united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

H4.SMR/1150 - 17

Fifth Workshop on Non-Linear Dynamics and Earthquake Prediction

4 - 22 October 1999

Pre-recession Pattern of Six Economic Indicators in the USA

V. Keilis-Borok

International Institute of Earthquake Prediction Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow 113556, Russian Federation

FOR730 10/3/99 16:50

Journal of Forecasting
J. Forecast. **18**, 00–00 (1999)

Pre-recession Pattern of Six Economic Indicators in the USA

V. KEILIS-BOROK,[1]* J. H. STOCK,[2] A. SOLOVIEV[1] and P. MIKHALEV[3]

[1] *International Institute for Earthquake Prediction Theory and Mathematical Geophysics, Moscow, Russia*
[2] *Kennedy School of Government, Harvard University, USA*
[3] *McKinsey & Co., Los Angeles, USA*

ABSTRACT

This paper applies a tightly parameterized pattern recognition algorithm, previously applied to earthquake prediction, to the problem of predicting recessions. Monthly data from 1962 to 1996 on six leading and coincident economic indicators for the USA are used. In the full sample, the model performs better than benchmark linear and non-linear models with the same number of parameters. Subsample and recursive analysis indicates that the algorithm is stable and produces reasonably accurate forecasts even when estimated using a small number of recessions. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS leading indicators; pattern recognition; business cycle forecasting

INTRODUCTION

A central problem of macroeconomic forecasting has been the prediction of the onset and end of recessions using time series data on economic indicators; recent contributions include Diebold and Rudebusch (1989), Fair (1993), Stock and Watson (1989, 1993), Sims (1993), Berk and Bikker (1995), and Mostaghimi and Rezayat (1996). At least four features of this problem make this particularly difficult. First, recessions are rare; in the United States, for example, since 1959 (the earliest date for which many monthly series are available) to date, there have been only five completed expansions. Second, recessions are discrete events, so forecast of recessions naturally entail some degree of non-linearity. Third, recessions are complicated events and their accurate prediction arguably requires the use of many series at once. Indeed, the presence of an economy-wide decline is an important factor in the *ex-post* identification of recessions, and statistical models that focus on a single series miss this aspect. Fourth, economic institutions, demographics, and economic policy are constantly evolving, and it is natural to expect that

* Correspondence to: V. Keilis-Borok, International Institute for Earthquake Prediction Theory and Mathematical Geophysics, Warshavskoye sh., 79, kor.2, Moscow 113556 Russia.

CCC 0277–6693/99/000000–00$17.50

Received ??
Revised September 1998

reduced-form relations will also evolve over time. In short, the problem requires fitting non-linear, high-dimensional models to a handful of observations generated by a possibly non-stationary economic environment.

In this paper we address these problems by considering recession forecasts produced by a pattern-recognition algorithm that was previously developed for the prediction of infrequent events (e.g. Gelfand *et al.*, 1976; Press and Allen, 1995, and the references therein). The general methodology was developed by the artificial intelligence school of IM. Gelfand for analysis of rare phenomena of complex origin. The specific algorithm used here has previously been applied to earthquake prediction (e.g. Keilis-Borok and Kossobokov, 1990; Vorobieva and Levshina, 1994) and to prediction of the results of American elections (Lichtman and Keilis-Borok, 1989; Keilis-Borok and Lichtman, 1990). The motivation for this application to recession forecasting is that, from a time series perspective, recessions and large earthquakes bear similarities: both are discrete, relatively rare events, and both have a rich set of observable indicators that can potentially help in their prediction. This raises the question of whether the empirical methods developed for handling many geophysical series to predict earthquakes can be useful for the analogous task of predicting recessions.

Applied to this macroeconomic problem, this algorithm warns of a recession when negative signals are widespread across a group of economic indicators. In the work presented here, six indicators (listed in the next section) are used; if four or more signal a recession, then a recession warning is signalled. The result is a binary warning signal, with a value of one indicating that the economy is in a time of increased probability of a recession. Formally, the pattern recognition algorithm uses a tightly parameterized non-linear model that produces a 0/1 alarm, A_t. Let $\{y_{kt}\}$, $t = 1, \ldots, T$, denote the time series of observations on the kth of N economic indicators. With this notation, the model can be written as

$$A_t = \begin{cases} 1 & \text{if} \sum_{k=1}^{N} \Psi_{kt} \geq N - b \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\Psi_{kt} = 1$ if $y_{kt} \geq c_k$ and $=0$ otherwise where $(b, c_1, \ldots, c_N)$ are free parameters. The time series $\{y_{kt}\}$ typically are transformations of the original economic indicators, for example a series might be transformed by taking quarterly growth rates, and it might be 'inverted' (that is, the series is multiplied by minus one) so that large values of the series are associated with an impending recession. (Details of such transformations are discussed in the next section.) The pattern-recognition interpretation of the model stems from noting that

$$N - \sum_{k=1}^{N} \Psi_{kt}$$

is the (Hamming) distance from the ideal prerecessionary pattern of $\Psi_{kt} = 1$, $k = 1, \ldots, N$.

From the perspective of the problems listed in the introductory paragraph, this model has several advantages. It uses multiple economic indicators (in our application, $N = 6$); it is highly non-linear and by construction produces 0/1 warnings; and it is tightly parameterized, with only $N + 1$ parameters (given the choice of $\{y_{kt}\}$). Because the economic indicator series are transformed to binary variables, the model also has the desirable feature of being robust to outliers in the data. Arguably, this transformation to binary variables could in addition make the model less sensitive to drift in the relationship between the individual indicators and the overall economy, so that this model would be robust to some types of structural instability.

