

# Using Classification Function to integrate Discriminant Analysis, Logistic Regression and Backpropagation Neural Networks for Interest Rates Forecasting

Kyong Joo Oh and Ingoo Han\*

\*Graduate School of Management  
Korea Advanced Institute of Science and Technology

## Abstract

This study suggests integrated neural network models for interest rate forecasting using change-point detection, classifiers, and classification functions based on structural change. The proposed model is composed of three phases with two-staged learning. The first phase is to detect successive and appropriate structural changes in interest rate dataset. The second phase is to forecast change-point group with classifiers (discriminant analysis, logistic regression, and backpropagation neural networks) and their combined classification functions. The final phase is to forecast the interest rate with backpropagation neural networks. We propose some classification functions to overcome the problems of two-staged learning that cannot measure the performance of the first learning. Subsequently, we compare the structured models with a neural network model alone and, in addition, determine which of classifiers and classification functions can perform better. This article then examines the predictability of the proposed classification functions for interest rate forecasting using structural change.

**Keywords:** Backpropagation Neural Networks, Structural Change, Change-Point Detection, Pettitt Test, Discriminant analysis, Logistic Regression

## 1. Introduction

Interest rate forecasting has clear ramifications for cash management in any enterprise. Its movement also affects financing decisions such as capital budgeting and strategic investment. The prediction of interest rate is critical for managing risk in an investment portfolio as well as securing finance for corporate investment. Over the past several decades, statistical techniques and traditional software have been used extensively to model financial markets. Such statistical and software

models have been beneficial for understanding market behavior. This is especially true when the objective of the model is merely to describe and summarize the characteristics of a market rather than forecast its trajectory. For the task of interest rate forecasting, however, numerous studies in the past have underscored the inadequacy of statistical techniques and simulations based on traditional procedures.

Currently, several studies have demonstrated that artificial intelligence approaches, such as fuzzy theory

(Ju et al., 1997) and neural networks (Deboeck and Cader, 1994; Hong and Han, 1996), can be alternative methodologies for chaotic interest rate data (Larrain, 1991; Peter, 1991; Jaditz and Sayers, 1995). Previous work in the interest rate forecasting has tended to emphasize statistical techniques and artificial intelligent (AI) techniques in isolation over the past decades. However, an integrated approach, which makes full use of statistical approaches and AI techniques, offers the promise of increasing performance over each method alone (Chatfield, 1993). This article explores the ways in which such technologies may be combined synergistically, and illustrates the approach through the use of DA, LR and BPN as a data mining classifier. Up to date, it has been proposed that the integrated neural network model combining two or more models have a potential to achieve a high predictive performance in interest rate forecasting (Kim and Kim, 1996; Kim and Noh, 1997).

The movement of Interest rate is more fluctuated sensitively by government's monetary policy than other financial data (Gordon and Leeper, 1994; Strongin, 1995; Bernanke and Mihov, 1995; Christiano et al., 1996; Leeper et al, 1996; Bagliano and Favero). Especially, banks play a very important role in determining the supply of money: Much regulation of these financial intermediaries is intended to improve its control. One crucial regulation is reserve requirements, which make it obligatory for all depository institutions to keep a certain fraction of their deposits in accounts with the Federal Reserve System, the central bank in the United States (Mishkin, 1995). It is supposed that government take an intentional action to control the currency flow which has direct influence upon interest rate. Therefore, we can conjecture that the movement of interest rate has a series of change points occurred by the planned monetary policy of government.

To reflect these inherent characteristics of interest rate, Oh and Han (2000) propose a structured model three phases as follows: The first phase is to detect successive and appropriate structural changes in interest rate dataset. The second phase is to forecast change-

point group with classifiers. The final phase is to forecast the interest rate with BPN. This model has a disadvantage not to evaluate whether the change-point group detection is appropriately established in the second phase. In order to overcome these problems, we suggest the classification functions to integrate discriminant analysis (DA), logistic regression (LR), and backpropagation neural networks (BPN). Subsequently, we determine which of three classifiers (DA, LR and BPN) can perform better and furthermore examine the predictability of the proposed classification functions.

