輔仁管理評論
中華民國 86 年 9 月，第四卷第二期，101-112

Time Series Forecasting of Kaohsiung Unemployment Rate Using Neural Network Model

CHIH-CHOU CHIU* AND DEBORAH F. COOK**

Department of Business Administration, Fu-Jen Catholic University

***Management Science Department, Virginia Tech, USA*

Received Jun. 21 1996; First Revised Sep. 2 1996; Second Revised Oct. 23 1996; Accepted Nov. 11 1996

ABSTRACT

A backpropagation neural network was designed for short term unemployment rate forecasting. The forecasting was performed on Kaohsiung unemployment rate to demonstrate the predictive capability of the network. Extensive studies were performed on the effect of various factors such as learning rate and the number of hidden nodes. The monthly unemployment rate from June 1983 to Feb. 1992 was evaluated by neural network model and alternative methods, such as space-time series analysis, univariate ARMA model and state space model, the utilization of neural model significantly provide better forecasting of unemployment rate than any other alternative methods. Generally, the prediction precision of the neural network is 40% higher than the prediction made by the other models.

Keywords: Unemployment rate, Space-Time series analysis, Artificial neural network, Backpropagation learning algorithm

I. INTRODUCTION

It is well known that the measure of labor market performance can provide useful information in determining the government development policy. Many measurements can be used as a basis for the prediction. Unemployment rate is an important indicator of labor market performance and can be used to guide predictive techniques. Various statistical models have been proposed for the prediction of economical behavior and these methodologies has been criticized by Sims (1980). On of Sim's major concerns is the often arbitrary exclusion restrictions used to identify the models. To decide the exact relationship between considered variables, Sim's has suggested the utilization of vector autoregressive methodology (VAR) to build the empirical model. He claims that the using of VAR procedure can avoid imposing incredible restrictions and the estimation procedure is simple and consistent. Zellner (1979) and Zellner and Palm (1974) also shown that the structural system can be written in the form of a VARMA multivariate time series model. Note that univariate time-series models (ARIMA) which use only the past history of a single variable to model the behavior of that same series can be regarded as a special case of a VARMA model which uses simply a smaller information set. Although the time series techniques can be used to build models of unemployment rates, these models have been reported to lack sufficient precision (Granger, 1969; Funke, 1992). Moreover, some additional data processing works, such as differencing and transformation, might needed right before these statistical models were employed to avoid the problems of non-stationarity in time and space (Bronars et al. 1987). To overcome these inconvenient works, several new approaches, such as model-free forecasting and neural network with model free concept, were proposed recently to do the forecasting (Wu, 1994; Chiu et al. 1995).

In this paper the use of neural network theory in predicting the unemployment rate was explored to solve the problem caused by the lack of precision and the extra works. In additions, several statistical methods, such as space-time series analysis, univariate ARIMA model and state space model, were used to build the forecasting models. To demonstrate the effectiveness of the proposed approach, short-term forecasting is performed on Kaohsiung unemployment rate. The backpropagation learning technique with various learning rates is extensively studied to determine the connection weights between neurons. In addition, the number of hidden neurons is also varied to see the effect on the convergence rate. Results from this study indicate that the utilization of neural model significantly provide better forecasting of unemployment rate than other alternative methods. Generally, the prediction precision of the neural network is 40% higher than the prediction made by the other models.

II. NEURAL NETWORKS

A neural network is a massively parallel system comprised of highly interconnected, interacting processing elements, or units, that are based on neurobiological models (Brainmaker, 1989). Neural networks process information through the interactions of a large number of simple processing elements or units, also known as neurons. Knowledge is not stored within individual processing units, but is represented by the strength between units (Bauer, 1988). Each piece of knowledge is a pattern of activity spread among many processing elements, and each processing element can be involved in the partial representation of many pieces of information.

