Kiel
Working Papers

Kiel Institute
for the World Economy

PREDICTING GROWTH RATES AND RECESSIONS

ASSESSING U.S. LEADING INDICATORS UNDER REAL-TIME CONDITIONS

by Jonas Dovern and Christina Ziegler

No. 1397 | January 2008

Web: www.ifw-kiel.de

Kiel Institute for the World Economy, Düsternbrooker Weg 120, 24105 Kiel, Germany

Kiel Working Paper No. 1397 | January 2008

PREDICTING GROWTH RATES AND RECESSIONS
ASSESSING U.S. LEADING INDICATORS UNDER REAL-TIME CONDITIONS

Jonas Dovern and Christina Ziegler

Abstract: In this paper we analyze the power of various indicators to predict growth rates of aggregate production using real-time data. In addition, we assess their ability to predict turning points of the economy. We consider four groups of indicators: survey data, composite indicators, real economic indicators, and financial data. Almost all indicators are found to improve short-run growth forecasts whereas the results for four-quarter-ahead growth forecasts and the prediction of recession probabilities in general are mixed. We can confirm the result that an indicator suited to improve growth forecasts does not necessarily help to produce more accurate recession forecasts. Only composite leading indicators perform generally well in both forecasting exercises.

Keywords: leading indicators, forecasting, recessions

JEL classification: C25,C32,E32,E37

Jonas Dovern
Kiel Institute for the World Economy
24100 Kiel, Germany
Telephone: +49-(0)431-8814-266
E-mail: Jonas.Dovern@ifw-kiel.de

Christina Ziegler
University of Leipzig
04109 Leipzig, Germany
Telephone: +49-(0)341-9733565
E-mail: ziegler@wifa.uni-leipzig.de

The responsibility for the contents of the working papers rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular working paper about results or caveats before referring to, or quoting, a paper. Any comments on working papers should be sent directly to the author.

Coverphoto: uni_com on photocase.com

PREDICTING GROWTH RATES AND RECESSIONS
ASSESSING U.S. LEADING INDICATORS UNDER REAL-TIME CONDITIONS

JONAS DOVERN AND CHRISTINA ZIEGLER

January 25, 2008

Abstract. In this paper we analyze the power of various indicators to predict growth rates of aggregate production using real-time data. In addition, we assess their ability to predict turning points of the economy. We consider four groups of indicators: survey data, composite indicators, real economic indicators, and financial data. Almost all indicators are found to improve short-run growth forecasts whereas the results for four quarter ahead growth forecasts and the prediction of recession probabilities in general is mixed. We can confirm the result that an indicator suited to improve growth forecasts does not necessarily help to produce more accurate recession forecasts. Only composite leading indicators perform generally well in both forecasting exercises.

Keywords: leading indicators, forecasting, recessions

JEL Classification: C25,C32,E32,E37

1. Introduction

The analysis and forecasting of business cycle fluctuation has a long tradition in economic research. In the forecasting context, the objective of the econometrician is to find a model that produces the best possible forecasts for a measure of economic activity. In most cases the target variable is the real growth rate of the gross domestic product (GDP). There are, however, also situations in which one might be interested in other target variables. Currently for instance, it is hotly debated whether the US economy will slide (or has already slid) into a recession following the turbulences on the US housing market and the crisis on international credit markets. Thus forecasts of turning points in the economy are considered to be much harder to forecast than simply accelerations or decelerations of growth rates. In this study we present a rich comparison of different leading indicators for predicting the real growth rate of the US economy as well as the occurrence of recessions.

A fact that complicates predictions about the economic outlook is that business and policy agents cannot observe the current state of the economy due to publication lags of the official statistics. Information on the national accounts of the US is published with considerable time lag. The business cycle indicators which are considered in this study help policy makers monitoring and forecasting the state of economy as they are available much earlier than the national accounts data.

This problem of non-observability of the current stance of the economy is amplified by the fact that the first data releases constitute rough estimates only which are usually heavily revised when more information is collected and processed by the statistical offices. This has also implications for the ex-post analysis of forecasting models. Recently, it has

Jonas Dovern, The Kiel Institute for the World Economy (IfW), jonas.dovern@ifw-kiel.de. Christina Ziegler, University of Leipzig, ziegler@wifa.uni-leipzig.de. The second author worked on this paper during a research visit to the IfW. The views presented in this paper reflect the authors' opinion, and do not necessarily coincide with those of the IfW. We are grateful to Kai Carstensen, Robinson Kruse, Carsten-Patrick Meier, and Joachim Scheide for their helpful comments.

become more and more standard to use so called real-time data sets for studies like this one. The basic idea is to use exactly those historical data vintages for out-of-sample analysis which were known at each point of time in the past. This applies to the estimation of models (Diebold and Rudebusch, 1991, Swanson, 1996) as well as to the evaluation of their forecast performance (Fair and Shiller, 1990).[1] In this way, the researcher is able to replicate as close as possible the informational conditions, under which a forecast would have had to be made in the past.

Against this background, the paper contributes to the literature along three dimensions. First, we analyze the predictive power of a bunch of business cycle indicators under real-time conditions which has not been done for the US in the literature up to now. Second, we analyze those indicators using identical sample periods which makes the results across indicators easily comparable. This condition is not necessarily given when comparing results from different contributions to the literature that consider only single indicators (like e.g. the slope of the yield curve) or a small number of indicators (like e.g. only survey indicators). In our analysis we evaluate the forecasting performance of survey indicators, indicators that can be attributed to the real side of the economy, and financial indicators. Finally, we analyze growth and recession forecasts simultaneously which enables us to draw some conclusion about which indicator performs better or worse in the two different kinds of forecasting exercises. Existing research indicates that it is not necessarily true that an indicator which helps reducing prediction errors when forecasting growth rates performs also well when it comes to forecasting turning points of the economy (Fritsche and Kuzin, 2005).

The remainder of the paper is structured as follows. Section 2 discusses the empirical approaches that we use to assess the ability of indicators to forecast growth rates of real GDP and to predict recessions. Section 3 elaborates on which indicators we include in the analysis and state which data sources we use. Section 4 presents the results of the two empirical analysis. The paper is concluded by section 5.

2. Empirical Approach

2.1. **Predicting Growth.** Forecasting based on vector autoregressive (VAR) models (Sims, 1980) has become the standard tool for predicting economic activity during the last decade. In contrast to factor models (Geweke, 1977), VAR models do include a rather limited number of different time series. Since we are interested in the relative performance of various indicators in small scale models, we stick to the VAR approach for forecasting the growth rates of real GDP in this paper. Our paper is related to the extensive literature that assesses the forecasting properties of various leading indicators for the business cycle such as Stock and Watson (2003), Banerjee and Marcellino (2006), Duarte et al. (2005), Davis and Fagan (1997), Schumacher (2007), Camacho (2004), Forni et al. (2003), Breitung and Jagodzinski (2001) or Kholodilin and Siliverstovs (2006).

A main feature of VAR models is their a-theoretic nature and the fact that all variables are endogenous. In both respects those attributes are convenient for our purpose. On the one hand, we don't want to make strict structural assumptions on how exactly the real economy interacts with the analyzed indicators; on the other hand it is also clear that the indicators are not exogenous in our models but rather influenced by the economic development. VAR models are the multivariate extension of an AR model. The evolution

[1]Recent examples of this procedural method include among others Bernanke and Boivin (2003), Schumacher and Breitung (2006), or Altissimo et al. (2006).

of a vector of time series is modeled in such a way that the dynamic interdependencies between the variables are accounted for. In our case, a bi-variate VAR model with lag order p can be formally expressed as

$$(1) \qquad z_t = \sum_{i=1}^{p} A_i z_{t-i} + \epsilon_t \ ,$$

where $z_t = [G_t, I_t]'$ is the 2×1 vector of time series G_t (GDP growth) and I_t (indicator series). A_i denotes the 2×2 coefficient matrix for lag i. The vector of innovations, ϵ_t, is assumed to follow an iid multivariate normal distribution with zero mean and covariance Σ. The lag length p is usually chosen by some information criteria like the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) or may be chosen ad hoc. Throughout this paper, we will rely on the BIC.[2]

Based on the VAR model, one may conduct either one-step or multi-step forecasts. Multi-step forecasts can generated in two different ways. At the iterated forecasting approach in each forecasting step a one-period ahead model is used to iterated forward for h periods. In comparison, direct forecasts use a horizon-specific estimated model, where the dependent variable is the multi-period ahead value being forecasted (Marcellino et al., 2005). To limit the length of the paper, we concentrate in what follows on iterated forecasts only. The choice of the forecasting scheme potentially influences the forecast performance as well. Either a recursive or a rolling window forecasting scheme can be chosen to obtain different indicator forecasts.

A forecast based on a recursive scheme relies on an estimation period of increasing length, where the starting point remains fixed. In contrast, a rolling scheme relies on a fixed-length window which is shifted every period. In the recursive forecasting scheme the sample size is increased by one period and the model is re-estimated. For the rolling window scheme, the initial sample length is equal to that in the recursive scheme. When an additional period of data is added after the first forecast step, however, the first period of the initial estimation sample is deleted. Hence, the estimation sample size in the rolling scheme remains constant, whereas the estimation sample size in the recursive scheme increases every period. However, the number of forecasts that can be compared with the data is equal for both methods. It is unclear a priori which scheme yields more precise forecasts. On the one hand, the recursive scheme allows us to exploit more information since estimation tends to be based on a larger sample. On the other hand, if the relation between growth and indicators are subject to structural breaks, a rolling scheme should deliver better forecasts compared to a recursive scheme (Giacomini and White, 2006).

2.2. **Predicting Recessions.** Several approaches to predict turning points of economic activity can be found in the literature. Most commonly, discrete choice models in which recessions are coded as one regime and booms as the other are employed as e.g. in Estrella and Hardouvelis (1991) or Estrella and Mishkin (1998) among others. Following Hamilton (1989), however, more recently also Markov switching models have been used to model and forecast recession probabilities. Examples include Bandholz and Funke (2003), Kholodilin (2005), Bengoechea et al. (2006), or Chauvet and Hamilton (2006).