The case study performed in this article consists of the Treasury bill rate of 3 month's maturity in the U.S. from Jan., 1961 to May, 1999. Input variable selection is based on the causal model of interest rate presented by the econometricians. To explore the predictability, we divided the interest data into the training data over one period and the testing data over the other period. The predictability of interest rate is examined using the metrics of the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE).

We review the development of change-point detection and its application to the financial economics in Section 2. Section 3 describes the proposed integrated neural network model details through the various data mining classifiers. Section 4 reports the processes and the results of applied study. Finally, the concluding remarks are presented in Section 5.

## 2. Change-Point Problems

### 2.1. Structural Change in Financial Economics

The detection and estimation of a structural or parametric change in forecasting is an important and difficult problem. In particular, financial analysts and econometricians have frequently used piecewise-linear models which also include change-point models. They are known as models with structural breaks in the economics literature. In these models, the parameters are assumed to shift - typically once - during a fixed sample period and the goal is to estimate the two sets of

parameters as well as the change point or structural break.

In order to detect the structural change, change-point detection methods have been applied to macroeconomic time series. Rappoport and Reichlin (1989) and Perron (1989, 1990) conduct the first study in this field. From then on, several statistics have been developed which work well in a change-point framework, all of which are considered in the context of breaking the trend variables (Banerjee et al., 1992; Christiano, 1992; Zivot and Andrews, 1992; Perron, 1995; Vogelsang and Perron, 1995). In those cases where only a shift in the mean is present, the statistics proposed in the papers of Perron (1990) or Perron and Vogelsang (1992) stand out.

In spite of the significant advances by these works, we should bear in mind that some variables do not show just one change point. Rather, it is common for them to exhibit the presence of multiple change points. Thus, it seems advisable to introduce a large number of change points in the specifications of the models that allow us to obtain the abovementioned statistics. For example, Lumsdaine and Papell (1997) have considered the presence of two or more change points in trend variables. Based on this fact, we also assume the Treasury bill rates have two or more change points in our research model.

Up to date, there are few artificial intelligence models for financial applications to represent the change-point detection problems. Most of the previous research has a focus on the finding of unknown change points for the past, not to forecast for the future (Volkenhauer and Edmunds, 1997; Li and Yu, 1999). Our model finds change points in the learning phase and forecasts change points in the testing phase. It is demonstrated that the introduction of change points to our model will make the predictability of interest rate greatly improve. In this article, a series of change points will be detected by Pettitt test, a nonparametric change-point detection method since nonparametric statistical property is a suitable match for a neural network model that is a kind of nonparametric method

(White, 1992).

## 2.2. The Pettitt Test

In this study, a series of change points will be detected by the Pettitt test (Pettitt, 1979, 1980a), a nonparametric change-point detection method, since nonparametric statistical property is a suitable match for a neural network model that is a kind of nonparametric method (White, 1992). In this point, the introduction of the Pettitt test is fairly appropriate for the analysis of chaotic time series data. The Pettitt test is explained as follows.

Consider a sequence of random variables  $X_1, X_2, \dots, X_T$ , then the sequence is said to have a change-point at  $\tau$  if  $X_t$  for  $t = 1, 2, \dots, \tau$  have a common distribution function  $F_1(x)$  and  $X_t$  for  $t = \tau + 1, \tau + 2, \dots, T$  have a common distribution  $F_2(x)$ , and  $F_1(x) \neq F_2(x)$ . We consider the problem of testing the null hypothesis of *no-change*,  $H_0: \tau = T$ , against the alternative hypothesis of *change*,  $H_A: 1 \leq \tau < T$ , using a non-parametric statistic.

An appealing non-parametric test to detect a change would be to use a version of the Mann-Whitney two-sample test. A Mann-Whitney type statistic has remarkably stable distribution and provides a robust test of the change point resistant to outliers (Pettitt, 1980b). Let

$$D_{ij} = \text{sgn}(X_i - X_j) \quad (1)$$

where  $\text{sgn}(x) = 1$  if  $x > 0$ ,  $0$  if  $x = 0$ ,  $-1$  if  $x < 0$ , then consider

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T D_{ij} \quad (2)$$

The statistic  $U_{t,T}$  is equivalent to a Mann-Whitney statistic for testing that the two samples  $X_1, \dots, X_t$  and  $X_{t+1}, \dots, X_T$  come from the same population. The statistic  $U_{t,T}$  is then considered for values of  $t$  with  $1 \leq t < T$ . For the test of  $H_0: \text{no change}$  against

$H_A$  : change, we propose the use of the statistic

$$K_T = \max_{1 \leq t < T} |U_{t,T}|. \quad (3)$$

The limiting distribution of  $K_T$  is  $\Pr \equiv 2 \exp\{-6k^2/(T^2 + T^3)\}$  for  $T \rightarrow \infty$ .