Neural networks can be classified into two different categories, feedforward networks and feedback networks (Rumelhart et al. 1986). The feedback networks contain neurons that are connected to themselves, enabling a neuron to influence other neurons and itself. Examples of this type of network are Kohonen self-organizing network and the Hopfield network. Neurons in feedforward networks (as shown in Figure 1) take inputs only from the previous layer and send outputs only to the next layer. The ADALINE and backpropagation neural network are two typical examples of this kind of network.

As shown in Figure 1, the neural net consists of a number of nodes or neurons connected by links. The nodes in the neural network can be divided into three layers: the input layer, the output layer and one or more hidden layers. The nodes in the

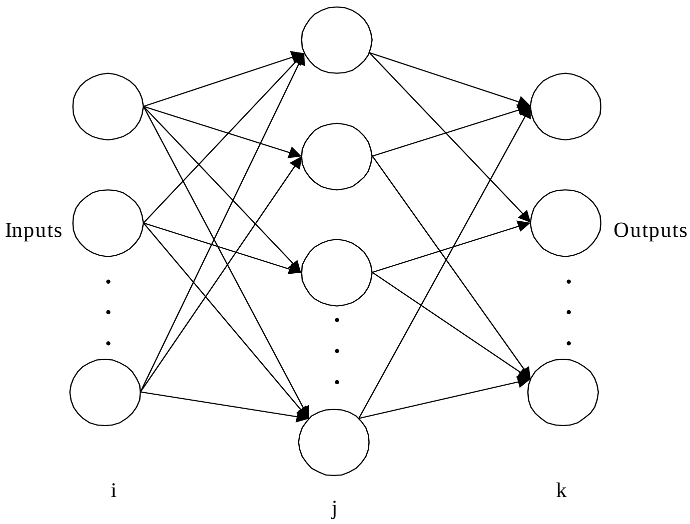

Figure 1. A backpropagation Neural Network

input layer receive input signals from an external source and the nodes in the input layer receive input signals from an external source and the nodes in the output layer provide the target output signals.

The output of each neuron in the input layer is the same as the input to that neuron. For each neuron j in the hidden layer and neuron k in the output layer, the net inputs are given by

$$net_j = \sum_i w_{ji} * o_i \quad \text{and} \tag{1}$$

$$net_k = \sum_j w_{kj} * o_j \tag{2}$$

where i (j) is a neuron in the previous layer, o_i (o_j) is the output of node i (j) and w_{ji} (w_{kj}) is the connection weight from neuron i (j) to neuron j (k). The neuron outputs are given by

$$o_i = net_i \tag{3}$$

$$o_j = \frac{1}{1+\exp-(net_j+\theta_j)} = f_j(net_j, \theta_j) \quad \textbf{(4)}$$

$$o_k = \frac{1}{1+\exp-(net_k+\theta_k)} = f_k(net_k, \theta_k) \quad \textbf{(5)}$$

where net_j (net_k) is the input signal from the external source to the node j (k) in the input layer and θ_j (θ_k) is a bias.

The generalized delta rule is the conventional technique used to derive the connection weights of the feedforward network (Rumelhart et al. 1986). Initially, a set of random numbers are assigned to the connection weights. Then for a presentation of a pattern p with target output vector $t_p=[t_{p1}, t_{p2},..., t_{pM}]^T$, the sum of squared error to be minimized is given by

$$E_p = \frac{1}{2}\sum_{j=1}^{M}(t_{pj} - o_{pj})^2 \quad \textbf{(6)}$$

where M is the number of output nodes. By minimizing the error E_p using the technique of gradient descent, the connection weights can be updated by using the following equations(Rumelhart et al.1986 1986) as

$$\Delta w_{ji}(p) = \eta\delta_{pj}o_{pi} + \alpha\Delta w_{ji}(p-1) \quad \textbf{(7)}$$

where for output nodes

$$\delta_{pj} = (t_{pj} - o_{pj})o_{pj}(1-o_{pj}) \quad \textbf{(8)}$$

and for other nodes

$$\delta_{pj} = (\sum_k \delta_{pk} * w_{kj})o_{pj}(1-o_{pj}) \quad \textbf{(9)}$$

