The application of neural networks model in forecasting oil production based on statistical inference: A comparative study with VAR and GSTAR Models

Dhoriva Urwatul Wutsqa[1] , Subanar[2] ,
Suryo Guritno[2], and Zanzawi Soejoeti[2]

[1]Mathematics Department, Yogyakarta State University, Indonesia
PhD Student, Mathematics Department, Gadjah Mada University, Indonesia

[2]Mathematics Department, Gadjah Mada University, Indonesia

Abstract

This article aims to investigate an appropriate space and time series model for oil production forecasting. We propose a new method using neural networks (NN) model based on the inference of R^2 incremental contribution to solve that problem. The existing studies in NN modeling usually use descriptive approach and focus on univariate case, whereas our method has employed statistical concepts, particularly hypothesis test and has accommodated multivariate case including space and time series. This method is performed on bottom up or forward scheme, which starts from empty model to gain the optimal neural networks model. The forecast result is compared to those from the linear models GSTAR (Generalized Space-Time Autoregressive) and VAR (Vector Autoregressive). We show that our method outperforms to these statistical techniques in forecasting accuracy. Thus, we suggest that the NN model is the proper model for oil production forecasting.

Keyword: Neural networks, Generalized Space-Time Autoregressive, Vector Autoregressive, inference of R^2 incremental contribution

1. Introduction

Recently, there has been a growing interest in nonlinear modeling. Neural network is a relatively new approach for modeling nonlinear relationship. Numerous publications reveal that neural networks (NN) has effectively applied in data analysis, including in time series analysis (see e.g. Chen *et.al..* (2001)., Dhoriva *et.al.* (2006),. Suhartono (2005). Suhartono and Subanar, (2006). NN model becomes popular because of its flexibility, by means that it needs not a firm prerequisite and that it can approximate any Borel-measureable function to an arbitrary degree of accuracy (see e.g. Hornik, *et.al.* (1990), White (1990)). However, this flexibility leads to a specification problem of the suitable neural network model. A main issue related to that problem is how to obtain an optimal combination between number of input variables and unit nodes in hidden layer (see Haykin (1999).

Many researchers have started developing strategy based on statistical approach to model selection for modeling neural network. The concepts of hypothesis testing have been introduced by Granger and Terasvirta (1993). Swanson and White (1995,1997) applied a criterion of model selection, SIC, on "bottom-up" procedure to increase number of unit nodes in hidden layer and select the input variables until finding the best FFNN model. This procedure is also recognized as "constructive learning" and one of the most popular is "cascade correlation" (see e.g. Fahlman and Lebiere (1990). Prechelt (1997), and it can be seen as "forward" method in statistical modeling. The current result is from Kaashoek and Van Dijk (2002) proposing backward method, which is started from simple model and then carry out an algorithm to reduce number of parameters based on R^2 increment of network residuals criteria until attain an optimal model. The works of Kaashoek and Van Dijk implement the criterion descriptively. All their works are focused in univariate case.

Appealed to their methods, we put forward the new procedure by using the inference of R^2 increment (as suggested in Kaashoek and Van Dijk (2002). Different from which of Kaashoek and Van Dijk, our work is focused on multivariate time series case and the method does not presented descriptively but by including the concept of hypothesis testing. The discussion associated to that subject consist of two issues, i.e. the construction of Feed forward Neural Network Model for Multivariate time series and the method of NN model building by forward procedure. They are explained in Section 2 and Section 3, respectively.

In this paper, we implement our method to gain the suitable model for oil production forecasting in three wells. The existing study of model for oil production forecasting is approach by using GSTAR. It is introduced by Borovkova et.al. (2002) to manage the multivariate data that depend not only on time (with past observations) but also depend on location or space. A typical linear model VARMA is another possible model that can be applied to space time data. In Section 4 and Section 5, respectively, we give briefly discussions about those models. The study produces three models of oil production forecasting in three wheels and provide the comparison of their forecast performances intended to select the best model due to the yielded accuracy. These results are discussed in Section 6. Finally, Section 7 has same concluding remarks.

.

