

## Prediction of the Start of Next Recession

Ah-Hin Pooi\* and You-Beng Koh\*\*

*The future value of the binary recession variable is modeled to be dependent on the present and  $l - 1$  past values of a set of  $m$  US economic variables selected from a pool of 14 variables via a conditional distribution which is derived from an  $[ml + 1]$  -dimensional power-normal distribution. The mean together with the 2.5% and 97.5% points of the conditional distribution are used to predict the start of the next US recession. When  $l = 2$  and  $1 \leq m \leq 2$ , some of the models can provide fairly good indicators for the start of the next US recession.*

**JEL Codes:** C1 and E3

### 1. Introduction

The prediction of the start of next recession is important to central bankers, investors and government policy makers who make their current decisions and planning based on the prediction of the economy in the future. However this prediction has so far been found to be difficult.

The classification of the US past recession periods is being carried out by the Business Cycle Dating Committee of the US National Bureau of Economic Research (NBER). The announcements by NBER on the peak or trough in economic activity are usually made long after the corresponding incident has occurred.

So far, the main approach in the literature focuses on finding the recession probability as a function of time, and use the probability to predict recession. As the recession variable will take the value of one if there is recession and zero otherwise, the mean of the binary variable exceeding half becomes an indication that the economy will tend to go towards recession. Thus the performance of the models which use recession probability for predicting recession may be assessed by the following criteria:

- (A) The predicted recession probability preceding the actual recession period is rarely less than half.
- (B) The predicted recession probability during the non-recession period is rarely larger than half.

Apart from the above two criteria, the other criteria which may be used include:

- (C) The actual recession period covers the time of occurrence of a local maximum of the predicted recession probability
- (D) The predicted recession probability gives early warning signals for the weakening of the economic conditions which may eventually develop into a recession.

---

\*Dr. Ah-Hin Pooi, Sunway University Business School, Sunway University, Malaysia. Email: ahhinp@sunway.edu.my

\*\*Dr. You-Beng Koh, Institute of Mathematical Sciences, University of Malaya. Email: kohyoubeng@um.edu.my

## Pooi & Koh

Some of the methods in the literature can achieve items (A), (B) and (C) above quite well. As for (D), it appears that so far none of the methods in the literature provides a satisfactory answer. Thus there is a need to come up with other versatile methods.

The objective of the present paper is to find a method which can produce functions of time for predicting the start of next recession and providing early warning signals for the possible adverse economic conditions in the near future.

The functions produced by the method in this paper are the recession probability together with the lower and upper limits of the prediction interval for the recession variable. The recession probability based on the proposed method satisfies, to a certain extent, items (A), (B) and (C) above and thus can provide indicator for recession for some of the recession periods. However the distinctive contribution of the present paper is the introduction of the indicator based on the lower limit of the prediction interval to provide results which are far more reliable than those given by the recession probability. Furthermore, the lower limit also provides a fairly reliable early warning signal which is lacking in the results based on the methods introduced so far in the literature.

Although the indicator for the start of next recession is important, the early warning signal provided by the present method is also important because it enables early preparation and precaution to be initiated to reduce the impact of the possible future recession.

The layout of the paper is as follows. After giving a brief literature review on the prediction of the US recessions in Section 2, we state the main contributions of this paper in Section 3. In Section 4 we give a brief description of the method based on multivariate power-normal distribution for finding prediction intervals. Section 5 describes the data and presents the empirical results. Section 6 concludes the paper.

## 2. Literature Review

A popular approach for predicting recession is the incorporation of economic variables into a time series probit model (Estrella and Mishkin 1998, Filardo 1999, Chauvet and Potter 2005, Wright 2006, Silvia, et al 2008, and Kauppi and Saikkonen 2008). The economic variables which have been incorporated into the models include, among others, the yield spread, change in the S&P 500 index, federal funds rate, change in the Conference Board's Leading Economic Index (LEI), and the Chicago PMI employment index.

The models which incorporate economic variables typically achieve (A) (B) and (C) in Section 1 only for some of the actual US recession periods. The achievement in (D) is also not obvious because the large variability of the predicted recession probabilities makes it difficult to obtain a clear upward movement of the predicted recession probabilities.