Three variants of model (1) are explored in this paper. The first follows the earthquake prediction methodology and involves judgemental choice of the free parameters $(c_1, \ldots, c_N, b)$ and series transformations that are not usually found in econometric applications. This variant is interesting because of its foundation in the geophysical applications but it has the drawback of requiring significant user judgement to implement. The other two variants incorporate more conventional econometric methods: the parameters are estimated by non-linear least squares, and in one variant standard linear filters are applied to the series. We also explore a modification of the alarm in model (1). Our analysis uses monthly data, and because model (1) is based on data only from the current month, the alarm could have high-frequency noise. We therefore also consider a smoothed alarm which issues a recession warning at date t if A_t signals an alarm in the current month or in any of the three preceding months.

Interestingly, it turns out that the predictions made by these three variants are quantitatively similar. For the full sample of cyclical peaks since 1961, we find that this model provides a recession warning in the months before each of the five cyclical peaks from 1961 to 1990, and, depending on the specific model, it provides few or no false alarms. This pattern of reliable performance holds up in a recursive (pseudo out-of-sample) forecasting experiment, and in this sense the model is found to be robust.

The data and preliminary transformations are described in the next section. Estimation methods and results based on the full sample are presented in the third section. The judgemental model is examined in more detail in the fourth section. The fifth section focuses on the two pattern-recognition models with econometrically estimated parameters, and subjects them to a simulated out-of-sample comparison with some more conventional multivariate models for recession probabilities based on the same leading indicators. The final section presents conclusions.

THE DATA AND PRELIMINARY TRANSFORMATIONS

We use monthly macroeconomic data for the United States for the period from January 1959 to April 1996. Over this period, there were six complete recessions (cyclical peaks followed by cyclical troughs). The cyclical peaks and troughs (NBER dates) are listed in Table I. The

Table I. NBER-dated cyclical turning points

Episode	Peak	Trough
1	1960:4	1961:2
2	1969:12	1970:11
3	1973:11	1975:3
4	1980:1	1980:7
5	1981:7	1982:11
6	1990:7	1991:3

Source: National Bureau of Economic Research.

focus of this exercise is predicting recessions, and we therefore consider the five complete expansionary episodes: W_1: 1961: 08–1969:12 (101 months); W_2: 1971: 05–1973:11 (31 months); W_3: 1975: 09–1980:01 (53 months); W_4: 1981:01–1981:07 (7 months); and W_5: 1983:05–1990:07 (87 months). The beginning dates of these episodes were set at six months after the previous cyclical trough. The six-month lag was chosen to allow for recognition lags in *ex post* identification of peaks and troughs, and because by convention turning points are dated so that expansions and recessions have a minimum length of six months. In all, these episodes contain 279 months.

The forecasting models use five leading economic indicators and a coincident index. The purpose of this paper is to explore model (1), not to compare the forecasting properties of different leading indicators. The indicators used here were selected prior to performing any of the calculations in this project. In particular no further variable selection was performed to improve the performance of a particular model. The series were selected judgementally based on their coverage of different aspects of the economy, the consistency of data definitions and availability, and their past performance as leading indicators. The five leading indicators used are:

G10FF. Difference between the interest rate on 10-year US Treasury bond and the federal funds interest rate, on an annual basis.
FYGM3. Interest rate on 90-day US Treasury bills at an annual rate (in per cent).
INVMTQ. Total manufacturing and trade inventories, in real (1987) dollars.
LUINC. Average weekly initial claims for state unemployment insurance.
LHEL. Index of help wanted advertising in newspapers.[1]

Considerable research has demonstrated the historical role of the term structure spread as a leading indicator (see, for example, Stock and Watson, 1989; Estrella and Mishkin, 1996). The specific long/short spread used here, the 10-year Treasury bond/federal funds spread, is the spread currently used in the Composite Index of Leading Indicators (LEI) maintained by The Conference Board (this index was formerly maintained by the US Department of Commerce). The interest rate plays an important role in many macro models and also generally leads economic activity (cf. Stock and Watson, 1998). Inventories were included to provide a leading indicator from the production side of the economy. Average weekly unemployment insurance claims are included in The Conference Board's LEI. Help wanted advertising is another leading indicator that measures labor market tightness, which also leads the business cycle (Stock and Watson, 1998).

The final variable included in the model is a measure of overall economic activity, the Stock–Watson (1989) coincident index of overall monthly economic activity (a weighted average current and past values of non-agricultural employment, industrial production, manufacturing and trade sales, and real personal income), here referred to as the *XCI*. The *XCI* is close numerically to The Conference Board's Composite Index of Coincident Indicators. The *XCI* is included because past values of the *XCI* are useful in predicting its future values and future turning points in the *XCI* are close to NBER-dated turning points.

Many of these series exhibit high-frequency noise and/or trends, either stochastic or deterministic. These series were therefore subject to further transformation prior to inclusion in the model. Two alternative transformations were therefore used. The first entails transformation to quarterly

[1] Data sources: *XCI*: authors' calculations. The other series were obtained from CITIBASE. *IVMTQ*, *LHEL*, *LUINC*, *FYGM3* are CITIBASE mnemonics, and *G10FF* = *FYGT10* – *FYFF*.

growth rates (*QG*); this is a procedure based on standard linear filtering theory and as such is familiar in economic applications. The second is a procedure which we will refer to as local linear (*LL*) detrending which entails interactive user judgement and was developed for the original geophysical applications of the model.