2. The Feed Forward Neural Network Model for Multivariate Time Series

Feed forward neural network (FFNN) is the most widely used NN model in performing time series prediction. Typical FFNN with one hidden layer for univariate case generated from Autoregressive model is called Autoregressive Neural Network (ARNN). In this model, the input layer contains the preceding lags observations, while the output gives the predictive future values. The nonlinear estimating is processed in the hidden layer (layer between input layer and output layer) by a transfer function. Here, we will construct FFNN with one hidden layer for multivariate tame series case. The structure of the proposed model is motivated by the generalized space-time autoregressive (GSTAR) model from Lopuhaa and Borokova (2005). The following are the steps of composing our FFNN model.

Supposed the time series process with *m* variables $Z_t = (Z_{1,t}, Z_{2,t} \;\dots , Z_{m,t})'$ is influenced by the past p lags values and let *n* as the number of the observations. Set design matrix $\underset{\sim}{X}$ = diag($\mathbf{X_1}$, $\mathbf{X_2}$, **…**, $\mathbf{X}_m$), output vector $Y = (Y_1', Y_2' \;\dots , Y_m')'$, parameter vector $\underset{\sim}{\gamma} = (\gamma_1, \gamma_2 \;\dots , \gamma_m)$ with $\gamma_i = (\gamma_{1,1}^{(i)}, \dots, \gamma_{1,p}^{(i)}, \dots, \gamma_{m,1}^{(i)}, \dots, \gamma_{m,p}^{(i)})$ and error vector $u = \left(u_1', u_2', \dots, u_m'\right)'$,

where $\mathbf{X}_i = \begin{pmatrix} Z_{1,p} & \cdots & Z_{1,1} & \cdots & Z_{m,p} & \cdots & Z_{m,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1,n-1} & \cdots & Z_{1,n-p} & \cdots & Z_{m,n-1} & \cdots & Z_{m,n-p} \end{pmatrix}$, $Y_i = \begin{pmatrix} Z_{i,p+1} \\ Z_{i,p+2} \\ \vdots \\ Z_{i,n} \end{pmatrix}$, and $u_i = \begin{pmatrix} u_{i,p+1} \\ u_{i,p+2} \\ \vdots \\ u_{i,n} \end{pmatrix}$.

Then we have the FFNN model for multivariate time series, which can be expressed as

$$Y = \beta_0 + \sum_{h=1}^{q} \lambda_h \psi(X\gamma) + u \qquad (1)$$

where u is an iid multivariate white noise with E $\left(uu' \mid X\right)$ = $\sigma^2 \mathbf{I}$, E$\left(u \mid X\right) = 0$, $X = (1, \underset{\sim}{X})$, , and $\gamma = (\gamma_0, \underset{\sim}{\gamma})'$.

The functions ψ represents non linear form, where in this paper we use logistic sigmoid

$$\psi(X\gamma) = \{1 + \exp(-X\gamma)\}^{-1} \;. \qquad (2)$$

The architecture of this model is illustrated in Figure 1, particularly for bivariate case with input one previous lag (lag 1).

The notations used in Figure 1. are defined as

$\mathbf{Y_t} = \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix}$, $\mathbf{Z}_{11,t-1} = \begin{pmatrix} \mathbf{Z}_{1,t-1} \\ \mathbf{0} \end{pmatrix}$, $\mathbf{Z}_{12,t-1} = \begin{pmatrix} \mathbf{Z}_{2,t-1} \\ \mathbf{0} \end{pmatrix}$, $\mathbf{Z}_{21,t-1} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{1,t-1} \end{pmatrix}$,

and $\mathbf{Z}_{22,t-1} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{2,t-1} \end{pmatrix}$