Instead of using only financial variables, Chen et al (2011) applied a factor model in which a few latent factors drive the co-movement of a large vector of financial variables. Their model is able to achieve (A), (B) and (C) for the four US recession periods between 1980 and 2001. As for (D), there are no early warning signals in this period. The model was further made dynamic by requiring the latent factors to follow a multivariate time series process (Chauvet and Senyuz 2012 and Fossati 2015). Chauvet and Sunyuz (2012) extracted the latent factors from the term spread, level and curvature of the yield curve, and from macroeconomic variables. Their predicted recession probabilities near the US recession periods are mostly very large while those during the non-recession periods are all less than half. However the actual recession period may not cover the time when the local maximum value of the recession probability

occurs. Fossati (2015) extracted latent factors from three small panels of indicators which are (1) bond and exchange rates, (2) stock market indicators and (3) real activity indicators. Among the various models in Fossati (2015), the probit model based on the “3-month less 10-year term spread” and the latent factors from Panels (2) and (3) exhibits a better performance. The short-horizon predicted recession probability of this better model performs quite well in (A),(B) and (C) but not (D) for the US recession periods from 1990 to 2010

Machine learning has also been used to forecast business cycle turning points (Qi 2001, Giusto and Piger 2014 and Berge 2015).The models based on machine learning can predict the US recessions, but with leads and lags. As the predicted probabilities based on machine learning almost always take values of either 0 or 1, they do not give any early warning signal for the weakening of the economic conditions.

### 3. Contributions of the Present Paper

This paper attempts to find the signals of possible recession in the near future, and predict the start of next recession. The method used in this paper requires the data for the vector  $\mathbf{r}(t)$  of the time- $t$  values of  $m$  selected economic variables for a fairly large number of values of  $t$ . The time- $(t+1)$  value  $z(t+1)$  of the recession variable which specifies whether there is a recession via its binary value is modelled to be dependent on the present and  $l-1$  past values  $\mathbf{r}(t), \mathbf{r}(t-1), \dots, \mathbf{r}(t-l+1)$  via a conditional distribution which is derived from an  $[ml+1]$ -dimensional power-normal distribution. The  $100(\alpha/2)\%$  and  $100(1-\alpha/2)\%$  points of the conditional distribution are next used to form a prediction interval for the value of the recession variable at the next future time point. The empirical results show that the indicators for the start of recession may be taken to be given by the mean of the conditional distribution having a value close to half, together with the lower and upper limits of the prediction interval having values close to zero and one, respectively.

The empirical results further show that the signals for possible recession in the near future are given by the upward movement of the mean, lower limit and upper limit of the prediction interval towards their respective critical values given by 0.5, 0, and 1, and the narrowing of the width of the prediction interval.

The above indicators and signals are tested using a pool of 14 selected economic variables over the period from Feb 1959 to Nov 2010 in the United States. For  $l=2$  and  $1 \leq m \leq 2$ , the combinations of  $m$  selected variables which achieve relatively shorter average lengths of the prediction intervals are obtained. The models based on the resulting combinations of variables are found to be able to provide fairly good indicators for the start of the next US recession, and early warning signals for the weakening of the economic conditions.

It appears that all the above methods given in the literature do not investigate the variability of the estimated probability of recession or the results from the identification based on machine learning. The main difference between the method in this paper and those in the literature is that, apart from estimating the probability of recession, the present method also makes use of the extent of variability of the estimated probability of recession to form the lower and upper limits of the prediction interval for the value of recession variable. These limits yield more indicators for the start of the recession. With more indicators, we may find out the indicator which tends to give more reliable results. The empirical results in the later part of the paper show that the indicator given by the lower limit of the prediction interval tends to be more

reliable. The lower limit of the prediction interval also gives a quite reliable early warning signal for the weakening of the economic conditions.

As the indicators may sometimes give false prediction, it would be a good strategy to find out the consensus view given by the models with relatively shorter average lengths of the prediction intervals. If a fairly large proportion of these models give similar indications, then we should take the indications seriously.