In this paper, we follow the first approach and estimate probit models, for which a binary series, say Y_t, capturing recession/expansion periods is regressed on lags of GDP growth and lags of the indicator. In addition, we do not exclude the possibility of a

[2]The BIC is given by $-2\,(LL/T) + (k\,ln(T))/T$, where LL denotes the log-likelihood, T is the number of observations, and the model includes k estimated parameters.

dynamic specification a priori - in the sense that we allow for the inclusion of lags of the binary time series as explanatory variables in the regression. As shown by Dueker (1997), this is the correct specification if there is information in the autocorrelation structure of the dependent variable. Formally, we estimate variants of the following model

$$Pr\left[Y_t = 1 | L(\beta) X_t, Y_{t-k}\right] = \Phi\left(\gamma_0 + L(\beta) I_t + L(\alpha) G_t + \delta Y_{t-k} + \epsilon_t\right) \tag{2}$$

where $L(\beta)$ and $L(\alpha)$ denote lag-polynomials of order p and q respectively, I_t is the leading indicator, and G_t denotes the growth rate of real GDP. Φ denotes the cumulative density function of the normal distribution. We assume that the ϵ_t are independently distributed. The fitted values of this equation, $\Phi(\hat{\gamma}_0 + L(\hat{\beta}) I_t + L(\hat{\alpha}) G_t + \hat{\delta} Y_{t-k})$, can be seen as the estimated probabilities for a recession conditional on the known lagged values of the indicator and GDP growth (and potentially the lagged values of the recession indicator).

The exact lag structure of the models are determined according to the following rule. First of all, the maximum lag orders for both the indicators and GDP growth are set to 5. For the lagged endogenous variable we initially allow for the choice between inclusion of the first or the second lag or no inclusion at all.[3] This leaves us with $5 \times 5 \times 3 = 75$ possible specifications. We have chosen to select the model which minimizes the BIC as the best model. Note that below we allow for different specifications at each point in time during the recursive out-of-sample analysis, i.e. for each point in time we re-select the optimal model based on the appropriate data vintage.

3. Choice of Indicators and Data

It has become more and more standard to use real-time data for research that concerns itself with the analysis of forecasting performance of econometric models or institutions. The basic idea is to use exactly those historical data vintages for out-of-sample analysis which were known at each point of time in the past. The use of data as published ex-post (after multiple revisions) would give a flawed picture of the information that has been available in the past. Especially for national account data like for instance aggregate output growth the magnitude of revisions can be substantial (Mork, 1987, Fukuda, 2007). Large revisions can be observed especially around the turning points of economic activity. Figure 1 demonstrates this issue. The bars show quarterly growth rates of real GDP as implied by the published output data around recessionary times.[4] It is evident that revisions to the initially published figures can be significant. Especially at the beginning of a recession, growth estimates are revised downwards in the majority of cases, i.e. initially the data do not reflect the contractionary momentum to its full extend (or in case of the 2001 recession not at all). As GDP growth is the main indicator to determine a recession, this implies that it is much harder to predict recessions using real-time data than it is based on heavily revised ex-post data.

In what follows, we use real-time data for all macroeconomic time series and for all indicators whenever those data are available. For the national account data we use the real-time data set provided by the Federal Reserve Bank of Philadelphia (2007).[5] It contains real-time vintages for the main macroeconomic time series starting in 1965Q4 up to

[3] In this we follow the argumentation by Duarte et al. (2005), who argue that "although at 3 months an NBER turning point date may not have officially announced, people have acquired enough other information to infer with a reasonable degree of accuracy whether the economy is in recession or not" (p.272).

[4] We refer to the official dating of recessions provided by the NBER at this point. They are indicated by the shaded areas in the plots.

[5] For more details see also Croushore and Stark (2001).

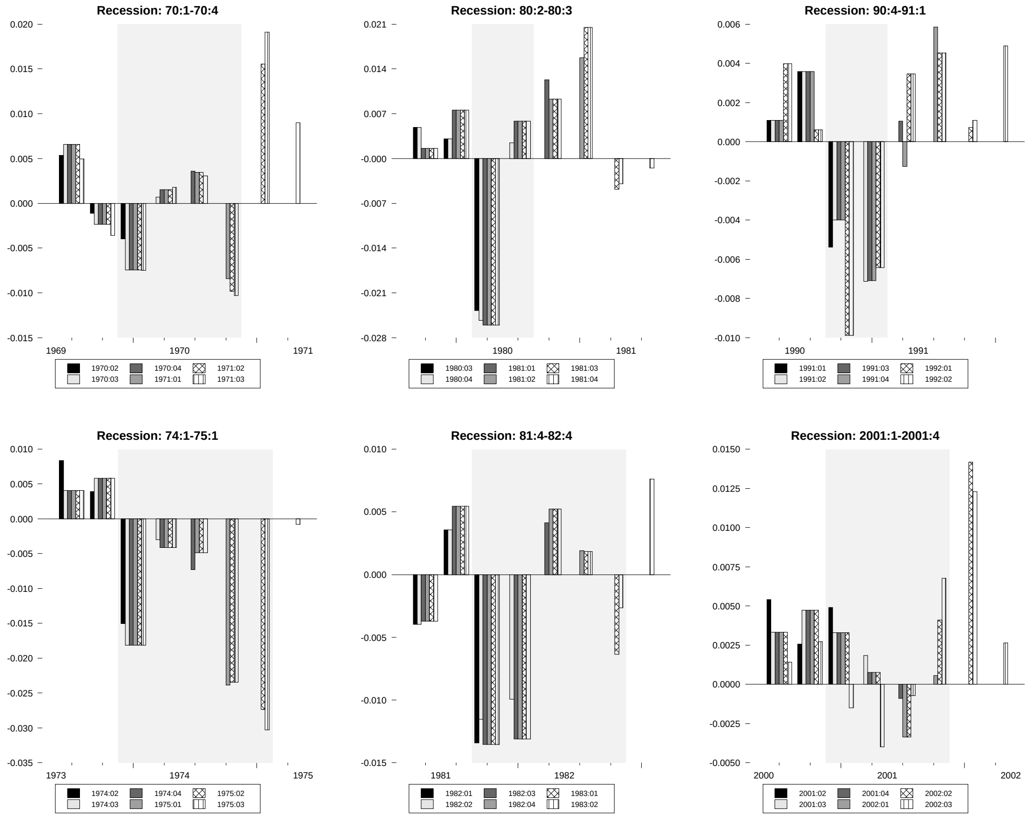

Figure 1. Real-Time Growth Estimates Around Recessions

the most recent release of the national account. For our out-of-sample analysis below we use the vintages from 1981Q3 to 2007Q2 for real output growth and the change of the real money supply.

The various indicators, which we test with respect to their ability to improve growth forecasts and predictions of turning points, can be grouped into four broad categories: survey indicators, composite indices (which combine different time series), measures of real economic activity (which are supposed to lead the aggregate business cycle), and financial indicators.[6] An overview about the considered indicators and the corresponding sources are given in Table 1. Figure 2 reveals that all indicators show some degree of co-movement with aggregate output growth. The plots show the indicators (dashed) together with the four-quarter growth rate of real GDP (solid). In what follows, we elaborate in more detail on the four groups of leading indicators.

First, survey indicators reflect beliefs about the future economic development as they are formed by different groups of economic agents. Presumingly, they are based on a very rich set of public information as well as multiple private information sets. Surveys among consumers are the most widely available kind of survey indicators (Curtin, 2007). For the US the most often considered indicators of this kind are the consumer confidence indices collected by the University of Michigan (*ICS*) and by the Conference Board (*CBCS*).

[6]Note that neither survey indicators nor financial indicators suffer from data revisions so that we do not face the real-time issue for those variables. Unfortunately, there do not yet exist real-time data sets for the different composite indices.