Note that the learning rate η affects the network's generalization and the learning speed to a great extent. The overall training (learning) process for the network using the gradient descent technique is summarized in Figure 2.

```
Initialize the weights between layers, w_ji and w_kj ;

for m=1 to output layer iteration number or error criterion do
        decrease the adjusting rate for w_ji and w_kj gradually;

    for n=1 to training sample size do
              calculate the output for each hidden node;
              calculate the output for each output node;
              accumulate the difference between actual and target outputs;
        calculate the modified gradient for w_kj;
              calculate the modified gradient for w_ji;
        modify w_kj;
        modify w_ji;
    end
end
```

Figure 2. Training Process for Networks

III. UNEMPLOYMENT RATE FORECASTING USING BACKPROPAGATION ALGORITHM

An artificial neural network was used to forecast short-term unemployment rate. This method is dynamic in the sense that the monthly unemployment rate is predicted sequentially using the previous value of the unemployment rate, along with two difference terms, *X(t-1)-X(t-2)* and *X(t-1)-X(t-3)*. The network investigated in this paper is illustrated in Figure 3.

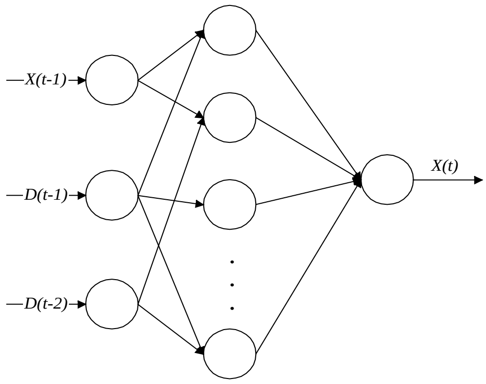

Figure 3. The Utilized Neural Network Topology

The network model consists of 3 layers. The input layer has 3 elements or nodes, one representing the actual unemployment rate at the previous time interval and the remains representing the differencing terms. The hidden layer consists of a number of nodes used for computational purposes. The number of hidden nodes must be determined experimentally and this determination is discussed later. The output layer consists of a single node representing the actual monthly unemployment rate at the next time interval. The terms in Figure 3 are defined as:

* X(t-1) = the actual unemployment rate at time t-1
* D(t-1) = the difference between the actual unemployment rate at time t-1 and the actual unemployment rate at time t-2 (ie. X(t-1)-X(t-2))
* D(t-2) = the difference between the actual unemployment rate at time t-1 and the actual unemployment rate at time t-3 (ie. X(t-1)-X(t-3))
* X(t) = the actual unemployment rate at time t

In this research, the differencing terms, *D (t-1)* and *D (t-2)* were utilized as inputs for the neural network because the unemployment rate time series data set has been identified as an autoregressive integrated moving average (ARIMA) model by Wu and Chen (1993). In Wu and Chen works, twice differencing operations have been used to get rid of the process non-stationarity. Thus, the differencing terms were considered as an impact factor and used as an input variable in the neural network model. An autoregressive integrated moving average model is a combination of regression with dependent variables related to past values of themselves at varying time lags and the weighted sum of previous deviations of the process. A general ARIMA process can be represented by the equation:

$$\phi(B)\,(1-B)^d\;X(t)=\theta(B)\,\varepsilon(t) \tag{10}$$

where $\phi(B)$ is the autoregressive operator of order p , such that $\phi(B)=1-\phi_1B-\phi_2B^2-\ldots\phi_pB^p$, B is the backward shift operator which can be defined by $BX(t)=X(t-1)$, d represents the difference time, $X(t)$ is value of the time series at time t, $\theta(B)$ is the moving average operator of order q , such that $\theta(B)=1-\theta_1B-\theta_2B^2-\ldots\theta_qB^q$, and $\varepsilon(t)$ is a normal, independently distributed error term.

