.947	—	—	—	0.403	0.241	—	0.947
CM01	0.677	0.103	0.342	n.a.	—	—	—	0.723	0.524	—	0.723
CM03	0.682	0.343	0.949	0.264	n.a.	0.065		0.803	0.601	0.053	0.949
CM05	0.637	0.375	0.910	0.412	0.251	n.a.	0.066	0.906	0.702	0.154	0.910
CM10	0.464	0.371	0.798	0.684	0.508	0.563	n.a.	0.818	0.976	0.380	0.976
EURO3	0.119	—	—	—	—	—	—	n.a.	0.240	—	0.240
CP6	0.272	—	—	—	—	—	—	0.659	n.a.	—	0.659
BAA	0.326	0.221	0.431	0.638	0.485	0.407	0.253	0.666	0.783	n.a.	0.783
					Real GDP (2 quarters)						
FF	n.a.	—	—	—	—	—	—	0.605	0.867	—	0.867
TB03	0.090	n.a.	0.925	0.310	—	—	—	0.340	—	—	0.925
TB06	0.337	0.448	n.a.	0.220	—	—	—	0.250	—	—	0.448
CM01	0.582	0.515	0.864	n.a.	—	—	—	0.293	—	—	0.864
CM03	0.617	0.959	0.443	0.109	n.a.	—	—	0.360	0.107	—	0.959
CM05	0.694	0.975	0.520	0.191	0.132	n.a.	—	0.450	0.197	0.137	0.975
CM10	0.665	0.763	0.418	0.210	0.096	—	n.a.	0.491	0.263	0.794	0.794
EURO3	0.231	—	—	—	—	—	—	n.a.	0.598	—	0.598
CP6	0.214	—	—	—	—	—	—	0.340	n.a.	—	0.340
BAA	0.574	0.302	0.635	0.837	0.429	0.228	—	0.989	0.742	n.a.	0.989
					Real GDP (4 quarters)						
FF	n.a.	—	—	—	—	—	—	—	—	—	0.044
TB03	0.963	n.a.	0.152	—	—	—	—	0.139	0.661	—	0.963
TB06	0.920	0.910	n.a.	—	—	—	—	0.255	0.662	—	0.920
CM01	0.596	0.373	—	n.a.	—	—	—	0.980	0.157	—	0.980
CM03	0.593	0.363	—	—	n.a.	—	—	0.623	0.166	—	0.623
CM05	0.541	0.302	—	—	—	n.a.	—	0.506	0.140	—	0.541
CM10	0.588	0.362	0.130	0.074	—	0.072	n.a.	0.555	0.211	0.419	0.588
EURO3	0.539	0.263	0.173	—	—	—	—	n.a.	0.785	—	0.785
CP6	0.746	—	—	—	—	—	—	0.052	n.a.	—	0.746
BAA	0.845	0.776	0.507	0.534	0.692	0.895	0.101	0.767	0.456	n.a.	0.895

NOTES: Values less than or equal to 0.05 are marked with a dash. Sample period is January 1962 - December 1991, quarterly data.

tors, we chose the federal funds rate as an example of how a one standard deviation increase in the interest rate today changes the growth rate of employment during the next 36 months. The forecasted growth rate of employment increases for approximately two months and then falls, plunging to very deep negative values especially during the first year. Eventually, the growth rate moves very close to zero as the horizon expands, indicating that the change in FF does not impact employment forecasting after approximately two years.

Finally, as shown in Table 1.5, our encompassing tests indicate that both FF and EURO3 contain significant information but neither of them dominates. This indicates that both interest rates are close substitutes, and using both would not improve the forecasting results since either interest rate contains all of the necessary information. For example, at the one quarter

FIGURE 1.1

Interest rate levels

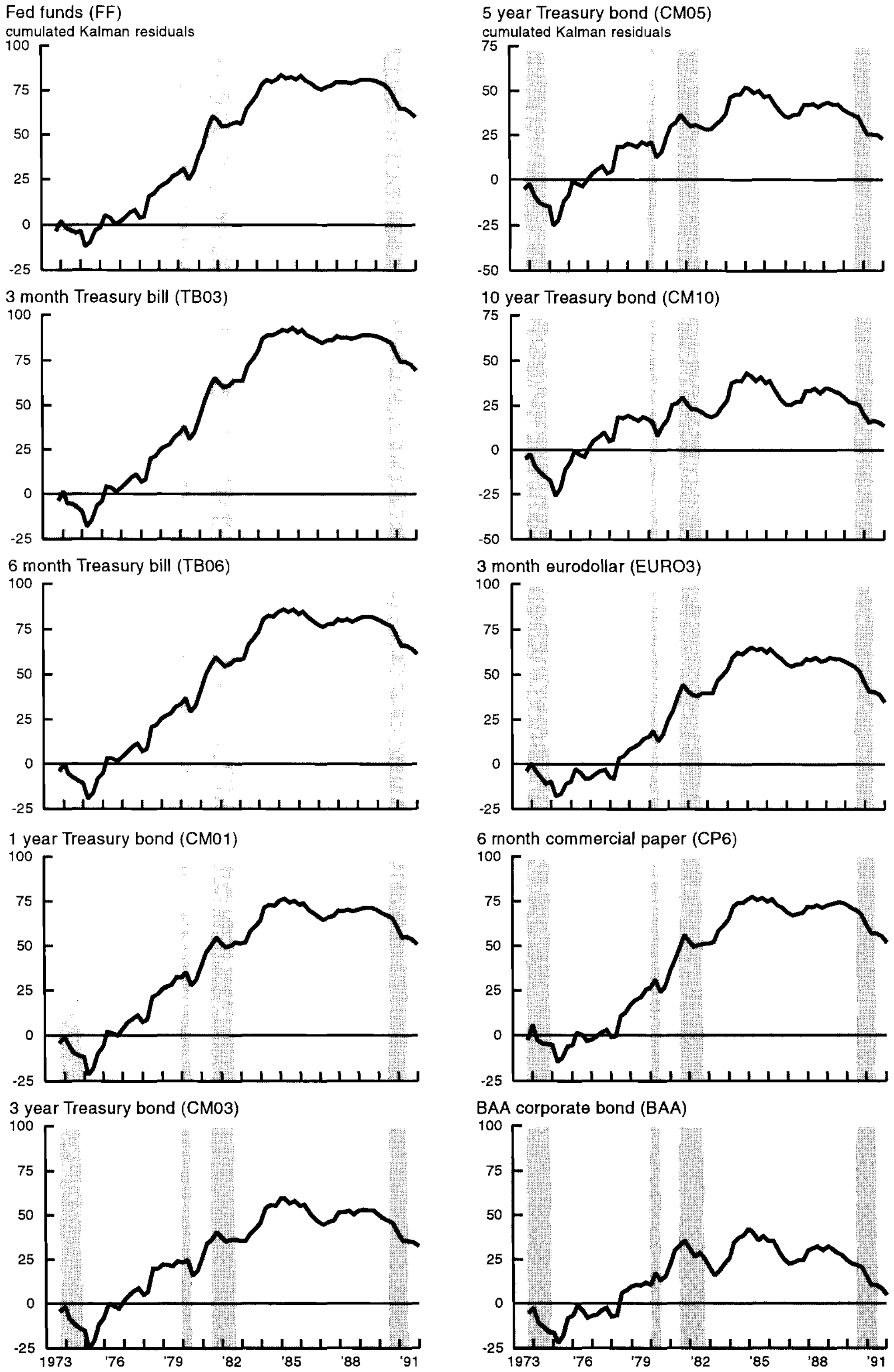


FIGURE 1.2

Dynamic response of employment to FF

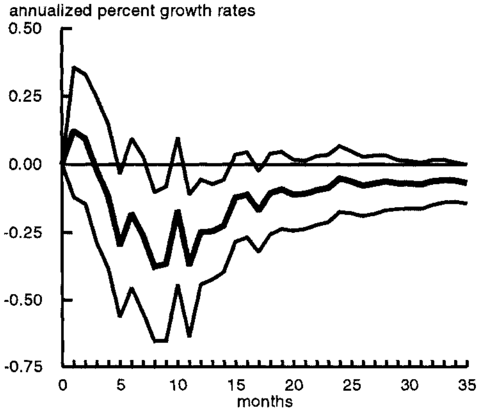


TABLE 2.1

Classical goodness-of-fit statistics

Indicator	R^2	Correlation with real GDP	P-value	Rank
MBSTL	0.166	0.034	0.1744	7
MB	0.145	0.013	0.4734	9
M1	0.172	0.157	0.1284	5
M2	0.219	0.236	0.0084	4
M3	0.169	0.246	0.1483	6
L	0.164	0.239	0.1993	8
DBTNF	0.124	0.180	0.9352	10
M1R	0.250	0.297	0.0012	2
M2R	0.346	0.353	0.0000	1
NBRX	0.249	0.154	0.0012	3

NOTE: Sample period is January 1962 - December 1991, quarterly data.

forecast horizon EURO3 encompasses all of the other indicators, but at the same time, EURO3 is encompassed by FF and CP6. However, because EURO3 ranked first in the in-sample forecasts at the one quarter horizon, and in the out-of-sample forecasts at the one and two quarter horizons, it is selected as our best indicator at both the one and two quarter forecast horizons. Similarly, FF is chosen as the best indicator at the one year forecast horizon for its strong performance in-sample and out-of-sample when the forecast horizon increases.

2. Money based measures

Table 2.1 lists the monetary indicators we selected for investigation: a measure of the monetary base developed by the Federal Reserve Bank of St. Louis (MBSTL); the Board of Governors' monetary base[6] (MB); M1; M2; M3; L;[7] long term debt of domestic nonfinancial institutions (DBTNF); real M1 (M1R) and real M2 (M2R) both deflated by the consumer price index; and NBRX, which is the ratio of nonborrowed reserves at time t to total reserves at time t–1. Strongin (1991) found that this normalized reserve aggregate (NBRX) contains much of the information about monetary policy actions which Sims (1991) attributes to innovations in the federal funds rate. Except for NBRX, all of the indicators in our family of money based measures are annualized log differences.

Goodness-of-fit statistics in Table 2.1 show that all of the money based indicators are positively correlated with real GDP. Not surprisingly, as the endogenous components of the monetary aggregate increase, the contemporaneous correlation with economic activity rises. Moreover, the broader monetary aggregates seem to impact real GDP more than the narrow-

TABLE 2.2

Multiperiod forecasts, in-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	R^2	Rank	R^2	Rank	R^2	Rank
MBSTL	0.166	7	0.154	8	0.102	8
MB	0.145	9	0.144	9	0.121	5
M1	0.172	5	0.183	7	0.096	10
M2	0.219	4	0.249	4	0.186	4
M3	0.169	6	0.189	5	0.107	7
L	0.164	8	0.184	6	0.097	9
DBTNF	0.124	10	0.133	10	0.121	6
M1R	0.250	2	0.288	3	0.244	3
M2R	0.346	1	0.447	1	0.514	1
NBRX	0.249	3	0.327	2	0.292	2
NONE	0.118	11	0.123	11	0.076	11

NOTE: Sample period is January 1962 - December 1991, quarterly data.

er measures of money. This is probably due to the fact that broader money measures consist of a larger number of components, each associated with movements in economic activity. M2R, M1R, M3, and L have the largest correlation coefficients with GDP, and M2R and M1R also show the strongest fit to the model, as their R^2s rank first and second, respectively. NBRX and M2 are also statistically significant, ranking third and fourth, respectively.