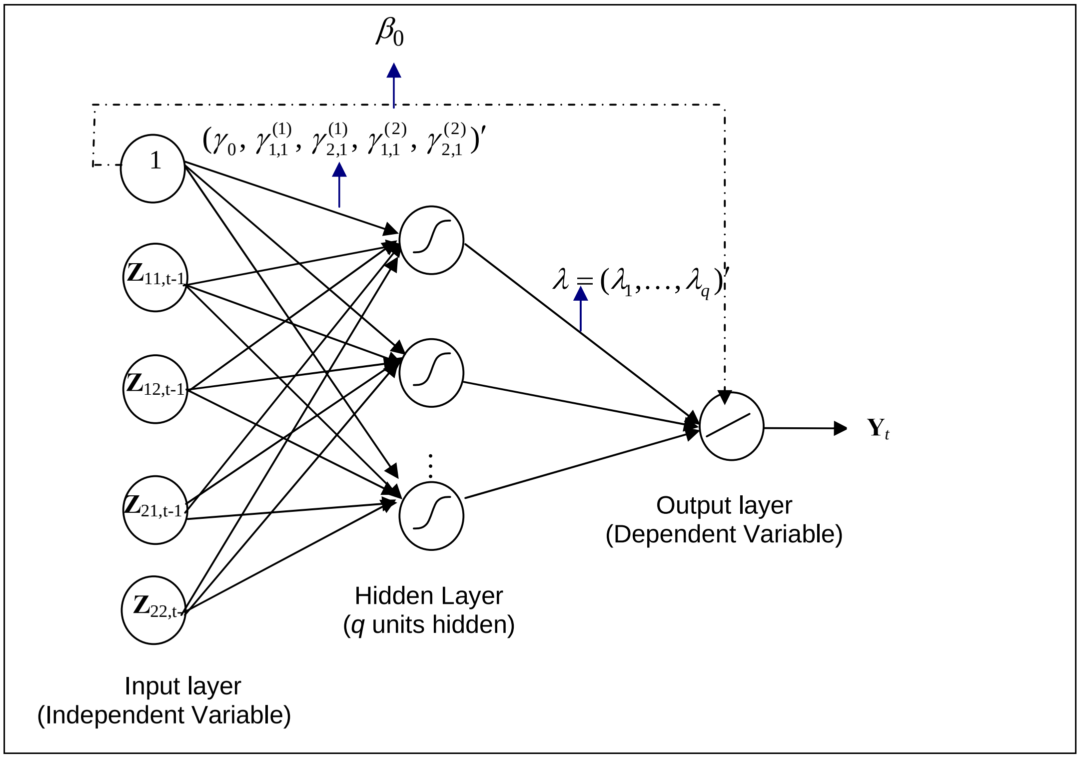

Figure 2. Architecture of neural network model with single hidden layer for bivariate case with input lag 1.

Notify that from the above expression (1), we have separated model

$$\mathbf{Z}_{i,t} = \sum_{h=1}^{q} \lambda_h \psi_h \left(\sum_{j=1}^{m} \sum_{k=1}^{p} \gamma_{j,k}^{(i)} \mathbf{Z}_{j,t-k} + \gamma_0 \right) + \mathbf{u}_{i,t} \quad , \tag{3}$$

for each site $i = 1, \ldots, m$.

The procedure of model selection will be constructed based on model (1); here the multivariate response is expressed in single response, but it still includes all the functional relationships simultaneously.

3. Forward Selection Procedure

The strategy to obtain the optimal neural network model correspond to the specifying a network architecture, where in multivariate time series case, it involves selecting the appropriate number of hidden units, the order (lags) of input variables included in the model, and the relevant input variables. All selection problems will be dealt with forward procedure through statistical approach.

The design of the forward procedure is adopted from general linear test approach that can be used for nonlinear model as stated by Kurtner *et.al* (2004). The procedure entails three basic steps. First, we begin with specification of the simple model from the data, which is also called the reduced or restricted model. In this study, the reduced model is a FFNN model with one hidden unit, i.e.

$$Y = \beta_0 + \lambda_1 \psi(X\gamma) + u \,. \tag{4}$$

To decide whether the model needs hidden unit extension we use the criteria suggested by Kaashoek and Van Dijk, i.e. square of the correlation coefficient. Therefore, we need to compute this value of reduced model, which is formulated as

$$R_R^2 = \frac{(\hat{y}_R' y)^2}{(y'y)(\hat{y}_R' \hat{y}_R)} \tag{5}$$

where $\hat{y}_R$ is the vector of network output points of reduced model. Next, we successively add the hidden unit. The extension model is considered as the complex or full model, i.e. FFNN model (2), starting from hidden units $q = 2$. Then fit the full model and obtain the square of the correlation coefficient R_F^2 , with the same formula as (5).