In the development of ARIMA model, Wu and Chen employed twice differencing operations to filter out the process non-stationarity and used autocorrelation diagram and partial autocorrelation diagram to identify the model order and relative coefficients. In addition to the above univariate ARIMA model, a couple of alternative forecasting methods, space time series analysis (STARMA) and state space model were also included in Wu and Chen's paper. The objective of their research is to investigate the forecasting performance for the unemployment rate by means of three different statistical approaches. A detailed description of all these builded models can be found in (Wu and Chen, 1993). Since there are only two input nodes and one output node in the neural network model, the initial number of hidden nodes to test was chosen to be 3, 4, 5, and 6 for this example. There is not a commonly accepted method for determining the number of hidden nodes to use in a backpropagation neural network model. Consequently, experimentation and rules of thumb were used to arrive at the numbers to use. Too few hidden nodes limit a network generalization capabilities, while too many hidden nodes can result in overtraining or memorization by the network.

An initial evaluation of the learning rate was also conducted. Learning rates of 0.01, 0.2, 0.4, and 0.6 were used with the networks. Large step sizes in the learning process can cause the network to oscillate and not accomplish the required minimization of the error term.

IV. NUMERICAL EXAMPLES

To demonstrate the effectiveness of the proposed neural technique, short-term unemployment rate forecasting is performed on the Kaohsiung unemployment rate. The neural network simulator NETS, developed by NASA (Baffes, 1989), was used to develop the unemployment rate forecasting networks. NETS was implemented on a PC with Pentium 75 MHz CPU. It is a C based simulator that provides a system for developing various neural network configurations using the generalized delta backpropagation learning algorithm.

The output node in this network corresponds to a one step ahead prediction. Each training pattern consists of one previous actual monthly unemployment rate, the differencing term, and the actual current unemployment rate. That is, the training patterns can be expressed as {*X(2), D (2), D (1), X(3)*}, { *X(3), D (3), D (2), X(4)* },..., and { *X(t-1), D (t-1), D (t-2), X(t)* } where *t=3, 4, ..., N*.

The predictions of the proposed approach and alternative methods are presented below as tables and as graphs of predicted vs. actual data. A calculation of RMSE is also included. The normalized outputs of the Kaohsiung unemployment rate from June 1983 to Feb. 1992 is shown in Figure 4 for values of *t* from 1 to 105. Of the 105 values shown in the figure, the first 96 values were used to train the network and the remaining 9 values were used for testing. The convergence criteria used for training is a mean square error less than or equal to 0.01 or a maximum of 3000 iterations.

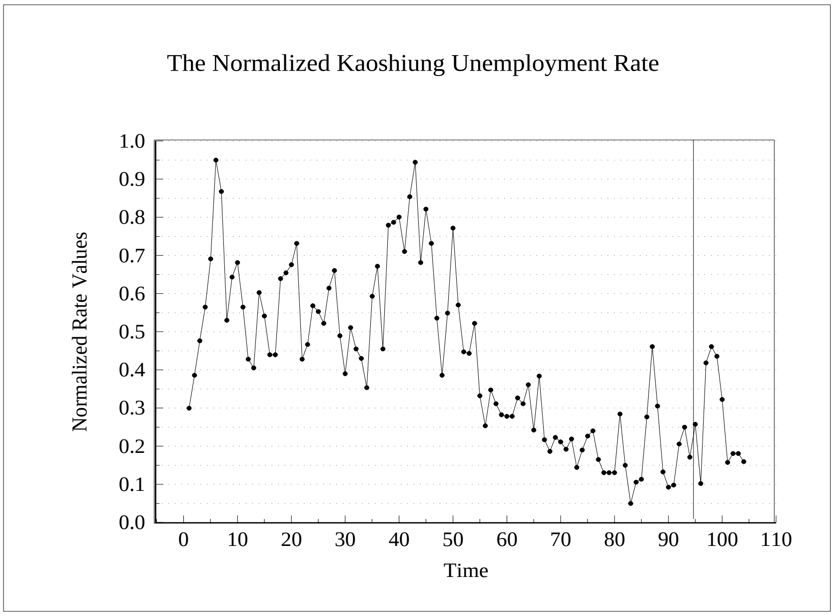

Figure 4. Normalized Kaohsiung Unemployment Rate From June 1983 to Feb. 1992

The analysis results of the neural network with different combinations of hidden nodes number and decreasing gradients in learning rates are comprised in Table 1.