The predictive power of our money based indicators is then tested at different forecast horizons, and in-sample results shown in Table 2.2 indicate that M2R, M1R, NBRX, and nominal M2 all continue to perform well, providing additional information to the forecasts as the horizon increases. M2R, however, clearly has the strongest predictive power at all forecast horizons (ranking always first), while M1R's ranking slightly deteriorates as the forecast horizon increases. On the other hand, NBRX's performance improves at the two quarter and four quarter horizons, ranking second in both.

Once again, to see how the indicators would actually perform using only data prior to the forecasting period, we use Kalman filtering techniques. Out-of-sample Kalman forecast results in Table 2.3 show M2R and NBRX to have the strongest fit at all horizons, as shown by their individual RMSEs, while M1R's performance greatly improves in the long run. As shown in Table 2.4, M2R also consistently performs well under different circumstances, and especially during expansionary periods. On the other hand, while M1R is a good predictor during recessions, its performance considerably worsens during expansions. NBRX's performance is noticeably consistent during recessions and expansions, as it ranks third during both.

The cumulated residuals from the Kalman forecasts shown in Figure 2.1 provide another perspective of the out-of-sample performance of our family of money based measures. In our case, the best indicator is again M2R as its cumulated residuals' path clearly stays near zero values, except for isolated periods of large forecast errors in 1978 and 1981, when M2R underforecasted

TABLE 2.3

Kalman multiperiod forecasts, out-of-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
MBSTL	4.108	7	3.474	10	2.904	8
MB	4.114	8	3.426	7	2.840	7
M1	4.149	10	3.455	9	2.992	11
M2	3.944	3	3.252	3	2.809	4
M3	4.073	5	3.394	6	2.948	10
L	4.136	9	3.432	8	2.926	9
DBTNF	4.242	11	3.495	11	2.820	6
M1R	4.097	6	3.285	4	2.775	3
M2R	3.674	1	2.844	1	2.219	1
NBRX	3.799	2	3.003	2	2.550	2
NONE	4.015	4	3.358	5	2.819	5

NOTE: Sample period is July 1973 - December 1991, quarterly data.

TABLE 2.4

Kalman 1 quarter ahead forecasts in recessions and expansions

	Real GDP					
	Actual		Recession		Expansion	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
MBSTL	4.108	7	5.774	8	3.700	6
MB	4.114	8	5.534	7	3.777	7
M1	4.149	10	5.245	4	3.901	9
M2	3.944	3	6.011	11	3.402	2
M3	4.073	5	5.848	10	3.631	5
L	4.136	9	5.256	5	3.883	8
DBTNF	4.242	11	5.400	6	3.980	11
M1R	4.097	6	4.793	1	3.949	10
M2R	3.674	1	5.109	2	3.326	1
NBRX	3.799	2	5.228	3	3.454	3
NONE	4.015	4	5.817	9	3.563	4

NOTE: Sample period is July 1973 - December 1991, quarterly data.

FIGURE 2.1

Money based measures

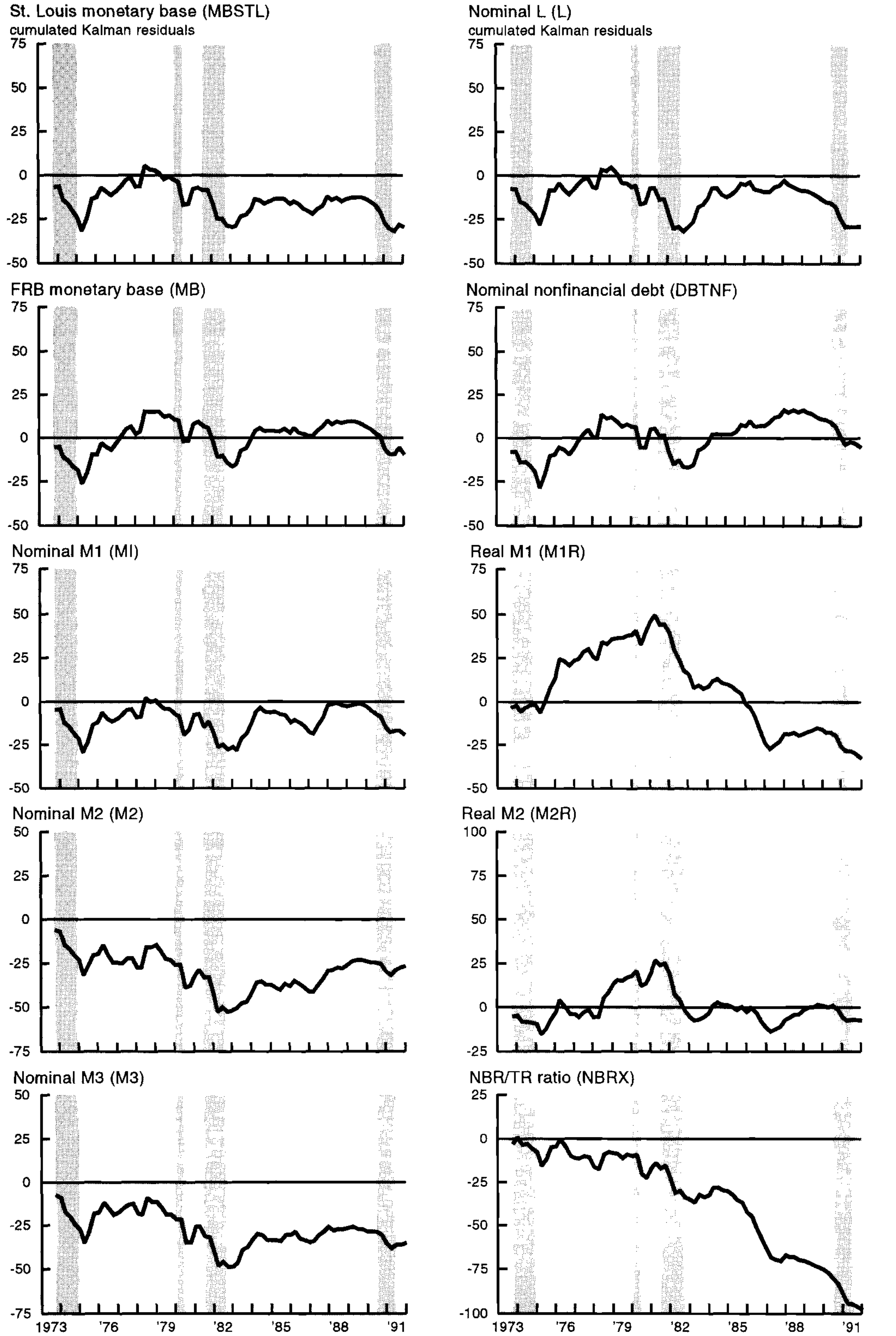


TABLE 2.5

Multiperiod encompassing tests

(Probability value for null hypothesis: X is encompassed by Y)

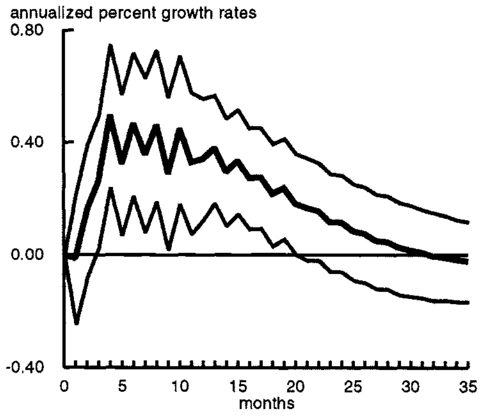
Y / X	MBSTL	MB	M1	M2	M3	L	DBTNF	M1R	M2R	NBRX	Maximum P-value
					Real GDP (1 quarter)						
MBSTL	n.a.	0.064	0.150	0.462	0.178	0.094	—	0.763	0.759	0.411	0.763
MB	0.726	n.a.	0.307	0.569	0.296	0.224	0.075	0.682	0.936	0.500	0.936
M1	0.055	—	n.a.	0.506	0.105	0.054	—	0.658	0.855	0.671	0.855
M2	—	—	—	n.a.	—	—	—	0.098	0.954	—	0.954
M3	0.138	—	0.135	0.733	n.a.	0.174	—	0.327	0.653	0.149	0.733
L	0.119	—	0.136	0.407	0.324	n.a.	—	0.322	0.449	0.286	0.449
DBTNF	0.694	0.755	0.669	0.771	0.716	0.825	n.a.	0.829	0.970	0.755	0.970
M1R	—	—	—	—	—	—	—	n.a.	0.924	—	0.924
M2R	—	—	—	—	—	—	—	—	n.a.	—	0.000
NBRX	—	—	—	—	—	—	—	—	0.286	n.a.	0.286
					Real GDP (2 quarters)						
MBSTL	n.a.	0.266	0.760	0.817	0.484	0.359	—	1.000	0.954	0.959	1.000
MB	0.595	n.a.	0.516	0.654	0.477	0.445	0.167	0.686	0.994	0.722	0.994
M1	—	—	n.a.	0.667	0.112	—	—	0.803	0.970	0.845	0.970
M2	—	—	—	n.a.	—	—	—	0.173	0.833	0.119	0.833
M3	—	—	0.064	0.603	n.a.	0.197	—	0.323	0.560	0.193	0.603
L	—	—	—	0.258	0.294	n.a.	—	0.274	0.284	0.333	0.333
DBTNF	0.490	0.715	0.604	0.691	0.533	0.697	n.a.	0.774	0.973	0.745	0.973
M1R	—	—	—	—	—	—	—	n.a.	0.752	0.101	0.752
M2R	—	—	—	—	—	—	—	—	n.a.	—	0.000
NBRX	—	—	—	—	—	—	—	—	0.133	n.a.	0.133
					Real GDP (4 quarters)						
MBSTL	n.a.	0.930	0.341	0.604	0.525	0.344	0.659	0.840	0.782	0.896	0.930
MB	0.336	n.a.	0.126	0.248	0.228	—	0.330	0.362	0.817	0.464	0.817
M1	0.658	0.693	n.a.	0.914	0.669	0.439	0.517	0.987	0.841	0.958	0.987
M2	—	—	—	n.a.	—	—	—	0.263	0.430	0.400	0.430
M3	0.452	0.392	0.424	0.612	n.a.	0.442	0.375	0.776	0.918	0.746	0.918
L	0.521	0.523	0.396	0.523	0.626	n.a.	0.652	0.802	0.824	0.975	0.975
DBTNF	0.196	0.331	0.072	0.230	0.209	0.089	n.a.	0.300	0.836	0.334	0.836
M1R	—	—	—	—	—	—	—	n.a.	0.257	0.305	0.305
M2R	—	—	—	—	—	—	—	—	n.a.	—	0.000
NBRX	—	—	—	—	—	—	—	—	0.473	n.a.	0.473

NOTES: Values less than or equal to 0.05 are marked with a dash. Sample period is January 1962 - December 1991, quarterly data.

economic activity. M2R's performance was again noticeably good between 1990 and 1991, when most of the other money based indicators clearly failed to predict the recession. NBRX was relatively stable from 1973 to 1981, but has shown a consistent pattern of overforecasting output growth since 1982. This deterioration may be due to increasing reluctance on the part of banks to borrow from the discount window. The performance of other monetary aggregates is less reliable and clearly more volatile than the behavior of M2R and NBRX. For example, the two measures of the monetary base and M1 consistently underforecasted real GDP between 1974 and 1977, as shown by their upward sloping paths. Overall, the path of nominal aggregates plunged during the credit control program of 1980, overpredicting output growth during the mild recession. From 1983 to 1988, these nominal aggregates performed fairly well, exhibiting uncharacteristic stability, except for M1 which did substantially worse between

FIGURE 2.2

Dynamic response of employment to M2R



1983 and 1984. Finally, between 1990 and 1991, there was a considerable deterioration in the performance of M1, L, and the two measures of the monetary base, as they consistently overpredicted economic growth.

