Applying Auto-Regressive Binomial Model to Forecast Economic Recession in U.S. and Sweden

Submitted by: Chunshu Zhao

Yamei Song

Supervisor: Md. Moudud Alam

D-level essay in Statistics, June 2010.

School of Technology & Business Studies, Dalarna University

Contents

ABSTRACT

In this paper, we construct an auto-regressive (AR) logit model which can be applied in the prediction of the U.S. and Sweden economic recessions. The predictive variables of this model vary between U.S. and Sweden. The predicted power of this AR logit model is checked in terms of in-sample and out-of-sample performance. We find that the goodness-of-fit of this model to data is remarkable and this model can give reasonable degree of accuracy in the predicting U.S. and Sweden business cycle.

Keywords: Auto-regressive logit model, Business cycle, Recession

1 Introduction

1.1 Background

The latest recession in U.S. occurred ever since the bankruptcy of Lehman Brothers Holding Inc. due to the sub prime mortgage crisis in 2008 and lasted more than two years. The consequence of this recession is quite significant which is viewed through several economic performance indicators, e.g. 10% unemployment rate and a deeply decline of real GDP. Although nowadays the government of U.S. declares that the recession in U.S. has end, the economy in Europe is experiencing a new round of sharply decline, such as the 2010 Euro crisis including Greece, Ireland, Spain, and Portugal. Thus, how to find an efficient ways to model and forecast economic recession is nowadays a popular topic among policy makers and researchers. A binary AR model has been considered as a useful model to forecast economic recession in U.S., thus we want to investigate whether this kind of model is also applicable in European countries such as Sweden.

Before going through the definition of recession, it is necessary to introduce a general idea about business cycle, since recession is regarded as a part of business cycle. In the view of macroeconomics, business cycle is a well-known financial variable to measure the economy-wide fluctuations in production over a long period. These fluctuations exit when the economic environment changes from an expansion to a recession and vice versa. Generally, people are more concerned with recession rather than expansion, since a majority of papers are focus on discussing recession. The definition of recession varies between different research institutes and therefore no exact definition of recession exists. One commonly used definition is provided by the National Bureau of Economic Research (NBER) in US. In their words, "a recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales"[1]. From this definition, we can see that being able to describe and predict a recession is of great interest for the people. More specifically, policy makers who work on macroeconomic issues are more concern with the trend of real GDP and real income, since a wrong prediction of the future of economic state leads to the ineffective of policy itself. And the employees in industry are mostly interested in the state of employment, and so on. Thus, a reasonable assessment of economic activities such as business cycle is of great importance to the whole society.

In modeling recession, AR binomial model has been found very useful. Kauppi and Saikkonen (2005) predict U.S. recessions with dynamic binary response models. They develop a unified model framework that accommodates most previously analyzed dynamic binary time series models as special cases (Kauppi and Saikkonen, 2005). The previously analyzed works are mainly focus on how to improve the predict model from static one to dynamic one. Startz (2006) gives an empirical study on U.S. recession by using binomial AR moving average (BARMA) models. The distinct advantage of BARMA model is to eliminate the curse of dimensionality found in long lag Markov models. Although, the BARMA (2, 2) model fit well with real data, comparing to traditional Markov model and other BARMA models with different lagged values, the author did not report the predictive power of this model. Thus, we lack the evidence that how well this model perform when applying to predict the process of recession. Nyberg (2009) extend the AR binomial model from univariate case to bivariate case. Unlike some typical cases in dynamic models based on latent variables (Chauvet and Potter, 2005), he measures the state of the economy in terms of recession periods defined by the business cycle and the growth rate cycle indicators. Besides the dependent variable and its AR term, the model also takes into account other financial variables as predictor. The empirical application in U.S. suggests the bivariate model outperform the univariate model. Since the most empirical results are only available for the U.S. recessions, how the AR binomial model performance in other homogeneous countries is of great interest. Yasuhiro Omori (2003) use discrete

[1] http://www.nber.org/cycles/cyclesmain.html

duration model with AR random effects to analyse Japanese diffusion index. Ulf Hamberg and David Verständig (2009) apply static logistic regression model to forecast business cycle under Swedish conditions.

Unlike what Hamberg and Verständig (2009) did in their thesis, first we apply the AR logit model in U.S. to examine whether the model works well. Second we extend the static logistic regression model to a binomial AR model and then applying the model with real data set from Sweden. By using our model, we forecast the status of economy in three quarters ahead. To the best of our knowledge, these have not been done previously.

This essay generally includes four sections. Section one introduces the background and basic definitions then proposes the aim of this essay. Section two describes the data we employed and the method we use. Section three presents the study results which we divided into two parts, in-sample forecasting results and out-of-sample forecasting results. Section four concludes.

1.2 Aim

The aim of this essay is to examine whether the AR logit model which is found useful for forecasting U.S. recession is also applicable to forecast the business cycle of the European countries, like Sweden. If yes, how well the performance of this model does.

2 Data and Methods

2.1 Data description

All our models are based on binomial AR model, thus the recession indicator has a binary form which means it can take only one of two values, “1” or “0”.Since modeling recession is of interest rather than modeling expansion, we let “1” denote economic recession and let “0” denote economic expansion. All the data applied in our analysis is quarterly data. Estrella and Mishkin (1998) suggest that quarterly data are more preferable since the time series become less variable, thus reducing noise and usually leading to a better goodness-of-fit in the model.

2.1.1 The recession indicator

The recession indicator applied in U.S. case is the state of business cycle reported by NBER[1]. According to the definition given by NBER (Moore, 1967), business cycle peak dates mark the end of a period of expansion and the beginning of a period of contraction; trough dates mark the end of a period of contraction and the beginning of a period of expansion. Thus, we label the period from a peak to the following trough as “1” which indicates a process of recession and label the period from a trough to the following peak as “0” which means an expansion of economy.

The recession indicator applied in Sweden is the state of business cycle reported by Economic Cycle Research Institute (ECRI)[2]. Since this institute also reports the business cycle of other counties besides U.S. and Sweden, the identification of peak and trough between different countries is consistent when applying them to models.