Table 1. The Analysis Results for the Neural Network Parameters

Number of Hidden Nodes	Learning Rate η	Training RMSE	Testing RMSE
3	0.01	0.107740	0.014026
	0.20	0.112956	0.013573
	0.40	0.152855	0.019264
	0.60	0.142933	0.014211
4	0.01	0.108201	0.013286
	0.20	0.104999	0.013050
	0.40	0.116101	0.013279
	0.60	0.143005	0.025124
5	0.01	0.103859	0.016121
	0.20	0.101363	0.015776
	0.40	0.114950	0.014347
	0.60	0.123409	0.035368
6	0.01	0.108150	0.013334
	0.20	0.099940	0.017094
	0.40	0.115776	0.013378
	0.60	0.139372	0.037368

It is observed that the results for the 3-4-1 network with the learning rate of 0.20 provides the best forecasting RMSEs for unemployment rate. In other words, the 3-4-1 network is the best forecasting neural network. To examine the convergence characteristics of the proposed combined approach, the RMSE in the learning process for the 3-4-1 network with the learning rate of 0.20 are depicted in Figure 5. The excellent convergence characteristic of the proposed approach can be easily observed.

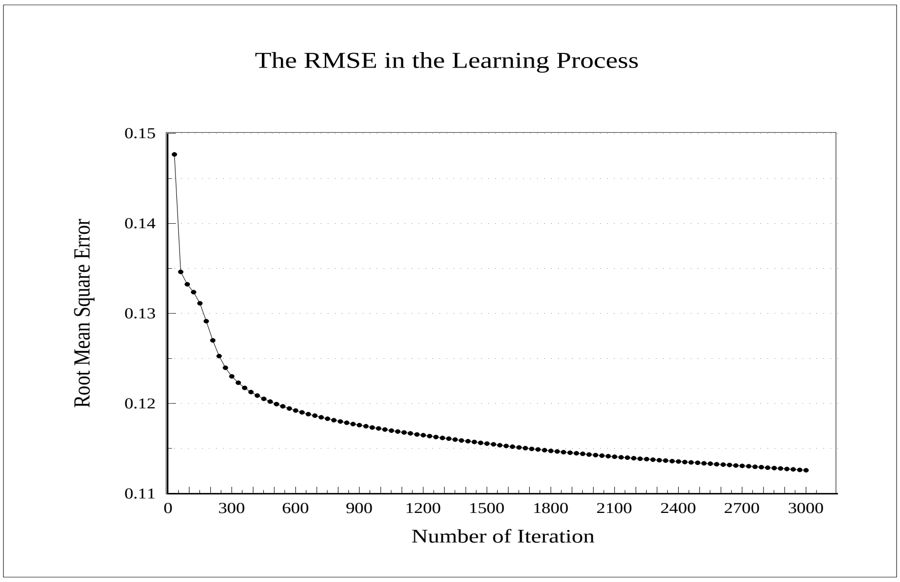

Figure 5. The RMSE in the Learning Process for Kaohsiung Unemployment Rate

For comparison, the predictive results for the Kaohsiung unemployment rate from June 1983 to Feb. 1992 using the proposed approach described in this research and the alternative methods proposed by Wu and Chen are summarized in Figure 6 and Table 2. In addition, several different comparison criteria were utilized to illustrate the difference between the statistical models and the proposed network in the original scale. The results are summarized in Table 3. It is observed that the neural network performed better in almost all cases than all alternative models. Generally, the prediction precision of the neural network is 40% higher than the prediction made by the other models.