To see how changes in money based measures affect the forecaster's expectations over time, we look at the dynamic response of employment to our strongest indicator, M2R. Figure 2.2 shows the response of the forecasted growth rate of employment when M2R increases by one standard deviation. In general, a positive impulse in a money based indicator leads to an increase in employment growth rates. In our case, the response to a one standard deviation increase in M2R is quick and persistent over a period of approximately 15 months, with the maximum impact occurring within the first year. These results indicate that the impact of changes in M2R on real economic activity is very strong, although somewhat short lived.

Finally, we test our family of money based measures to determine the degree of independent information they contribute to the model individually. As shown in Table 2.5, our encompassing tests show that M2R is clearly the dominant indicator within our family of monetary aggregates. In fact, M2R is not encompassed by any of the other indicators at all forecast horizons. The row labeled M2R in the table has dashes, indicating that the hypothesis that M2R is encompassed by any of the other indicators is consistently rejected. Similarly, the high significance levels in the column labeled M2R indicate that M2R encompasses all of the other indicators at all forecast horizons. This suggests that M2R contains unique information and that adding another money based indicator to the model would add no additional information.

3. Interest rate spreads

Recent research on financial market indicators of economic activity has brought renewed attention to interest rate spreads. Laurent (1988), Bernanke (1990), Estrella and Hardouvelis (1991), Friedman and Kuttner (1992), Kashyap, Stein, and Wilcox (1991), and Stock and Watson (1989b) all have suggested and tested various interest rate spreads as predictors of economic activity with significant success. The idea behind most of these spreads is that the difference in yields between two different debt instruments has a greater informational content than interest rate levels. The two primary types of interest rate spreads that have been used are risk spreads which measure the difference in yield between a private debt instrument and a government bond of equivalent maturity, and term spreads which measure the difference in yield between two government debt instruments of different maturities.

Typically, risk spreads contain information useful to the forecaster because the return on the private debt instrument is a measure of the market's assessment of the near term risk in the relevant business environment, and higher returns are usually associated with higher perceived business risk. Friedman and Kuttner (1992) have argued that this interpretation is probably flawed since the spreads are typically too large to be explained by any reasonable estimate of the risk inherent in the private debt instruments. Therefore, they suggest that liquidity considerations play a significant role in the pricing of private/public spreads. Following their lead, we will also refer to these spreads as private/public spreads.

Term spreads seek to measure the market's perception of the relative availability of credit through time. The convention is that the yield on the debt instrument with the shorter maturity is subtracted from the yield on the instrument with the longer maturity. Thus, a positive spread would indicate that short term funding is cheaper than long term funding, therefore boosting current economic activity. An alternative explanation is that the higher long term yields may signal expectations of higher future credit demand resulting from increased economic activity. An additional interpretation is that by taking the difference between long and short term interest rates, the short term rate is corrected for changes in inflationary expectations and taxes, leaving a better measure of short run credit conditions. In any case, all of these term spread regressions have the counterintuitive implication that a rise in long term interest rates is good for the near term outlook of the economy. Estrella et al. (1991) and Strongin (1990) attempt to reconcile the term spread results with current theory, however with limited success.

As shown in Table 3.1, we tested seven term spreads and three private/public spreads.[8] Five of the seven term spreads are based on the federal funds rate (FF), and they are: the 3 month Treasury bill rate less FF (TB3FF); the 6 month Treasury bill rate less FF (TB6FF); the 12 month Treasury bill rate less FF (TB12FF); the 5 year Treasury constant maturity bond rate less FF (CM05FF); and the 10 year Treasury constant maturity bond rate less FF (CM10FF). TB3FF is a short term spread; TB6FF is a medium term spread; and TB12FF, CM05FF, and CM10FF are all long term spreads. Our term spreads also include two intermediate spreads: the difference between the 12 month and the 3 month Treasury bill rates (TB12TB3), and the difference between the 10 year and the 1 year Treasury constant maturity bond rates (CM10CM1).

The three private/public spreads we investigated are: the 3 month eurodollar rate less the 3 month Treasury bill rate (EUROTB3); the 6 month commercial paper rate less the 6 month Treasury bill rate (CP6TB6); and the BAA corporate bond rate less the 10 year Treasury constant maturity bond rate (BAACM10).[9]

Goodness-of-fit statistics in Table 3.1 indicate that all of our term spreads are positively associated with real GDP, with the short and medium spreads showing the strongest correlation coefficients. The positive association is not surprising given that short term interest rates tend to be more volatile than long term interest rates, and that a decline in short term interest rates is typically associated with a steepening of the yield curve. On the other hand, private/public spreads are negatively correlated with GDP, with CP6TB6 having the strongest correlation coefficient in absolute terms. An increase in the yield on private debt instruments may signal a riskier economic environment, which is then associated with a decline in investment and a drop in output. In this case, if the return on public instruments is unchanged, the private/public spread increases while economic activity declines. CP6TB6 has also the strongest fit to the model, as shown by its R^2, followed by TB12FF and TB3FF.

TABLE 3.1

Classical goodness-of-fit statistics

Indicator	R^2	Correlation with real GDP	P-value	Rank
TB3FF	0.327	0.449	0.0000	3
TB6FF	0.321	0.442	0.0000	4
TB12FF	0.330	0.425	0.0000	2
CM05FF	0.302	0.321	0.0000	6
CM10FF	0.309	0.309	0.0000	5
TB12TB3	0.238	0.225	0.0026	9
CM10CM1	0.284	0.170	0.0001	8
EUROTB3	0.294	-0.378	0.0001	7
CP6TB6	0.339	-0.431	0.0000	1
BAACM10	0.234	-0.297	0.0033	10

NOTE: Sample period is January 1962 - December 1991, quarterly data.

TABLE 3.2

Multiperiod forecasts, in-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	R^2	Rank	R^2	Rank	R^2	Rank
TB3FF	0.327	3	0.446	3	0.437	5
TB6FF	0.321	4	0.459	2	0.490	4
TB12FF	0.330	2	0.470	1	0.518	1
CM05FF	0.302	6	0.428	6	0.498	2
CM10FF	0.309	5	0.435	4	0.491	3
TB12TB3	0.238	9	0.333	9	0.383	7
CM10CM1	0.284	8	0.374	7	0.396	6
EUROTB3	0.294	7	0.364	8	0.230	9
CP6TB6	0.339	1	0.429	5	0.289	8
BAACM10	0.234	10	0.175	10	0.138	10
NONE	0.118	11	0.123	11	0.076	11

NOTE: Sample period is January 1962 - December 1991, quarterly data.

The predictive power of our family of interest rate spreads is next tested at different forecast horizons, and in-sample results in Table 3.2 show a strong deterioration in the performance of CP6TB6 at the two and four quarter forecast horizons, while the strength of TB12FF improves considerably in the long run. In general, the predictive power of medium and long term spreads seems to improve as the forecast horizon increases. Also, term spreads perform better than private/public spreads across horizons, except for CP6TB6, which is the strongest indicator at the one quarter forecast horizon. This scenario is virtually unchanged in the out-of-sample Kalman forecasts shown in Table 3.3. As we test the actual performance of our indicators using only data available prior to the forecasting period, we see that CP6TB6 remains very strong in the short run, although its ranking somewhat deteriorates when compared to in-sample results. Although the out-of-sample performance of TB12FF at short term horizons considerably worsens, its strength increases at the four quarter forecast horizon, as its RMSE ranks first. Under different circumstances, we see that overall, private/public spreads, such as CP6TB6 and EUROTB3, perform better during expansionary periods than our term spreads, as shown in Table 3.4. On the other hand, term spreads outperform private/public spreads during recessions, as TB3FF and TB12FF rank first and second, respectively, according to their individual RMSEs.

The cumulated residuals from the Kalman forecasts in Figure 3.1 show some striking similarities in the overall forecasting performance of our family of interest rate spreads. Except for TB3FF, TB6FF, and TB12FF, all of our spreads tend to overforecast real GDP, as shown by their consistently negative residuals. While TB3FF, TB6FF, and TB12FF performed fairly well from 1973 to 1980, they clearly failed during the last three recessions. In fact, they all underforecasted economic activity between 1980 and 1982, and then overpredicted real GDP between 1990 and 1991. Between 1982 and 1989, their path was conspicuously flat. This suggests that these spreads do well in

TABLE 3.3

Kalman multiperiod forecasts, out-of-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
TB3FF	3.609	1	2.674	1	2.253	5
TB6FF	3.691	3	2.691	2	2.081	2
TB12FF	3.753	6	2.754	3	2.015	1
CM05FF	3.745	5	2.811	6	2.111	3
CM10FF	3.763	7	2.785	5	2.161	4
TB12TB3	4.197	11	3.187	9	2.370	6
CM10CM1	3.857	8	2.970	8	2.389	7
EUROTB3	3.698	4	2.886	7	2.721	8
CP6TB6	3.656	2	2.760	4	2.744	9
BAACM10	3.983	9	3.485	11	2.846	11
NONE	4.015	10	3.358	10	2.819	10

NOTE: Sample period is July 1973 - December 1991, quarterly data.

TABLE 3.4

Kalman 1 quarter ahead forecasts in recessions and expansions

	Real GDP					
	Actual		Recession		Expansion	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
TB3FF	3.609	1	4.353	1	3.447	3
TB6FF	3.691	3	4.634	3	3.479	4
TB12FF	3.753	6	4.599	2	3.566	9
CM05FF	3.745	5	4.714	5	3.527	6
CM10FF	3.763	7	4.823	6	3.521	5
TB12TB3	4.197	11	5.172	8	3.980	11
CM10CM1	3.857	8	4.707	4	3.670	10
EUROTB3	3.698	4	4.987	7	3.393	2
CP6TB6	3.656	2	5.727	10	3.099	1
BAACM10	3.983	9	5.698	9	3.557	7
NONE	4.015	10	5.817	11	3.563	8

NOTE: Sample period is July 1973 - December 1991, quarterly data.