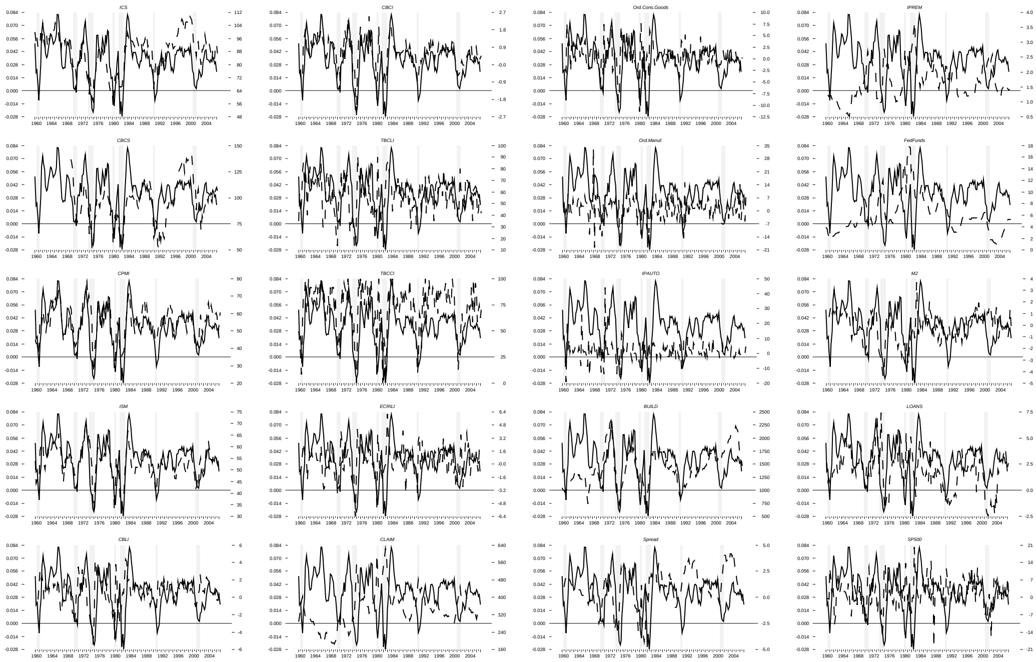

FIGURE 2. Indicators and Real GDP Growth

	Transf.	Correlation [7]	GC-Test	Source
Survey Indicators				
ICS	Level	0.452 (0)	11.7 (0.00)	University of Michigan
ICS	LogDiff	0.273 (0)	5.4 (0.02)	University of Michigan
CBCS	Level	0.357 (-2)	18.7 (0.00)	Conference Board
CPMI	Level	0.545 (-1)	2.3 (0.12)	Chicago Nat. Ass. of Purch. Managers
ISM	Level	0.584 (0)	28.8 (0.00)	Institute for Supply Management
Composite Indicators				
CBLI	LogDiff	0.586 (1)	74.2 (0.00)	The Conference Board
CBCI	LogDiff	0.764 (0)	27.0 (0.00)	The Conference Board
TBCLI	Level	0.534 (1)	77.4 (0.00)	The Conference Board
TBCCI	Level	0.639 (0)	21.4 (0.00)	The Conference Board
ECRILI	LogDiff	0.460 (1)	34.1 (0.00)	Economic Cycle Research Institute
Real Economic Indicators				
CLAIMS	Level	0.122 (8)	27.6 (0.00)	The Conference Board
OCONS	LogDiff	0.636 (0)	22.6 (0.00)	The Conference Board
OMANU	LogDiff	0.339 (0)	7.0 (0.01)	The Conference Board
IPAUTO	LogDiff	0.514 (0)	0.6 (0.45)	FED. St. Louis
BUILD	Level	0.274 (0)	42.4 (0.00)	Bureau of the Census
Financial Indicators				
SPREAD	Level	0.399 (2)	18.9 (0.00)	FED. St. Louis
IPREM	Level	0.423 (-3)	15.4 (0.00)	FED. St. Louis
FEDF	Level	0.364 (2)	17.5 (0.00)	FED. St. Louis
M2	LogDiff	0.402 (2)	23.0 (0.00)	The Conf. Board/B. of Econ. Anal.
LOANS	LogDiff	0.369 (-4)	1.1 (0.30)	FED. St. Louis
SP500	LogDiff	0.210 (0)	23.7 (0.00)	Standard & Poors'

TABLE 1. Indicators Used in the Analysis

The evidence on their predictive power is mixed, however. While some studies report significant predictive power of the *ICS* to forecast output growth or recessions (Matsuasaka and Sbordone, 1995, Howrey, 2001) other studies come to an opposite conclusion (Carroll et al., 1994). Other indices are based on surveys conducted among managers of business enterprizes. We include the index constructed by the Institute for Supply

[7]Number in parenthesis indicate the lead/lag for which the maximal correlation that is shown in the table is obtained. Positive numbers indicate that the series is leading GDP growth.

Management (*ISM*) as well as the index provided by the Chicago National Association of Purchasing Managers (*CPMI*).

Second, composite indices combine information from multivariate time series data sets. In this way they should draw a picture of the overall stance of the economy. Usually, composite *leading* indicators and composite *coincidence* indicators are distinguished. Whereas the latter aggregate information from time series that closely follow the business cycle, the former include information from time series which show a lead with respect to the business cycle. The indicators that receives most attention are the composite leading (*CBLI*) and coincidence (*CBCI*) indicators calculated by the Conference Board. Other indicators which we consider in this paper and that fall into this category are the less used leading diffusion (*TBCLI*) and coincidence diffusion (*TBCCI*) indicators provided by the Conference Board[8] and the leading indicator computed by the Economic Cycle Research Institute (*ECRILI*).

Third, also single time series that are leading the business cycle are used as indicators to predict business cycle fluctuations. Some of these series have an additional advantage by the fact that they are published more timely than other macroeconomic data. We consider the following indicators: Initial claims for unemployment benefits (*CLAIM*), new orders for consumption goods (*OCONS*), new orders in the manufacturing industry (*OMANU*), industrial production in the automotive sector (*IPAUTO*), and issued building permits to private households (*BUILD*).

Finally, agents on the financial markets base decisions on their beliefs about the future economic development. And, hence, prices should reflect those expectations under the assumption of optimizing agents. The financial indicator that is most often used to predict aggregate growth is the spread between the short-term interest rate and the yield on ten year treasury bonds (*SPREAD*) (Estrella and Hardouvelis, 1991, Stock and Watson, 1989). While there is some evidence that the predictive relationship between the yield curve and output growth shows a significant degree of instability (Stock and Watson, 2003, Giacomini and Rossi, 2006), the use of the yield curve in models with binary recession indicators seems to be more promising (Estrella et al., 2003). In addition we consider the Fed-Funds rate (*FEDF*) as an indicator. Recent evidence suggests that it might have superior predictive power compared to the slope of the yield curve (Ang et al., 2006). But also the change of stock prices should include some information about expectations on the economic outlook as the stock price reflects beliefs about future dividend payments or seen more broadly output (Stock and Watson, 2003). The existing empirical evidence, however, indicates that stock returns have little predictive power for output growth (Estrella and Mishkin, 1998, Campbell, 1999). We, nevertheless, include the percentage change of the S&P500 as one potential leading indicator in our sample (*S&P500*). As additional financial indicators we include a risk premium measured by the spread between BAA-rated corporate bonds with a maturity of ten years and the yield on ten year treasury bonds (*IPREM*) and the growth rate of real money supply as measured by M2 deflated with the GDP deflator (*M2*).

4. Results

4.1. **Preliminary Analysis.** We start this section with an in-sample analysis of the predictive power of the indicators. Even if good in-sample properties do not necessary

[8]The *TBCLI* (*TBCCI*) measures the proportion of the ten (four) components of the *CBLI* (*CBCI*) that contribute positively to the overall *CBLI* (*CBCI*).

imply good out-of-sample performance, an indicator is usually required to have satisfying in-sample qualities. Therefore, we employ two popular tools to investigate the in-sample quality in our study, namely a cross-correlation analysis and a test for Granger causality (Granger, 1969).

In the cross-correlation analysis we consider the cross-correlation coefficients for lags $k = -12, \dots, 0, \dots, 12$ which are given by the correlation coefficient between the reference series G_t and the indicator I_{t-k}. If the indicator is leading the reference series then the cross-correlation should be as high as possible for positive values of k. The results are shown in the third column of Table 1. The *CBCS*, *CPMI*, *IPREM*, and *LOANS* indicators are not leading in the sense that their maximal cross-correlation coefficient is associated with a negative value of k. On the contrary, other indicators have satisfying properties. Most of the indicators have their maximum correlation with the reference series at a lag of $k = 1$, which implies that those indicators are leading the reference series by one quarter. However, the indicator with the highest correlation (0.764) is *CBCI* at lag zero, while *CLAIMS* reaches its highest correlation at $k = 8$. Except of *LOANS* and *IPREM* the group of financial indicators seems to be most promising for improving growth forecasts based on this correlation analysis.

The causality test suggested by Granger (1969) can be used to statistically test the leading properties of an indicator in a more rigorous manner. The test is used in the VAR framework to analyze whether lagged values of the indicator are able to explain the variation of the reference series. The results in the fourth column of Table 1 show the test statistics together with the corresponding p-values[9]. The null hypothesis of no causality in Granger's sense can be rejected for nearly all indicators except *CPMI*, *IPAUTO*, and *LOANS*. This suggests that those three indicators have no explanatory power for future output growth. In general, however, the results are promising for the subsequent out-of-sample analysis.

4.2. **Predicting Growth.** We start our out-of-sample forecasting experiment with the analysis of growth forecasts under real-time conditions. Based on the VAR model explained in the previous chapter, we conduct a sequence of dynamic out-of-sample forecasts using different leading indicators. Due to the fact that formal tests for equal forecasting ability require the model specifications to be constant over time we choose $p = 2$ in all cases. This is also the lag order which is suggested by the BIC in the wide majority of cases when re-selecting the optimal lag length for each model recursively for each forecasting time.

For each VAR model, we produce one-quarter ahead as well as iterative four-quarter ahead forecasts. We apply the recursive as well as the rolling window forecasting scheme in our analysis. The initial estimation sample covers the first twenty years of our whole sample, from $1961Q1$ to $1981Q2$. Forecasts are computed with a forecast horizon of $h = 1$ and 4 starting with $1981Q3$ and $1982Q2$ respectively and moving until the end of the sample in $2007Q2$. In the recursive forecasting scheme the sample size is consecutively increased by one period and the model is re-estimated for each time period. For the rolling window scheme the initial initial estimation sample is equal to the one in the recursive scheme. In contrast to the latter scheme, however, the respective first observation of the estimation sample is deleted whenever an additional period is added to the end of the

[9]The lag length was set equal to 2. Granger causality tests, where the optimal lag length is chosen accordingly to AIC or BIC information criteria deliver similar results

	h=1		h=4	
	recursive	rolling	recursive	rolling
Survey Indicators				
ICS	1.02	0.92	1.33	1.18
CBCS	1.02	0.97	1.06	0.99
CPMI	0.95	0.93	1.03	1.03
ISM	0.97	0.99	1.00	1.00
Composite Indicators				
CBLI	0.85	0.81	0.90	0.90
CBCI	0.92	0.92	0.99	0.98
TBCLI	0.87	0.84	0.94	0.93
TBCCI	0.88	0.92	1.02	1.02
ECRILI	0.89	0.88	1.01	1.01
Real Economic Indicators				
CLAIMS	0.90	0.91	0.94	0.99
OCONS	0.96	0.95	1.00	0.99
OMANU	0.98	1.02	1.01	0.99
IPAUTO	1.01	1.03	1.02	1.01
BUILD	0.87	0.90	0.97	1.03
Financial Indicators				
SPREAD	1.23	1.04	1.26	1.05
IPREM	0.99	0.99	0.97	0.97
FEDF	1.32	1.19	1.64	1.46
M2	1.20	1.09	1.00	0.93
LOANS	1.04	0.99	1.17	1.07
SP500	0.98	0.92	1.27	1.23
AR	1.00	1.00	1.00	1.00

TABLE 2. Relative MSEs

sample. Hence, the size of the estimation sample in the rolling scheme remains constant (82 observations) whereas it increases every period in the recursive scheme.