#### 4. Method Based on Multivariate Power-Normal Distribution

Let us begin with the power transformation introduced in Yeo and Johnson (2000):

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

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

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

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

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

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

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

When the values of  $y_1, y_2, \dots, y_{k-1}$  are given, we may find an approximation for the conditional probability density function (pdf) of the last component  $y_k$  of  $\mathbf{y}$  by using the numerical procedure given in Pooi (2012).

We may choose the variables  $y_1, y_2, \dots, y_k$  to be those given by the values of the components of  $\mathbf{r}(t - l + 1), \dots, \mathbf{r}(t - 1), \mathbf{r}(t)$  together with the value of  $z(t + 1)$  in the lag- $(l - 1)$  model.

From the data which spans over  $T$  units of time, we can form a table of  $T - l$  rows with each row representing an observed value of  $(y_1, y_2, \dots, y_k)$ . From the table, we can form the  $i_w$ -th moving window of size  $n_w$  from the  $i_w$ -th row till the  $(i_w + n_w - 1)$ -th row. We can form a total of  $T - l - n_w$  such windows of size  $n_w$ . We next find a  $k$ -dimensional power-normal distribution for  $(y_1, y_2, \dots, y_k)$  using the data in the  $i_w$ -th window.

Letting  $y_1, y_2, \dots, y_{k-1}$  be given by the first  $k - 1$  values in the  $(i_w + n_w)$ -th row which is immediately after the  $i_w$ -th window, we may now find a conditional distribution for  $y_k$  when  $y_1, y_2, \dots, y_{k-1}$  are given. The mean  $\hat{y}_k^{(i_w)}$  of the conditional distribution is then an estimate of the value of the last component at the next unit of time. On the other hand, the  $100(\alpha/2)\%$  and  $100(1 - \alpha/2)\%$  points of the conditional distribution may be regarded as the lower and upper limits of the nominally  $100(1 - \alpha)\%$  out-of-sample prediction interval for the value of the last component at the next unit of time.

The coverage probability of the prediction interval may be estimated by the proportion of prediction intervals which include the observed value of the last component at the next unit of time. Meanwhile, the expected length of the prediction interval may be estimated by the average length of the prediction intervals. When the estimated coverage probability is close to the target value  $1 - \alpha$ , a small value of the average length is indicative of good predictive power of the model.

## 5. Data and Empirical Results

Consider the monthly U.S. data on 14 economic variables in the period from Feb 1959 to Nov 2010. This set of data covers the recent recession which began in Jan 2008. A short description of the 14 variables is given in Table 1.

**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

Various combinations of  $m$  variables are chosen from the 14 economic variables. The time- $t$  values of a chosen combination of  $m$  variables are used to form the vector  $\mathbf{r}(t)$ . The vector  $\mathbf{y}$  of  $k = ml + 1$  variables is next formed:

$$\mathbf{y} = [\mathbf{r}(t - l + 1), \dots, \mathbf{r}(t - 1), \mathbf{r}(t), z(t + 1)]^T \quad (4)$$