The final step is calculating the test statistic:

$$F^* = \frac{R_{(F)}^2 - R_{(R)}^2}{df_R - df_F} \div \frac{(1 - R_F^2)}{df_F} \tag{6}$$

or

$$F^* = \frac{R_{(Increment)}^2}{df_R - df_F} \div \frac{(1 - R_F^2)}{df_F} \,.$$

Gujarati (2002) showed that equation (6) is equal to the following expression

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \tag{7}$$

For large *n*, this test statistic (7) and consequently the test statistic (6) are distributed approximately as $F(v1 = df_R - df_F, v2 = df_F)$ when H_0 holds, i.e. additional parameters in full model all equal to 0. This test (6) is applied to decide on the significance of the additional parameter.

Thus the model selection strategy is performed by following the entire steps. As starting point, we determine all the candidate input variables, and then we carry on all the steps sequentially until we find the additional hidden unit does not leads to be significant. Once the optimal number of hidden units is found out, we continue to find the relevant lags that influenced the time series process, begin from the lag that gives the largest R^2. Finally, we employ the forward procedure to decide on the significance of the single input, again starting from the input, which has the largest R^2. If the process of including the input in the model yields significant *p*-value, then it is included to the model, otherwise it is removed from the model.

4. Model (*Generalized Space-Time Autoregressive*)

Models that explicitly take into account spatial dependency are referred to as space-time models. In this model data indexed by time and location. Pfeifer dan Deutsch (1980) suggest *Space-Time Autoregressive* (STAR) to model such data. GSTAR model introduced by Borovkova et.al. (2002) is a more flexible model as a result of STAR model generalization. Mathematically, the notation of space-time autoregressive model of autoregressive order p and spatial order 1, GSTAR(p_1) is the same as STAR(p_1) model. The main difference is the parameters of GSTAR(p_1) model at the same space must not equal. In matrix notation, GSTAR(p_1) model could be written as

$$Z(t) = \sum_{k=1}^{p}\left[\Phi_{k0} + \Phi_{k1}W\right]Z(t-k) + e(t) \qquad (8)$$

where

- $\Phi_{k0} = \mathrm{diag}\left(\phi_{k0}^{1},\ldots,\phi_{k0}^{m}\right)$ and $\Phi_{k1} = \mathrm{diag}\left(\phi_{k1}^{1},\ldots,\phi_{k1}^{m}\right)$,
- W is the weights matrix measuring the spatial dependency among sites, which are chosen to satisfy $w_{ii} = 0$ and $\sum_{i \neq j} w_{ij} = 1$.

If we set $V(t) = WZ(t)$, we can expressed model (8) in linear model structure

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon\ , \qquad (9)$$

with response vector $Y = (Y_1', Y_2' \ \ldots\ , Y_m')'$, design matrix $\underset{\sim}{X}$ = diag(**X**$_1$, **X**$_2$, **…**, **X**$_m$**)**, parameter vector $\boldsymbol{\beta} = (\phi_{01}, \phi_{11}, \phi_{02}, \phi_{12}, \ldots, \phi_{0m}, \phi_{1m})'$ and error vector

$\varepsilon = \left(\varepsilon_1', \varepsilon_2', ..., \varepsilon_m'\right)'$, with

$$\mathbf{Y}_i = \begin{pmatrix} Z_i(2) \\ Z_i(3) \\ \vdots \\ Z_i(n) \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} Z_i(1) & V_i(1) \\ Z_i(2) & V_i(2) \\ \vdots & \vdots \\ Z_i(n-1) & V_i(n-1) \end{pmatrix}, \text{and } \varepsilon_i = \begin{pmatrix} \varepsilon_i(2) \\ \varepsilon_i(3) \\ \vdots \\ \varepsilon_i(n) \end{pmatrix}.$$

Parameter estimation of GSTAR model can be done by using Least Square Method. The theory and methodology about parameter estimation of GSTAR model can be read extensively in Borovkova et.al (2002) and Nurani(2002).