For the quarterly growth rate transformation, y_{k_t} is computed as $(1 - L)^3 x_{k_t}$, where x_{k_t} denotes the natural logarithm of the original series. Because $(1 - L)^3 = (1 + L + L^2)(1 - L)$, this has the interpretation as a three-quarter moving average of the first differences of the series. This constitutes a bandpass filter which filters out both linear deterministic trends and stochastic trends that are integrated of order one. It also filters out high-frequency noise (albeit imperfectly because of a side lobe in the transfer function of this filter at the highest frequencies). For results based on *QG* detrending, this transformation was applied to *XCI*, *LHEL*, *LUINC*, and *IVMTQ*. The term spread, *G10FF*, has little high-frequency noise and little or no trend. The 90-day Treasury Bill rate, *FYGM3*, does not exhibit a deterministic trend, and unit root tests are ambiguous about *FYGM3* exhibits a stochastic trend. In levels, it has little power at high frequencies. For results based on *QG* detrending, these two series were therefore not further transformed.

The local linear detrending methods were developed for use in earthquake prediction and were applied here using that methodology. Applying these transformations requires user judgement. The series *XCI*, *IVMTQ*, and *FYGM3* were transformed to be deviations from their trend value over the current expansion. Specifically, the series y_t was regressed on $(1, t)$ by recursive least squares over the period $t = \tau_s + 1, \ldots, s - 1$, where τ_s is the date of the most recent cyclical trough as of month s. This yields the estimated intercept and trend coefficients $\hat{\alpha}_s$ and $\hat{\beta}_s$. The transformed series, $\tilde{y}_s$, is the deviation from this local trend, $\tilde{y}_s = y_s - \hat{\alpha}_s - \hat{\beta}_s s$. The series so transformed are denoted *XCIR*, *INVR*, and *FYG3R*. The series *LHEL* and *LUINC* were transformed to be weighted averages of their changes, specifically, they were transformed to be the estimated slope coefficients of a local linear least squares regression of the series on $(1, t)$, for $t = s - m, \ldots, s$. For *LHEL*, the value $m = 5$ was used, and for *LUINC*, the value $m = 10$ was used. The transformed values of these two series are denoted *LHK5* and *LUK10*. No transformation was applied to *G10FF*. Because these transformations are recursive they maintain the information set of the original variables, augmented by the trough dates.[2] A representative locally linearly detrended series, *INVR*, is plotted in Figure 1 (the values of *INVR* are arbitrarily set to zero during recessions; these observations are not used).

The signs of the series are set so that, after series transformation, a one is indicated by values of *XCI*, *IVMTQ*, *LHEL*, and *G10FF* below the threshold and by values of *LUINC* and *FYGM3* above the threshold. For example, a recession is signalled by: a value of the coincident indicator series *XCI* sufficiently below its trend over the current expansion; a low value of the spread between the long- and short-term interest rates, *G10FF*; and a value of the 90-day Treasury bill rate series *FYGM3* that is far above its trend value over the current expansion.

[2] The NBER business cycle dating committee announces cyclical peaks and troughs with a variable lag, which can be a year or longer. Thus the most recent trough date is not known during the initial part of an expansion. This concern is mitigated here somewhat by starting the W_i episodes six months into the recession. An alternative would to be define these functionals on a sufficiently long sliding time window.

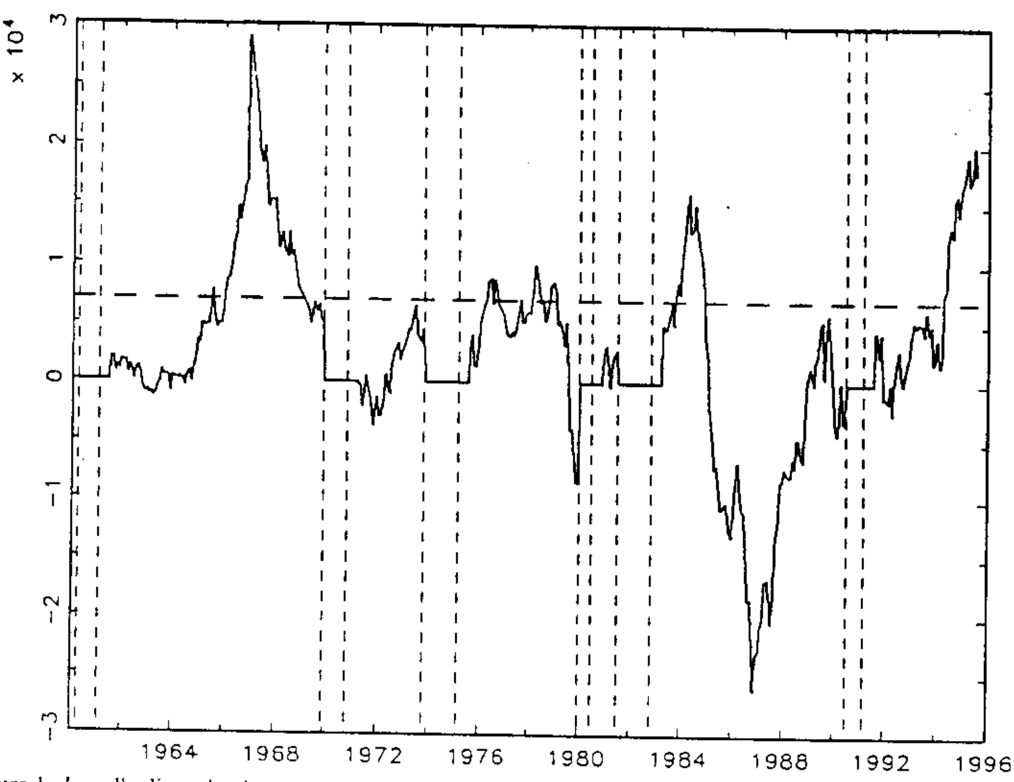

Figure 1. Locally linearly detrended investment (INVR, solid line), discretization threshold (horizontal dashed line), and NBER-dated cyclical turning points (vertical dashed lines), 1960–95