FIGURE 3.1

Interest rate spreads

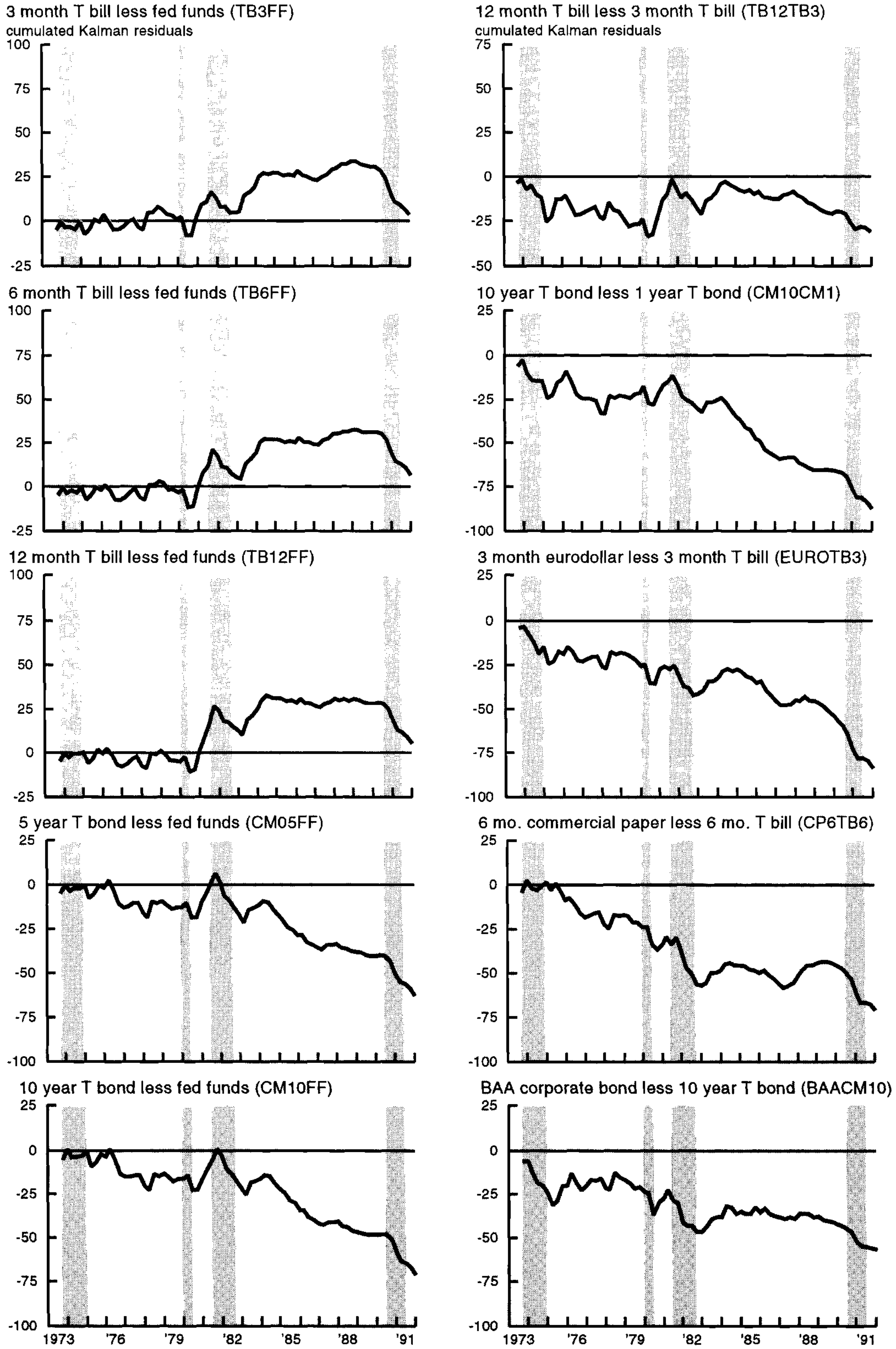
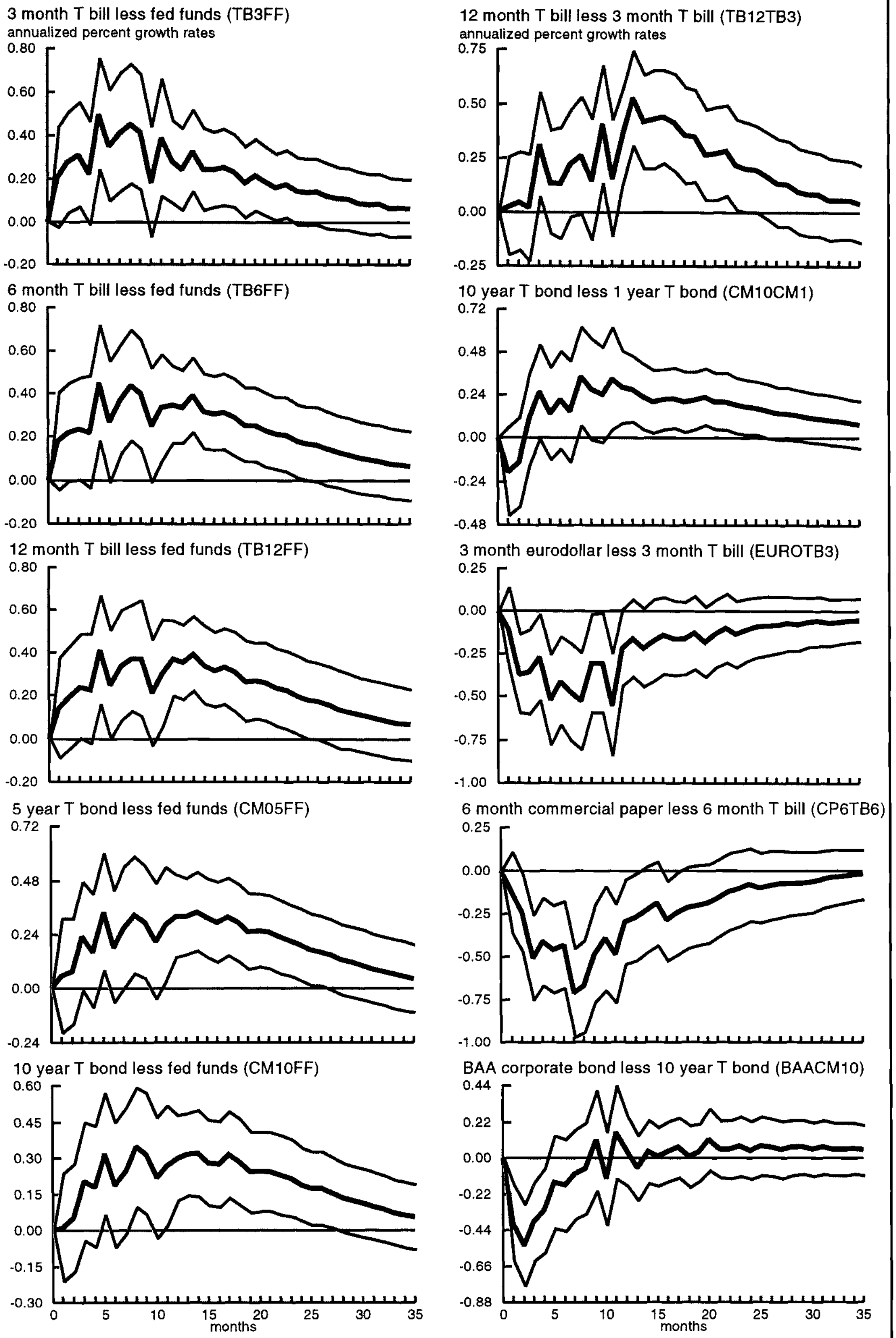


FIGURE 3.2

Dynamic response of employment to interest rate spreads



forecasting "normal" periods of economic activity, but periodically fail in predicting recessions. Although CM05FF and CM10FF follow a similar pattern between 1973 and 1981, after 1982 their cumulated residuals' path never stabilized but plunged to persistently negative values. Our intermediate term spreads (TB12TB3 and CM10CM1) failed during all of the recessions in our sample period (including the 1973-1975 recession), and developed a consistently negative bias after 1982, as they clearly overpredicted real GDP. All of the private/public spreads followed the same general pattern of mediocre performance from 1973 to 1981, and persistent overprediction of economic activity thereafter. In general, we conclude that, although a persistent bias in forecasting exists in all of the interest rate spreads we investigated, some of them did fairly well during most of our sample period, but failed during periods of large scale financial restructuring.

TABLE 3.5

Multiperiod encompassing tests

(Probability value for null hypothesis: X is encompassed by Y)

X \ Y	TB3FF	TB6FF	TB12FF	CM05FF	CM10FF	TB12TB3	CM10CM1	EUROTB3	CP6TB6	BAACM10	Maximum P-value
					Real GDP (1 quarter)						
TB3FF	n.a.	0.185	0.167	—	—	—	—	—	0.156	—	0.185
TB6FF	0.462	n.a.	0.999	—	—	—	—	—	0.109	—	0.999
TB12FF	0.105	0.227	n.a.	—	—	—	—	—	—	—	0.227
CM05FF	—	0.109	0.389	n.a.	0.540	—	0.098	—	0.053	—	0.540
CM10FF	—	0.062	0.231	0.215	n.a.	—	0.077	—	—	—	0.231
TB12TB3	0.093	0.333	0.797	0.430	0.413	n.a.	0.115	—	0.055	—	0.797
CM10CM1	—	—	0.106	0.454	0.699	—	n.a.	—	—	—	0.699
EUROTB3	0.125	—	—	—	—	—	—	n.a.	0.186	—	0.186
CP6TB6	0.066	—	—	—	—	—	—	—	n.a.	—	0.066
BAACM10	—	—	—	—	—	—	0.072	0.053	0.104	n.a.	0.104
					Real GDP (2 quarters)						
TB3FF	n.a.	0.569	0.467	—	—	—	—	—	—	—	0.569
TB6FF	0.070	n.a.	0.798	—	—	—	—	—	—	—	0.798
TB12FF	—	0.155	n.a.	—	—	—	—	—	—	—	0.155
CM05FF	—	0.092	0.337	n.a.	0.665	—	—	—	—	—	0.665
CM10FF	—	0.055	0.206	0.271	n.a.	—	—	—	—	—	0.271
TB12TB3	—	0.256	0.755	0.370	0.353	n.a.	0.071	—	—	—	0.755
CM10CM1	—	—	0.126	0.876	0.710	—	n.a.	—	—	—	0.876
EUROTB3	0.214	—	—	—	—	—	—	n.a.	0.222	—	0.222
CP6TB6	0.093	—	—	—	—	—	—	—	n.a.	—	0.093
BAACM10	0.545	0.459	0.580	0.908	0.991	0.436	0.935	0.807	0.936	n.a.	0.991
					Real GDP (4 quarters)						
TB3FF	n.a.	0.322	0.548	0.131	0.092	—	—	—	—	—	0.548
TB6FF	—	n.a.	0.197	0.056	—	—	—	—	—	—	0.197
TB12FF	—	—	n.a.	—	—	—	—	—	—	—	0.027
CM05FF	—	—	0.176	n.a.	0.170	—	—	—	—	—	0.176
CM10FF	—	—	0.144	0.720	n.a.	—	—	—	—	—	0.720
TB12TB3	—	0.142	0.576	0.333	0.261	n.a.	—	—	—	—	0.576
CM10CM1	—	—	0.062	0.593	0.230	—	n.a.	—	—	—	0.593
EUROTB3	0.752	0.989	0.979	0.899	0.965	0.094	0.428	n.a.	0.560	—	0.989
CP6TB6	0.883	0.870	0.840	0.774	0.783	—	0.115	—	n.a.	—	0.883
BAACM10	0.500	0.392	0.442	0.863	0.930	0.111	0.973	0.569	0.575	n.a.	0.973

NOTES: Values less than or equal to 0.05 are marked with a dash. Sample period is January 1962 - December 1991, quarterly data.