5. The VARMA (*Vector Autoregressive Moving Average*) Model

The standard model used in multivariate *time series* data is VARMA (*Vector Autoregressive Moving Average*) model. The VARMA model denotes the extention of the univariate ARIMA model, which is utilized to model multivariate *time series* data. In this model, the series not only depend on their past values, but also involve interdependency among different series of variables. Brockwell & Davis (1993) define that a process $\{\mathbf{X}_t, t = 0, \pm 1, \pm 2, ...\}$ is said to be an VARMA(*p*,*q*) if $\mathbf{X}_t$ stationary and if for every *t,*

$$\mathbf{\Phi}_p(B)\mathbf{X}_t = \mathbf{\Theta}_q(B)\boldsymbol{\varepsilon}_t \tag{10}$$

where

$$\mathbf{\Phi}_p(z) = \boldsymbol{I} - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots \mathbf{\Phi}_p z^p$$

and

$$\mathbf{\Theta}_q(z) = \boldsymbol{I} + \mathbf{\Theta}_1 z + \mathbf{\Theta}_2 z^2 + \cdots \mathbf{\Theta}_q z^q,$$

are polynomial matrix of autoregressive and moving average order p and q, respectively and B is backward shift operator. Here, the error vector is assumed to be white noise with men **0** and covariance $\mathbf{\Omega}$. The complete discussions of forecasting by VARMA model can be seen on Wei (1990) and Brockwell & Davis (1993, 1996). The application of the model in financial data forecast can be found in Tsay (2005).

6. Result and Discussion

Description of oil production in three wells is presented in time series plot of Figure 4. We employ standardization process on the observed data. We examine the data through three models; those are FFNN, GSTAR, and VARMA models. Here, the data of oil production consist of 60 observations. We divide data into 50 observations for in-sample set (training data) and the rest for out-of-sample set (testing data). Training process is intended for estimating model, and testing process is intended for evaluating the capability of the model estimated to predict next observations.

We begin the model building from FFNN model using the forward method that has been discussed above. In this study, we have three variables, considering the input of GSTAR model (9), we set input layer of FFNN model with six units, i.e.

$$\mathbf{Z}_{1,t-1} = \begin{pmatrix} \mathbf{Z}_{1,t-1} \\ 0 \\ 0 \end{pmatrix}, \mathbf{V}_{1,t-1} = \begin{pmatrix} w_{12}Z_{2,t-1} + w_{13}Z_{3,t-1} \\ 0 \\ 0 \end{pmatrix}, \mathbf{Z}_{2,t-1} = \begin{pmatrix} 0 \\ Z_{2,t-1} \\ 0 \end{pmatrix},$$

$$\mathbf{V}_{2,t-1} = \begin{pmatrix} w_{21}Z_{1,t-1} + w_{23}Z_{3,t-1} \\ 0 \\ 0 \end{pmatrix}, \mathbf{Z}_{3,t-1} = \begin{pmatrix} 0 \\ 0 \\ Z_{3,t-1} \end{pmatrix}, \text{and } \mathbf{V}_{3,t-1} = \begin{pmatrix} 0 \\ 0 \\ w_{31}Z_{1,t-1} + w_{32}Z_{2,t-1} \end{pmatrix}.$$

We continue the forward procedure starting with a FFNN with variable inputs $\left(\mathbf{Z}_{1,t-1}, \mathbf{V}_{1,t-1}, \mathbf{Z}_{2,t-2}, \mathbf{V}_{2,t-2}, \mathbf{Z}_{3,t-1}, \mathbf{V}_{3,t-1}\right)$ and one constant input to find the optimal unit hidden layer cells. The results of an optimization steps is provided in Table 1.

Table 1. The results of forward procedure to get the optimal number of hidden units

Number of hidden unit	R^2	$R^2_{\text{increment}}$	F test	p-value
1	0.7214076		-	-
2	0.7766328	0.0552252	3.387634	0.00145375
3	0.8189804	0.0423476	3.129492	0.002977309
4	0.8626274	0.04367	4.087592	0.0002605009
5	0.8755935	0.0129661	1.279813	0.2618043

We can see from Table 1 that after four hidden units, the optimization procedure shows the insignificant *p*-value; thus, we stop the process and determine four unit cells be the optimal result. Since we only consider lag 1 as a candidate input, we can directly go on to the final step that is selecting the appropriate inputs among inputs of lag 1 whose result is presented in Table 3. In this step, we optimize the FFNN models through each input. Their ordered coefficients $\boldsymbol{R}^2$ are given in the first part of Table 2. It is shown that the FFNN model with input $\mathbf{Z}_{3,t-1}$ has the highest square of the correlation coefficient, so it is chosen as the restricted model. Then we employ the forward procedure based on the ordered $\boldsymbol{R}^2$ values. Subsequently the inputs are entered the model.

