Prediction of the start of next recession using latent factors

Ah-Hin Pooi, Huei-Ching Soo, and Wei-Yeing Pan

Citation: AIP Conference Proceedings **1782**, 050013 (2016); doi: 10.1063/1.4966103
View online: http://dx.doi.org/10.1063/1.4966103
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1782?ver=pdfcov
Published by the AIP Publishing

Prediction of the Start of Next Recession Using Latent Factors

Ah-Hin Pooi[1, a)] , Huei-Ching Soo[2, b)] and Wei-Yeing Pan[3, c)]

[1] *Sunway University Business School, Sunway Universitiy, Malaysia.*
[2] *Heriot-Watt University Malaysia.*
[3] *Universiti Tunku Abdul Rahman, Malaysia.*

[a)] ahhinp@sunway.edu.my
[b)] H.Soo@hw.ac.uk
[c)] panwy@utar.edu.my

Abstract. The data on the binary recession variable and the latent factors extracted from a large set of economic variables are fitted with a multivariate power-normal distribution. A conditional distribution for the recession variable is obtained from the fitted multivariate distribution. The results based on the US economic data show that the 2.5% point of the conditional distribution provides a good indicator for the start of the next recession.

INTRODUCTION

Prediction of the start of next economic recession is of great importance to households, central bankers, investors and government policy makers.

A popular approach for predicting recession is given by the probit model which incorporates the economic variables ([1]-[6]). Chen el al ([7]) instead used a factor model which contains a few latent factors extracted from a large set of economic variables. The factor model becomes dynamic when the latent factors are modelled as a multivariate time series ([8]-[9]). Machine learning provides yet another approach for predicting recession ([10]-[12]).

Recently, Pooi and Koh ([13]) fitted the data on recession variable and selected economic variables with the multivariate power-normal distribution and used the mean together with the 2.5% and 97.5% points of the conditional distribution for the recession variable to form indicators for the start of the next recession.

In this paper we adopt the approach of [13] to predict recession. However, instead of using selected economic variables, we use a number of latent factors extracted from a large set of economic variables. The numerical results based on the US economic data reveal that with only three latent factors, the resulting model yields results which are comparable to those of the models based on the *optimal* economic variables chosen from a large set of economic variables.

The layout of the paper is as follows. In Section 2, we give a short introduction to the method for constructing prediction interval using multivariate power-normal distribution. In Section 3, we describe the prediction of recession using latent factors. Section 4 gives the numerical results based on the US economic data. Section 5 concludes the paper.

The 4th International Conference on Quantitative Sciences and Its Applications (ICOQSIA 2016)
AIP Conf. Proc. 1782, 050013-1–050013-6; doi: 10.1063/1.4966103
Published by AIP Publishing. 978-0-7354-1444-0/$30.00

METHOD BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION

Let us begin with the power transformation introduced in Yeo and Johnson ([14]):

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} \left[(z+1)^{\lambda^+} - 1\right]/\lambda^+, & (z \geq 0, \lambda^+ \neq 0) \\ \log(z+1), & (z \geq 0, \lambda^+ = 0) \\ -\left[(-z+1)^{\lambda^-} - 1\right]/\lambda^-, & (z < 0, \lambda^- \neq 0) \\ -\log(-z+1), & (z < 0, \lambda^- = 0) \end{cases} \quad (1)$$

If z in Equation (1) has the standard normal distribution, then $\tilde{\varepsilon}$ is said to have a power-normal distribution with parameters λ^+ and λ^-.

Let $\boldsymbol{y}$ be a vector consisting of k correlated random variables. The vector $\boldsymbol{y}$ is said to have a k-dimensional power-normal distribution with parameters $\boldsymbol{\mu}, \boldsymbol{H}, \lambda_i^+, \lambda_i^-, \sigma_i, 1 \leq i \leq k$ if

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{H}\boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{y})$, $\boldsymbol{H}$ is an orthogonal matrix, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ are uncorrelated,

$$\varepsilon_i = \sigma_i[\tilde{\varepsilon}_i - \mathbb{E}(\tilde{\varepsilon}_i)]/[var(\tilde{\varepsilon}_i)]^{1/2}, \quad (3)$$

$\sigma_i > 0$ is a constant, and $\tilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^-.

When the values of $y_1, y_2, \ldots, y_{k-1}$ are given, we may find an approximation for the conditional probability density function (pdf) of the last component y_k of $\boldsymbol{y}$ by using the numerical procedure given in Pooi ([15]).