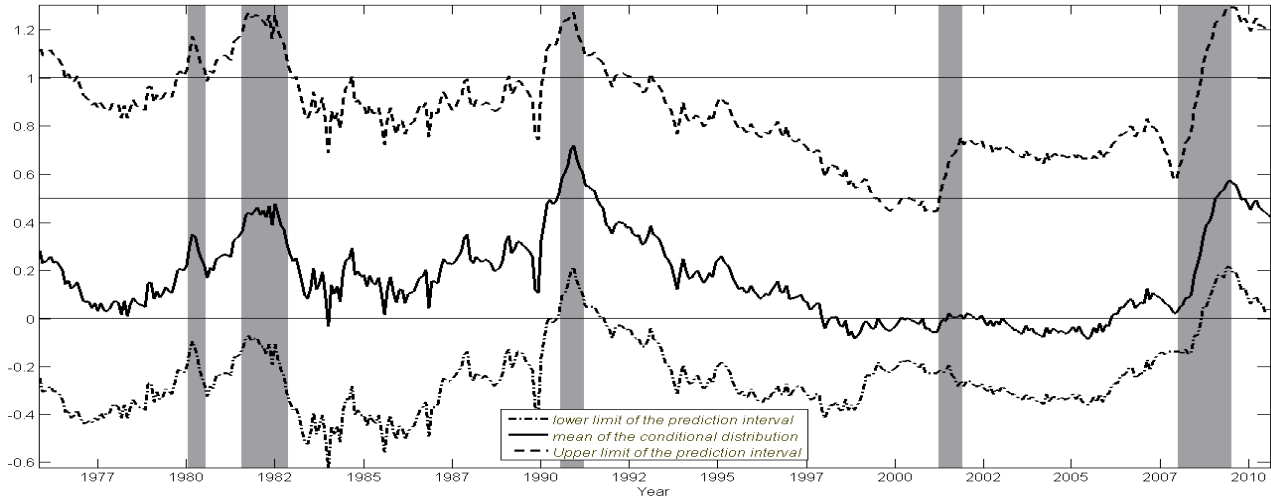
We fit the values for  $\mathbf{y}$  in the  $i_w$ -th moving window of size  $n_w = 200$  by a  $k$ -dimensional power-normal distribution, and find a nominally 95% out-of-sample prediction interval for the next future value of the recession variable when the values of  $y_1, y_2, \dots, y_{k-1}$  are given by the first  $k - 1$  values in the  $(i_w + n_w)$ -th row which is immediately after the  $i_w$ -th window.

The values of  $(l, m)$  may be chosen such that further increase of the chosen values of  $l$  and  $m$  does not decrease the average lengths of the prediction intervals very much. It is found that a suitable choice for  $(l, m)$  is  $(2, 2)$ . Among the models of which  $(l, m) = (2, 2)$ , we select the models which satisfy the two criteria given respectively by the relatively shorter average lengths of the prediction intervals and ability to provide fairly good predictions of the recession periods. The models with  $m$  less than 2 may also be chosen provided that, to a certain extent, the above two criteria are met.

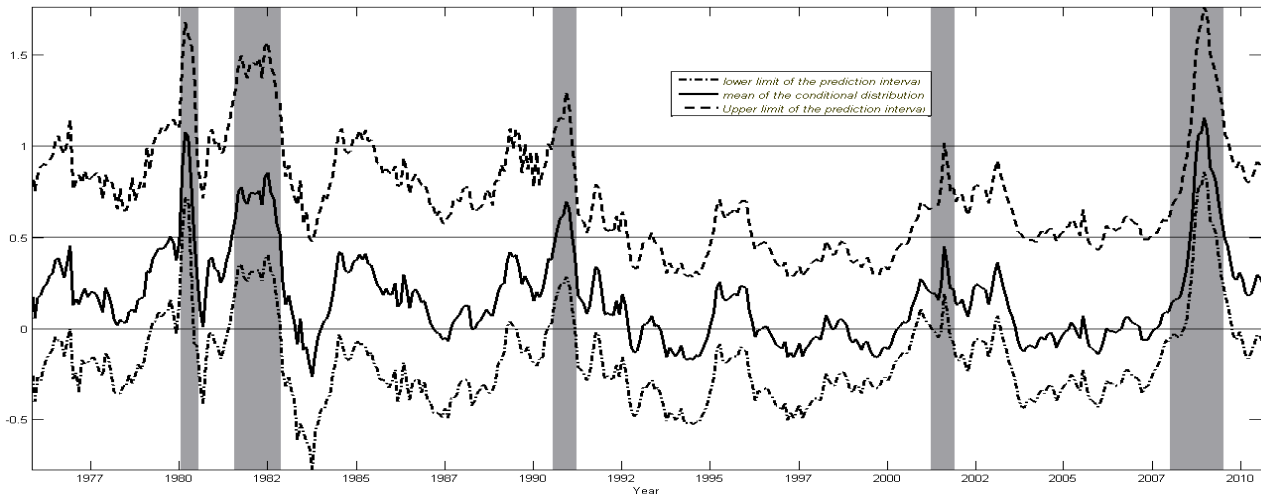
## Pooi & Koh

The out-of sample prediction intervals thus obtained for three chosen models are shown in Figures 1-3. These figures show that before the start of a recession, one or more of the three chosen models are able to signal that the economy is moving towards a recession. These figures also show that some of the models may be able to provide indicators for the start of the next recession on or before the actual commencement month of the recession.