To see how changes in interest rate spreads may affect a forecaster's analysis of real economic activity over time, we look at the dynamic response of forecasted employment growth rates to a one standard deviation increase in our family of spreads. The paths depicted in Figure 3.2 show substantial differences between the response to changes in the term spreads and changes in the private/public spreads. The response of forecasted employment growth rates when BAACM10 increases is a quick dip in the first two months followed by a fast jump which peaks after eight months and then dies quickly. The response to the two shorter term private/public spreads (EUROTB3 and CP6TB6) follows an exact opposite path, first declining rapidly for approximately ten months, and then rapidly flattening. With the exception of TB12TB3, the response paths of the term spreads are all very similar, with employment growth rates increasing slowly, peaking at approximately ten months, and then flattening thereafter. This means that either a decline in short term interest rates or a rise in long term interest rates would cause forecasters to increase their predictions of future economic activity. The scenario depicted thus far indicates that the strength of the BAACM10 spread is in very short forecast horizons, as its impact on real economic activity dies fairly quickly compared to other spreads. On the other hand, our analysis shows that the strength of CP6TB6 is in the short and medium forecast horizons, while term spreads' overall impact on real economic activity is extremely persistent.

The results of our encompassing tests shown in Table 3.5 are exactly what we would have expected, given our analysis thus far. That is, we need to look at both a private/public spread and a term spread to obtain all of the information necessary for forecasting economic activity using interest rate spreads. This is due to the fact that term spreads usually perform better at longer horizons, while private/public spreads have a stronger predictive power at shorter horizons. CP6TB6 and TB12FF dominate their respective groupings. At the four quarter horizon, CP6TB6 no longer contains additional information beyond that contained in TB12FF. Now, however, a longer horizon term spread such as CM10FF is also necessary to fully cover the information set. It is interesting to note that the analysis of all of the encompassing results indicates that the separation between the private/public spreads and the term spreads is not very clear. In fact, at some forecast horizons the results reverse. This indicates that there are common multiple driving forces in the determination of these spreads, and that the driving factors associated with longer horizons of economic activity predominate in the term spreads, while the common factors that drive short run performance dominate the private/public spreads.

4. Composite indicators

Table 4.1 lists the composite indicators we investigated: the National Bureau of Economic Research (NBER) Experimental Leading Index (XLI); the NBER Nonfinancial Experimental Recession Index[10] (XRI2); the Department of Commerce (DOC) Composite Index of Leading Indicators (LEAD); the Purchasing Managers' Index (PMI) of the National Association of Purchasing Management (NAPM); the Standard and Poor's 500 Stock Index (S&P); the percent change in sensitive materials prices (SMPS);[11] and the Kashyap-Stein-Wilcox "mix" (KSWMIX), which is the ratio of bank lending to the sum of bank lending and commercial paper lending [see Kashyap et al. (1991)]. Note that the NBER Experimental Leading Index includes the 10 year Treasury bond/1 year Treasury bond spread and the 6 month commercial paper/6 month Treasury bill spread, while the Department of Commerce Composite Index of Leading Indicators includes real M2, all of which have been discussed in previous sections. The two composite leading indicators and the NBER Nonfinancial Experimental Recession Index are designed to predict economic activity at a six month horizon, although the optimization for the Department of Commerce Index is not as specif-

TABLE 4.1

Classical goodness-of-fit statistics

Indicator	R^2	Correlation with real GDP	P-value	Rank
XLI	0.455	0.547	0.0000	1
XRI2	0.385	−0.649	0.0000	3
LEAD	0.405	0.600	0.0000	2
PMI	0.265	0.632	0.0005	4
S&P	0.205	0.185	0.0222	7
SMPS	0.232	0.278	0.0045	6
KSWMIX	0.243	0.316	0.0023	5

NOTE: Sample period is January 1963 - December 1991, quarterly data.

ic as either of the NBER indices. Except for S&P and LEAD, which are annualized log differences, all of the indicators in our family of composite indicators are used in levels. Also, because data on the XRI2 start in January 1962, our sample period for this family of indicators starts in January 1963.

Goodness-of-fit tests in Table 4.1 show that, except for XRI2, our composite indicators have a positive correlation with contemporaneous economic activity. XRI2 has the strongest correlation with real GDP in absolute terms, while XLI has the strongest fit to the model as it ranks first according to its R^2. LEAD and XRI2 also show considerable strength as their R^2s rank second and third, respectively. The predictive power of our family of composite indicators is then tested in-sample at different forecast horizons. The results reported in Table 4.2 show that XLI and LEAD continue to perform very well at all forecast horizons, while XRI2 loses strength at the four quarter horizon. PMI and S&P continue to show weakness, especially in the long run, while SMPS' performance slightly improves at the four quarter horizon.

TABLE 4.2

Multiperiod forecasts, in-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	R^2	Rank	R^2	Rank	R^2	Rank
XLI	0.455	1	0.568	1	0.401	1
XRI2	0.382	3	0.316	3	0.168	6
LEAD	0.405	2	0.341	2	0.247	2
PMI	0.265	4	0.203	7	0.173	5
S&P	0.205	7	0.216	5	0.152	7
SMPS	0.232	6	0.206	6	0.229	3
KSWMIX	0.243	5	0.249	4	0.193	4
NONE	0.117	8	0.117	8	0.072	8

NOTE: Sample period is January 1963 - December 1991, quarterly data.

The results of out-of-sample Kalman tests in Table 4.3 show a picture very similar to the in-sample results, as XLI continues to rank first across horizons. LEAD continues to rank second, except for a slight deterioration in the four quarter forecast horizon where it ranks third. XRI2 has again a strong predictive power in the short run, while its performance worsens at the four quarter horizon. XRI2's behavior is expected, however, as the indicator was created to forecast recessions with a six month horizon.

Under different circumstances we notice that XLI loses some of its strength outside of "normal" economic activity, as shown in Table 4.4. That is, XLI's predictive power is slightly weaker during both recessions and expansions. On the other hand, LEAD performs well during expansions, although its performance worsens during recessionary periods. As expected, XRI2 is our best performer during recessions.

The cumulated Kalman residuals in Figure 4.1 show some striking similarities and some differences in actual performance across these indicators. Except for KSWMIX, all of our composite indicators have overforecasted real GDP over time, as their cumulated residuals are consistently negative. This bias is clearly evident during recessions and becomes more dramatic after 1980. After 1982, while the negative bias is exacerbated in XLI and S&P, the path becomes somewhat more stable for most of our indicators. XRI2 is our best performer during this period, which is not surpris-

TABLE 4.3

Kalman multiperiod forecasts, out-of-sample

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
XLI	3.246	1	2.376	1	2.392	1
XRI2	3.427	3	3.026	3	2.758	5
LEAD	3.307	2	3.024	2	2.669	3
PMI	3.838	4	3.319	6	2.736	4
S&P	3.964	6	3.253	4	2.758	6
SMPS	3.914	5	3.306	5	2.612	2
KSWMIX	4.078	8	3.377	8	2.846	8
NONE	4.052	7	3.369	7	2.799	7

NOTE: Sample period is July 1973 - December 1991, quarterly data.

TABLE 4.4

Kalman 1 quarter ahead forecasts in recessions and expansions

	Real GDP					
	Actual		Recession		Expansion	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
XLI	3.246	1	4.657	2	2.895	2
XRI2	3.427	3	4.321	1	3.226	3
LEAD	3.307	2	5.148	4	2.814	1
PMI	3.838	4	5.879	6	3.300	4
S&P	3.964	6	5.919	8	3.460	6
SMPS	3.914	5	5.711	5	3.460	5
KSWMIX	4.078	8	4.724	3	3.941	8
NONE	4.052	7	5.894	7	3.588	7

NOTE: Sample period is July 1973 - December 1991, quarterly data.

ing since the index was originally developed in response to the failure of XLI to forecast the 1990-1991 recession.

The dynamic responses of forecasted employment growth rates to changes in our composite indicators in Figure 4.2[12] show somewhat similar patterns for XLI and LEAD, where the response peaks quickly within approximately five months. From the peak, both graphs exhibit significantly different behaviors. The path in the XLI graph stabilizes for four to five months and then drops off before the end of the year, while the path in the LEAD graph falls more quickly and more dramatically, until the impact of the indicator on real economic activity disappears. The path of XRI2 is inverted instead when compared to the path of the two leading indicators. In fact, as the graph shows, the response path plunges very rapidly during the first five months, then increases for another six months, and finally stabilizes thereafter. The

TABLE 4.5

Multiperiod encompassing tests

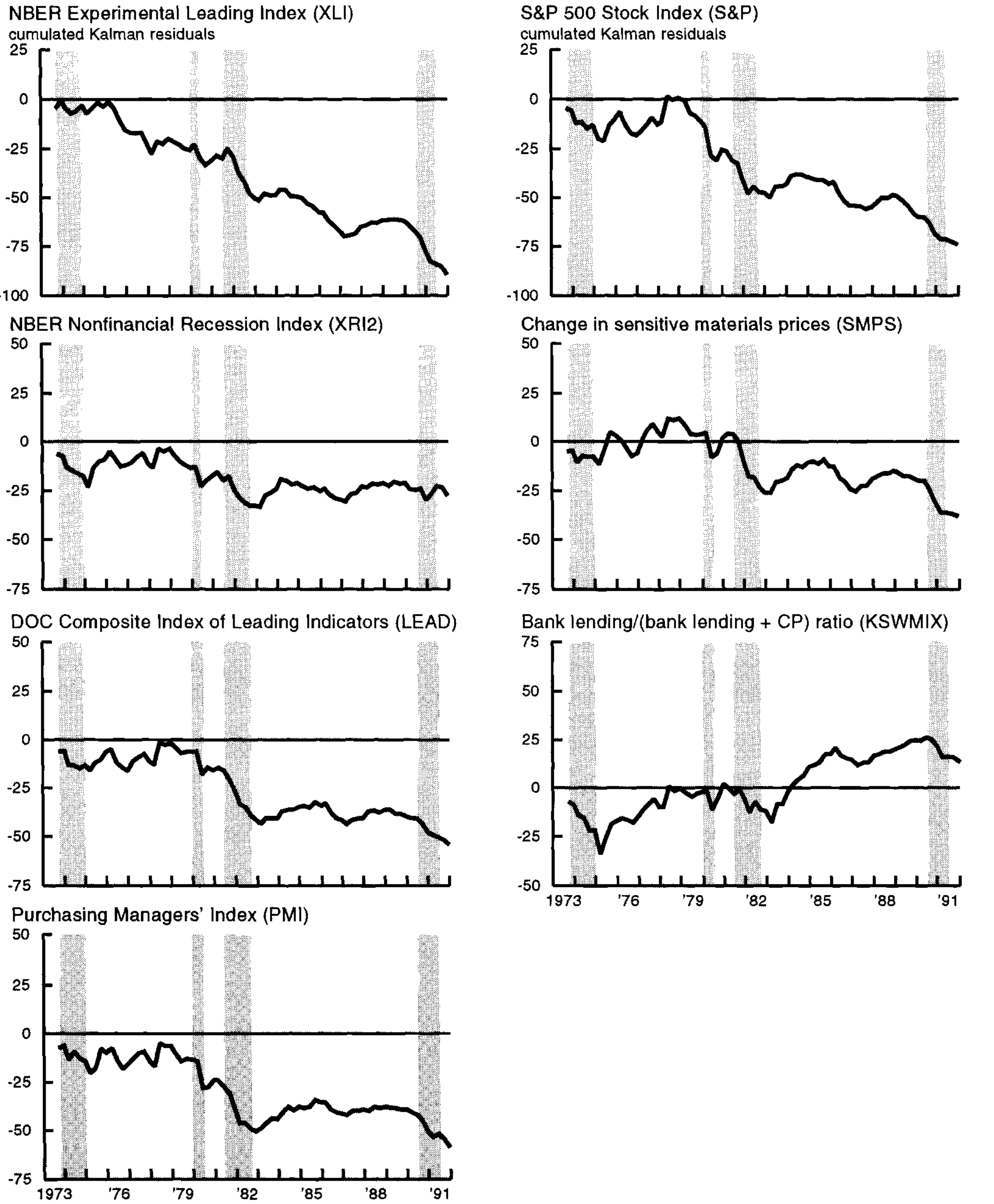
(Probability value for null hypothesis: X is encompassed by Y)