FULL-SAMPLE RESULTS

Parameter selection

The model depends on $N+1$ free parameters ($c_1, \ldots, c_N, b$). In this paper we consider two different approaches to the specification of these parameters. The first uses the methods developed in the original application of this model to earthquake prediction, and employs algorithms developed at the International Institute for Earthquake Prediction Theory and Mathematical Geophysics in Moscow. This entails judgemental selection of the parameters over a restricted set using interactive software. Because of its history in earthquake prediction, it is intriguing to apply it to the problem of recession prediction. This methodology is, however, difficult for other researchers to apply to new applications.

The second approach requires less user judgement because parameters are estimated by non-linear least squares. Specifically, the objective function is $\Sigma_t(I_t - A_t)^2$, where $I_t = 1$ if an NBER-dated cyclical peak will occur within nine months, and $= 0$ otherwise. The nine-month horizon was chosen arbitrarily as representative of short-term recession forecasting horizons.[3] Because

[3] Results have also been computed for the estimated models at 6- and 12-month horizons. The results are generally similar, although of course the accuracy of all the models is better at the 6-month horizon, and worse at the 12-month horizon, relative to the 9-month horizon reported here. To save space, only results for the 9-month horizon are reported.

the alarms are defined in terms of indicator functions, the objective function is not differentiable in any of the parameters, so estimation must proceed using random search methods. Here, a simulated annealing algorithm was used with 10,000 random draws.

Empirical results

This section summarizes results for three variants of the model: judgemental parameters with *LL* detrending; estimated parameters with *LL* detrending; and estimated parameters with *QG* detrending. The full sample consists of the combined expansion episodes, denoted by $W = W_1 \cup W_2 \cup W_3 \cup W_4 \cup W_5$.

The parameters for the three variants are given in Table II, along with some summary statistics for the models. The original parameters $\{c_k\}$ are quantiles in the units of the data; for ease of interpretation they are presented here as the fractions of the empirical distribution to which these quantiles correspond. In the judgemental model, the cutoff b was set to 2 based on inspection of the individual 0/1 indicators $\{y_{kt}\}$. A horizon of nine months was selected for evaluating these models. Accordingly, the false negative rate is the fraction of months in which an alarm is not signalled but a recession is in fact nine months or less away; the false positive rate is the fraction of months in which an alarm is issued but a recession is in fact more than nine months away; and the overall accuracy rate is the fraction of months in the prediction episode in which the 0/1 prediction was, in retrospect, correct.

Table II. Summary results, pattern-recognition models

Parameter	Judgemental/*LL*	Estimated/*LL*	Estimated/*QG*
Estimated parameters			
p_{XCI}	0.25 (1:3)	0.266	0.873
p_{IVMTQ}	0.75 (3:1)	0.764	0.839
p_{G10FF}	0.10 (1:9)	0.125	0.123
p_{LHEL}	0.333 (1:2)	0.845	0.019
p_{LUINC}	0.833 (5:1)	0.007	0.999
p_{FYGM3}	0.75 (3:1)	0.514	0.404
b	2	1	2
Summary statistics for prediction of a recession within 9 months:			
False neg. rate	0.311	0.111	0.222
False pos. rate	0.027	0.023	0.008
Accuracy	0.931	0.964	0.961

The entries in parentheses are the probabilities for the judgemental model expressed as odds ratios. The estimated parameters are non-linear least squares estimates computed by simulated annealing with 10,000 random draws.

For the models with *LL* detrending, the estimated and judgemental parameters are quite close for the *XCI*, inventories and the term spread, but differ for the other three variables. In the estimated model, the quantile for *LUINC* is effectively driven to zero, so that the *LUINC* always signals a recession which has the effect of dropping this variable from the model. For the *QG* model, *LUINC* is also effectively dropped, although in the opposite way: the cutoff is such that *LUINC* never delivers a recession signal, but the number of series for which a signal is required to sound an alarm drops from five to four in the estimated *QG* model. In both the estimated models, the term spread is not transformed, and essentially the same parameter values are estimated for