2.1.2 The explanatory variables

Estrella and Mishkin (1998) give a detailed discussion about the financial variables as leading indicators when predicting U.S. recessions. Series such as interest rates and spreads, stock prices, and monetary aggregates can play an important role in macroeconomic prediction. Among them, the term spread between a short-term and a long-term interest rate and stock prices indexes emerge as the most useful simple

[1] www.nber.org/cycles/cyclesmain.html
[2] www.businesscycle.com

financial indicators. Hence, it is of interest to examine whether the term spread and stock price indexes have predictive power in our models. Besides, we also employ the stock market return and German term spread, the same explanatory variables as Nyberg (2009), as the predictors in our models. The lagged values of all these variables are the same as in Nyberg (2009).

Specifically:

Term spread at time t (TS): $TS_t = R_t - i_t$

where R_t is ten-year treasury bond yield rate with constant maturity[1] and i_t is three-month treasury bill rate on secondary market[2]. The expected sign of this variable should be negative. Since for a given long-term yield, a large value of yield spread indicate a decreasing of short-term yield which associated with increased expectation of strong economic activity.

Stock market return (SMR): $SMR_t = \Delta \log(\text{S \& P500})$

where S&P 500 is the standard and poor 500 index[3]. To make sure the variable is stationary, which is the basic assumption of time series model, we take the log-difference value of S&P500 index. The expected sign of this variable should be negative. Since an increasing stock market return indicates a high expectation of investment return and therefore associated with an expansion of economic.

German term spread (GTS): $GTS_t = R_t^{GE} - i_t^{GE}$

German term spread is constructed as the difference between 10-year Federal security[4] and the three-month money market rate[5]. The expected sign of this variable should be negative as well. The reason for this assumption is the same as term spread.

We generally apply the same variables employed in U.S. case when considering the Swedish case. Hamberg and Verständig (2009) take into account six different explanatory variables as the independent variables; they are changes in a stock price

1 http://research.stlouisfed.org/fred2/data/GS10.txt
2 http://research.stlouisfed.org/fred2/data/TB3MS.txt
3 http://finance.yahoo.com/q/hp?s=%5EGSPC&a=00&b=3&c=1950&d=11&e=23&f=2009&g=m
4 http://www.bundesbank.de/statistik/statistik_zeitreihen.en.php?func=row&tr=WZ9826
http://stats.oecd.org/index.aspx?queryid=86 (1971M1-1972M8)
5 http://www.bundesbank.de/statistik/statistik_zeitreihen.en.php?lang=en&open=&func=row&tr=SU0107

index, oil price, a confidence indicator, European GDP, residential building as well as the spread between a short-term and a long-term interest rate. However, some variables used by Hamberg and Verständig (2009) do not work in our model, we just use the term spread, stock market return and term spread in German as the leading indicators.

Specifically:

We use OMX Stockholm All Share (OMXS)-PI[1] including all the shares listed on OMX Nordic Exchange Stockholm as the indicator of stock price. One advantage of OMX Stockholm PI index is to reduce the impact of firm-specific risk in accordance with normal portfolio theory (Bodie, Kane and Marcus, 2005), since other stock market index may just include a small number of stocks issued by specific companies. We take its log-difference transformation to make it consistence with what we did in U.S case. The expected sign of this variable should be negative.

$$SMR_t = \Delta \log(\mathrm{OMX} - \mathrm{PI})$$

Swedish term spread (STS): $STS_t = R_t^{SW} - r_t^{SW}$

Where R_t^{SW} is the long-term interest rate measured by Swedish government bond with 10 years maturity[2] and r_t^{SW} is short-term interest rate measured by Swedish Treasury Bill with 3 months maturity[3]. The expected sign of this variable should be negative.

2.2 Methods

2.2.1 AR logit model

Consider having a random process y_t, recession indicator, which is binary value, that is, 0-1 value and this can be expressed in indicator function form,

1 http://www.nasdaqomxnordic.com/indexes/historical_prices/?Instrument=SE0000744195
2 http://www.riksbank.com/templates/stat.aspx?id=17188
3 http://www.riksbank.com/templates/stat.aspx?id=17187

$y_t = \mathrm{I}(recession\ at\ period\ t)$

So, conditional on the information set $\Omega_t = \{(y_{t\text{-}1}, x_{t-h}) \mid h \geq 1\}$, where y_{t-1} is the lagged value of y_t and x_{t-h} is the explanatory variable with lag *h*, y_t can be fitted as a binary model with canonical link,

$$E(y_t \mid \Omega_t) = p_t$$
$$logit(\ p_t) = \eta_t = X_{t\text{-}h}\beta \qquad (1)$$

Where $X_{t\text{-}h}$ is the design matrix and simplifies the notation of lagged explanatory variables because different explanatory variables can have different lags and p_t is the conditional probability that economy is in recession at period *t*.

Model (1) has been used in former economic recession study (Hamberg and Verständig, 2009). A rational extension to model (1) is a dynamic specification by adding a lagged value of recession indicator y_t

$$logit(\ p_t) = \eta_t = \mu \cdot y_{t-1} + X_{t\text{-}h}\beta \qquad (2)$$

Based on former study (Kauppi and Saikkonen, 2005) and model (2), a first order AR logit model can be proposed

$$logit(\ p_t) = \eta_t = \alpha \cdot \eta_{t-1} + \mu \cdot y_{t-1} + X_{t\text{-}h}\beta \qquad (3)$$

2.2.2 Parameters estimation

Suppose, there are observed processes $y_t\ (t = \mathrm{T} \cdots 1)$, $x_{i,t}\ (i = \mathrm{n}, t = \mathrm{T} \cdots 1)$ where T is the number of observation and n is the number of explanatory variables. Define a corresponding parameter vector $\theta = (\alpha, \mu, \beta)$ where $\beta = (\beta_1 \cdots \beta_\mathrm{n})$. Then model (3) can be written as

$$E(y_t \mid \Omega_t) = p_t \qquad (4)$$

$$logit(p_t) = \eta_t = \alpha \cdot \eta_{t-1} + \mu \cdot y_{t-1} + \beta_0 + x_{1,t\text{-}i}\beta_1 + \cdots + x_{\mathrm{n},t\text{-}j}\beta_{\mathrm{n}}$$

Based on the binomial model (4), the conditional likelihood function has the following form by multiplying individual contribution

$$L(\theta) = \prod_{t=1}^{T} l_t(\theta) = \prod_{t=1}^{T} p_t^{y_t}(1 - p_t)^{1-y_t} = \prod_{t=1}^{T} (logit^{-1}(\eta_t(\theta)))^{y_t}(1 - logit^{-1}(\eta_t(\theta)))^{1-y_t} .$$