**Figure 1: Out-of sample prediction interval when the model is based on Variable 5**  
( $m = 1, l = 2, \alpha = 0.05, n_w = 200$ )



**Figure 2: Out-of sample prediction interval when the model is based on Variables 1 and 3**  
( $m = 2, l = 2, \alpha = 0.05, n_w = 200$ )



## Pooi & Koh

**Figure 3: Out-of sample prediction interval when the model is based on Variables 3 and 5 ( $m = 2, l = 2, \alpha = 0.05, n_w = 200$ )**

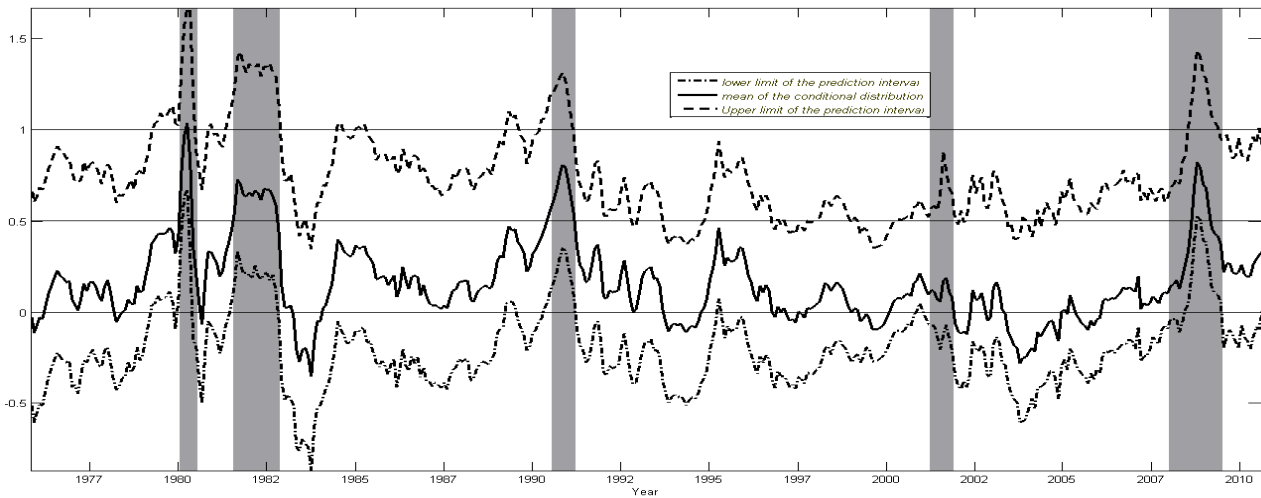


Table 2 gives a summary of the indicators and signals provided by the three chosen models. In this table, Indicators 1,2,3, are respectively the mean of the conditional distribution, together with the lower and upper limits of the prediction interval, which reach their respective critical values at a time not long before the actual start of the recession. Signals 1,2,3 in the same table are respectively the upward movements of the mean of the conditional distribution, together with the lower and upper limits of the prediction intervals towards their respective critical values, and Signal 4 is the narrowing of the width of the prediction interval.

**Table 2: Indicators and signals for the recessions**

Variable(s) in the model	Indicator	Signal
<b>Recession beginning in Feb1980</b>		
5	3	2,3
1,3	1,2,3	1,2,3
3,5	2,3	1,2,3
<b>Recession beginning in Aug 1981</b>		
5	None	2
1,3	None	1,2
3,5	2	1,2
<b>Recession beginning in Aug 1990</b>		
5	1,2,3	1,2,3
1,3	2	2
3,5	1,2	1,2
<b>Recession beginning in April 2001</b>		
5	None	4
1,3	2	2
3,5	2	2
<b>Recession beginning in Jan 2008</b>		
5	None	2,4
1,3	None	2
3,5	None	2

A closer look at Figures 1-3 reveals that among Indicators 1-3, Indicator 3 tends to give false alarms while Indicators 1 and 2 are more reliable. The figures also show that among the 4 signals, Signal 2 seems to prevail whenever a recession is going to occur.