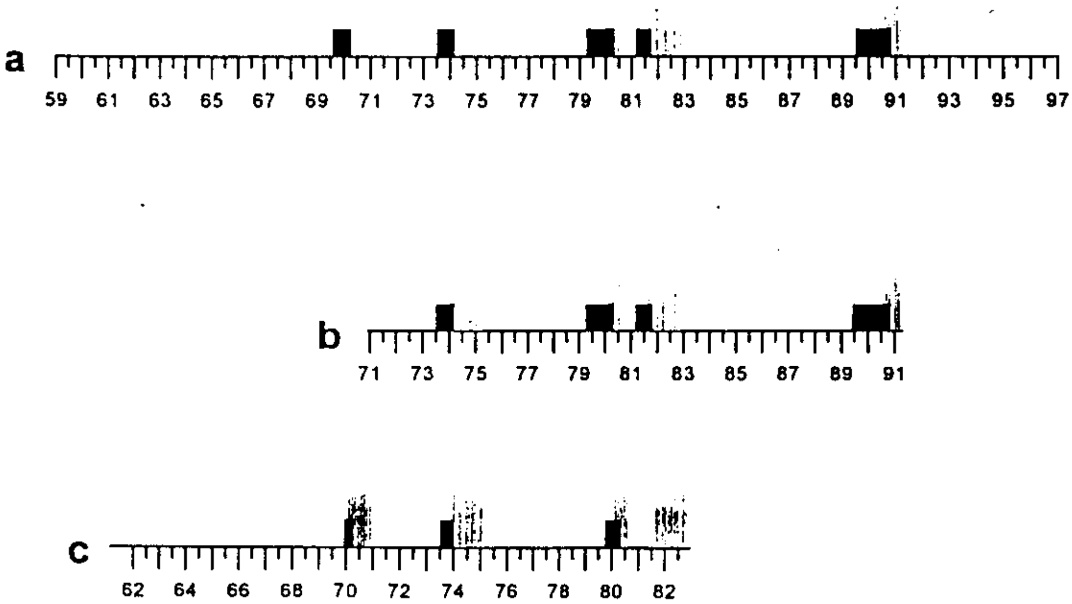

Figure 2. Alarms (shown by black bars): (a) main version; (b) prediction history; (c) reverse prediction history

its cutoff. As expected, when measured by overall accuracy the two estimated models have somewhat better performance than the judgemental model.

The four-month smoothed alarms are shown in Figure 2(a) by black bars. The periods covered by the alarms inside the set *W* are listed in Table III. Note that each alarm extends into the first two or three months of a recession. Evidently all five recessions were preceded by continuous alarms. The longest alarm lasted 13 months, one alarm lasted 10 months, and three alarms lasted 5 months. There were no continuous alarms not ending in a recession. The total duration of the alarms was 38 months, 13.6% of the time covered by the analysis (*W* set). There was no recession from 1991:3 to 1996:4, and the application of this algorithm to this period gives no alarms.

Although the pattern-recognition model is tightly parameterized relative to alternative non-linear models, nonetheless it has several free parameters that have been selected, either informally or formally, using a small number of recessions. We therefore turn to an analysis of the sensitivity of these models to the parameters. This is done in two parts. The next section performs a sensitivity analysis on the judgemental pattern-recognition model. In the fifth section, the

Table III. Alarms within the set *W*

#	Period	Duration (months)
1	1969:08–1969:12	5
2	1973:07–1973:11	5
3	1979:04–1980:01	10
4	1981:03–1981:07	5
5	1989:07–1990:07	13

estimated pattern-recognition model is subjected to a recursive, or simulated out-of-sample, comparison with some standard multivariate benchmark models.

SENSITIVITY ANALYSIS OF JUDGEMENTAL MODEL

The judgemental model involves considerable data fitting. (Dr W. Press, discussing an early example of such an approach, reminded one author of the words of E. Fermi: 'With four exponents I can fit an element.') This section therefore examines the sensitivity of this model to changes in some of the parameters, assesses the subsample stability of the model using the full-sample judgemental cutoff probabilities, and examines the effect of omitting one or more of the indicators from the judgemental model.

We first consider the pseudo out-of-sample performance of the algorithm by a recursive experiment. Given the odds ratios in Table II, the thresholds were first determined using data from episode W_1, and the prediction algorithm was then applied to the episode W_2. The thresholds were then recomputed using data from $W_1 \cup W_2$ (and the odds ratios in Table II) and the algorithm was then used to forecast episode W_3. This was repeated to produce recursive forecasts of episodes W_4 (using $W_1 \cup W_2 \cup W_3$) and W_5 (using $W_1 \cup W_2 \cup W_3 \cup W_4$). The four-month smoothed alarms obtained in the experiment are shown in Figure 2(b) and Table IV; the latter contains the thresholds and the alarms within the prediction set. The agreement with the main version is practically complete. The only difference is that the last alarm became a month longer.

The recursive experiments should not be interpreted literally as what would have happened had this approach been applied in real time, in particular the odds ratios $p{:}q$ are fixed here at their full-sample judgemental values. A recursive (simulated real-time) evaluation of the models with estimated parameters is presented in the next section.

We also performed a reverse recursive experiment, in which the data set was expanded in reverse chronological order: given the odds ratios, the thresholds were estimated on W_5, $W_4 \cup W_5$, $W_3 \cup W_4 \cup W_5$, and $W_2 \cup W_3 W_4 \cup W_5$ and the resulting model was then used to make predictions over W_4, W_3, W_2, and W_1 respectively. The alarms obtained in this experiment are shown in Figure 2(c). Table V lists the thresholds and alarms. Most of the results from this reverse recursive exercise are the same as from the full sample. However, the recession in 1981–2 was missed and the alarms before the recessions in 1969–70 and in 1980 are shorter than in the full-sample version. Technically this is due to the differences in the thresholds: comparing to the main version they are smaller for the functions *XCIR*, *INVR*, and *LHK5*, and larger for the functions *LUK10* and *FYG3R*.

In another set of experiments, we altered the odds ratio $p{:}q$, as indicated in Table VI; the table also shows the corresponding change in the alarms within the set W. Again, the results are similar to those in the main version. Note that the second part of the alarm before the 1973–5 recession (line 8) starts in the same month as the recession and has no intersection with the set W. Of course, this alarm still has some predictive value because recessions are formally identified only after a lag of at least several months.