Y / X	XLI	XRI2	LEAD	PMI	S&P	SMPS	KSWMIX	Maximum P-value
Real GDP (1 quarter)								
XLI	n.a.	—	—	—	—	—	—	0.001
XRI2	—	n.a.	—	—	—	—	—	0.012
LEAD	—	—	n.a.	—	—	—	—	0.030
PMI	0.300	0.195	0.889	n.a.	—	—	—	0.889
S&P	0.754	0.334	0.619	0.100	n.a.	0.090	—	0.754
SMPS	0.598	0.114	0.923	0.127	—	n.a.	—	0.923
KSWMIX	0.061	—	0.088	—	—	—	n.a.	0.088
Real GDP (2 quarters)								
XLI	n.a.	—	—	—	—	—	—	0.000
XRI2	0.370	n.a.	—	—	—	—	—	0.370
LEAD	0.761	—	n.a.	—	—	—	—	0.761
PMI	0.609	0.603	0.314	n.a.	0.065	0.197	0.143	0.609
S&P	0.897	0.211	0.861	—	n.a.	0.097	0.064	0.897
SMPS	0.644	0.305	0.728	0.162	0.160	n.a.	0.179	0.728
KSWMIX	0.087	—	0.060	—	—	—	n.a.	0.087
Real GDP (4 quarters)								
XLI	n.a.	—	—	—	—	—	—	0.000
XRI2	0.939	n.a.	0.690	0.282	—	0.310	0.113	0.939
LEAD	0.420	—	n.a.	0.076	—	0.087	—	0.420
PMI	0.903	0.244	0.829	n.a.	0.121	0.636	0.181	0.903
S&P	0.616	0.104	0.748	0.303	n.a.	0.329	0.142	0.748
SMPS	0.377	0.055	0.153	—	—	n.a.	0.090	0.377
KSWMIX	0.166	0.056	0.056	0.113	—	0.196	n.a.	0.196

NOTES: Values less than or equal to 0.05 are marked with a dash. Sample period is January 1963 - December 1991, quarterly data.

FIGURE 4.1

Composite indicators

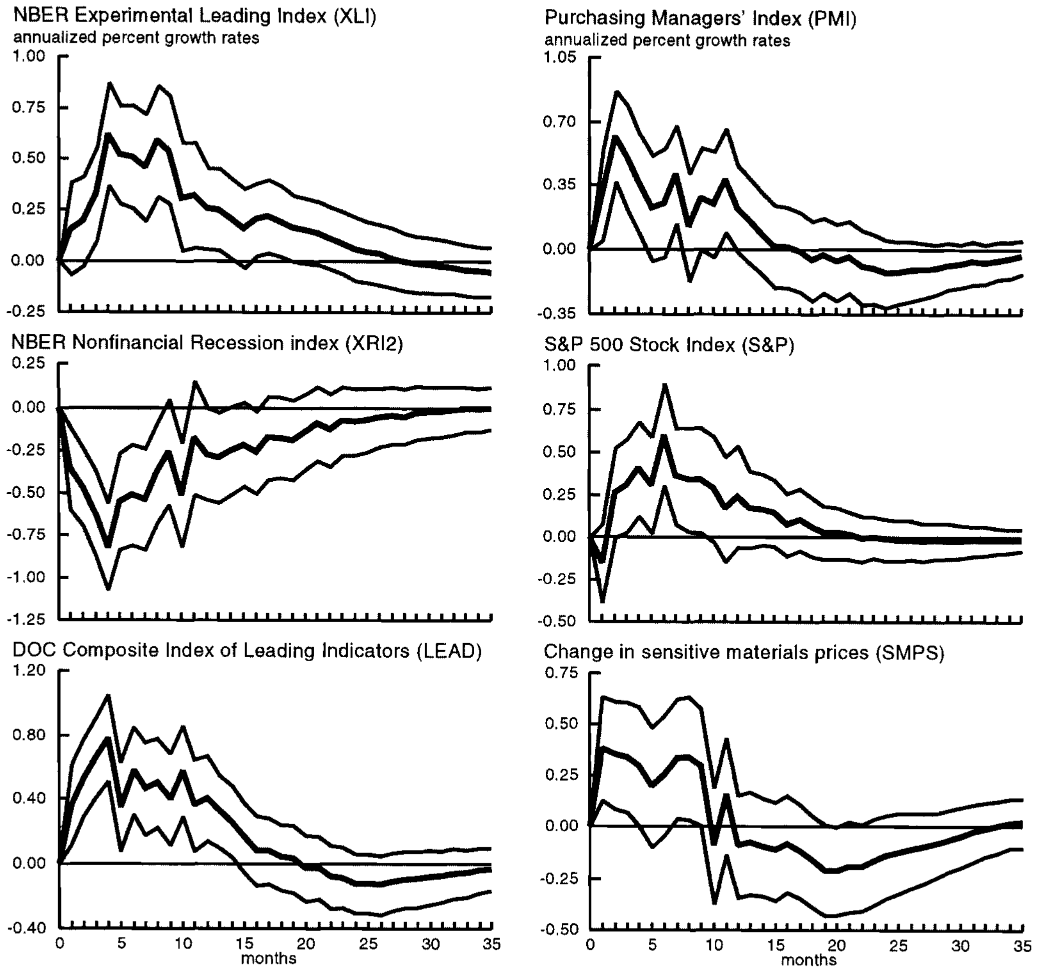


response path of employment to changes in PMI and SMPS shows dramatic jumps in forecasted growth rates within the first two months. Employment growth then steadily falls in PMI while it flattens in SMPS. The S&P graph shows a path similar to that depicted in the PMI graph, except for a rapid drop in the first month. It is interesting to note that all of these dynamic response paths are virtually insignificant at the one year mark, although the initial impact on real economic activity is fairly strong and well defined. Finally, as a group, these indicators seem to hold a lot of information about short run changes in economic activity, with most of that information centered at the three to nine month horizon.

The encompassing results in Table 4.5 show that XLI strongly dominates this entire

FIGURE 4.2

Dynamic response of employment to composite indicators



family of indicators, especially at the two and four quarter forecast horizons. At the one quarter horizon, both LEAD and XRI2 are not encompassed by any of the other indicators. These results are not surprising in light of the statistical results discussed earlier and the fact that XLI was designed to provide the "best" forecast of economic activity at a six month horizon, using virtually all of the macroeconomic data available.

5. Mixing models for real GDP

This section analyzes those indicators drawn from the previous sections that contain independent information and did well in the out-of-sample Kalman rankings. The indicators are subjected to another round of encompassing tests and rankings. Finally, the usefulness of these final indicators is assessed in the context of a time varying forecast mixing model.

Table 5.1 presents the Kalman forecast RMSEs for the one, two, and four quarter horizon forecasts of real GDP. For the one quarter horizon the best indicators are the NBER composite indicators (XLI and XRI2), and the Department of Commerce Composite Index of Leading Indicators (LEAD). The spreads and real M2 (M2R) do the worst at this short horizon, but all of the remaining indicators do contribute information beyond the own past history of GDP (NONE). At the two quarter horizon, the best indicator is the NBER Experimental Leading Index (XLI) with the 12 month Treasury bill/federal funds spread (TB12FF) coming in a distant second: XLI is 14 percent more accurate than TB12FF. This is not surprising since XLI was constructed by Stock and Watson to produce the "best" fore-

TABLE 5.1

Kalman residuals for surviving indicators

	Real GDP					
	1 quarter		2 quarters		4 quarters	
Indicator	RMSE	Rank	RMSE	Rank	RMSE	Rank
EURO3	3.622	4	2.754	3	n.a.	n.a.
FF	n.a.	n.a.	n.a.	n.a.	2.160	2
M2R	3.674	6	2.844	5	2.219	4
CP6TB6	3.656	5	2.760	4	n.a.	n.a.
TB12FF	3.753	7	2.751	2	2.002	1
CM10FF	n.a.	n.a.	n.a.	n.a.	2.161	3
XLI	3.246	1	2.376	1	2.392	5
XRI2	3.427	3	n.a.	n.a.	n.a.	n.a.
LEAD	3.307	2	n.a.	n.a.	n.a.	n.a.
NONE	4.052	8	3.369	6	2.799	6

NOTES: n.a.: The indicator is not an initial survivor at this forecast horizon.
Sample period is July 1973 - December 1991, quarterly data.

cast of the growth in economic activity over the six month horizon considered here. Turning to the four quarter horizon, it seems surprising that XLI comes in last after TB12FF, the federal funds rate (FF), the 10 year Treasury bond/ federal funds spread (CM10FF), and M2R. This demonstrates again that the choice of economic indicators depends critically upon the horizon being forecasted: at the four quarter growth horizon, a different collection of interest rate spreads than the ones selected by Stock and Watson is useful.

New encompassing results are displayed in Table 5.2. At this point, the purpose of these tests is to narrow the list of indicators in a structured manner. However, a rigid adherence to a statistical significance level is not maintained if an indicator is relatively useful and of independent interest. At the one quarter horizon, XLI, XRI2, and LEAD are each undominated and together sufficient. The two quarter horizon is more interesting. Three indicators are clearly necessary. XLI is undominated, and TB12FF is undominated at the 10 percent level. The 3 month eurodollar rate (EURO3) is not covered by these two indicators, and it is not dominated at the 11 percent significance level. M2R is also included in this final cut for two reasons: it is only covered by XLI at the 14 percent significance level and it is of inherent interest as the best monetary aggregate considered here. Finally, notice that the 6 month commercial paper/6 month Treasury bill spread (CP6TB6) did not make the final list at the two quarter forecast horizon, but it is a component of XLI.