The estimated parameters can be carried out directly by Newton-Raphson method. In order to simplify the calculation, the log-likelihood function should be considered as

$$LL(\theta) = log\,(L(\theta)) = \sum_{t=1}^{T} y_t log(logit^{-1}(\eta_t(\theta))) + (1 - y_t) log(1 - logit^{-1}(\eta_t(\theta))) .$$

The score function of log-likelihood function is

$$S(\theta) = \frac{\partial LL(\theta)}{\partial \theta} = (\frac{\partial LL(\theta)}{\partial \alpha}, \frac{\partial LL(\theta)}{\partial \mu}, \frac{\partial LL(\theta)}{\partial \beta_0}, \frac{\partial LL(\theta)}{\partial \beta_1}, \cdots, \frac{\partial LL(\theta)}{\partial \beta_{\mathrm{n}}})'$$

The large sample theory can be applied to the parameters estimated and its practical use is the calculation of information matrix,

$$I(\theta) \xrightarrow{p} lim_T\, T^{-1} \sum_{t=1}^{T} S_t(\theta) S_t'(\theta)$$

According to Newton-Raphson method, the formal iterative function is

$$\theta_{\mathrm{n}} = \theta_{\mathrm{n}-1} + \frac{1}{I(\theta)} S(\theta)$$

And assume there is an initial value θ_0 of parameter vector θ, the iterative value θ_1 can be calculated out by

$$\theta_1 = \theta_0 + \frac{1}{I(\theta)} S(\theta)$$

When $\theta_{\mathrm{n}} - \theta_{\mathrm{n}-1} \to 0$ as $n \to \infty$, or $|S(\theta)| \to 0$ as $n \to \infty$, the process of iteration can be over. And finally, the parameters are estimated out.

2.2.3 Parameters test

In order to test the significance of correlation coefficient under the null hypothesis $H_0: \theta = 0$, Wald test can be used in large sample, and its has form

$$z = \frac{\hat{\theta} - 0}{SE(\hat{\theta})},$$

Where $SE(\hat{\theta})$ is the standard error of parameter estimates, $\hat{\theta}$, calculated from the non-null model.

The Standard Error of $\hat{\theta}$ can be estimated by Fisher's Information matrix, I, so the Wald test can be written as

$$z = \frac{\hat{\theta} - 0}{diag(I^{-1})},$$

Where $diag(\cdot)$ is the diagonal element of the parameter matrix.

If the p-value, calculated by quantile z under normal distribution, is larger than the significant level under two side test, the null hypothesis θ equals to zero is accepted, otherwise, the estimated value of θ is accepted.

2.2.4 Forecasting

Kauppi and Saikkonen (2005) mentioned that one-period or multi-period forecast can be made based on the explicit formula.

Based on the given information set $\Omega_t = \{(y_{t\text{-}1}, x_{t-h}) \mid h \geq 1\}$, the forecast conditional recession probability given as

$$p_t = logit^{-1}(\eta_t) = logit^{-1}(\alpha\eta_{t-1} + \mu y_{t-1} + X_{t\text{-}h}\beta) \tag{5}$$

where h is the forecast horizon.

At time t-1, the forecasting function (5) can give one period ahead forecast directly. But it will be a little complicated but still straight forward, if more than one period ahead forecast is needed. For simplicity, suppose the forecast horizon *h* equals to two, the iterative process of two-period ahead forecast is given as following.

The linear predictor based on information up to time *t*-2 is

$$\begin{aligned}\eta_t &= \alpha_1\eta_{t-1} + \mu_1 \mathrm{y}_{t-1} + X_{t\text{-}2}\beta \\ &= \alpha_1(\alpha_1\eta_{t-2} + \mu_1 \mathrm{y}_{t-2} + X_{t\text{-}3}\beta) + \mu_1 \mathrm{y}_{t-1} + X_{t\text{-}2}\beta \\ &= \alpha^2\eta_{t-2} + \sum_{i=1}^{2}(\alpha^{i-1}\mu \cdot y_{t-i} + \alpha^{i-1}X_{t\text{-}i\text{-}1}\beta)\end{aligned}$$

The information of y_{t-1} can not be obtained at forecasting time *t*-2, but the conditional probability p_{t-1} equals $logit^{-1}(\alpha\eta_{t-2} + \mu y_{t-2} + X_{t\text{-}3}\beta)$, which will be used to calculate the predictive recession probability p_t

$$\begin{aligned}p_t &= E(logit^{-1}(\eta_t)) \\ &= p_{t-1}logit^{-1}(\eta_t \mid y_{t-1} = 1) + (1 - p_{t-1})logit^{-1}(\eta_t \mid y_{t-1} = 0)\end{aligned}$$

So, the right hand side of forecasting function (5) can make two-period ahead forecast iteratively at forecasting time *t*-2.

Thus, in order to make *h*-period ahead forecast, the AR term should be tailored to match information available at forecasting time *t-h*. The process of calculation is little complicated but can be solved by repetitively.

The general formula is

$$\begin{aligned}\eta_t &= \alpha^{\mathrm{h}}\eta_{t-\mathrm{h}} + \sum_{i=1}^{\mathrm{h}}(\alpha^{i-1}\mu \cdot y_{t-i} + \alpha^{i-1}X_{t\text{-}i\text{-}1}\beta) \\ p_t &= E(logit^{-1}(\eta_t)) \\ &= \sum_{\mathrm{y_t}\in(0,1)}\left[\left(\prod_{i=1}^{h-1} p_{t-\mathrm{i}}{}^{y_{t\text{-}\mathrm{i}}}(1 - p_{t-\mathrm{i}})^{1-y_{t-\mathrm{i}}}\right)logit^{-1}(\eta_t \mid y_{t-1}\cdots y_{t-\mathrm{h}+1})\right]\end{aligned} \tag{6}$$

In conclusion, the general form of forecasting function (5) can give one period ahead forecast directly for the conditional probability $p_t = P(y_t = 1)$ at time *t*-1, and *h* period ahead forecast iteratively at forecasting time *t-h*, using information up to time *t-h*.