We also performed experiments in which we dropped, one at a time, each of the six economic indicators, and computed forecasts using the thresholds estimated over the full sample. The alarms are declared using $b = 1$ instead of $b = 2$ in the main version. The major change is failure to predict the 1981–2 recession after the elimination of *XCIR*, *INVR*, *G10FF*, or *LHK5*. The smallest changes are caused by elimination of *LUK10*. When *INVR*, *G10FF*, *LHK5*, or

Table IV. Results of the historical experiment

#	Set for definition of thresholds	Thresholds for functions						Results		
		XCIR	*INVR*	*G10FF*	*LHK5*	*LUK10*	*FYG3R*	Examined set	Alarms within the set *W*	Duration of alarms (months)
1	W_1	−0.77	11531	−0.61	1.38	12.15	0.26	W_2	1973:07–1973:11	5
2	$W_1 \cup W_2$	−0.74	8548	−1.32	2.47	11.17	0.51	W_3	1979:04–1980:01	10
3	$W_1 \cup W_2 \cup W_3$	−0.78	8036	−1.59	3.04	20.41	0.48	W_4	1981:03–1981:07	5
4	$W_1 \cup W_2 \cup W_3 \cup W_4$	−0.86	7991	−1.99	2.78	20.16	0.92	W_5	1989:06–1990:07	14

Table V. Results of the reverse historical experiment

#	Set for definition of thresholds	Thresholds for functions						Results		
		XCIR	*INVR*	*G10FF*	*LHK5*	*LUK10*	*FYG3R*	Examined set	Alarms within the set *W*	Duration of alarms (months)
1	W_5	−3.65	5305	−0.61	−0.69	29.87	1.27	W_4	No	0
2	$W_5 \cup W_4$	−3.61	4511	−0.95	−1.03	27.89	1.18	W_3	1979:10–1980:01	4
3	$W_5 \cup W_4 \cup W_3$	−3.30	7024	−1.38	0.97	44.47	1.29	W_2	1973:07–1973:11	5
4	$W_5 \cup W_4 \cup W_3 \cup W_2$	−3.04	6320	−1.59	2.33	29.87	1.29	W_1	1969:12	1

Table VI. Results of changing $p:q$

#	Function	New $p:q$	Threshold	Alarms, different from those in the main version (Table III)	Total duration of alarms in % of total time
1	*XCIR*	1:2	−0.98	1989:06–1990:07	14.0
2		1:4	−2.20	1981:04–1981:07	13.3
3	*INVR*	2:1	6167	1969:08–1969:10, 1969:12	13.3
4		4:1	8870	1979:03–1908:01	14.0
5	*G10FF*	1:7	−0.99	No	13:6
6		2:23	−1.90	1969:09–1969:12, 1973:10–1973:11	12.2
7	*LHK5*	1:1	5.38	1985:03–1985:08 (added)	15.8
8		1:3	−0.33	1973:07–1973:09, alarm after 1973:11	12.9
9	*LUK10*	4:1	19.78	No	13:6
10		6:1	29.87	1979:05–1980:01	13:3
11	*FYG3R*	2:1	0.54	1969:07–1969:12	14:0
12		4:1	1.17	1969:08–1969:10, 1969:12	13:3

FYGR3 is eliminated, the alarm before the recession in 1973–5 starts in the same month as the recession.

We also considered prediction using a single indicator, in which case an alarm is declared if that indicator passes its full-sample threshold. The performance of these single-indicator models was worse than the multivariate models, in some cases because recessions were missed, in others, because predicted recessions did not occur.

Performance of the different variations of the prediction algorithm is juxtaposed in the error diagram (Molchan, 1994), shown in Figure 3.

COMPARISON WITH BENCHMARK MODELS

This section compares the pattern-recognition forecasts to those of two conventional, simple models in a recursive, or simulated out-of-sample, experiment. All three versions of the pattern recognition model in Table II are considered here. The benchmark comparison models are a linear probability model and a probit model, which predict the probability of a recession beginning some time in the next nine months, using current values of the six economic indicators. Like the pattern-recognition model, the linear probability model has seven free parameters, while the probit model has eight. For both the linear probability and probit models, an alarm was indicated if a recession probability exceeded 50%. These models were estimated using both *LL* and *QG* detrended data.[4]

[4] Attempts were also made to estimate Hamilton (1989) switching models using these indicators. However, when more than two indicators were included as predictors, these models failed to converge because the parameters were essentially unidentified. This is not surprising because these models contain a large number of parameters to be estimated and there are only six recessions in this data set. This underscores the problem raised in the Introduction faced by empirical models generally, and non-linear models in particular, of predicting recessions using many leading indicator and only a handful of recessions.

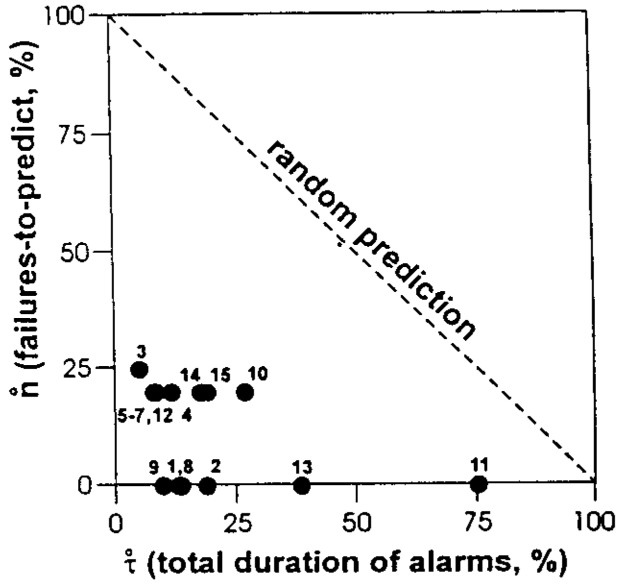

Figure 3. Error diagram: (1) main version; (2) prediction history; (3) reverse prediction history; (4)–(9) exclusion of variables: (4) *XC1R*, (5) *INVR*, (6) *G10FF*, (7) *LHK5*, (8) *LUK10*, (9) *FYG3R*, (10)–(15), prediction using a single variable: (10) *XC1R*, (11) *INVR*, (12) *G10FF*, (13) *LHK5*, (14) *LUK10*, (15) *GYG3R*