For all models, we compute the mean square errors (MSE) relative to an univariate autoregressive (AR) benchmark model[10] as $\theta = MSE^{Ind}/MSE^{AR}$. Here MSE^{Ind} denotes the MSE of the respective indicator model and MSE^{AR} is the MSE of the benchmark model. If $\theta < 1$ the indicator forecasts show a superior forecasting performance than the benchmark model and vice versa. The results are given in Table 2. In the first two columns the relative $MSEs$ of the one-step forecasts are reported. Whereas the relative $MSEs$ in the first column are computed via the recursive forecasting scheme, the second column shows the results for the rolling window forecasting scheme. For the one-step forecasts all *Composite Indicators* perform better than the benchmark model. The predictive power within the other class of indicator groups remains mixed as well as for the recursive or the rolling window forecasting scheme. The *CBLI* indicator delivers the smallest relative MSE in each column; this indicator outperforms the benchmark model also in four quarter ahead predictions. In the class of *financial indicators* the *SP500* and the *IPREM* are the only indicators with a lower MSE than the benchmark model as well as in forecasting with a rolling and with a recursive forecasting scheme. When looking at the predictive power of the different indicator series at forecasting four quarters ahead, one sees that the prediction with the naive AR model frequently delivers lower $MSEs$ than the indicator forecasts do. However, the performance of the indicator forecasts seems to be quite superior at forecasting in the short run.

To assess the statistical significance of those differences in forecasting performance we use a test in the spirit of the test suggested by Diebold and Mariano (1995) (DM-test). It is, however, a well-known fact that the asymptotic results on which the original DM-test relies are not valid in cases where one model is nested by the other model. Therefore, we use an F-test suggested by Clark (1999) (with critical values for small samples provided by

[10]Analogous to our VAR-approach, the order of the autoregressive lag polynomial is set equal to two.

	recursive	rolling
Survey Indicators		
ICS	-1.99	8.71***
CBCS	-1.71	2.73**
CPMI	5.58***	7.24***
ISM	3.18**	1.34*
Composite Indicators		
CBLI	17.92***	24.24***
CBCI	9.11***	9.48***
TBCLI	16.24***	20.37***
TBCCI	14.28***	8.91***
ECRILI	13.58***	13.80***
Real Economic Indicators		
CLAIMS	11.07***	10.99***
OCONS	4.87***	4.98***
OMANU	2.56**	-1.78
IPAUTO	-0.78	-2.75
BUILD	15.01***	11.98***
Financial Indicators		
SPREAD	-19.17	-3.85
IPREM	1.78*	1.47**
FEDF	-24.90	-16.68
M2	-17.19	-9.012
LOANS	-4.44	0.49*
SP500	2.14**	9.03***
Critical Values		
10%	0.748	0.271
5%	1.791	1.398
1%	4.338	4.280

Notes: *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

Table 3. Tests for equal forecast accuracy

Clark and McCracken (2001)) to test for equal forecast accuracy for the one step ahead predictions.[11] In this setup, the null hypothesis of equal $MSEs$ is tested against the alternative that the richer model (which nests the other alternative) produces superior forecasts. Test statistics and critical values are given in Table 3. They show that those indicators with a relative $MSEs$ smaller than one (compare Table 2) produce indeed significantly better forecasts than the AR-benchmark model. For *IPREM* this is, however, only true at the 10% level for the recursive forecasting scheme; the same holds for *ISM* and *LOANS* when the rolling window scheme is applied.

Finally, we want to compare the predictive power of the different indicators pairwise with each other. To this end, we can make use of the DM-test. An analysis of outcomes from the pairwise DM-tests confirms the previous conclusions that were drawn from the comparison of relative $MSEs$. Now, we can make statements about the significance of pairwise differences in computed $MSEs$. Results are presented in Tables 4-7. A significant and negative value of the test statistic suggests that the model in the corresponding row outperforms the model in the corresponding column. A significant and positive value indicates that the column model performs better than the row model.

[11]The asymptotic distribution of tests for equal forecast accuracy for multi-step forecasts ($h > 1$) generally depend on the data generating process and critical values are not available.

h=1	ICS	CBCS	CPM	ISM	CBLI	CBCI	TBCLI	TBCCI	ECLIRI	CLAIMS	OCONS	OMANU	IPAUTO	BUILIP	SPREAD	IPREM	FEDF	;M2	LOANS
CBCS	-0.03																		
CPM	-0.67	-0.74																	
ISM	-0.39	-0.45	0.27																
CBLI	-1.23	-1.77**	-1.19	-1.40*															
CBCI	-0.94	-1.12	-0.41	-0.60	0.72														
TBCLI	-0.98	-1.41*	-0.83	-0.98	0.17	-0.49													
TBCCI	-1.21	-1.15	-0.51	-0.78	0.18	-0.41	0.08												
ECLIRI	-0.88	-1.26	-0.68	-0.92	0.37	-0.35	0.28	0.03											
CLAIMS	-0.90	-1.08	-0.46	-0.80	0.68	-0.19	0.40	0.19	0.20										
OCONS	-0.56	-0.62	0.11	-0.21	1.22	0.57	0.82	0.67	0.69	0.58									
OMANU	-0.43	-0.32	0.35	0.05	0.94	0.61	0.73	0.75	0.63	0.56	0.27								
IPAUTO	-0.14	-0.09	1.02	0.37	1.32*	1.16	1.04	1.06	0.93	0.85	0.75	0.61							
BUILIP	-1.12	-1.03	-0.65	-1.23	0.19	-0.39	0.06	-0.04	-0.08	-0.28	-0.78	-0.75	-1.07						
SPREAD	1.18	1.25	1.94**	1.88**	2.99***	2.04**	2.37***	1.90**	2.16**	2.72***	1.92**	1.61*	1.32*	2.34***					
IPREM	-0.33	-0.37	0.60	0.13	1.43*	0.80	1.25	0.68	1.29*	0.84	0.35	0.07	-0.28	0.86	-1.67**				
FEDF	1.41*	1.18	1.57	1.45*	1.68**	1.69**	1.54*	2.03**	1.51*	1.56*	1.50*	1.69**	1.54*	1.78**	0.36	1.32*			
M2	0.99	1.06	1.65**	1.16	2.02**	1.52*	2.05**	1.25	1.94**	1.59*	1.30*	1.20	1.12	1.48*	-0.13	1.59*	-0.40		
LOANS	0.19	0.17	0.98	0.60	1.46*	0.97	1.18	0.98	1.08	1.12	0.91	0.83	0.36	1.27	-1.42*	0.56	-1.20	-0.81	
SP500	-0.38	-0.27	0.28	0.09	0.86	0.54	0.71	0.87	0.65	0.54	0.21	0.04	-0.30	0.90	-1.36*	-0.03	-1.78**	-1.01	-0.49
Crit. Val.:																			
10%	1.28																		
5%	1.65																		
1%	2.32																		

Notes: *, **, and *** indicate significance levels of 10, 5, and 1%.

TABLE 4. Diebold-Mariano-Test (1995), recursive forecasting scheme

h=4	ICS	CBCS	CPM	ISM	CBLI	CBCI	TBCLI	TBCCI	ECLIRI	CLAIMS	OCONS	OMANU	IPAUTO	BUILIP	SPREAD	IPREM	FEDF	;M2	LOANS
CBCS	-0.93																		
CPM	-1.38*	-0.15																	
ISM	-1.53*	-0.41	-2.52***																
CBLI	-1.77**	-1.26	-2.08**	-1.53*															
CBCI	-1.81**	-0.51	-0.94	-0.13	1.28														
TBCLI	-1.70**	-0.92	-1.39*	-0.83	0.93	-0.79													
TBCCI	-2.01**	-0.25	-0.19	0.25	1.13	0.65	0.89												
ECLIRI	-1.90**	-0.32	-0.33	0.20	1.20	0.70	0.94	-0.35											
CLAIMS	-1.68**	-0.94	-2.28**	-1.41*	0.79	-0.87	-0.03	-0.85	-0.89										
OCONS	-2.00**	-0.40	-0.47	0.05	1.05	0.32	0.75	-1.27	-0.93	0.74									
OMANU	-2.07**	-0.33	-0.32	0.14	1.03	0.44	0.77	-0.94	-0.16	0.76	0.65								
IPAUTO	-2.02**	-0.28	-0.23	0.22	1.09	0.61	0.85	-0.32	0.21	0.83	1.35*	1.02							
BUILIP	-1.78**	-0.66	-1.44*	-0.54	1.26	-0.77	0.56	-0.78	-0.80	0.66	-0.57	-0.62	-0.73						
SPREAD	-0.25	1.10	1.36*	1.62*	2.89***	1.692**	2.72***	1.38*	1.49*	2.30**	1.54*	1.44*	1.40*	1.84**					
IPREM	-1.65**	-0.66	-1.46*	-0.63	1.11	-0.40	0.36	-0.55	-0.54	0.66	-0.40	-0.46	-0.53	-0.12	-1.88**				
FEDF	1.33*	1.59*	1.83**	1.92**	2.13**	2.11**	2.20**	2.27**	2.23**	2.10**	2.27**	2.29**	2.27**	2.08**	1.18	1.98**			
M2	-1.57*	-0.33	-0.53	0.04	1.39*	0.12	0.81	-0.19	-0.13	0.80	-0.01	-0.09	-0.16	0.37	-1.78**	0.53	-1.93**		
LOANS	-0.82	0.57	1.41*	1.95**	2.71***	1.91**	2.42***	1.39*	1.55*	2.34***	1.63*	1.48*	1.41*	2.10**	-0.61	2.04**	-1.54*	1.84**	
SP500	-0.69	0.83	1.14	1.30*	1.63*	1.63*	1.60	1.88**	1.78**	1.50*	1.87**	1.92**	1.90**	1.60*	0.03	1.40*	-1.93**	1.29*	0.51
Crit. Val.:																			
10	1.28																		
5	1.65																		
1	2.32																		

Notes: *, **, and *** indicate significance levels of 10, 5, and 1%.