Table 2. The results of forward procedure to get the significant input variables

Input	R^2	$R^2_{incremental}$	F test	p-value
$\mathbf{Z}_{1,t-1}$	0.4639163	–	–	–
$\mathbf{V}_{1,t-1}$	0.37432	–	–	–
$\mathbf{Z}_{2,t-1}$	0.4364234	–	–	–
$\mathbf{V}_{2,t-1}$	0.361592	–	–	–
$\mathbf{Z}_{3,t-1}$	0.5131752	–	–	–
$\mathbf{V}_{3,t-1}$	0.2950832	–	–	–
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$	0.6791248	0.1659496	11.93499	2.732036e-008
$\mathbf{Z}_{3,t-1}$, $\mathbf{V}_{1,t-1}$	0.628855	0.1156798	7.102067	0.00003334902
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{2,t-1}$	0.650147	0.1369718	8.970331	1.982919e-006
$\mathbf{Z}_{3,t-1}$, $\mathbf{V}_{2,t-1}$	0.5981331	0.0849579	4.777615	0.00125372
$\mathbf{Z}_{3,t-1}$, $\mathbf{V}_{3,t-1}$	0.5356073	0.0224321	1.072226	0.3729841
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{V}_{1,t-1}$	0.6990568	0.019932	1.692395	0.1558273
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$	0.7843107	0.1051859	12.59932	1.183296e-008
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{V}_{2,t-1}$	0.724445	0.0453202	4.216886	0.003081562
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{V}_{3,t-1}$	0.6817995	0.0026747	0.2143747	0.9300397
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{1,t-1}$	0.8107789	0.0264682	3.758174	0.006422455
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{2,t-1}$	0.8105914	0.0262807	3.727795	0.006737111
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{3,t-1}$	0.7900439	0.0057332	0.732504	0.5714752
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{1,t-1}$, $\mathbf{V}_{2,t-1}$	0.8290159	0.018237	2.362428	0.05706716
$\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{1,t-1}$, $\mathbf{V}_{3,t-1}$	0.8251057	0.0143268	1.716774	0.1508185

We also can see from the result that the inputs $\mathbf{V}_{2,t-1}$ and $\mathbf{V}_{3,t-1}$ produce insignificant *p*–value, thus they are not included in the model. Hence, we get the optimal FFNN model with three hidden units and input variables ($\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{1,t-1}$)

Next, we derive GSTAR forecasting model for the oil production data. We use uniform weight on spatial lag 1 to proceed the GSTAR(1_1) model for oil production. This weight choice is due to condition of the three wells that located nearby in the same group. The model estimation of GSTAR is done by implementing MINITAB software. We attain the best GSTAR(1_1) model as follows

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{bmatrix} = \begin{bmatrix} 0,61 & 0 & 0 \\ 0 & 0,61 & 0 \\ 0 & 0 & 0,78 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{bmatrix} + \begin{bmatrix} 0,32 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0,5 & 0,5 \\ 0,5 & 0 & 0,5 \\ 0,5 & 0,5 & 0 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix} \quad (11)$$

where $z_i(t)$ is time series of oil production at i^{th} well and t^{th} month. The above GSTAR(1_1) model can be written as

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{bmatrix} = \begin{bmatrix} 0,61 & 0,16 & 0,16 \\ 0 & 0,61 & 0 \\ 0 & 0 & 0,78 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix}. \quad (12)$$

Finally, we process the model building of VARMA by implementing PROC STATESPACE program in part of SAS package. The steps of the model estimation is done through the following procedures: the identification by using MACF, MPACF, and AIC value, parameter estimation, check diagnostic to verify that the residual of the model satisfy the white noise requirement. The complete result is provided in appendix. After following all those steps, we obtain the best VARMA model for this case is VAR (1), that is

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{bmatrix} = \begin{bmatrix} 0,64 & 0,19 & 0 \\ 0 & 0,60 & 0 \\ 0 & 0 & 0,76 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix} \quad (13)$$

.