In the time sequence dataset, the Pettitt test detects a possible change point in which the structural change is occurred. Once the structural change is detected through the test, the dataset is divided into two intervals. The intervals before and after the change point form homogeneous groups which take heterogeneous characteristics from each other. This process becomes a fundamental part of the binary segmentation method explained in Section 3.

### 3. Description of the Proposed Model

#### 3.1. The Integrated Neural Network Model

Classifiers, change-point detection method and neural network learning methods have been integrated to forecast the Treasury bill rate of 3 month's maturity in the U.S. The advantages of combining multiple techniques to yield synergism for discovery and prediction have been widely recognized (Gottman, 1981; Kaufman et al., 1991). This section provides the architecture and the characteristics of our research model to include the change-point detection and BPN. The proposed model is composed of three phases based on the Pettitt test as follows:

##### Phase 1: Constructing homogeneous groups

Pettitt test is a method to find a change-point in longitudinal data (Pettitt, 1979). It is known that interest rate at time  $t$  are more important than fundamental economic variables in determining interest rate at time  $t+1$ . (Larrain, 1991). Thus, we apply Pettitt test to Treasury bill rates at time  $t$  to generate a forecast for  $t+1$  in the leaning phase. The Pettitt test mentioned in Section 2 is method for finding just one change point in time series data. Based on this method, multiple change points can be obtained under the binary segmentation method (Vostrikova, 1981). With

$H_0$  as in Section 2, under the alternative hypothesis we now assume that there are  $R$  changes in the parameters, where  $R$  is a known integer. The alternative can be formulated as

$H_A^{(R)}$  : there are integers  $1 < k_1 < k_2 < \dots < k_R < n$  such that  $\theta_1 = \dots = \theta_{k_1} \neq \theta_{k_1+1} = \dots = \theta_{k_2} \neq \theta_{k_2+1} = \dots = \theta_{k_R} \neq \theta_{k_R+1} = \dots = \theta_n$  for the parameter  $\theta$ s.

We note that the test statistics under the null hypothesis will remain consistent against  $H_A^{(R)}$  as well, despite the fact that they were derived under the assumption that  $R=1$ . Without the loss of generality, we can deduce that the tests mentioned in Section 2 are extended to the form for “no change” against the “ $R$  changes” alternative  $H_A^{(R)}$ .

Vostrikova (1981) suggested a binary segmentation method as follows. First, use the change-point detection test. If  $H_0$  is rejected, the find  $\hat{k}_1$  that is the time where Equation (3) is satisfied. Next divide the random sample into two subsamples  $\{X_i : 1 \leq i \leq \hat{k}_1\}$  and  $\{X_i : \hat{k}_1 < i \leq n\}$ , and test both subsamples for further changes. One continues this segmentation procedure until no subsamples contain further change points. If exactly  $R$  changes are found, then one rejects  $H_0$  in favor of  $H_A$ .

This process plays a role of clustering that constructs groups as well as maintains the time sequence. In this point, Phase 1 is distinguished from other clustering methods such as the k-means nearest neighbor method and the hierarchical clustering method. They classify data samples by the Euclidean distance between cases without considering time sequence.

##### Phase 2: Group forecasting with classifiers

The significant intervals by Phase 1 are grouped to detect the regularities hidden in them and to represent the homogeneous characteristics of them. Such groups represent a set of meaningful trends encompassing the significant intervals. Since those trends help to find

regularity among the related output values more clearly, the neural network model can have a better ability of generalization for the unknown data. This is indeed a very useful point for sample design. In general, the error for forecasting may be reduced by making the subsampling units within groups homogeneous and the variation between groups heterogeneous (Cochran, 1977). After Phase 1 detects the appropriate groups hidden in the significant intervals, various classifiers (DA, LR and BPN) are applied and integrated to the input data samples at time  $t$  with group outputs for  $t+1$ . In this sense, Phase 2 is a model that is trained to find an appropriate group for each given sample.