The mean of the conditional distribution is then an estimate of the value of the last component. On the other hand, the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may be regarded respectively as the lower and upper limits of the nominally $100(1-\alpha)\%$ prediction interval for the value of the last component.

When the estimated coverage probability of the prediction interval is close to the target value$1-\alpha$, a small value of the average length of the prediction interval is indicative of good predictive power of the prediction interval.

PREDICTION OF RECESSION USING LATENT FACTORS

Let $\mathbf{x}_t$ be an N-dimensional vector of a large number N of economic time series. The static representation of the factor model is given by

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{e}_t$$

where $\mathbf{e}_t$ is an N x 1 vector of idiosyncratic disturbances, $\mathbf{\Lambda}$ is an N x r matrix of factor loadings and $\mathbf{F}_t$ is an r x 1 vector of common latent factors underlying $\mathbf{x}_t$.

From the data which span over T units of time, we can form a table of T rows with the i-th row representing the observed values of the N economic variables recorded at the i-th unit of time. We next can form the j_w-th sub-table from the j_w to $j_w + n_t - 1$ rows of the table of T rows.

We perform a principal component analysis of the N columns of the observations in the j_w- th sub-table, and from the original set of N principal components, we obtain r principal components $f_1, f_2, \ldots, f_r$ of which f_i has the i-th largest variance.

Let f_{ij} be the value of f_i extracted from the j-th row of economic data from the j_w- th sub-table. We now construct the j_w- th window from the f_{ij} extracted from the j_w- th sub-table such that the j-th row of the j_w- th window consists of the values $\left(f_{1\,j-l+1} f_{2\,j-l+1} \cdots f_{r\,j-l+1} \cdots f_{1j} f_{2j} \cdots f_{rj}, r_{j+1}\right)$ where $1 \leq j \leq n_w + l$, $n_w = n_t - l$ and r_{j+1} is the value of the recession variable observed at the immediate future time point.

An $[lr+1]$-dimensional power-normal distribution is fitted to the data in the initial n_w rows of the j_w- th window. Given the value of the initial lr entries in the last row of the j_w- th window, we use the fitted multivariate power-normal distribution to find a conditional distribution for the recession variable.

The mean of the conditional distribution is then an estimate of the probability of recession at the immediate future time point. The $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may be considered respectively as the lower and upper limits of the nominally $100(1-\alpha)\%$ out-of-sample prediction interval for the recession variable at the immediate future time point.

As in [13], an indicator for the start of recession may be taken to be given by the lower limit of the prediction interval having a value close to zero ,and a signal for possible recession in the near future may be given by the upward movement of the lower limit of the prediction interval towards the value zero.

DATA AND EMPIRICAL RESULTS

The monthly U.S. data on 14 economic variables in the period from Feb 1959 to Nov 2010 are presently used to test the performance of the indicator and signal based on the lower limit of the prediction interval for the recession variable.

Table 1 gives a short description of the 14 U.S. economic variables.

TABLE 1: Description of the US economic variables

Variable No.	Variable name
1	Employment situation
2	Average weekly initial claims for unemployment insurance
3	Consumer goods and materials
4	ISM diffusion index of new orders
5	Nondefense capital goods
6	Building permits for new private housing unit
7	S&P 500 stock price indexes
8	Leading credit index
9	Treasury bond less federal fund rate
10	Average consumer expectation on business and economic conditions
11	Index of leading indicators
12	Index of leading indicators, change from previous month
13	Money supply
14	Index of consumer expectation

The out-of-sample prediction intervals for the model based on r $(1 \le r \le 4)$ latent variables are shown in **FIGURE 1** until **FIGURE 4** along with the value 0 or 1 of the recession variable. Figures 1 and 2 show that when one or two latent factors are used, the upward movement of the lower limit of the prediction interval for the recession variable towards the critical value zero is observed for only some of the recessions after Oct 1978. When three latent factors are used, the link between the upward movement of the lower limit and the imminent recession becomes fairly obvious for almost all the recessions. Furthermore the times when the lower limits cross the zero line are fairly close to the starting times of the recession periods.

When four latent factors are used, the inference based on the lower limit is about the same as that given by the lower limit based on the model with three latent factors.

The inference given by the lower limit based on the model with three latent factors is also quite similar to that given by the models (see [13]) with one or two carefully selected economic variables.

FIGURE 1. Out of sample prediction interval when the model is based on one latent factor ($r = 1$, $l = 2$, $\alpha = 0.05$, $n_w = 200$).