## Pooi & Koh

Apart from the above three chosen models, there are several other models which give fairly good predictions. These models include the lag-1 models based respectively on Variables 3 and 12, Variables 5 and 12, and Variables 6 and 12.

As the indicators and signals may give false alarms, it would be a good strategy to compare the indicators and signals provided by a set of models, and look for the indicators and signals which are common in a fairly large subset of the models.

## 6. Conclusions

The premise of the analysis is that the multivariate power-normal distribution can be used to fit the data given by a set of financial variables of which the initial variables are continuous while the last variable is binary. When the values of the initial variables are given, a continuous conditional distribution is found for the last variable by using the fitted multivariate power-normal distribution. The continuous conditional distribution is not converted to the corresponding binary distribution. Instead the mean together with the 2.5% and 97.5% points are found from the continuous conditional distribution and later used to form indicators and signals for the next recession. It is found that the indicators based on the mean and 2.5% point of the distribution are able to predict fairly well the start of the next recession. On the other hand, the 2.5% point is able to provide a fairly good signal for the imminent recession.

The starting months of a recession given by the indicators very often differ somewhat from the actual month for the start of a recession. However, as it is not difficult to get a strong signal for an imminent recession, early preparation and precaution may be initiated to overcome the difficulty in predicting the actual commencement of the next recession.

The method in this paper may be applied to countries other than USA provided that there are sufficient amount of data over evenly spaced time intervals, and the number of recession periods covered by the data is not too small.

## Acknowledgement

The authors would like to thank Liu Yang of University at Albany, SUNY, USA, for sharing his copy of the US data on economic and recession variables.

## References

- Berge, T. 2015, 'Predicting recessions with leading indicators: Model averaging and selection over the business cycle', *Journal of Forecasting*, Vol 34, No 6, pp 455-471.
- Chauvet, M and Potter, S 2005, 'Forecasting recessions using the yield curve', *Journal of Forecasting*, Vol 24, No 2, pp 77-103.
- Chauvet, M & Senyuz, Z 2012, 'A dynamic factor model of the yield curve as a predictor of the economy', Technical Report 2012-32, Federal Reserve Board.
- Chen, Z, Iqbal, A, and Lai, H 2011, 'Forecasting the probability of US recessions: a Probit and dynamic factor modelling approach', *Canadian Journal of Economics/Revue canadienne d'économique*, Vol 44, No 2, pp 651-672.
- Estrella, A and Mishkin, FS 1998, 'Predicting US recessions: financial variables as leading indicators', *Review of Economics and Statistics*, Vol 80, No 1, pp 45-61.
- Filardo, AJ 1999, 'How reliable are recession prediction models?' *Economic Review-Federal Reserve Bank of Kansas City*, Vol 84, pp 35-56.



## Pooi & Koh

- Fossati, S 2015, 'Dating US business cycle with macro factors'. working paper.
- Giusto, A and Piger, J 2013, 'Nowcasting US business cycle turning points with vector Quantization'. Working paper, University of Oregon.
- Kauppi, H, and Saikkonen, P 2008, 'Predicting US recessions with dynamic binary response models', *The Review of Economics and Statistics*, Vol 90 ,No 4, pp 777-791.
- Pooi, AH 2012, 'A Model for Time Series Analysis', *Applied Mathematical Sciences*, Vol 6 ,No 115, pp. 5733-5748.
- Qi, M 2001, 'Predicting US recessions with leading indicators via neural network models'. *International Journal of Forecasting* 17, pp. 383-401.
- Silvia, J, Bullard, S, and Lai, H 2008, 'Forecasting US recessions with probit stepwise regression models', *Business Economics*, Vol 43, No 1, pp 7-18.
- Wright, JH 2006, 'The yield curve and predicting recessions', *Finance and Economics Discussion Series. Federal Reserve Board, February*.
- Yeo, IK., and Johnson, RA 2000, 'A new family of power transformations to improve normality or symmetry', *Biometrika*, 87 ,pp 954-959.