### Phase 3: Forecasting the output with BPN

Phase 3 is built by applying the BPN model to each group. Phase 3 is a mapping function between the input sample and corresponding desired output (i.e. Treasury bill rate). Once Phase 3 is built, then the sample can be used to forecast the Treasury bill rate.

## 3.2. Classification Models

In Phase 2, we apply several classification models (DA, LR, and BPN) to interest rate forecasting. Based on these classifiers, we suggest new classification functions to integrate them. In this study, 3-month T-bills are assumed to be just one change-point. For multiple change-points, the proposed classification functions can be adjusted without the loss of generality.

### 3.2.1. Discussion of DA, LR, and BPN

The neural network methodology has been applied extensively to solve practical problems following the publication of the backpropagation algorithm for the multi-layer perceptron (Rumelhart, 1986). The algorithm was developed for the perceptron model, a simple structure to simulate a neuron (Rosenblatt, 1957). The backpropagation algorithm is based on artificial neural networks, where the neuron input path ( $i$ ) has a signal on it ( $X_i$ ) and the strength of the path is characterized by a weight ( $w_i$ ) for  $k$  input

variables. The neuron is modeled as summing the path weight times the input signal over all paths and adding the node bias ( $\Theta$ ). The output ( $Y$ ) is usually a sigmoid shaped logistic function that is expressed as follows:

$$Y = f(X) = \frac{1}{1 + \exp\left(-\sum_{i=1}^k w_i X_i + \Theta\right)} \quad (4)$$

Note that this S-shaped function reduces the effect of extreme input variables on the performance of the network. Today, BPN is the most widely used neural networks algorithm suitable for nonlinear data analysis in science, engineering, finance and other fields (Patterson, 1996). Thus, we introduce BPN to our model as a classification tool and a forecasting tool.

Discriminant analysis is used to classify individuals into one of two or more alternative groups on the basis of a set of measurements. Theoretically, this method is based on the Fisher's linear discriminant function by maximizing the ratio of between-groups and within-groups variances as follows (Fisher, 1936). The linear discriminant function is as follows:

$$D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (5)$$

where  $D$  is a discriminant score,  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  are estimated coefficients, and  $X_1, X_2, \dots, X_k$  are independent variables. The probability that a case with a discriminant score of  $D$  belongs to group  $i$  among  $g$  groups is estimated by the following equation

$$P(G_i | D) = \frac{P(D | G_i)P(G_i)}{\sum_{i=1}^g P(D | G_i)P(G_i)} \quad (6)$$

The prior probability, represented by  $P(G_i)$ , is an estimate of the likelihood that a case belongs to a particular group. The groups are known to be distinct, and each individual belongs to one of them. In addition, DA can be used to identify which variables contribute to making the classification (McLachlan, 1992; Hair, et al., 1995). This study applies DA to forecast the change-point group in the second stage.

Logistic regression can be used whenever an observation is to be classified into one of two populations. Thus, it is an alternative to the DA. For a binary response  $Y$  and a quantitative explanatory variable  $X$ , let  $\pi(X)$  denote the “success” probability when  $X$  takes value  $x$ . This probability is the parameter for the binomial distribution. The logistic regression model has linear form as like equation (4) for the logit transformation  $\log[\pi/(1-\pi)]$  of  $\pi$ , symbolized by  $\text{logit}(\pi)$ :

$$\begin{aligned}\text{logit}[\pi(X)] &= \log\left[\frac{\pi(X)}{1-\pi(X)}\right] \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k\end{aligned}\quad (7)$$

where  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  are estimated coefficients, and  $X_1, X_2, \dots, X_k$  are independent variables. Formula (7) implies that  $\pi(x)$  increases or decreases as an S-shaped function of  $x$ . This shows that LR represents nonlinear characteristic based on the statistical concept. Therefore, LR can be regarded as a mixed model to take both linear attribute of DA and nonlinear attribute of BPN.