2.2.5 Model selection criteria

Model comparison can be carried out by evaluating the goodness-of-fit measures, one of which is Schwarz-Bayesian information criterion, BIC (Schwarz, 1978),

$$BIC = -log\mathrm{L} + K\frac{log(T)}{2}$$

Where $log\mathrm{L}$ is the log-likelihood function, K is the number of estimated parameters, and T is the number of observations. The model is fitting better when the value of this statistic is declining.

A facilitated measure of evaluating the performance of predictive model is the percentage of correct prediction. The goodness-of-fit of model to data can be checked visually by a cross table as table 4.

Finally, standard errors and p-values of estimated parameters are also considered when checking the goodness-of-fit of specification models.

3 Results

In this part, the performance of AR logit model (3) is checked by forecasting business cycle recession periods for U.S. and Sweden respectively. The predictive power of AR logit model can be examined by corresponding figures and cross tables. We are mainly interested in out-of-sample performance of the model, but it is also very helpful to check its in-sample forecasting first in choosing the explanatory variables and corresponding lags.

3.1 Model selection and in-sample results

3.1.1 Empirical analysis to U.S.

The in-sample performance of AR logit model can be evaluated by the goodness-of-fit to the data and the accuracy of recession prediction.

Based on data description section, explanatory variable candidates are U.S. term spread, TS_t, stock market return, SMR_t, and German term spread, GTS_t. So, the design matrix in model (3) has the form,

$$X_{t\text{-}h} = (1,\ TS_{t\text{-}i},\ SMR_{t\text{-}j},\ GTS_{t\text{-}k})$$

And the proposed model is

$$\begin{aligned} &logit(\ p_t) = \eta_t \\ &\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 TS_{t-i} + \beta_2 SMR_{t-j} + \beta_3 GTS_{t\text{-}k} \end{aligned} \tag{7}$$

After the selection of explanatory variables, the next step is to fix the lags of the variables of model (7) that fit the data[1] best. Part of the in-sample results of model (7) for directly forecasting U.S. economy status one quarter ahead and corresponding selection criteria values are listed in table 1.

Table 1 the summaries of model (7) with different lags

Predictor	Lag (*i, j, k*)			
	(1,1,1)	(1,1,2)	(2,1,1)	(2,2,2)
(intercept)	-1.77	-1.69	-1.47	-0.55
η_{t-1}	0.39	0.38	0.42	0.48
y_{t-1}	5.49	5.15	4.27	1.37
$TS_{t\text{-}i}$	-0.72	-0.63	-0.72	-0.51
$SMR_{t\text{-}j}$	25.79	25.14	24.19	13.97
$GTS_{t\text{-}k}$	-0.18	-0.37	-0.16	-0.46
Log-likelihood	-17.72	-17.04	-16.80	-23.14
BIC	20.24	19.56	19.32	25.66

From table 1, the coefficients of $SMR_{t\text{-}j}$ in the four models are positive which does not make economic sense, and the positive sign of the parameter goes through all the models with different lags not just limited to the models refer to table 1. So, the explanatory variable, SMR_t, should be removed from the final model.

[1] Using quarterly data from the first quarter of 1971 to the fourth quarter of 2009

The modified model of (7) is

$$
\begin{aligned}
&logit(p_t) = \eta_t \\
&\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 TS_{t-i} + \beta_2 GTS_{t\text{-}j}
\end{aligned}
\qquad (8)
$$

And a part of summaries of different models, explanatory variables with different lags, is listed in table 2.

Table 2 the summaries of model (8) with different lags

Lag (i, j)	*Log-likelihood*	*BIC*	Lag (i, j)	*Log-likelihood*	*BIC*
(1,1)	-29.99	32.51	(2,3)	-23.54	26.06
(1,2)	-28.00	30.52	(3,1)	-29.27	31.79
(1,3)	-25.85	28.37	(3,2)	-27.87	30.39
(2,1)	-27.70	30.22	(3,3)	-24.92	27.44
(2,2)	-26.25	28.77			

Based on the summary, explanatory variables with lags (2, 3) have the biggest predicted power in recession because of the smallest value of *BIC*. And a summary of the best model is in table 3.

Table 3 the summary of model (8), explanatory variables with lags (2, 3)

Predictor	Lag (2, 3)	Criteria	
(intercept)	-2.46	*Log-likelihood*	-23.54
η_{t-1}	-0.30	*BIC*	26.06
y_{t-1}	5.81		
$TS_{t\text{-}i}$	-0.94		
$GTS_{t\text{-}j}$	-1.15		

And the sign of the two explanatory variables are negative, that is, when U.S. term spread or German term spread is going down, the probability of U.S. economy will be in recession goes up. So, the better AR logit model used to forecast the U.S. economy is

$$
\begin{aligned}
&logit(p_t) = \eta_t \\
&\eta_t = \text{-}2.46 \text{ - } 0.30\eta_{t-1} + 5.81 y_{t-1} \text{ - } 0.94 TS_{t-2} \text{ - } 1.15 GTS_{t-3}
\end{aligned}
\qquad (9)
$$

In order to illustrate the in-sample performance of model (9), figure 1 shows the estimated probability, the solid line, that the U.S. economy will be in recession in one quarter ahead and the actual U.S. economy recession period, the shadow areas.

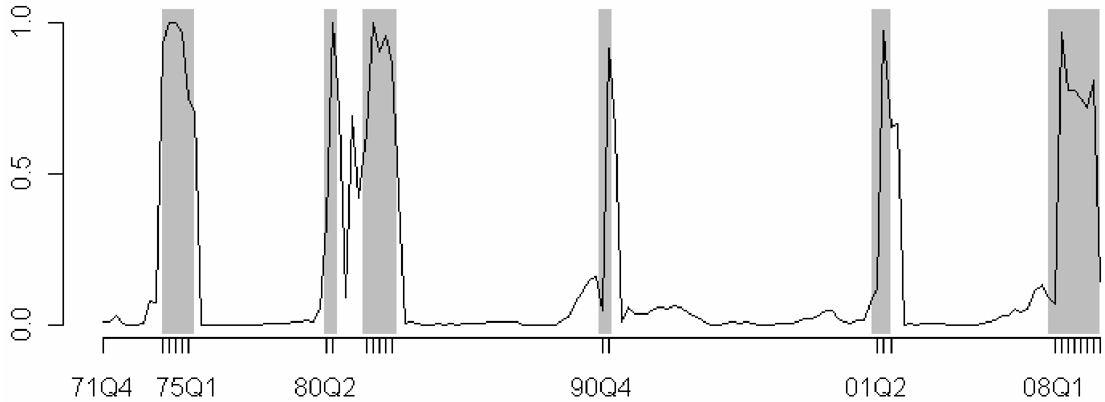

Figure 1 probability of recessions and actual recession periods of U.S.