TABLE 5. Diebold-Mariano-Test (1995), recursive forecasting scheme

h=1	ICS	CBCS	CPM	ISM	CBLI	CBCI	TBCLI	TBCCI	ECLIRI	CLAIMS	OCONS	OMANU	IPAUTO	BUILIP	SPREAD	IPREM	FEDF	;M2	LOANS
CBCS	0.66																		
CPM	0.14	-0.44																	
ISM	0.58	0.13	0.66																
CBLI	-0.96	-1.82**	-1.72**	-2.08**															
CBCI	-0.07	-0.65	-0.27	-0.82	1.25														
TBCLI	-0.60	-1.25	-0.98	-1.30*	0.37	-0.69													
TBCCI	-0.02	-0.44	-0.10	-0.54	0.75	0.04	0.47												
ECLIRI	-0.28	-0.82	-0.57	-1.14	0.87	-0.34	0.57	-0.22											
CLAIMS	-0.18	-0.76	-0.35	-0.99	1.32*	-0.17	0.67	-0.13	0.24										
OCONS	0.39	-0.23	0.35	-0.48	1.89**	0.61	1.05	0.29	0.72	0.64									
OMANU	1.25	0.38	0.99	0.26	1.65**	1.06	1.19	0.74	0.94	0.95	0.87								
IPAUTO	1.56*	0.55	1.42*	0.39	1.93**	1.47*	1.35*	0.85	1.12	1.10	1.11	0.16							
BUILIP	-0.21	-0.59	-0.29	-0.90	0.69	-0.16	0.39	-0.17	0.10	-0.07	-0.53	-0.86	-1.04						
SPREAD	0.89	0.48	0.95	0.43	2.10**	1.04	1.53*	0.72	1.18	1.32*	0.81	0.17	0.08	0.95					
IPREM	0.68	0.15	0.85	-0.01	2.28**	1.03	1.58*	0.43	1.37*	1.07	0.45	-0.31	-0.49	0.71	-0.48				
FEDF	1.40*	0.91	1.13	0.86	1.41*	1.24	1.23	1.34*	1.11	1.14	1.06	0.94	0.87	1.17	0.66	0.84			
M2	1.04	0.81	1.14	0.58	1.87**	1.12	1.77**	0.72	1.45*	1.19	0.86	0.47	0.44	0.95	0.33	0.93	-0.34		
LOANS	0.61	0.13	0.56	0.06	1.35*	0.62	0.99	0.42	0.73	0.70	0.42	-0.27	-0.29	0.65	-0.39	0.08	-0.91	-0.56	
SP500	-0.03	-0.42	-0.15	-0.69	0.83	0.04	0.54	-0.01	0.27	0.13	-0.36	-0.99	-1.29*	0.18	-0.85	-0.54	-1.53*	-0.88	-0.62
Crit. Val.:																			
10%	1.28																		
5%	1.65																		
1%	2.32																		

Notes: *, **, and *** indicate significance levels of 10, 5, and 1%.

Table 6. Diebold-Mariano-Test (1995). rolling window forecasting scheme

h=4	ICS	CBCS	CPM	ISM	CBLI	CBCI	TBCLI	TBCCI	ECLIRI	CLAIMS	OCONS	OMANU	IPAUTO	BUILIP	SPREAD	IPREM	FEDF	;M2	LOANS
CBCS	-0.84																		
CPM	-0.79	0.44																	
ISM	-0.93	0.05	-1.43*																
CBLI	-1.29*	-1.21	-2.05**	-1.59*															
CBCI	-1.23	-0.19	-1.11	-0.35	1.14														
TBCLI	-1.22	-0.68	-1.35*	-0.90	0.75	-0.67													
TBCCI	-1.34*	0.21	-0.16	0.20	1.07	0.76	0.85												
ECLIRI	-1.23	0.20	-0.24	0.17	1.17	0.90	0.92	-0.18											
CLAIMS	-0.92	-0.07	-0.71	-0.14	1.38*	0.15	0.75	-0.29	-0.28										
OCONS	-1.37*	0.02	-0.46	-0.02	0.98	0.47	0.71	-1.22	-1.11	0.09									
OMANU	-1.47	0.03	-0.39	-0.01	0.88	0.37	0.64	-1.33*	-0.62	0.09	0.07								
IPAUTO	-1.37*	0.18	-0.20	0.15	1.01	0.67	0.79	-0.34	0.02	0.25	1.09	1.53*							
BUILIP	-0.90	0.42	0.04	0.50	1.72**	1.28*	1.26	0.23	0.30	0.49	0.61	0.52	0.27						
SPREAD	-0.58	0.55	0.20	0.45	1.76**	0.66	1.45*	0.28	0.33	0.62	0.46	0.43	0.31	0.18					
IPREM	-0.99	-0.29	-0.75	-0.36	0.94	-0.12	0.40	-0.42	-0.42	-0.32	-0.27	-0.26	-0.39	-0.76	-0.75				
FEDF	1.54*	1.42*	1.40*	1.45*	1.71**	1.67**	1.73**	1.78**	1.72**	1.50*	1.78**	1.83**	1.80**	1.46*	1.35*	1.50*			
M2	-1.39*	-0.55	-1.37*	-1.01	0.36	-0.69	-0.03	-0.88	-0.87	-0.60	-0.72	-0.68	-0.81	-1.89**	-1.02	-0.44	-1.72**		
LOANS	-0.64	0.64	0.51	0.93	1.98**	1.18	1.67**	0.65	0.71	0.85	0.90	0.83	0.66	0.46	0.18	0.98	-1.36*	1.63*	
SP500	0.65	1.16	1.08	1.18	1.57*	1.64*	1.52*	1.91**	1.76**	1.23	1.91**	2.06**	1.98**	1.23	0.81	1.25	-1.26	1.63*	0.86
Crit. Val.:																			
10%	1.28																		
5%	1.65																		
1%	2.32																		

Notes: *, **, and *** indicate significance levels of 10, 5, and 1%.

TABLE 7. Diebold-Mariano-Test (1995), rolling window forecasting scheme

Regarding the one-step ahead forecasts generated by the recursive forecasting scheme we find that survey, composite and real economic indicators, especially the *CBLI*, are significantly better in terms of MSE than financial indicators. Additional evidence for superior predictive power of composite indicators is found in the rolling window setting. Again, the *CBLI* indicator is significantly better performing than all real economic indicators except of BUILD and all survey indicators except *ICS*. Furthermore, *IPAUTO* is significantly worse than most composite and even two survey indicators. Interestingly, we do not find much evidence for superiority of real economic and survey indicators.

When interpreting the results for four-step ahead forecasts we find a similar picture for one-step ahead recursive forecasts with the exception of a bad performing *ICS* indicator. Financial indicators, especially the *FEDF*, are clearly outperformed by composite and real economic indicators. The last statement is true for rolling window forecasts as well. Moreover, composite indicators, especially *CBLI*, perform significantly better than indicators from other groups.

4.3. **Predicting Recessions.**

4.3.1. *Probability Forecasts.* We use versions of the model given in equation (2) to predict turning points of the business cycle. First, we are concerned with the in-sample performance of different models. The results of the model selection are given in Table 8. The second to fourth columns show the lag orders of the selected models. The fifth column contains the BIC. As an additional measure of goodness-of-fit we report the pseudo R^2 as proposed in Estrella (1998).[12] The last two columns show two different accuracy measures frequently used to evaluate probability forecasts. We report the Quadratic Probability Score (QPS) (Brier, 1950) and the Logarithmic Probability Score (LPS).[13] Small values of both measurements indicate a high predictive power of a model with the lower bound of 0 being reached for perfect predictions in both cases.

The results show that in general the models seem to be able to indicate recessions well without giving too many false signals. All measures indicate that *IPREM* is by far the worst performing indicator. This is also revealed in Figure 3. The set of best indicators - with very similar QPS and LPS - contains *CBLI*, *ECRILI*, and *S&P500*. The performance of the last indicator is relativized, however, with a look at the number of false signals the indicator gave. In total, *S&P500* pointed towards a recession in four cases where no recession could be observed later on, if one is willing to accept the unconditional probability for a recession as a threshold to transform the probability estimates into binary signals.[14] Comparing the different groups of indicators it is evident that on average the composite indices perform best while the survey indicators show the highest probability scores. The two other groups show QPSs in the same range of values (leaving out *IPREM* for the moment). The real economic indicators have, however, higher corresponding LPSs compared to the financial indicators which indicates that the former tend to produce fewer but large probability errors.[15]

[12] It is given by $pseudo - R^2 = 1 - (LL_u/LL_c)^{(-2/TLL_c)}$, where LL_u denotes the log-likelihood of the unconstrained model and LL_c is the log-likelihood of the constrained model which includes a constant only.

[13] The measures are given by $QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - R_t)^2$ and $LQS = -\frac{1}{T}\sum_{t=1}^{T}(1 - R_t)\ln(1 - P_t) + R_t \ln(P_t)$, where P_t denotes the probability forecast made for time t and R_t is the realization for the same sample point.

[14] See also discussion below for this point.

[15] Note that especially the LPS should be treated with caution if one is interested in the binary signals only. Imagine a model that perfectly predicts every recessionary quarter by signaling a probability of say

	$Lags^{Ind}$	$Lags^{GDP}$	Lag^{RecInd}	BIC	ps. R^2	QPS	LPS
Survey Indicators							
ICS	2	2	1	0.47	0.478	0.096	0.151
CBCS	2	2	-	0.55	0.532	0.121	0.191
CPMI	1	2	1	0.51	0.398	0.113	0.183
ISM	2	2	1	0.50	0.439	0.108	0.166
Composite Indicators							
CBLI	1	1	1	0.36	0.547	0.077	0.124
CBCI	2	2	1	0.47	0.480	0.101	0.150
TBCLI	1	1	1	0.36	0.555	0.078	0.121
TBCCI	2	2	1	0.44	0.512	0.091	0.137
ECRILI	1	2	-	0.36	0.554	0.079	0.122
Real Economic Indicators							
CLAIMS	2	2	-	0.47	0.447	0.096	0.163
OCONS	2	2	1	0.50	0.445	0.105	0.164
OMANU	1	1	1	0.51	0.365	0.109	0.197
IPAUTO	1	2	1	0.51	0.398	0.115	0.183
BUILD	2	1	1	0.46	0.451	0.091	0.161
Financial Indicators							
SPREAD	1	1	1	0.40	0.502	0.088	0.141
IPREM	3	2	-	0.45	0.065	0.195	0.334
FEDF	1	1	1	0.44	0.450	0.100	0.162
M2	1	1	1	0.48	0.402	0.107	0.181
LOANS	1	2	1	0.50	0.411	0.111	0.177
SP500	1	2	-	0.36	0.549	0.075	0.123