The success probability can be directly obtained from the formula (7) using the exponential function  $\exp(x) = e^x$ , in the form

$$\pi(X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)} \quad (8)$$

The Fisher’s linear discriminant function gives rise to the logistic posterior probability when the multivariate normal model is assumed. LR represents an alternative method of classification when the multivariate normal model is not justified (Agresti, 1990, 1995; Christensen, 1997). In general, DA estimators are superior to LR estimators for the classification problems if the populations are normal with identical covariance matrices. Under non-normality, LR provides the improvement of prediction performance with maximum likelihood estimators for solving classification problems (Press and Wilson, 1978).

### 3.2.2. Proposed Classification Functions

Predicted values of DA, LR, and BPN are used as input in order to find the optimal predicted group. We suggest three classification functions to combine MDA, LR, and BPN: (1) the voting method (VM) for predicted group of three classifiers, (2) the arithmetic mean-assisted methods (AM) for predicted probability of three classifiers, and (3) the geometric mean-assisted methods (GM) for predicted probability of three classifiers, which are defined as follows.

$$VM_i = \frac{g[f(X_i)] + g[P(G_i | D)] + g[\pi(X_i)]}{3} \quad (9)$$

where  $g$  is the threshold function which generates 0 or 1 for  $i$ th observation.

$$AM_i = \frac{f(X_i) + P(G_i | D) + \pi(X_i)}{3} \quad (10)$$

$$GM_i = \sqrt[3]{f(X_i) P(G_i | D) \pi(X_i)} \quad (11)$$

Finally, the optimal predicted group for each method is obtained by the threshold function  $F(X)$  which is defined as follows:

$$F(X_i) = \begin{cases} 0, & VM_i \text{ (or } AM_i \text{ or } GM_i) < 0.5 \\ 1, & VM_i \text{ (or } AM_i \text{ or } GM_i) \geq 0.5 \end{cases} \quad (12)$$

## 4. Empirical Results

The input variables used in this study are M2, consumer price index, expected real inflation rates and industrial production index. They are used in both Phase 2 and Phase 3. The lists of variables used in this study are summarized in Table 1. They are those which were found significant in interest rates forecasting by previous study (Oh and Han, 2000). To obtain stationary and thereby facilitate forecast, the input data were transformed by a logarithm and a difference operation. Moreover, the resulting variables were standardized to eliminate the effects of units.

The training phase included observations from January 1961 to December 1986 while the testing phase runs from January 1987 to May 1999. The interest rate data are presented in Figure 1. Figure 1 shows that the movement of interest rates is highly fluctuated during the last forty years.

Table 1. Description of Variables.

Variable Name	Description	Attribute
TBILL	Treasury Bill with 3 month's maturity	Output
M2	Money Stock	Input
CPI	Consumer Price Index	Input
ERIR	Expected Real Interest rate	Input
IPI	Industrial Production Index	Input

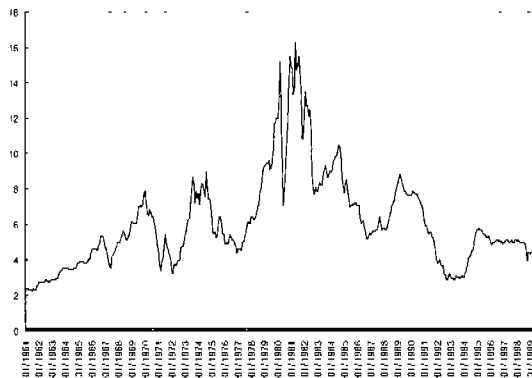


Figure 1. Yield of the U.S. Treasury bills with a maturity of 3 months from Jan. 1961 to May 1999

The study employed seven neural network models. One model, labeled PureNN, involve input variables at time  $t$  to generate a forecast for  $t+1$ . The input variable is M2, CPI, ERIR and IPI. The second type has two-step forecasting models which consist of three phases mentioned in section 3. The first step is Phase 2 that forecasts the change-point group while the next step is Phase 3 that forecasts the output. Table 2 shows classifiers and classification functions used in the proposed model. For validation, seven learning models were also compared.

Table 2. Models and their associated classifiers and classification function for the U.S. Treasury bill rate forecasting.