Figure 1 shows six periods of recessions in shadow areas. The probability line matches the first three shadow areas well, that is, model (9) can forecast the recessions happened from the first quarter of 1974 to the first quarter of 1975, from the second quarter of 1980 to the third quarter of 1980 and from the fourth quarter of 1981 to the fourth quarter of 1982 well. But model (9) forecasts the later three recessions about one quarter later than recession has happened. And we will notice that model (9) forecasts a recession happens at about the first quarter of 1981 if the threshold is 0.5, but actually no recession happens at that period.

Overall, the in-sample forecasting of model (9) can give clear signals of recessions, which can be proved in the following method. The percentage of correct prediction can be checked by cross table 4, while the threshold is fixed at 0.5.

Table 4 the percentage of correct prediction of model (9)

Observed Economy	Predicted Economy: Expansion	Predicted Economy: Recession	Correct Percentage
Expansion	123	5	96.1%
Recession	5	20	80.0%
Overall Percentage			93.5%

The overall percentage of correct prediction is high, so the fitted AR logit model (9) to the U.S. data is good, and model (9) can give clear signals in the predicting the

U.S. recessions.

3.1.2 Empirical analysis to Sweden

We use the same process and criteria to implement the selection of the models fitted to Swedish data.

Explanatory variables includes Stockholm PI index (SPI), Swedish term spread (STS) and German term spread (GTS) according to the former section of data description. Thus the design matrix of model (3) has the form

$$X_{t\text{-}h} = (1,\ SPI_{t-i},\ STS_{t-j},\ GTS_{t-k})$$

The proposed AR logit model should be

$$\begin{aligned} &logit(p_t) = \eta_t \\ &\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 SPI_{t-i} + \beta_2 STS_{t\text{-}j} + \beta_3 GTS_{t\text{-}k} \end{aligned} \tag{10}$$

We fit the data[1] using different lagged values of explanatory variables and then select the most satisfied one which has the smallest BIC value. The values of different criteria are shown as following.

Table 5 the summaries of model (10) with different lags

Predictor	Lag (*i, j, k*)			
	(1,1,1)	(1,1,2)	(2,1,1)	(2,2,2)
(intercept)	-2.56	-2.58	-8.14	-2.01
η_{t-1}	0.16	0.15	-0.73	0.35
y_{t-1}	5.43	5.37	24.82	4.79
$SPI_{t\text{-}i}$	6.82	6.59	14.81	6.98
$STS_{t\text{-}j}$	-1.25	-0.97	-4.75	0.74
$GTS_{t\text{-}k}$	-0.63	-0.66	-4.46	-0.52
Log-likelihood	-10.91	-10.90	-7.67	-10.64
BIC	13.16	13.15	9.92	12.88

[1] Swedish data including from the first quarter of 1971 to the fourth quarter of 2009

From the above results, the sign of coefficient of Stockholm PI index is positive which does not make economic sense. Meanwhile the coefficient is a little larger than expected. So, Stockholm PI index can not be used as explanatory variable and deleted from design matrix.

And the new proposed AR logit model is

$$logit(p_t) = \eta_t$$
$$\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 STS_{t\text{-}i} + \beta_2 GTS_{t\text{-}j} \tag{11}$$

The next step is to decide the lagged values of explanatory variables, *STS* and *GTS*. Since when employed long lags, the degree of freedom will be sacrificed correspondingly, we only compare the models that explanatory variables with lags up to four for simplicity.

After simple calculation, we find that model with lagged values $(STS_{t\text{-}3},\ GTS_{t\text{-}4})$ has the smaller BIC value and the signs of explanatory variables' coefficients make sense. And a good result is that the BIC is not declining when lags increase, that is to say, the influence power of *STS* and *GTS* on the current Swedish economy status is increasing within one year and then reducing as time passes.

The summary of the final model is shown as below.

Table 6 the summary of model (11), explanatory variables with lags (3, 4)

Predictor	Lag (3, 4)	Criteria	
(intercept)	-2.40	*Log-likelihood*	-10.43
η_{t-1}	-0.18	*BIC*	12.67
y_{t-1}	5.54		
STS_{t-i}	-1.29		
$GTS_{t\text{-}j}$	-0.63		

Then the final model is

$$logit(p_t) = \eta_t$$
$$\eta_t = -2.40 - 0.18\eta_{t\text{-}1} + 5.54 y_{t-1} - 1.29 STS_{t-3} - 0.63 GTS_{t\text{-}4} \tag{12}$$

To illustrate the in-sample performance of model (12), figure 2 shows the probability that the Sweden economy would be in a recession in one quarter ahead. The shadow area shows the actual Sweden economy status.

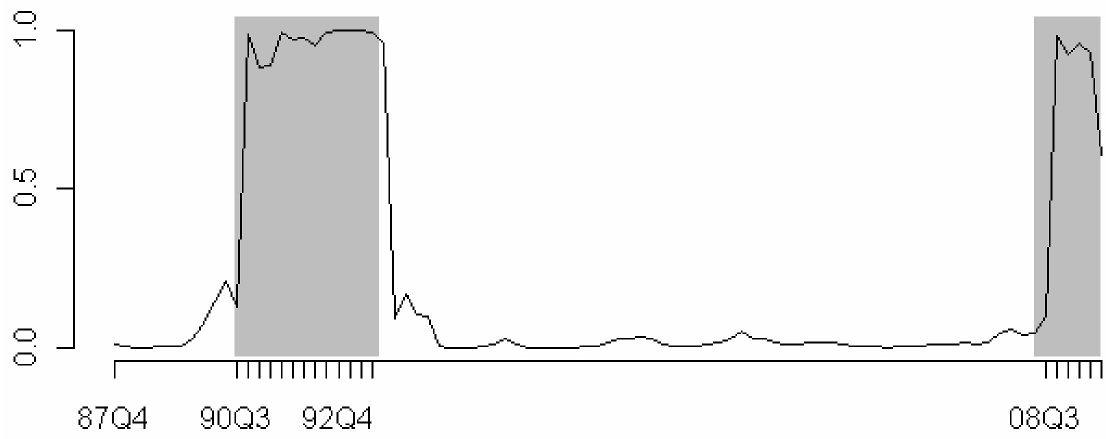

Figure 2 probability of recessions and actual recession periods of Sweden

Figure 2 shows two periods of recessions in shadow areas, one from the third quarter of 1990 to the third quarter of 1993 and another from the third quarter of 2008 to the fourth quarter of 2009[1].