The performance of these models in predicting recessions at the nine-month horizon is summarized in Table VII, panel A for the full sample. For comparison purposes, summary measures of the performance of the judgemental model are repeated there as well. All models have overall accuracy rates that exceed 92%. The estimated pattern-recognition models outperform the linear probability and probit models based on this criterion. Interestingly, the choice of the non-standard *LL* or the more conventional *QG* detrending method does not have a large qualitative impact on the results or on the rankings across models.

An interesting question is whether the non-linear transformations $\{y_k\}$ contain information not present in the linear regression. To address this question $\{y_{kt}\}$ were included as six additional regressors in the linear probability model. For the *QG* detrended regression, the *F*-statistic for these six non-linear variables is 20.18; for the *LL* detrended regression, the *F*-statistic is 25.33. The conventional *F* distribution is suspect here because the $\{y_{kt}\}$ variables contain estimated parameters (the cutoffs) and because of overlapping horizons in the dependent variable, and we do not provide any distribution theory for these statistics. Nonetheless, the values of these *F*-statistics far exceeds all critical values, which suggests that these non-linear transformations have important marginal predictive content beyond the linear terms.

Panels B–E report the results of four recursive experiments. The experiments differ by the estimation and evaluation data sets. For example, in panel C, the parameters were estimated on the sample $W_1 \cup W_2$, but the reported accuracy rates refer to the episodes $W_3 \cup W_4 \cup W_5$. In each experiment, the parameters of the linear model were estimated by ordinary least squares; the parameters of the probit model were estimated by maximum likelihood; and the seven parameters

Table VII. Comparison of pattern recognition, linear probability, and probit models

Algorithm	Parameters	Trans-formation	Smoothed alarms?	False negative rate	False positive rate	Overall accuracy rate
(A) Full sample						
Pattern rec.	Estimated	*QG*	N	0.222	0.008	0.961
Linear prob.	Estimated	*QG*	N	0.400	0.019	0.924
Probit	Estimated	*QG*	N	0.267	0.039	0.928
Pattern rec.	Estimated	*LL*	N	0.111	0.023	0.964
Linear prob.	Estimated	*LL*	N	0.333	0.015	0.938
Probit	Estimated	*LL*	N	0.222	0.027	0.944
Pattern rec.	Judgemental	*LL*	N	0.311	0.027	0.931
Pattern rec.	Judgemental	*LL*	Y	0.200	0.027	0.947
(B) Estimation: W_1, prediction: $W_2 \cup W_3 \cup W_4 \cup W_5$						
Pattern rec.	Estimated	*QG*	N	0.611	0.074	0.828
Linear prob.	Estimated	*QG*	N	0.333	0.056	0.894
Probit	Estimated	*QG*	N	0.528	0.012	0.894
Pattern rec.	Estimated	*LL*	N	0.389	0.086	0.859
Linear prob.	Estimated	*LL*	N	0.056	0.086	0.919
Probit	Estimated	*LL*	N	0.111	0.198	0.818
(C) Estimation: $W_1 \cup W_2$, prediction: $W_3 \cup W_4 \cup W_5$						
Pattern rec.	Estimated	*QG*	N	0.185	0.185	0.815
Linear prob.	Estimated	*QG*	N	0.000	0.178	0.852
Probit	Estimated	*QG*	N	0.000	0.489	0.593
Pattern rec.	Estimated	*LL*	N	0.296	0.089	0.877
Linear prob.	Estimated	*LL*	N	0.278	0.160	0.827
Probit	Estimated	*LL*	N	0.037	0.267	0.772
(D) Estimation: $W_1 \cup W_2 \cup W_3$, prediction: $W_4 \cup W_5$						
Pattern rec.	Estimated	*QG*	N	0.333	0.174	0.798
Linear prob.	Estimated	*QG*	N	0.444	0.105	0.837
Probit	Estimated	*QG*	N	0.500	0.070	0.856
Pattern rec.	Estimated	*LL*	N	0.778	0.163	0.731
Linear prob.	Estimated	*LL*	N	0.389	0.140	0.817
Probit	Estimated	*LL*	N	0.278	0.360	0.654

of the pattern-recognition model were estimated by non-linear least squares using simulated annealing (10,000 random draws).

Several results are apparent. The other non-linear model, the probit model, is unstable, and produces forecasts that have quite low accuracy in several of the experiments. The ranking between the linear probability model and the estimated pattern-recognition model is less clear; generally speaking, the estimated pattern-recognition model does better than the linear model when the estimation data set is longer. The estimated pattern-recognition model appears to be more robust than the probit model in the sense of being less prone to significant deteriorations in accuracy.

CONCLUSIONS

Taken together, the evidence presented here suggests that these simple binary transformations of these economic indicators have significant predictive content for recessions. The pattern-recognition model has the interesting feature of being highly non-linear yet only having the same number of free parameters as a linear regression model. The sensitivity analysis of the judgemental pattern-recognition model and the simulated out-of-sample analysis of the estimated pattern-recognition model suggests that this model is stable to outliers and changes in regimes. It is striking that these models, in which the information in the data is reduced to binary indicators, has predictive content comparable to or, in many cases, better than that of more conventional models.