Model	Classifier and classification function used in the model
PureNN	None
LRNN	Logistic regression
BPNN	Backpropagation Neural Network
DANN	Discriminant analysis
VMNN	Formula (9) by voting method
AMNN	Formula (10) based on Arithmetic Mean
GMNN	Formula (11) based on Geometric Mean

The Pettitt test is applied to the training set of

interest rates. It is assumed that there exist two structural groups in 3-month T-bills dataset in this study. The first interval runs from January 1961 to January 1973 while the second interval is from February 1973 to December 1986. Thus, we obtain two significant intervals as like the result of Table 3. Table 3 also presents descriptive statistics including the mean and the variance. Group 1 is the stable interval that has low variance. Group 2 is more fluctuated intervals than Group 1 in term of the variance. We conjecture that two groups have dissimilar attributes from each other.

Table 3. Period and descriptive statistics of groups for the learning phase, Jan. 1961 ~ Dec. 1986

	Group 1	Group 2
Periods	Jan., 1961 ~ Jan., 1973	Feb., 1973 ~ Dec., 1986
Minimum	2.24	4.35
Maximum	7.87	16.3
Range	5.63	11.95
Mean	4.44	8.31
Variance	1.77	7.90
Standard Deviation	1.35	2.82
Skewness	0.63	0.91
Kurtosis	-0.14	0.17

Numerical values for the performance metrics by predictive model are given in Table 4. Where the forecasts are not statistically independent and not always normally distributed, the comparison using the forecast's APEs is commonly used (Carbone and Armstrong, 1982). According to MAPE, therefore, the outcomes indicate that the models with classifiers and classification functions are superior to the pure BPN model. In particular, GMNN is the best model among all of the models.

Table 4. Performance results in the case of US Treasury bill rate forecasting based on RMSE, MAE and MAPE

Model	RMSE	MAE	MAPE (%)
PureNN	0.3775	0.3178	6.493
DANN	0.4346	0.3289	6.066
LRNN	0.4044	0.3098	5.744
BPNN	0.3948	0.3024	5.615
VMNN	0.3992	0.3036	5.679
AMNN	0.3465	0.2696	5.326
GMNN	0.3242	0.2449	4.741

Our approach to integration involves a multistrategy technique which may be called two-staged learning. The integrated models using classifiers and classification functions provide the good results with the two-staged learning process. Especially, the models with the proposed classification functions perform well in this study. In two-staged learning, the forecast from the superior method is selected on a case-by-case basis to determine the output of overall model. In other words, the second learning (BPN in this study) serves as a metalevel process to determine which of three elementary modules (LR, BPN, DA, and their combined classification functions in this study) perform better. In this point, GM is a good metalevel predictive method. Thus, we will choose geometric mean-assisted method as the elementary module for real application of model.

The pairwise t-test is used to examine whether there exist the differences in the predicted values of models according to the absolute percentage error (APE), where is highly robust (Armstrong and Collopy, 1992; Makridakis, 1993). Iman and Conover (1983) demonstrate that this test is robust to the distribution of the data, to nonhomogeneity of variances, and to statistical dependence, where sample sizes are reasonably large. Table 5 shows t-values when the prediction accuracies of the left-vertical methods are compared with those for the right-horizontal methods. Mostly, classification function-assisted methods perform significantly better than the other models at 1% or 10% significant level. Specially, the geometric mean-assisted method is demonstrated to obtain the best improved performance.

The integrated neural network models using structural change turn out to have a high potential in interest rate forecasting. This is attributable to the fact that it categorizes the input data samples into homogeneous group and extracts regularities from each homogeneous group. Therefore, the proposed network models using structural change can cope with the noise or irregularities more efficiently than the pure BPN model. In addition, geometric mean-assisted model and

arithmetic mean-assisted model perform very well as a tool in interest rate forecasting.

Table 5. Pairwise t-tests for the differences in residuals for US interest rate prediction based on the absolute percentage error (APE).

Model	DA NN	LR NN	VM NN	BP NN	AM NN	GM NN
PureNN	0.81	1.56	1.59	1.79	3.34	5.42
				*	***	***
DANN		4.14	4.11	4.51	2.54	3.68
		***	***	***	**	***
LRNN			0.67	1.64	1.70	3.17
					*	***
VMNN				1.18	1.24	2.76
						***
BPNN					1.06	2.66
						***
AMNN						2.61
						***

\*\*\* Significant at 1%; \*\* Significant at 5%;

\*Significant at 10%

## 5. Concluding Remarks

This article has suggested the integrated neural network models in the interest rate forecasting using structural change. The basic concept of proposed model is to obtain significant intervals by change-point detection, to identify them as change-point groups, and to involve them in interest rate forecasting. We propose integrated neural network models which consist of three phases. In the first phase, we conduct the nonparametric statistical test for the change-point detection to construct the homogeneous groups. In the second phase, we apply several classifiers and classification functions to forecast the change-point group. In the final phase, we apply BPN to forecast the output.