The probability line matches the first shadow area well, that is to say, model (12) forecasts the first recession exactly. But model (12) forecasts the second recession is a little later than recession has happened about one quarter.

The percentage of correct prediction of model (12) can be checked by cross table 7. The threshold is fixed at 0.5.

Table 7 the percentage of correct prediction of model (12)

Observed Economy	Predicted Economy		Correct Percentage
	Expansion	Recession	
Expansion	69	1	98.6%
Recession	2	17	89.5%
Overall Percentage			96.6%

When the observed economy is in expansion, the model gives a 98.6% correct fitness of the expansion. Meanwhile, when the observed economy is in recession, only two predicted values of economy are not the same as the observed ones. Overall, the percentage of correct prediction of this model reaches a notably high level of 96.6%.

1 The data of interest rate can only be used from the first quarter of 1987

3.2 Out-of-sample results

Out-of-sample performance of the specified models will be mainly concerned, and this section will check the predicted power of the models specified in the previous section.

3.2.1 Out-of-sample results of U.S.

Based on previous section, the model (9) can predict well in one quarter ahead in-sample U.S. recession forecasts. So, consider using the same model, explanatory variables with same lags, to check the out-of-sample performance of AR logit model in U.S. recession forecasting.

$$logit(p_t) = \eta_t$$
$$\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 TS_{t-2} + \beta_2 GTS_{t-3} \tag{13}$$

Because forecast horizon is two, model (13) can make one quarter ahead and two quarters ahead forecasts.

Using data dated before forecasting quarter to estimate the parameters of model (13), and the out-of-sample performance of model (13) one quarter ahead forecast is shown as figure 3.

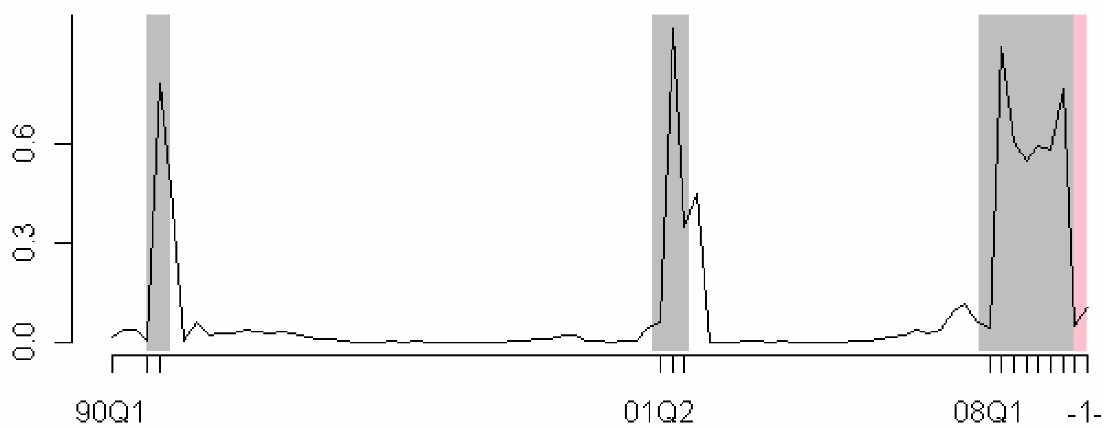

Figure 3 out-of-sample forecasts of U.S. one quarter ahead

One quarter ahead recession forecasts are made from the first quarter of 1990 to the first quarter of 2010. Because the actual status the first quarter of 2010, in recession or not, is not known, that means it is not included in the using data set, so using pink background to make it different from the others.

From figure 3, there are three recession periods. The first recession period from the fourth quarter of 1990 to first quarter of 1991 can be predicted well by the model. But the later two recession period starting from the second quarter of 2001 and the first quarter of 2008 can not be predicted exactly about one quarter late than the actual starting quarters, respectively. And the economic recession predicted by model (13) will end at the first quarter of 2010.

And out of sample performance of model (13) two quarters ahead forecast is shown as figure 4.

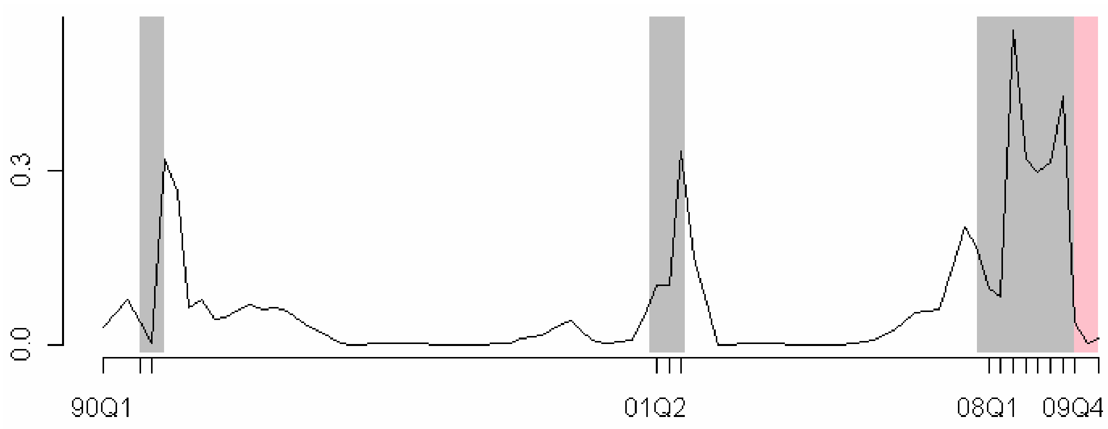

Figure 4 out-of-sample forecasts of U.S. two quarters ahead

Two quarters ahead recession forecasts are made from the first quarter of 1990 to the second quarter of 2010. The unknown actual economy status, in recession or not, is marked with pink background.