TABLE 8. Selected Models in In-Sample Analysis

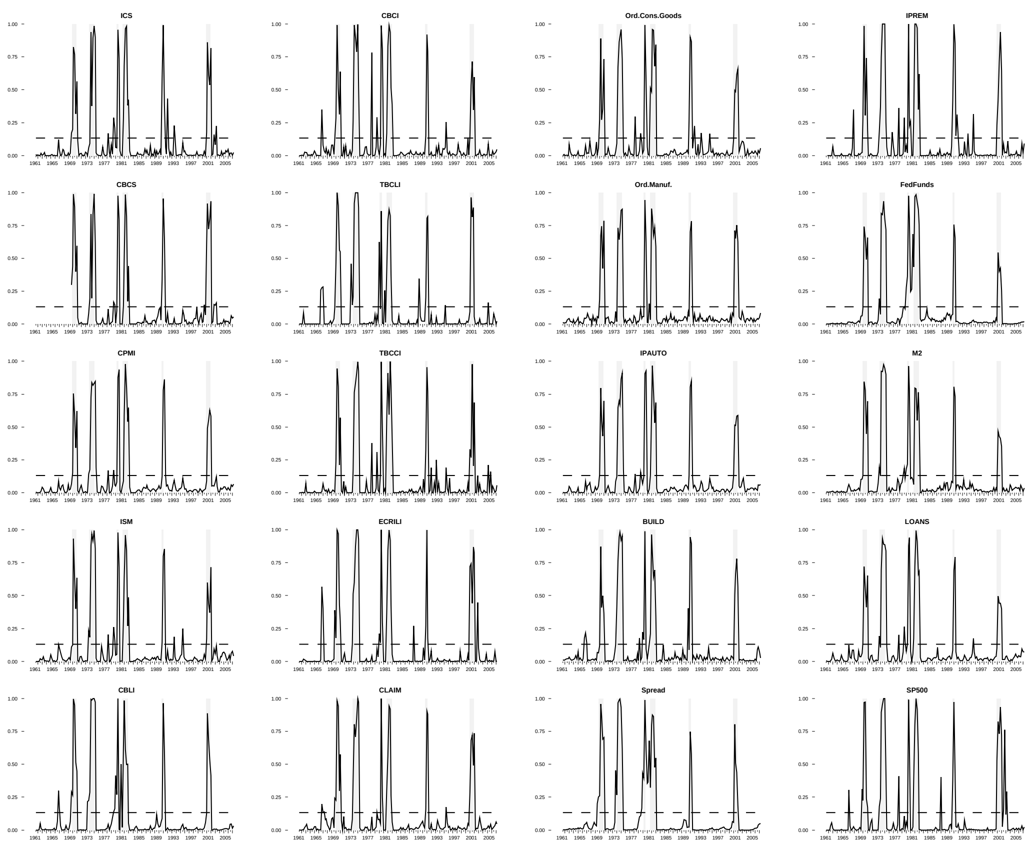

FIGURE 3. Fitted Probabilities of a Recession Based on In-Sample Analysis

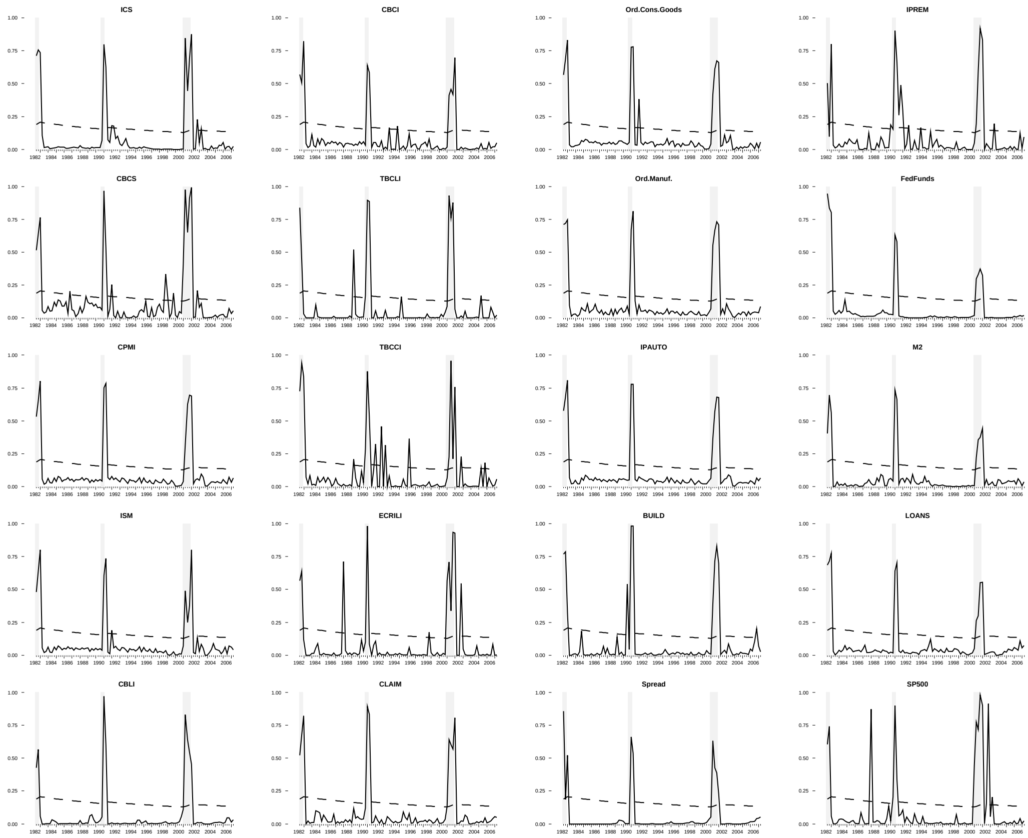

FIGURE 4. Predicted Probabilities of a Recession Based on Out-of-Sample Analysis

The in-sample performance of an indicator or a specific model specification gives a first hint on which indicators are promising with regards to their predictive power. Eventually, however, the forecast performance of models should be evaluated according to their out-of-sample performance (Ashley et al., 1980, Meese and Rogoff, 1983).[16] Therefore, we now turn to the out-of-sample analysis of the recession prediction models. In what follows we concentrate on one-step ahead forecasts.[17]

As mentioned above, we re-select the optimal model specifications in each of the recursive estimation and forecasting exercises. More formally, for each point in time t (between 1981Q3 and 2007Q2) we (i) estimate equation 2 (for all possible different lag structures) based on the sample 1960Q1 to time $t-1$ using the data as it is available in the vintage that had been published at time t, (ii) select the best model according to the BIC,[18] and (iii) use the results to estimate the probability for the economy being in a recession at time t. Through this procedure, we obtain a series of out-of-sample predictions for recession probabilities.

0.50 while signaling a probability of 0.00 for all other quarters. The corresponding QPS and LPS would be equal to 0.06 (not much smaller as the best observed QPSs) and 0.69 (much higher than any observed LPS here) respectively.

[16]Admittedly, this conventional view has been criticized recently by Inoue and Kilian (2004).

[17]In practice - due to data publication lags - this boils down to what is sometimes referred to as "nowcasting", i.e. predicting a contemporaneous event conditional on past data because current data is not yet known to the forecaster.

[18]The detailed sequences of specifications are not reported here for spacial reasons. They are, however, available from the authors upon request.

Also these out-of-sample results indicate that most of the indicators do a good job predicting recessions. All models are able to indicate all recessions and there are only very few cases of false signals (Figure 4). Concerning the latter it is again *S&P500* which shows many and very high “false alarms”. But also *TBCCI* and *CBCS* signal recessions during actually expansionary periods at multiple times. In contrast to the *S&P500*, however, those signals lie only marginally above the unconditional recession probability. Admittedly, it is striking that most of the models so not signal a recession in the first quarter of most of the downswings while they do so from most of the second quarter onwards. Thus, one might wonder how much of the good performance is only due to the included lagged dependent variable term. Since the assumption that already the first lag of the recession indicator is observable without any uncertainty by the forecaster is somewhat *ad-hoc* and might not hold in practise, it is probably adequate to judge the performance of indicators based on a slightly restricted model.

Therefore, we repeated the same out-of-sample forecasting exercise restricting the coefficient of the lagged recession indicator to zero.[19] Figure 5 shows these new probability forecasts for all indicators together with the (time varying) unconditional probability for a recession as it was observed at each point in time. The differences to the forecasts shown in Figure 4 are striking. The plots indicate that the performances have deteriorated considerably relative to those of the unrestricted (but probably unrealistic) models. The signals for recessions are on average far weaker, i.e. the indicated probabilities for recessions are lower. A second observation is that although for most indicators the probability forecasts increase already before or with the beginning of recessions the highest recession probabilities are actually forecasted towards the end or shortly after a recession. Exceptions are two of the explicit leading indicators, *TBCLI* and *ECRILI*. This confirms that leading indicators do actually provide some value added for forecasting recessions. Two other indicators which show partly good results with respect to their performance in the first quarters of recessions are *BUILD* and *S&P500*. *BUILD* correctly points towards a recession in the first quarters of the recessions in 1981 and 1990 while missing initially the “double-dip” recession in 2001. *S&P500* performs well in all first quarters but on the other hand signals again multiple “false alarms”.

The same impression is also given by a comparison of probability scores of the unrestricted and the restricted models. They are displayed in Table 9. While the scores for the unrestricted models are naturally already higher than those based on the in-sample analysis because much less information is used, we find that the probability scores generally increase further considerably when restricting δ to zero.

Finally, we would like to test the statistical significance of the gain that is delivered by the different indicators in terms of forecast precision. Like in section 4.2 we base this analysis on the critical values for different tests for equal forecast accuracy provided by Clark and McCracken (2001). In a recent contribution Clements and Harvey (2007) show, that those tests can be applied to probability forecasts generated by binary choice models when the test statistics are based on quadratic loss functions like the QPS. We follow this approach to judge the forecast performance of the various indicators. Because the critical values for the tests are non-standard and depend on several parameters (most importantly the number of restrictions for the nested model), we have to keep the specification of both models fixed over time (in contrast to the out-of-sample analysis presented so far). We

[19]Note, of course, that this zero-restriction is already imposed by our model selection rule for some of the models at some points in time automatically.