The neural network models to represent structural change perform significantly better than the pure BPN model at a 1% or 10% significant level. Experimental results showed that the proposed models outperform the pure BPN model significantly, which implies the high potential of involving the change-point detection in the model. The geometric mean-assisted model and arithmetic mean-assisted model performed very well as classification functions. Our integrated neural network models are demonstrated to be useful intelligent data analysis methods with the concept of structural change.



In conclusion, we have shown that the proposed models improve the predictability of interest rate significantly.

The proposed models have the promising possibility of improving the performance if further studies are to focus on the decision of the optimal number of change. In final phase of the model, other intelligent approaches can be used to forecast the final output besides BPN. In addition, the proposed models may be applied to other chaotic time series data, such as stock market prediction and exchange rate prediction. By the extension of these points, future research is expected to provide more improved neural network models with superior performances.

## References

- Agresti, A. (1990). *Categorical Data Analysis*, New York: Wiley.
- Agresti, A. (1995). *An Introduction to Categorical Data Analysis*, New York: Wiley.
- Armstrong, J.S. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons, *International Journal of Forecasting*, 8, 69-80.
- Bagliano, F.C. & Favero, C.A. (1999). Information from financial markets and VAR measures of monetary policy, *European Economic Review*, 43, 825-837.
- Banerjee, A., Lumsdaine, R. & Stock, J. (1992). Recursive and sequential tests of the unit root and trend break hypothesis: Theory and international evidence, *Journal of Business and Economic Statistics*, 10, 271-287.
- Bernanke, B.S. & Mihov, I. (1995). Measuring monetary policy, *Working Paper No. 5145*, NBER, Cambridge, MA.
- Carbone, R. & Armstrong, J.S. (1982). Evaluation of extrapolative forecasting methods: Results of academicians and practitioners, *Journal of Forecasting*, 1, 215-217.
- Chatfield, C. (1993). Neural networks: Forecasting breakthrough or passing fad?, *International Journal of Forecasting*, 9, 1-3.
- Christensen, R. (1997). *Log-Linear Models and Logistic Regression*, New York: Springer.
- Christiano, L.J., Eichenbaum, M. & Evans, C.L. (1996). The effects of monetary policy shocks: Evidence from the flow of funds, *Review of Economics and Statistics*, 78, 16-34.
- Cochran, W.G. (1977). *Sampling Techniques*, Third edition, New York: Wiley.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, 7, 179-188.
- Gordon, D. & Leeper, E.M. (1994). The dynamic impacts of monetary policy: An exercise in tentative identification, *Journal of Political Economy*, 102, 1228-1247.
- Gottman, J.M. (1981). *Time-Series Analysis*, New York: Cambridge University Press.
- Hair, J.F. Jr., Anderson, R.E., Tatham, R.L., & Balck, W.C. (1995). *Multivariate Data Analysis with Readings*, Fourth edition, New York: Prentice Hall.
- Hong, T. & Han, I. (1996). The prediction of interest rate using artificial neural network models, *First asia-pacific decision sciences institute conference*, 975-984.
- Iman, R. & Conover, W.J. (1983). *Modern Business Statistics*, New York: Wiley.
- Jaditz, T. & Sayers, C.L. (1995). Nonlinear in the interest rate risk premium, In: R.R. Trippi (Ed.), *Chaos and Nonlinear Dynamics in the Financial Markets*. Chicago: Irwin, pp.335-357.
- Ju Y.J., Kim, C.E., & Shim, J.C. (1997). Genetic-based fuzzy models: interest rate forecasting problem, *Computers and Industrial Engineering*, 33(3-4), 561-564.
- Kaufman, K.A., Michalski, R.S. & Kerschberg, L. (1991). Mining for knowledge in databases: Goals and general description of the INLEN system. In: G. Piatetsky-Shapiro and W.J. Frawley (Eds.), *Knowledge Discovery in Databases*, Cambridge, MA: AAAI/MIT Press, pp. 449-462.
- Kim S.H. & Noh, H.J. (1997). Predictability of interest rate using data mining tools: a comparative analysis of Korea and the US, *Expert Systems with Application*, 15, 85-95.
- Kim, S.H. & Joo, J. (1997). Second-order learning for complex forecasting tasks: case study of Video-On-Demand, *Korean Journal of Expert System*, 3, 31-45.
- Kim, S.H. & Kim, K.M. (1996). Integrating multivariate statistics and neural networks for financial prediction: case study in interest rate forecasting, *First Asia-Pacific Decision Sciences Institute Conference*, 995-1003.
- Larain, M. (1991). Empirical tests of chaotic behavior in a nonlinear interest rate model, *Financial Analysts Journal*, Sep.-Oct., 51-62.
- Leeper, E.M. (1997). Narrative and VAR approaches to monetary policy: Common identification problems, *Journal of Monetary Economics*, 40, 641-657.
- Li, H.L. & Yu, J.R. (1999). A piecewise regression analysis with automatic change-point detection, *Intelligent Data Analysis*, 3, 75-85.
- Lumsdaine, R.L. & Papell, D.H. (1997). Multiple trends and the unit root hypothesis, *The Review of Economics and Statistics*, 79, 212-218.
- McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*, New York: Wiley.
- Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns, *International Journal of Forecasting*, 9, 527-529.
- Mishkin, F.S. (1995). *The economics of money, banking, and financial markets*, Fourth edition, New York: Harper Collins.
- Oh, K.J. & Han, I. (2000). Using change-point detection to support artificial neural networks for