At the four quarter horizon, three indicators are undominated: FF, M2R, and TB12FF. The NBER Experimental Leading Index (XLI) does not contain independent in formation beyond these indicators. CM10FF is included in the final list for three reasons: it is undominated at the 15 percent significance level, it covers the NBER Experimental Leading Index better than the shorter end of the term structure (TB12FF), and it is interesting to include a long term spread at this horizon since Stock and Watson found a long term spread useful at the two quarter horizon.

The next step is to combine these forecasts into a forecasting model (for each horizon) which allows the weights on the indicators to vary over time depending upon their recent performance. Essentially we would like the model to take the following form:

(3) $F_t = \phi_{1t} for(A)_t + \phi_{2t} for(B)_t + \phi_{3t} for(C)_t$;

where *for(A)* represents a forecast based upon indicator *A* and F_t is the combined forecast. The weights ϕ_{it} should be nonnegative and sum to one: in this case, the indicator's weight is a direct measure of its importance for the forecast. When the weights vary over time according to their forecast accuracy, the time path of the weights provide a direct measure of the indicators' reliability over time. We implement this model in the following way. Let ε_{it}^2 be the sum of (recent) squared forecast errors based upon indicator *i*'s model. In this paper, we take "recent" to be one year of known forecast errors (4 quarters). Let $avg_t(\varepsilon_{it}^2)$ be the average of the ε_{it}^2s at time *t* and μ_i be the average of $\varepsilon_{it}^2 - avg_t(\varepsilon_{it}^2)$ over time. Then ϕ_{it} is defined to be:

(4) $\phi_{it} = \alpha_i - \beta_i (\varepsilon_{it}^2 - avg_t(\varepsilon_{it}^2) - \mu_i), \quad \alpha_i, \beta_i \geq 0$;

TABLE 5.2

Mixed multiperiod encompassing tests

(Probability value for null hypothesis: X is encompassed by Y)

Y → / X ↓	EURO3	FF	M2R	CP6TB6	TB12FF	CM10FF	XLI	XRI2	LEAD	Maximum P-Value
Real GDP (1 quarter)										
EURO3	n.a.	0.100	—	—	—	—	0.107	—	—	0.107
FF	0.958	n.a.	—	—	0.067	—	0.144	—	—	0.958
M2R	—	—	n.a.	—	—	—	0.168	—	0.055	0.168
CP6TB6	—	—	—	n.a.	—	—	0.288	—	—	0.288
TB12FF	0.186	0.193	—	—	n.a.	—	0.453	—	—	0.453
CM10FF	0.168	0.098	—	—	0.260	n.a.	0.809	—	—	0.809
XLI	—	—	—	—	—	—	n.a.	—	—	0.001
XRI2	—	—	—	—	—	—	—	n.a.	—	0.012
LEAD	—	—	—	—	—	—	—	—	n.a.	0.030
Real GDP (2 quarters)										
EURO3	n.a.	0.110	—	—	—	—	—	—	—	0.110
FF	0.868	n.a.	—	—	0.161	—	—	—	—	0.868
M2R	—	—	n.a.	—	—	—	0.139	—	—	0.139
CP6TB6	—	—	—	n.a.	—	—	0.304	—	—	0.304
TB12FF	0.064	0.082	—	—	n.a.	—	0.062	—	—	0.082
CM10FF	0.076	—	—	—	0.228	n.a.	0.514	—	—	0.514
XLI	—	—	—	—	—	—	n.a.	—	—	0.000
XRI2	—	—	0.066	—	—	—	0.370	n.a.	—	0.370
LEAD	—	—	0.230	0.088	—	—	0.761	—	n.a.	0.761
Real GDP (4 quarters)										
EURO3	n.a.	0.609	—	—	—	—	—	—	—	0.609
FF	—	n.a.	—	—	—	—	—	—	—	0.023
M2R	—	—	n.a.	—	—	—	—	—	—	0.007
CP6TB6	0.270	0.327	0.420	n.a.	0.850	0.779	0.401	—	—	0.850
TB12FF	—	—	—	—	n.a.	—	—	—	—	0.011
CM10FF	—	—	—	—	0.147	n.a.	—	—	—	0.147
XLI	—	—	0.105	—	0.157	0.298	n.a.	—	—	0.298
XRI2	0.791	0.817	0.959	0.364	0.839	0.711	0.939	n.a.	0.690	0.959
LEAD	0.102	0.122	0.960	—	0.240	0.300	0.420	—	n.a.	0.960

NOTES: Values less than or equal to 0.05 are marked with a dash. Sample period is January 1963 - December 1991, quarterly data.

where the parameters α and β can be estimated by a linear regression model if the nonnegativity constraints are ignored, or nonlinear methods if the constraints are imposed. Since $\varepsilon_{it}^2 - avg_t(\varepsilon_{it}^2) - \mu_i$ is mean zero by construction, the time variation due to the βs nets out to zero over time. Consequently, the α estimates represent the average weight associated with each indicator forecast. However, over short periods of time when an indicator's forecast misbehaves, its errors ε_{it}^2 will be larger than the average errors; this will lead to the indicator's forecast receiving a temporarily smaller weight.

Table 5.3 displays the estimated α weights for these models. The one quarter results indicate that XLI is the most reliable, having an average weight of .533 in the combined forecast. The other indices (XRI2 and LEAD) received about equal shares of the remaining weight. The βs in this case are estimated to be zero; that is, there is no significant contribution to the forecast accuracy by allowing the weights to vary over time.

The two quarter results are more interesting. As was expected from the encompassing results, XLI receives the bulk of the weight in the final forecast (62 percent). This agrees with the analysis of Stock and Watson who constructed the NBER Experimental Leading In-

TABLE 5.3

Relative weights in mixing regressions

Indicator	Real GDP 1 quarter	2 quarters	4 quarters
EURO3	*	0.093 (0.260)	n.a.
FF	n.a.	n.a.	0.105 (0.209)
M2R	*	0.187 (0.227)	0.414 (0.178)
CP6TB6	*	*	n.a.
TB12FF	*	0.103 (0.238)	0.368 (0.259)
CM10FF	n.a.	n.a.	0.114 (0.212)
XLI	0.533 (0.174)	0.617 (0.197)	*
XRI2	0.214 (0.155)	n.a.	n.a.
LEAD	0.253 (0.206)	n.a.	n.a.

NOTES: Numbers in parenthesis are standard errors.
n.a.: The indicator is not an initial survivor at this forecast horizon.
(*): The indicator is encompassed by other indicators at this horizon.

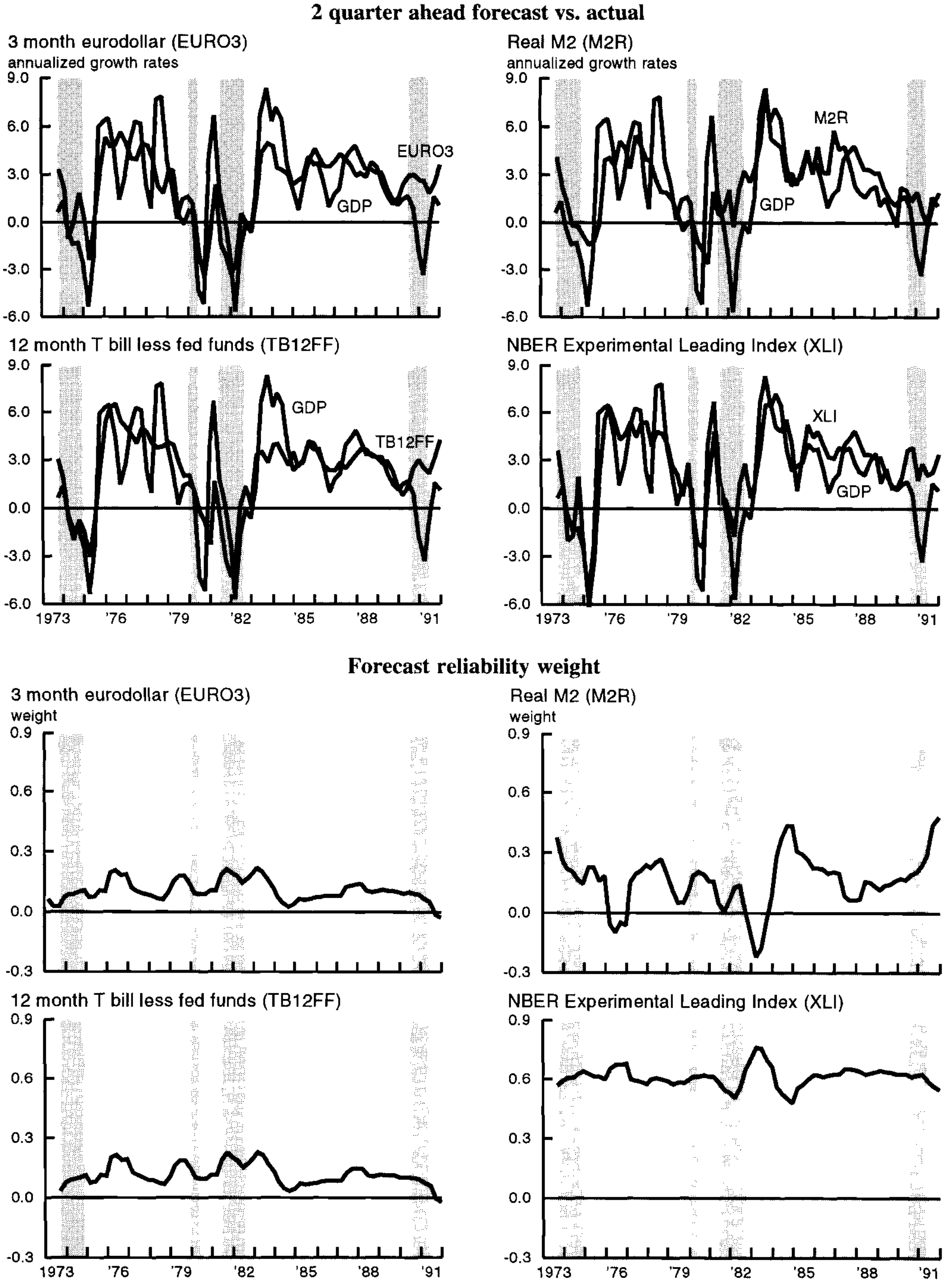
dex explicitly for its ability to forecast at this two quarter horizon. We do find that M2R receives a substantial weight (19 percent), while the TB12FF spread is at 10 percent and EURO3 is at 9 percent. Figure 5.1 graphs the time path of the ϕ weights for these four indicators, as well as the two quarter GDP forecast and actual. Notice first that the NBER Experimental Leading Index forecasts have been quite reliable, only once dropping below a 50 percent weight in the combined forecast. M2R, however, has varied dramatically in its usefulness, going negative on two occasions: in 1976 and immediately following the 1981-82 recession. During that recession, M2R did not forecast negative growth at any time (although it did in the 1980 recession), whereas EURO3, TB12FF, and XLI did forecast negative growth during some portion of this recession.[13] This poor performance is captured in the time varying model by decreasing the weight on the M2R forecast temporarily until it begins to improve. On the other hand, during the most recent recession M2R has gone above a 50 percent weight (keep in mind that the average weight for M2R is .19). During this time, M2R has grown only slowly and this led to a forecast of slow growth during 1991 (see Figure 5.1). At this same time, EURO3, TB12FF, and XLI signalled substantially higher growth than was realized. Each of these indicators is currently receiving less than its average weight. Consequently, the time varying mixing model finds that M2R has been an unusually useful indicator during the recent recession, despite its generally erratic performance at this horizon versus its relative failure at the twelve month horizon.