	In-Sample	Out-of-Sample$^{unrestr.}$	Out-of-Sample$^{restr.}$
Survey Indicators			
ICS	0.080	0.103	0.127
CBCS	0.082	0.101	0.112
CPMI	0.104	0.115	0.139
ISM	0.103	0.129	0.150
Composite Indicators			
CBLI	0.075	0.080	0.127
CBCI	0.094	0.116	0.143
TBCLI	0.079	0.075	0.130
TBCCI	0.071	0.116	0.126
ECRILI	0.057	0.087	0.095
Real Economic Indicators			
CLAIMS	0.085	0.107	0.119
OCONS	0.086	0.112	0.140
OMANU	0.102	0.104	0.144
IPAUTO	0.107	0.110	0.134
BUILD	0.097	0.101	0.135
Financial Indicators			
SPREAD	0.103	0.123	0.173
IPREM	0.080	0.135	0.136
FEDF	0.098	0.127	0.143
M2	0.111	0.141	0.189
LOANS	0.104	0.112	0.132
SP500	0.056	0.089	0.095

Table 9. Comparison of Quadratic Probability Scores

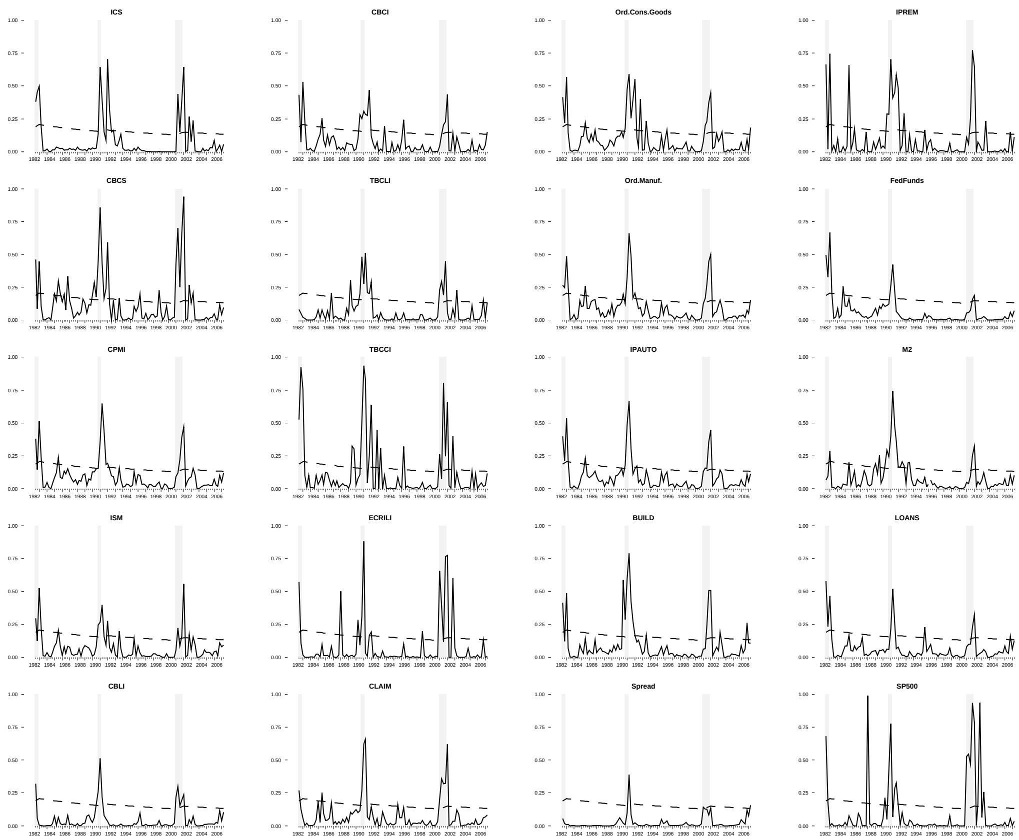

Figure 5. Predicted Probabilities of a Recession Based on Out-of-Sample Analysis (Restricted Model)

Indicator	MSE-F	MSE-t	ENC-F	ENC-t
ICS	23.86***	1.25***	25.32***	2.22***
CBCS	17.52***	0.75**	23.24***	1.7**
CPMI	-4.61	-0.96	-1.27	-0.55
ISM	-7.52	-0.8	-0.16	-0.03
CBLI	2.29**	0.13	12.65***	1.57**
CBCI	5.93***	0.72**	5.62***	1.3*
TBCLI	21.9***	1.6***	20.63***	2.42***
TBCCI	26.95***	1.04**	38.44***	2.38***
ECRILI	35.27***	1.31***	47.27***	2.65***
CLAIMS	17.61***	1.35***	16.69***	2.07**
OCONS	2.41**	0.26	6.25***	1.3*
OMANU	-0.94	-0.25	0.05	0.02
IPAUTO	5.78***	1.01**	4.06**	1.34*
BUILD	2.65**	0.25	6.71***	1.2*
SPREAD	-18.39	-0.95	8.36***	0.98
IPREM	-5.4	-0.55	1.25	0.27
FEDF	0.75*	0.04	13.65***	1.51**
M2	-18.45	-1.84	-4.11	-0.95
LOANS	0.66	0.08	3.14**	0.8
SP500	50.36***	1.3***	68.33***	3.25***
Critical Values				
10%	0.74	0.27	1.63	1.05
5%	1.79	0.61	2.54	1.43
1%	4.33	1.24	4.55	2.16

Notes: *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

TABLE 10. Tests for equal forecast accuracy

decided to use the most frequently selected specification conditional on $\delta = 0$, namely $p = 2$ and $q = 2$. The results are given in Table 10.[20] While the significance by which we can conclude that the well performing indicators help to produce superior forecasts differs of course considerably across indicators, there are several indicators which do not help to improve the probability forecasts at all. Those indicators are the two purchasing manager indices *CPMI* and *ISM*, new orders in the manufacturing sector *OMANU*, and a set of financial indicators, namely *SPREAD*, *IPREM*, *M2*, and *LOANS*.

4.3.2. *Directional Forecasts.* While probability forecasts and the probability scores give a first assessment of the goodness of our models for prediction purposes, we are actually dealing with a problem that involves a binary decision (between recession and expansion). Some practitioners might not be interested in the exact probability for a recession but rather the binary decision: *recession - yes or no*? To transform our probability forecasts into binary outcomes we use the following rule

$$P_t^{bin} = \begin{cases} 1 & if\ P_t > R_t^{uc}, \\ 0 & if\ P_t \leq R_t^{uc}, \end{cases} \tag{3}$$

where as above P_t denotes the predicted probability, P_t^{bin} stands for the binary prediction outcome, and R_t^{uc} is the unconditional probability computed over the sample observations up to time $t-1$.[21]

[20] We use the same abbreviations as in Clark and McCracken (2001) to refer to the four tests.

[21] One could argue for other thresholds (as e.g. 25% or 50%). We believe, however, that those would be more of an arbitrary assumption than our rule. Besides, the following results are fairly robust to the choice of sensible thresholds.

	Actual Boom	Actual Recession	Sum
Predicted Boom	O_{bb}	O_{br}	$O_{b.}$
Predicted Recession	O_{rb}	O_{rr}	$O_{r.}$
Sum	$O_{.b}$	$O_{.r}$	O

TABLE 11. Summary of directional forecast performance

A test for the quality of such binary forecasts is described in Diebold and Lopez (1996). It is based on the realized and expected values of the number of wrongly and rightly predicted outcomes. Let the quantities O_{ij} $(i, j \in \{b, r\})$ denote the number of forecast outcome as described in Table 11. The informational content of the forecasts can be condensed in a measure $I = \frac{O_{bb}}{O_{.b}} + \frac{O_{rr}}{O_{.r}}$ which is asymptotically bound between 1 and 2 and converges to 1 in a "coin-flip"-like experiment and to 2 if we are dealing with perfect forecasts. The significance of its informational content can be tested using Pearson's χ^2 statistic which is in this case given by $C = \sum_{i\in\{b,r\}} \sum_{j\in\{b,r\}} \frac{(O_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}} \sim \chi^2(1)$, where $\hat{E}_{ij}$ is estimated by $O_{b.}\, O_{r.}\, /O$. The various estimates for I as well as C and the corresponding p-value can be found in Table 12. To concentrate on the potential of the indicators to predict the binary outcome, we focus on the models which include only lags of the indicators and which restrict the coefficients of lagged GDP growth and the lagged recession indicator to zero. Most indicators show significant informational content at high significance levels; the only exceptions being *IPAUTO*, *M2*, and *LOANS*.

	I	C	p-val
Survey Indicators			
ICS	1.19	6.86	0.009
CBCS	1.32	16.95	0.000
CPMI	1.32	13.56	0.000
ISM	1.31	11.94	0.001
Composite Indicators			
CBLI	1.31	38.70	0.000
CBCI	1.33	15.41	0.000
TBCLI	1.33	29.18	0.000
TBCCI	1.42	29.86	0.000
ECRILI	1.32	26.37	0.000
Real Economic Indicators			
CLAIMS	1.37	24.61	0.000
OCONS	1.27	12.89	0.000
OMANU	1.21	6.18	0.013
IPAUTO	1.07	0.90	0.344
BUILD	1.18	9.25	0.002
Financial Indicators			
SPREAD	1.17	17.51	0.000
IPREM	1.19	5.26	0.022
FEDF	1.15	5.36	0.021
M2	0.95	0.55	0.457
LOANS	0.93	1.60	0.206
SP500	1.38	26.43	0.000

TABLE 12. Informational Content of Out-of-Sample Binary Forecasts

5. Conclusion

In this paper we have presented a thorough assessment of a battery of potential leading indicators for US business cycle forecasting. In the analysis, we made use of so-called real-time data taking into account frequent revisions of macroeconomic data. First, we analyzed their ability to improve forecasts of quarterly growth rates of GDP. Second, we have shown which indicators help to form superior recession forecasts, i.e. assessed the indicators' ability to predict business cycle turning points. The results can be summarized as follows.