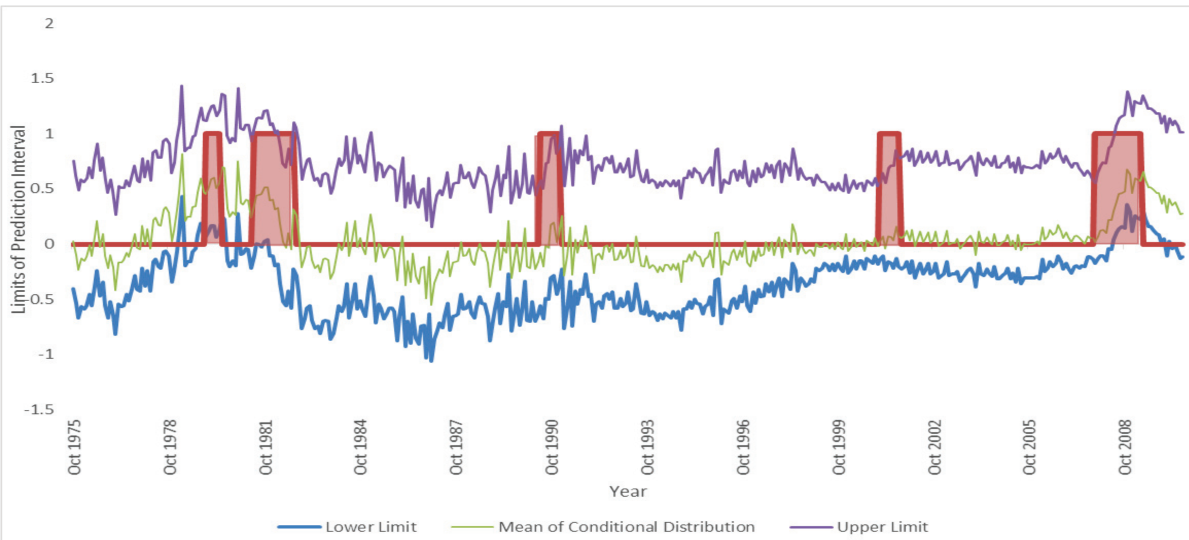

FIGURE 2. Out of sample prediction interval when the model is based on two latent factors ($r = 2$, $l = 2$, $\alpha = 0.05$, $n_w = 200$).

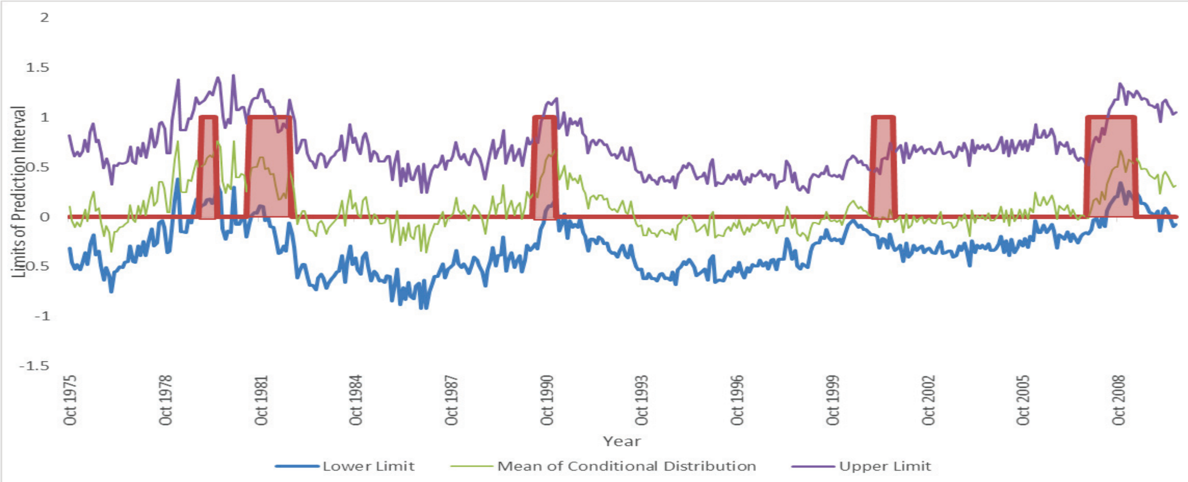

FIGURE 3. Out of sample prediction interval when the model is based on three latent factors ($r = 3$, $l = 2$, $\alpha = 0.05$, $n_w = 200$).