During the last two quarters' period starting from the fourth quarter of 1990, U.S. economy is in recession, but the estimated recession probability is lower than the threshold value, 0.5, and a little lag on the time. The same situation of lower lagged estimated probability happens at the other two recession periods.

Actually, there is only one period in recession, if the probability threshold is chosen at 0.5. That will be a kind of inconsistency with the fact. But the volatile probability can be checked directly, that is, when the recession probability goes up enough, economy recession happens in near periods. So, alternative choice is decrease the threshold to 0.3, and the prediction power of the model has been changed higher.

Figure 4 also shows that the recession staring from the first quarter of 2008 will end at the first quarter of 2010 finally.

3.2.2 Out-of-sample results of Sweden

Based on model (12), consider using following model to check the out-of-sample performance of AR logit model in Sweden recession forecasting.

$$logit(p_t) = \eta_t$$
$$\eta_t = \alpha\eta_{t-1} + \mu y_{t-1} + \beta_0 + \beta_1 STS_{t-3} + \beta_2 GTS_{t-4} \qquad (14)$$

Model (14) can make one quarter ahead up to three quarters ahead forecasts due to its forecast horizon is three. We still use data dated before forecasting quarter to estimate the parameters of model (14), and then make the out-of-sample forecast.

One quarter ahead forecast of model (14) is shown as figure 5.

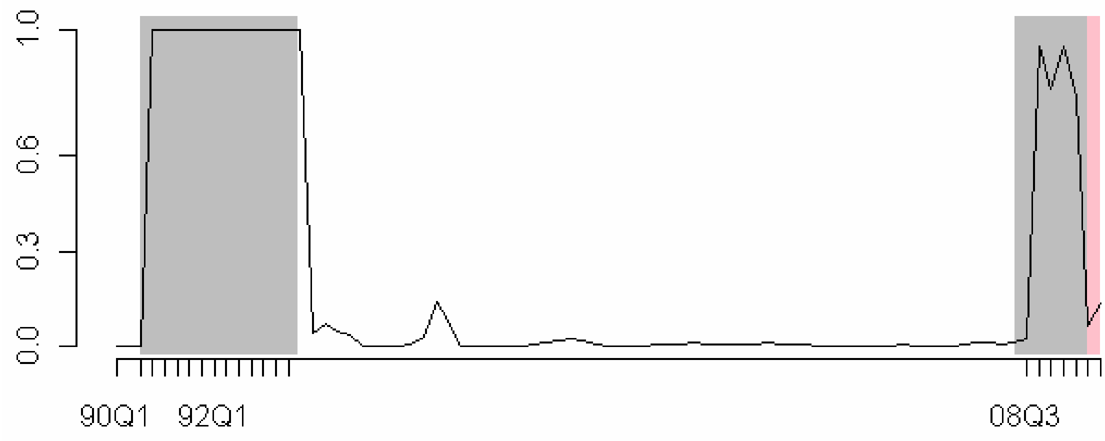

Figure 5 out-of-sample forecasts of Sweden one quarter ahead

Figure 5 shows the out-of-sample predictive performance one quarter ahead. The forecast is dated from the first quarter of 1990 and end up to the first quarter of 2010.

This model fits well when the first recession occurs. However, as for the second recession happens at the third quarter of 2008, the predicted result given by this model is about one quarter later than the real recession happens. In addition to the second recession forecasting, the model matches well with the real data.

Since the status of economy at the first quarter of 2010 is unknown in the data set, we use pink highlight to distinguish this period from others. Model (14) predicts that the recession starting from the third quarter of 2008 will end at the first quarter of 2010.

And the two quarters ahead forecast of model (14) is shown as figure 6.

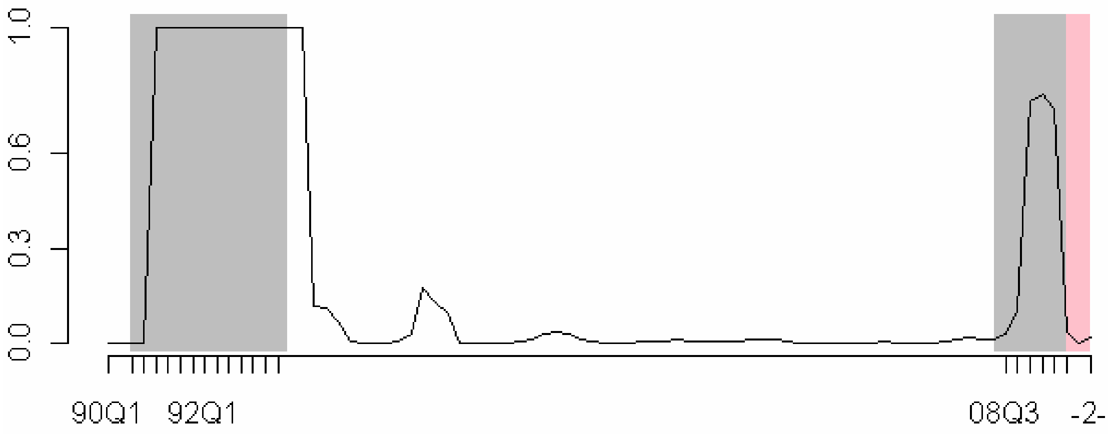

Figure 6 out-of-sample forecasts of Sweden two quarters ahead

Figure 6 illustrates the out-of-sample predictive result when model (14) gives two quarter ahead forecast. The forecast is dated from the first quarter of 1990 and end up to the second quarter of 2010.

When the first recessions occur at the third quarter of 1990, the corresponding predictive period is about one quarter lagged behind. When the first recession ends, the predictive value is one quarter lagged behind as well. The same situation appears when the second recession occurs. Since we do not know when the second recession end, we could hardly say whether the predictive value is lag or ahead to the real value.

The predictive result which is highlighted by pink color background suggests that the recession are going to end at the first quarter of 2010 same as one quarter ahead forecast.

As referred before, forecast horizon of model (14) is three, so the three quarters ahead forecast for Sweden business cycle is shown as figure 7.

From figure 7, we can see that the forecast is not very good which can be proved further by cross table 8.

Table 8 the percentage of correct prediction three quarters ahead

Observed Economy	Predicted Economy		Correct Percentage
	Expansion	Recession	
Expansion	58	3	95.1%
Recession	7	12	63.2%
Overall Percentage			87.5%

When the observed economy is in recession, the model gives a 63.2% correct estimation. But when observed economy is in expansion, three predicted quarters are not consistent with the observed economy. Overall, the percentage of correct prediction of this model is declining to 87.5%.