Several caveats should be emphasized. First, despite the subsample and pseudo-out-of-sample analysis, the only true test of this algorithm is to track the performance of the forecasts from the pattern-recognition models during the current expansion and beyond.[5] Second, only two methods for parameter selection (judgemental and full non-linear least squares) have been investigated here, and it might be that estimation methods could be refined to further reduce the effective number of parameters of the pattern-recognition model. Third, although the six economic indicators used here were selected because of their quality (documented elsewhere) as coincident or leading economic indicators, there is no guarantee that these indicators will continue to be good predictors in the future. Fourth, the assessment here has relied on pseudo-out-of-sample forecasting exercises to assess the validity of this algorithm, and it would be useful (but difficult) to augment this with formal statistical measures of sampling uncertainty. Despite these cautionary notes, however, we consider these results encouraging and supportive of additional investigation of this approach.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous referees and the editor for helpful suggestions.

REFERENCES

Berk, J. M. and Bikker, J. A., 'International interdependence of business cycles in the manufacturing industry: the use of leading indicators for forecasting and analysis', *Journal of Forecasting*, **14** (1995), 1–23.

Diebold, F. X. and Rudebusch, G. D., 'Scoring the leading indicators', *Journal of Business*, **62** (1989), 000–000.

Estrella, A. and Mishkin, F. S., 'The yield curve as a predictor of U.S. recessions', *Current Issues in Economics and Finance*, Federal Reserve Bank of New York, **2** (1996) No. 7, June.

Fair, R., 'Estimating event probabilities from macroeconomic models using stochastic simulation', in Stock, J. H. and Watson, M. W. (eds), *Business Cycles, Indicators and Forecasting*, (NBER Studies in Business Cycles, Vol. 28) 1993, pp. 157–78.

[5] We can provide some evidence on this. The analysis in this paper used data through 1996:4. In a late revision of this paper we had the opportunity to update the signals using data through 1997:3. This constitutes a brief, truly out-of-sample experiment using the judgemental algorithm. During the ten months from May 1996 through March 1997 (inclusive), no recession alarm has been sounded by the judgemental model. As of May 1998, no cyclical turning point had been declared by the NBER, nor is one suspected to have occurred, during this period. This brief episode provides a test of the judgemental model, which it passes.

Gelfand, I., Keilis-Borok, V. I., Knopoff, L., Press, F., Rantsman, E., Rotwain, I. and Sadovsky, A., 'Pattern recognition applied to earthquake epicenters in California', *Phys. Earth Planet. Inter.*, **11** (1976), 227–83.

Hamilton, J. D., 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57 (1989), 357–84.

Keilis-Borok, V. I. and Kossobokov, V. G., 'Premonitory activation of seismic flow: algorithm M8', *Physics of the Earth and Planetary Interiors*, **61** (1990), 73–83.

Keilis-Borok, V. I. and Lichtman, A., 'Understanding and prediction of large and complex unstable systems in the absence of basic equations: concepts and selforganization and similarity', in Costa, G., Calucci, G. and Giorgi, M. (eds), *Proceedings of the First International Symposium on Conceptual Tools for Understanding Nature*, Trieste 26–28 September, 1990, p. 145.

Lichtman, A. and Keilis-Borok, V. I., 'Aggregate-level analysis and prediction of midterm senatorial elections in the United States', *Proc. Natl. Acad. Sci. USA*, 86, December 1989, pp. 10176–80.

Molchan, G. M., 'Models for optimisation of earthquake prediction', *Computational Seismology and Geodynamics*, (AGU, Washington, DC) **2** (1994) 1–10.

Mostaghimi, M. and Rezayat, F., 'Probability forecast of downturn in U.S. economy using classical statistical decision theory', *Empirical Economics*, **21** (1996), 255–79.

Press, F. and Allen, C., 'Pattern of seismic release in the Southern California region', *Journal of Geophysical Research*, **100** (1995), 6421–30.

Sims, C. A., 'A nine-variable probabilistic macroeconomic forecasting model', in Stock, J. H. and Watson, M. W. (eds), *Business Cycles, Indicators, and Forecasting*, NBER Studies in Business Cycles, Vol. 28, 1993, pp. 179–212.

Stock, J. H. and Watson, M. W., 'New indexes of leading and coincident economic indicators', *NBER Macroeconomics Annual*, 1989, pp. 351–94.

Stock, J. H. and Watson, M. W., 'A procedure for predicting recessions with leading indicators', in Stock, J. H. and Watson, M. W. (eds), *Business Cycles, Indicators, and Forecasting*, NBER Studies in Business Cycles, Vol. 28, 1993, pp. 95–156.

Stock, J. H. and Watson, M. W., 'Business cycle fluctuations in U.S. macroeconomic time series', in Taylor, J. and Woodford, M. (eds), *Handbook of Macroeconomics*, forthcoming, 1998, 000–000.

Vorobieva, I. A. and Levshina, T. A., 'Prediction of a second large earthquake based on aftershock sequence', *Computational Seismology and Geodynamics*, (AGU, Washington, DC), **2** (1994), 27–36.

Authors' biographies:
(To be supplied)

Authors' addresses:
V. Keilis-Brook, and **A. Soloviev**, International Institute for Earthquake Prediction Theory and Mathematical Geophysics, Warshavskoye sh., 79, kor. 2, Moscow 113556, Russia.

J. H. Stock, Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA.

P. Mikhalev, McKinsey & Co., 400 South Hope Street, Suite 700, Los Angeles, CA 90071-2890, USA.