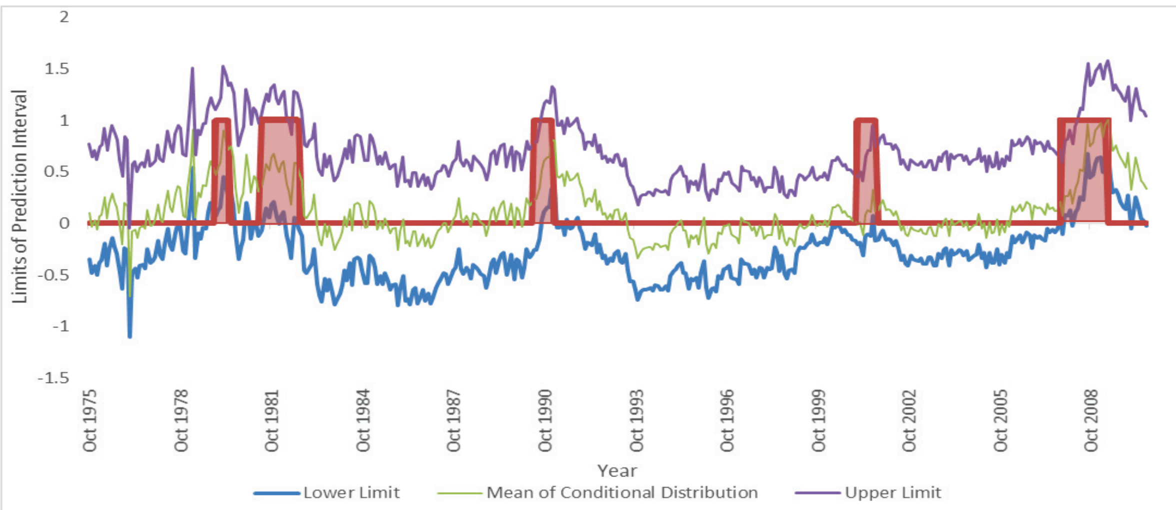

FIGURE 4. Out of sample prediction interval when the model is based on four latent factors ($r = 4$, $l = 2$, $\alpha = 0.05$, $n_w = 200$).

CONCLUSIONS

Generally the signals provided by the models with latent factors or selected economic variables are fairly consistent. However while the indicators provided by some models may lead by a few months in predicting the start of a recession, those given by the other models may predict the same event with some degrees of lag. The problem posed by these slight differences in the predicted times for the start of a recession may be overcome by examining the results provided by a fairly wide range of models, and initialize an early preparation and precaution for the possible start of the next recession as soon as there is a model which yields an indication for recession.

REFERENCES

1. A. Estrella and F. Mishkin, Review of Economics and Statistics, **80**, 45-61 (1998).
2. A. Filardo, Economic Review- Federal Reserve Bank of Kansas City, **84**, 35-56 (1999).
3. M. Chauvet and S. Potter, **24**, 77-103 (2005).

4. J. Wright, "The yield curve and predicting recessions," *Finance and Economics Discussion Series.* Federal Reserve Board, February (2006).
5. J. Silvia, S. Bullard and H. Lai, Business Economics, **43**, 7-18 (2008).
6. H. Kauppi and P. Saikkonen, Review of Economics and Statistics, **90**, 777-791 (2008).
7. Z. Chen, A. Iqbal and H. Lai, Canadian Journal of Economics/Revue canadienne d'économique, **44**, 651-672 (2011).
8. M. Chauvet and Z. Senyuz, International Journal of Forecasting, **32**, 324-343 (2016).
9. S. Fossati, "Dating US business cycles with macro factors. Studies in Nonlinear Dynamics & Econometrics," Forthcoming
10. M. Qi, International Journal of Forecasting, **17**, 383-401 (2001).
11. A. Giusto and J. Piger, **"**Nowcasting US business cycle turning points with vector Quantization," Working paper, University of Oregon, (2013).
12. T. Berge, Journal of Forecasting, **34**, 455-471 (2015).
13. A. H. Pooi and Y. B. Koh, Journal of Accounting, Finance and Economics, **6**, pp 21 – 29, (2016).
14. I. K. Yeo and R. A. Johnson, Biometrika, **87**, 954-9 (2000).
15. A. H. Pooi, Applied Mathematical Sciences, **6,** 5735-48 (2012).