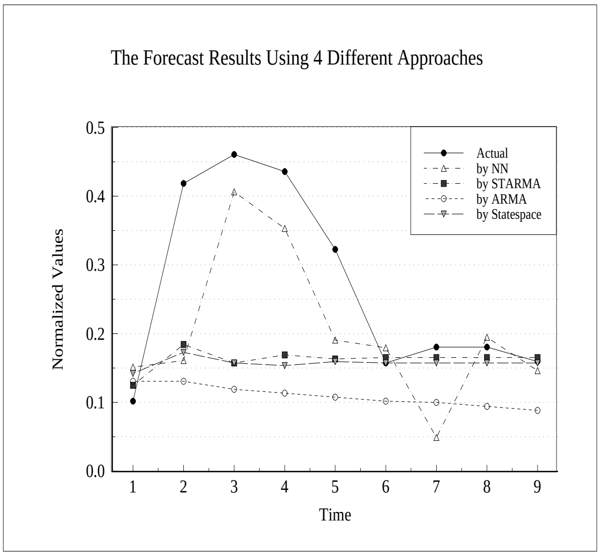

Figure 6. Kaohsiung Unemployment Rate Forecast Results

Table 2. Kaohsiung Unemployment Rate Forecast Results in Normalized Scale

Observation	Actual	Predicted by Neural Network	Predicted by STARMA	Predicted by ARIMA	Predicted by Statespace
1	0.101812	0.151392	0.124840	0.130597	0.142111
2	0.418443	0.160896	0.184328	0.130597	0.172814
3	0.460661	0.406405	0.157463	0.119083	0.157463
4	0.435714	0.352975	0.168977	0.113326	0.153625
5	0.322495	0.190526	0.163220	0.107569	0.159382
6	0.157463	0.179559	0.165139	0.101812	0.157463
7	0.180490	0.048864	0.165139	0.099893	0.157463
8	0.180490	0.194650	0.165139	0.094136	0.157463
9	0.159382	0.146255	0.165139	0.088380	0.157463
RMSE	---	0.112635	0.164773	0.203464	0.170337

Table 3. Kaohsiung Unemployment Rate Forecast Results in Actual Scale

Observation	Actual	Predicted by Neural Network	Predicted by STARMA	Predicted by ARIMA	Predicted by Statespace
1	1.34	1.60	1.46	1.49	1.55
2	2.99	1.65	1.77	1.49	1.71
3	3.21	2.93	1.63	1.43	1.63
4	3.08	2.65	1.69	1.4	1.61
5	2.49	1.80	1.66	1.37	1.64
6	1.63	1.75	1.67	1.34	1.63
7	1.75	1.06	1.67	1.33	1.63
8	1.75	1.82	1.67	1.30	1.63
9	1.64	1.57	1.67	1.27	1.63
SSE*		3.100653	6.6355	10.1176	7.0912
ABS(Error)**		3.945332	5.3700	7.7600	5.6400
Ratio(Error)**		1.692399	1.9087	3.0641	2.0389

*SSE= Σ (Predicted$_i$ - Actual$_i$)^2

**ABS(Error)= Σ | Predicted$_i$ - Actual$_i$ |

***Ratio(Error)= Σ (| Predicted$_i$ - Actual$_i$ | / Actual$_i$)

V. CONCLUSIONS

The Kaohsiung Monthly unemployment rate from June 1982 to Feb. 1986 were analyzed using a neural network and some alternative statistical techniques (space-time series analysis, univariate ARIMA model and state space model). Analysis of this data set were conducted to determine if the utilization a neural network would perform better predictive capability than the other statistical methods. Generally, the prediction precision of the neural network is 40% higher than the prediction made by the other models. The effect of the learning rate and the number of hidden nodes on the efficiency of the neural network learning algorithm were extensively studied to identify the learning rate and the number of hidden nodes that resulted in the best predictions of unemployment rate. The further exploration of the complete parameter space builded by the number of hidden nodes, the value of learning rates and the initial values of connection weights can be done by the traditional response surface methodology. The detail description of this approach can be found in Chiu et al. (1995).