- interest rates forecasting, *Expert Systems with Applications*, 19, 105-115.
- Patterson, D.W. (1996). *Artificial Neural Networks*. Englewood Cliffs, NJ: Prentice Hall.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, 57, 1361-1402.
- Perron, P. (1990). Testing for a unit root in time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153-162.
- Perron, P. (1995). Further evidence on breaking trend functions in macroeconomic variables," manuscript, Université de Montreal, Canada.
- Perron, P. & Vogelsang, T. (1992a). Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics*, 10, 301-320.
- Peters E.E. (1991). *Chaos and Order in the Capital Markets*, New York: Wiley, 1991.
- Pettitt, A.N. (1979). A non-parametric approach to the change-point problem, *Applied Statistics*, 28(2), 126-135.
- Pettitt, A.N. (1980a). A simple cumulative sum type statistic for the change-point problem with zero-one observations, *Biometrika*, 67, 79-84.
- Pettitt, A.N. (1980b). Some results on estimating a change-point using nonparametric type statistics, *Journal of Statistical Computation and Simulation*, 11, 261-272.
- Press, S.J., and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of American Statistical Associations*, 73(364), 699-705.
- Rapport, P. & Reichlin, L. (1989). Segmented trends and non-stationary time series," *The Economic Journal*, 99, 168-177.
- Rosenblatt, F. (1962). *Principles of Neurodynamics*, New York: Spartan.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by back propagation. In: D.E. Rumelhart, J.L. McClelland & PDP research group (Eds.), *Parallel Distributed Processing*, Vol. 1, Cambridge, MA: MIT Press.
- Strongin, S. (1995). The identification of monetary policy disturbances. Explaining the liquidity puzzle, *Journal of Monetary Economics*, 35, 463-497.
- Vogelsang T. & Perron, P. (1995). Additional tests for a unit root allowing for a break in the trend function at an unknown time, Manuscript, Department of Economics, Ithaca, New York.
- Vostrikova, L.J. (1981). Detecting disorder in multidimensional random process, *Soviet Mathematics Doklady*, 24, 55-59.
- White, H. (1992). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. In: H. White (Ed.), *Artificial Neural Networks: Approximations and Learning Theory*, Oxford, UK: Blackwell.
- Wolkenhauer, O. & Edmunds, J.M. (1997). Possibilistic testing of distribution functions for change detection, *Intelligent Data Analysis*, 1, 119-127.
- Zivot, E. & Andrews, D.W.K. (1992). Further evidence on the great crash, the oil-price shocks, and the unit-root hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.