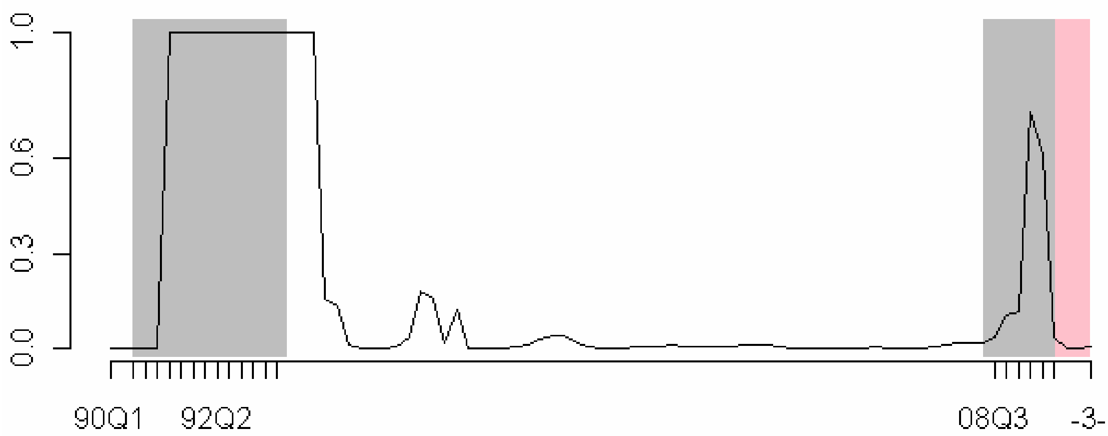

Figure 7 out-of-sample forecasts of Sweden three quarters ahead

So, based on the above analysis, although model (14) can make three quarters ahead forecast for Sweden economy, it is good choice to omit the three quarters ahead forecast of model (14).

4 Conclusion

In this paper, we examine that the AR logit model can be used to forecast not only U.S. economic recessions but also Sweden economic recessions. The processes of selecting design matrix for U.S. and Sweden are same, but the results are different. The lagged value of AR term is fixed at one, but the lagged values of explanatory variables are flexible. Generally, the determination of the lagged values of explanatory variables is through the BIC value of each model. After selecting the best fitted model, we are more concern about the in-sample and out-of-sample forecasting performance. The in-sample forecasting of U.S. with one quarter ahead shows a good fitness of the model. Specifically, this model gives an overall 93.5% correction of prediction. Thus, we can believe that the model will also suggest a good result for out-of-sample forecasting. The in-sample forecasting performance of AR logit model when applying it to Sweden shows an overall 96.6% correction of prediction. This result indicates that the AR logit model is applicable in Sweden, what's more, the percentage correction of prediction even better than U.S. Thus, we can conclude that the AR logit model is quite suitable when modeling the economy recession in Sweden.

After comparing the in-sample forecasting performance of AR logit model, the out-of-sample forecasting is more concerned when applying the model in practice. Since the forecast horizon in U.S. model is two, we can implement the out-of-sample forecasting up to two quarters ahead. We can get the forecasting result with one quarter ahead directly, but when the step of forecasting is larger than one, we use iterative method to get the forecasting results. Generally, the results of out-of-sample forecasting are not as good as the ones of in-sample forecasting. To get a better visual figure, we must limit the threshold to smaller value than 0.5 sometimes. Both the forecasting results with one quarter ahead and two quarters ahead show a small probability of recession which indicates the economy in U.S. will recover from a deep recession. In other words, the economy in U.S. will experience an expansion at least until second quarter in 2010. Since the forecast horizon in Sweden model is three, we can implement the out-of-sample forecasting up to three quarters ahead. The forecasting with one quarter ahead shows a better performance than the others. Nevertheless, all these forecasting results show a small probability of recession at the starting quarter of 2010 which means that the economy in Sweden will experience an expansion at least up to the third quarter in 2010.

Although the AR logit model can be considered as a powerful tool to model and predict economic recessions, we can observe from the figures that the predicted recessions mostly occur later than the real ones. This phenomenon partly can be explained as the lag effect of models. Since all the models we proposed include an AR term, it makes sense when we observe the predicted results react later than the real ones. Besides, it maybe interest to add more explanatory variables to improve the performance of AR logit model or consider applying the AR logit model to describe the economy recession in other homogeneous countries e.g. EEC is another possible way for further study.

References

[1] Bodie Z., Kane A. & Marcus A. J., 2004. *Investments*. 6th ed. U.S.: McGraw-Hill/Irwin.

[2] Chauvet M. & Potter S., 2005. Forecasting Recessions Using the Yield Curve. *Journal of Forecasting*, 24(2), pp.77-103.

[3] Estrella A. & Mishkin F. S., 1998. Predicting U.S. Recessions: Financial Variables as Leading Indicators. *The Review of Economics and Statistics*, 80(1), pp.45-61.

[4] Hamberg U. & Verständig D., 2009. Applying Logistic Regression Models on Business Cycle Prediction. Unpublished master thesis, *Stockholm School of Economics*.

[5] Kauppi H. & Saikkonen P., 2005. Predicting U.S. Recessions with Dynamic Binary Response Models. *Helsinki Center of Economic Research*, Discussion Paper No. 79.

[6] Moore.G.H, 1967. What is a Recession? *The American Statistician*, 21(4), pp. 16-190.

[7] Nyberg H., 2008. Testing an Autoregressive Structure in Binary Time Series Models. *Helsinki Center of Economic Research*, Discussion Paper No. 243.

[8] Nyberg H., 2009. A Bivariate Autoregressive Probit Model: Predicting U.S. Business Cycle and Growth Rate Cycle Recessions. *Helsinki Center of Economic Research*, Discussion Paper No. 272.

[9] Omori Y., 2003. Discrete Duration Model Having Autoregressive Random Effects with Application to Japanese Diffusion Index. *Journal of the Japan statistical society*, 33(1), pp.1-22.

[10] Schwarz G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), pp.461-464.

[11] Startz R., 2006. Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Center for Statistics in the Social Sciences University of Washington*, Working Paper No. 56.