By contrast the four quarter horizon results in Figure 5.2 appear to be a picture of stability. M2R and TB12FF receive the largest unconditional weights, 41 percent and 37 percent respectively. FF and CM10FF receive considerably less (around 10 percent each). The graphs of the time varying weights indicate that, at this horizon, M2R and TB12FF have been reasonably reliable indicators, always staying near their unconditional weight. On the other hand, CM10FF has been extremely unreliable, going to zero or negative in 1987-88 and during the recent recession.

The contrast between the dominance of XLI at the two quarter forecast horizon and its submissiveness at the four quarter horizon demonstrates strongly the need for a different set of indicators for each forecast horizon. The usefulness of TB12FF and M2R for forecasting real GDP at the one year horizon indicates that a different index would be constructed if this forecast horizon was the relevant objective. A note on standard errors is in order. Examination of Table 5.3 indicates that the standard errors associated with the parameters of these mixing models are fairly large. This is not surprising in light of the high degree of collinearity that would be expected of a set of reasonably successful forecasts. In fact, it is typically the case that only the strongest indicator at a given horizon is statistically significant. All this is saying is that the relative weights among successful indicators are subject to substantial uncertainty and that the marginal information after the first one or two indicators quickly drops toward 0. Nevertheless, the point estimates and time paths of these relative weights provide a useful bench mark, even though the precision with which they are estimated would not change strongly held prior beliefs.

FIGURE 5.1

Mixing results

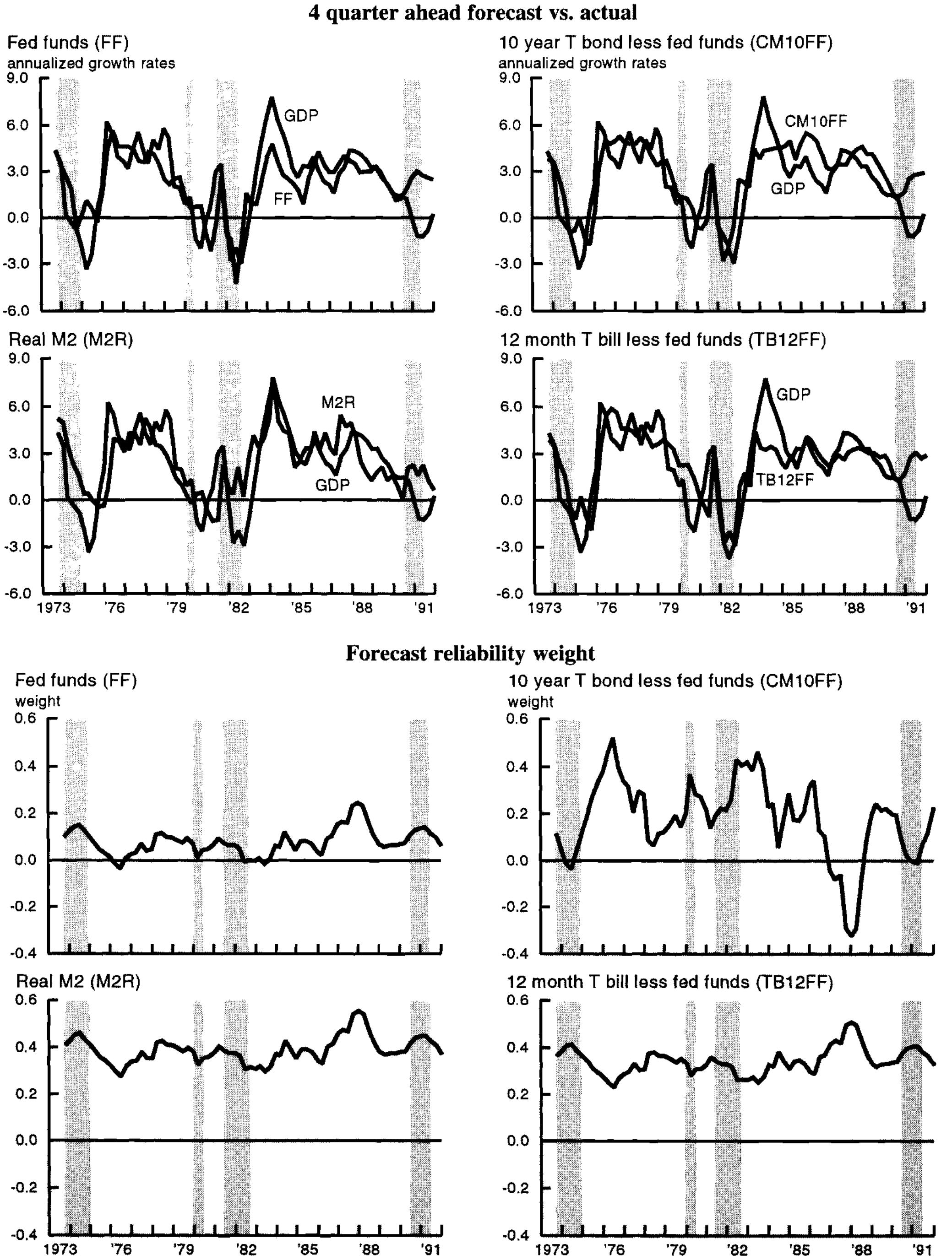


Conclusion

Four things became clear as the preceding analysis developed. First, the forecast horizon is an essential aspect of choosing and evaluating indicators. Second, substantial information resides in the term and private/public spreads and both of these seemingly very different types of spreads seem to include common as well as independent information. Third, while composite indicators may be extremely useful,

FIGURE 5.2

Mixing results



they are only as good as their design allows. The NBER Experimental Leading Index does very well at precisely what it was designed for, that is, forecasting economic activity at a six month horizon. Its usefulness beyond this horizon is far more limited than prior analysis would have suggested. Fourth, the analysis also suggests that the type of general purpose target variable that the old monetary targeting literature sought probably does not exist, at least in terms of real economic activity. Policymakers will continue to need to mix information according to their current focus. Mixing models of the sort used in this article are meant to be preliminary work in this regard.

FOOTNOTES

[1]The NBER Experimental Leading Index (XLI) developed by James Stock and Mark Watson is a clear exception, since it was created as a single "best" indicator of economic activity [see Stock and Watson, (1989b)].

[2]The following examples illustrate the notation we will use in the Methodology section to indicate different classes of tables: Table _.1 refers to the first table in each family of indicators, Table _.2 refers to the second table in each family, and so forth.

[3]It should be noted that these are not iterated VAR forecasts, rather, the forecast parameters are chosen to maximize performance at the forecast horizon specified. This can be thought of either as a state space estimation minimizing the $t+k$ forecast variance or as a simple OLS regression with the $t+k$ growth rate as the dependent variable. This avoids any problem that might result from an indicator that performs poorly at high frequencies interfering with longer frequency forecasting.

[4]The standard deviation measure used is the one from a bivariate VAR for the indicator and the measure of economic activity. This is used to approximate the average size of the movement in the indicator series.

[5]This is basically the same as an impulse response function except that the identifying assumption is not derived from a specific decomposition of the error matrix, but from the assumed path of the actual series, that is, the indicator changes given the level of current activity. This is arithmetically equivalent to an impulse response function using a Choleski decomposition with the indicator ordered last.

[6]The monetary base is the sum of reserve balances at the Federal Reserve Banks and currency in circulation.

[7]L is the broadest monetary aggregate, consisting of M3 plus the nonbank public holdings of U.S. savings bonds, short term Treasury securities, commercial paper, and bankers' acceptances, net of money market mutual fund holdings of these assets.

[8]These are the only commonly used spreads available for the entire sample period.

[9]We used the 10 year Treasury constant maturity bond rate because the 7 year bond rate, which might be preferred, is not available for the entire sample period.

[10]The NBER Nonfinancial Experimental Recession Index, which estimates the probability that the economy will be in a recession six months later, is based on a set of nonfinancial leading indicators. (See NBER Press Release, January 30, 1991.)

[11]SMPS is calculated as the quarterly average of the monthly changes in sensitive materials prices, smoothed. The sources for the monthly data are: U.S. Department of Commerce, U.S. Department of Labor, and the Commodity Research Bureau, Inc.

[12]The dynamic response graph for KSWMIX is not shown because data on the mix are available only on a quarterly basis, while employment data are monthly.

[13]It is useful to remember that the primary components of the NBER Experimental Leading Index are the 6 month commercial paper/6 month Treasury bill spread and the 10 year Treasury bond/1 year Treasury bond spread. So it should not be surprising that the NBER Experimental Leading Index misbehaved during this period when the 3 month eurodollar rate and the 12 month Treasury bill/ federal funds spread also misbehaved.

REFERENCES

Bernanke, Ben S., "On the predictive power of interest rates and interest rate spreads," *New England Economic Review*, November-December, 1990, pp. 51-68.

Chong, Y. and D. Hendry, "Econometric evaluation of linear macro-economic models," *Review of Economic Studies*, 53, 1986, pp. 671-690.

Estrella, A. and G. Hardouvelis, "The term structure as a predictor of real economic activity," *Journal of Finance*, 46, 1991, pp. 555-576.

Friedman, B. and K. Kuttner, "Why does the paper-bill spread predict real economic activity?" forthcoming in James H. Stock and Mark W. Watson eds., New Research in Business Cycles, Indicators and Forecasting, University of Chicago Press and the NBER, 1992.

Kashyap, A., J. Stein and D. Wilcox, "Monetary policy and credit conditions: evidence from the composition of external finance," Federal Reserve Board, Working Paper No. 154, 1991.

Laurent, Robert D., "An interest rate-based indicator of monetary policy," *Economic Perspectives*, Federal Reserve Bank of Chicago, January/February, 1988, pp. 3-14.

National Bureau of Economic Research, Press release, January 30, 1991.

Sims, Christopher A., "Interpreting the macroeconomic time series facts: the effects of monetary policy," manuscript, 1991.

Stock, J. and M. Watson, "Interpreting the evidence on money-income causality," *Journal of Econometrics*, Vol. 40, 1989a, pp. 161-182.

__________, "New indexes of coincident and leading economic indicators," in *NBER Macroeconomics Annual*, edited by O. Blanchard and S. Fischer, The MIT Press, 1989b, pp. 351-409

Strongin, Steven, "Macroeconomic models and the term structure of interest rates," Federal Reserve Bank of Chicago, Working Paper No. 90-14, 1990.

__________, "The identification of monetary policy disturbances: explaining the liquidity puzzle," Federal Reserve Bank of Chicago, Working Paper No. 91-24, 1991.