First, in the context of growth forecasts most of the indicators help to improve short-run forecasts for GDP growth while they do not add to the predictability of growth rates four quarters ahead in the majority of cases. Second, composite indicators, especially *CBLI*, are performing quite well for both one- and four-step horizons whereas financial indicators are often outperformed by composite, real economic and survey indicators. Second, the results in the context of prediction of turning points are much less optimistic. There is a wide range of indicators which do not significantly improve recession probability forecasts at all (although with the exception of three indicators they all show some degree of value added when it comes to making the binary choice between *recession* and *no recession*). On the other hand, there are other indicators (especially the composite leading indicators) which improve the forecast performance by a large extend. It has to be mentioned, however, that there is no indicator that seems to be "always working"; all models perform worse for some recessions on the one hand and indicate one or more "false alarms" on the other hand. Finally, we confirm the result that in general it is not true that an indicator that helps to improve forecasts for growth rates is also necessarily useful for predicting recessions. The group of composite leading indicators seems to be the only class of indicators for which this conclusion holds. The last result should, however, be relativized somewhat because there are no real-time vintages available for the composite indicators. Therefore, the performance might partly be driven by the fact that some of the input series for those indicators are actually revised time series.

In our view, future research should focus on three issues. First, compiling real-time data sets also of composite leading indicators, which are usually revised once revised macroeconomic data becomes available, would help to judge them in an even more realistic setup. Second, non-linearities could be potentially important. Paap et al. (2007) show evidence that there might be differences in the "leading properties" of indicators before troughs as opposed to times before peaks. Therefore, it should be thoroughly analyzed if any non-linear relationships play an important role in forecast models with leading indicators. Finally, we haven't taken into account yet the different frequencies, at which indicators are actually available. While we concentrated on quarterly observations for all data series (and this is commonly done in the literature on forecasting recessions), there is a growing literature on mixed data-frequency sampling (MIDAS), which uses information from time series with different frequencies to forecast e.g. GDP growth rates. It would be interesting to extend this approach to the issue of forecasting recessions by binary choice models.

References

Altissimo, Filippo, Riccardo Cristadoro, Mario Forni, Marco Lippi, and Giovanni Veronese (2006), "New EuroCOIN: Tracking Economic Growth in Real Time," CEPR Discussion Paper 5633.

Ang, Andrew, Monika Piazzesi, and Min Wei (2006), "What does the yield curve tell us about GDP growth?" *Journal of Econometrics*, 131, 359–403.

Ashley, Richard A., Clive W. J. Granger, and Richard Schmalensee (1980), "Advertising and aggregate consumption: An analysis of causality," *Econometrica*, 48, 1149–1167.

Bandholz, Harm, and Michael Funke (2003), "In search of leading indicators of economic activity in Germany," *Journal of Forecasting*, 22(4), 277–297.

Banerjee, Anindya, and Massimiliano Marcellino (2006), "Are there any reliable leading indicators for US inflation and GDP growth?" *International Journal of Forecasting*, 22, 137–151.

Bengoechea, Pilar, Maximo Camacho, and Gabriel Perez-Quiros (2006), "A useful tool for forecasting the Euro-area business cycle phases," *International Journal of Forecasting*, 22(4), 735–749.

Bernanke, Ben S., and Jean Boivin (2003), "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50, 525–546.

Breitung, Jörg, and Doris Jagodzinski (2001), "Prognoseeigenschaften alternativer Indikatoren für die Konjunkturentwicklung in Deutschland," *Konjunkturpolitik*, 47, 292–314.

Brier, Glenn W. (1950), "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 75, 491–504.

Camacho, Maximo (2004), "Vector Smooth Transition Regression Models for US GDP and the Composite Index of Leading Indicators," *Journal of Forecasting*, 23, 173–196.

Campbell, John Y. (1999), "Asset Prices, Consumption and the Business Cycle," in John B. Taylor and Micheal Woodford, editors, *The Handbook of Macroeconomics, Vol. 1*, 1231–1303, Elsevier, Amsterdam.

Carroll, Christopher D., Jeffrey C. Fuhrer, and David W. Wilcox (1994), "Does Consumer Sentiment Forecast Household Spending?" *American Economic Review*, 84, 1397–1408.

Chauvet, Marcelle, and James D. Hamilton (2006), "Dating business cycle turning points," in Costas Milas, Philip Rothman, and Dick van Dijk, editors, *Nonlinear time series analysis of business cycles*, Elsevier, Amsterdam.

Clark, Todd E. (1999), "Finite-sample properties of tests for equal forecast accuracy." *Journal of Forecasting*, 18, 489–504.

Clark, Todd E., and Michael W. McCracken (2001), "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, 105, 85–110.

Clements, Michael P., and David I. Harvey (2007), "Forecast Encompassing Tests and Probability Forecasts," mimeo, University of Warwick.

Croushore, Dean, and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111–130.

Curtin, Richard T. (2007), "Consumer sentiment surveys: Worldwide review and assessment," *Journal of business cycle measurement and analysis*, 3, 7–42.

Davis, E. Philip, and Gabriel Fagan (1997), "Are Financial Spreads Useful Indicators of Future Inflation and Output Growth in EU Countries?" *Journal of Applied Econometrics*, 12(6), 701–714.

Diebold, Francis X., and Jose A. Lopez (1996), "Forecast Evaluation and Combination," in G.S. Maddala and C.R. Rao, editors, *Statistical Methods in Finance*, North-Holland, Amsterdam.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–263.

Diebold, Francis X., and Glenn D. Rudebusch (1991), "Forecasting Output With the Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association*, 86, 603–610.

Duarte, Augustin, Ioannis A. Venetis, and Ivan Paya (2005), "Predicting real growth and the probability of recession in the Euro area using the yield spread," *International Journal of Forecasting*, 21, 261–277.

Dueker, Michael (1997), "Strengthening the case for the yield curve as a predictor of U.S. recessions," *Federal Reserve Bank of St. Louis, Review*, 2, 41–51.

Estrella, Arturo (1998), "A New Measure of Fit for Equations with Dichotomous Dependent Variables," *Journal of Business & Economic Statistics*, 16, 198–205.

Estrella, Arturo, and Gikas A. Hardouvelis (1991), "The term structure as predictor of real economic activity," *Journal of Finance*, 46, 555–576.

Estrella, Arturo, and Frederic S. Mishkin (1998), "Predicting U.S. recessions: financial variables as leading indicators," *Review of Economics and Statistics*, 1, 45–61.

Estrella, Arturo, Anthony P. Rodrigues, and Sebastian T. Schich (2003), "How stable is the predictive power of the yield curve? Evidence from Germany and the United States," *The review of Economics and Statistics*, 85, 629–644.

Fair, Ray C., and Rober J. Shiller (1990), "Comparing Information in Forecasts from Econometric Models," *American Economic Review*, 80, 375–389.

Federal Reserve Bank of Philadelphia (2007), "Real-Time Data Set," Retrieved from the World Wide Web at October 16, 2007.
`http://www.philadelphiafed.org/econ/forecast/real-time-data/index.cfm`

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003), "Do financial variables help forecasting inflation and real activity in the euro area?" *Journal of Monetary Economics*, 50, 1243–1255.

Fritsche, Ulrich, and Vladimir Kuzin (2005), "Prediction of Business Cycle Turning Points in Germany," *Jahrbücher für Nationalökonomie und Statistik*, 225, 22–42.

Fukuda, Kosei (2007), "Forecasting real-time data allowing for data revisions," *Journal of Forecasting*, 26, 429–444.

Geweke, John (1977), "The dynamic factor analysis of economic time series," in D.J. Aigner and A.S. Goldberger, editors, *Latent variables in socio-economic models*, Edward Elgar, North Holland Amsterdam.

Giacomini, Raffaella, and Barbara Rossi (2006), "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?" *Oxford Bulletin of Economics and Statistics*, 68, 783–795.

Giacomini, Raffaella, and Halbert White (2006), "Tests of conditional predictive ability," *Econometrica*, 74(6), 1545–1578.

Granger, Clive W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.

Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Non Stationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Howrey, Philip (2001), "The Predictive Power of the Index of Consumer Sentiment," *Brooking Papers on Economic Activity*, 2001:1, 175–216.

Inoue, Atsushi, and Lutz Kilian (2004), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews*, 23(4), 371–204.

Kholodilin, Konstantin A. (2005), "Measuring and predicting turning points using a dynamic bi-factor model," *International Journal of Forecasting*, 21(3), 525–537.

Kholodilin, Konstantin A., and Boriss Siliverstovs (2006), "On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence," *Jahrbücher für Nationalökonomie und Statistik*, 226(3), 234–259.

Marcellino, Massimiliano, James Stock, and Mark Watson (2005), "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, 135, 499–526.

Matsuasaka, John G., and Argia M. Sbordone (1995), "Consumer Confidence and Economic Fluctuations," *Economic Inquiry*, 33, 296–318.

Meese, Richard A., and Kenneth Rogoff (1983), "Empirical Exchange Rate Models of the Seventies. Do They Fit Out of Sample?" *Journal of International Economics*, 14, 3–24.

Mork, Knut A. (1987), "Ain't Behavin': Forecast Errors and Measurement Errors in Early GNP," *Journal of Business and Economic Statistics*, 5, 165–175.

Paap, Richard, Rene Segers, and Dick van Dijk (2007), "Do Leading Indicators Lead Peaks More Than Troughs?" Report EI 2007-08, Econometric Institute.

Schumacher, Christian (2007), "Forecasting German GDP Using Alternative Factor Models Based on Large Datasets," *Journal of Forecasting*, 26, 271–302.

Schumacher, Christian, and Jörg Breitung (2006), "Real-time forecasting of GDP based on large factor model with monthly and quarterly data," Discussion Paper 33, Deutsche Bundebank.

Sims, Christopher A. (1980), "Macroeconomics and Reality," *Econometrica*, 1, 1–48.

Stock, James H., and Marc W. Watson (1989), "New indexes of coincident and leading indicators," in Oliver Blanchard and S. Fisher, editors, *NBER Macroeconomics Annual 1989*, 351–393, MIT Press, Cambridge.

Stock, James H., and Marc W. Watson (2003), "Forecasting Output and Inflation. The Role of Asset Prices," *Journal of Economic Literature*, 41, 788–829.

Swanson, Norman R. (1996), "Forecasting using first-available versus fully revised economic time-series data," *Studies in nonlinear dynamics and econometrics*, 1(1), 47–64.