Moreover, the selected size of the training and testing patterns were 96 and 9. If additional data were available, further improvement in prediction accuracy may result. Additional information, such as impacted variables, along with predicted values provided by others techniques, such as time series model and recursive forecasting algorithm, could be used as additional input information. These inclusions may further increase the forecasting accuracy.

REFERENCES

Brainmaker, Version 2.3, Sierra Madre, Ca. Salifronic Scientific Software, 1989.

B. Bauer, "parallel distributed processing models using the back-propagation rule for studying analytic and holistic modes of processing in category learning," Master Thesis, 1988, Texas A&M University, College Station.

Baffes, NETS User Guide. Software Technology Branch, Lyndon B. Johnson Space Center, Clear Lake, Texas, 1989.

Bronars, S. G., and Jansen, D. W., (1987), "The geographic distribution of unemployment rates in the U.S.--a spatial-time series analysis," Journal of Econometrics, Vol. 36, pp. 251-279.

Chiu, C. C., Cook, D. F., Pignatiello, J. J., Jr., and Whittaker, A. D., (1997), "Design of a Radial Basis Function Neural Network with a Radius Modification Algorithm Using Response Surface Methodology," Journal of Intelligent Manufacturing, Vol. 8, No. 2 pp. 117-124.

Funke, M., (1992), "Time-series forecasting of the German unemployment rate," Journal of forecasting Vol. 11, pp. 111-125.

Granger, C. W. J., (1969), "Spatial data and time series analysis," London Papers in Regional Science 1, pp. 1-24.

Rumelhart, E., Hinton, G. E., and Williams, R. J., (1986), Learning internal representations by error propagation in parallel distributed processing, Vol.1, MIT Press, Cambridge, MA, pp. 318-362.

Sims, C., (1980), "Macroeconomics and reality," Econometrica, Vol. 48, pp. 1-48.

Wu, B., (1994), "Identification environment and robust forecasting for nonlinear time series," Computational Economics, Vol. 7, pp. 37-53.

Wu, B., (1995), "Model-free forecasting for nonlinear time series: with application in exchange rates," Computational Statistic and Data Analysis, Vol. 19, pp. 433-459.

Wu, B. and Chen, Y. M., (1993), "Space-time series analysis for the unemployment rate of Taiwan," Journal of Human Resources Development, Vol. 3, pp. 1-34.

Zellner, A., (1979), "Statistical analysis of econometric models," Journal of the American Statistical Association, Vol. 81, pp. 628-643.

Zellner, A. and Palm, F., (1974), "Time series analysis and simultaneous equation econometric models," Journal of Econometrics, Vol. 2, pp. 17-54.

類神經網路模式在時間序列預測上之應用以高雄市失業率為例

邱志洲*．DEBORAH F. COOK**

*輔仁大學企業管理學系

**美國維吉尼亞理工學院管理科學學系

（收稿日期：85 年 6 月 21 日；第一次修正：85 年 9 月 2 日；
第二次修正：85 年 10 月 23 日；接受刊登日期：85 年 11 月 11 日）

摘要

本文針對高雄市失業率，利用倒傳遞類神經網路模式，進行預測並加以討論。其中，對網路模式之相關參數，例如：學習率及隱藏層之神經元數目等，皆有進一步之探討。文中所使用之資料數據為，自民國 72 年 6 月至民國 81 年 2 月之高雄市月失業率資料。除類神經網路模式之應用外，時空數列分析，單變量時間序列模式及狀態空間分析模式等方法，亦被引用來評估預測之好壞。根據結果顯示，類神經網路模式能提供高於其它方法，約 40%之預測準確度。

關鍵詞彙：失業率，預測，類神經網路，時間序列模式