To select the proper model for oil production forecasting, we need to compare their accuracy relied on MSE values. Table 3. gives the comparison of MSE values of those models both in training and testing set. The result generally delivers least MSE values on FFNN model. Therefore, we recommend FFNN model as the suitable model for oil production forecasting. Additionally, we can conclude that the significant input variables included in the model are $\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{\mathbf{1},t-1}$. Fortunately, all the models attempted in this study yield the same result in selecting the relevant input variables.

Tabel 3. The result of forecasting performance among FFNN, GSTAR, and VAR on oil production data

Model	MSE of Training Data			MSE of Testing Data		
	Y1	Y2	Y3	Y1	Y2	Y3
1. FFNN	0.2342391	0.5248447	0.2169494	0.1895563	0.2989587	0.05983502
2. GSTAR(1_1)	0,4967	0,6485	0,3097	0,1772	0,2554	0,0917
3. VAR(1)	0,4982	0,6334	0,4017	0,2070	0,2580	0,0972

7. Conclusions

In this study, we implement model selection for FFNN by to real data that involve space-time. The building blocks of FFNN is not in black box, but developed on the basis of statistical concept, i.e. hypothesis testing. We apply forward procedure to determine the optimal number of hidden units and the input variables included in the model by utilizing R^2 incremental inference. The empirical study finds that all model regarded in this study obtain the same significant input variables, i.e. $\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{\mathbf{1},t-1}$, but FFNN model generally show the highest accuracy. Hence we conclude that the appropriate model for oil production forecasting is FFNN with four hidden units and input variables $\mathbf{Z}_{3,t-1}$, $\mathbf{Z}_{1,t-1}$, $\mathbf{Z}_{2,t-1}$, $\mathbf{V}_{\mathbf{1},t-1}$.

REFERENCES

Brockwell, P.J. and Davis, R.A. (1993). *Time series: Theory and Methods.* 2nd edition, New York: Springer Verlag.

Brockwell P.J. and Davis R.A. (1996). *Introducion to Time Series and Forecasting*.. New York: Springer Verlag,

Borovkova, S.A., Lopuhaa, H.P., and Nurani, B. (2002). Generalized STAR model with experimental weights. In M Stasinopoulos & G Touloumi (Eds.), *Proceedings of the 17th International Workshop on Statistical Modelling* (pp. 139-147).

Chen X., Racine J., and Swanson N. R. (2001). Semiparametric ARX Neural-Network Models with an application to Forecasting Inflation, *IEEE Transaction on Neural Networks*, Vol. 12, p. 674-683.

Dhoriva U.W., Subanar, Suryo G., and Zanzawi S. (2006). Forecasting Performance of VAR-NN and VARMA Models. Proceeding of Regional Conference on Mathematics, Statistics and Their Applications, Penang, Malaysia.

Fahlman, S. E. and C. Lebiere (1990), The Cascade-Correlation Learning Architecture, in Touretzky, D. S. (ed.), *Advances in Neural Information Processing Systems 2*, Los Altos, CA: Morgan Kaufmann Publishers, pp. 524-532.

Granger, C. W. J. and Terasvirta, T. (1993). *Modeling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Gujarati, D.N. (2002), *Basic Econometrics*, 6th edition, McGraw Hill International, New York.

Haykin, H. (1999), *Neural Networks: A Comprehensive Foundation*, Second edition, Prentice-Hall, Oxford..

Hornik, K., M. Stinchombe and H. White (1990), Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, **3**, 551–560.

Kaashoek, J.F. and H.K. Van Dijk (2002), Neural Network Pruning Applied to Real Exchange Rate Analysis, *Journal of Forecasting*, **21**, pp. 559–577.

Kutner, M.H., C.J. Nachtsheim and J. Neter (2004), *Applied Linear Regression Models*, McGraw Hill International, New York.

Lopuhaa H.P.and Borovkova S (2005). Asymptotic properties of least squares estimators in generalized STAR models. *Technical report*. Delft University of Technology.

Nurani, B. (2002). Pemodelan Kurva Produksi Minyak Bumi Menggunakan Model Generalisasi S-TAR. *Forum Statistika dan Komputasi*, IPB, Bogo.

Pfeifer, P.E. and Deutsch, S.J. (1980b). Identification and Interpretation of First Order Space-Time ARMA Models. *Technometrics*, Vol. 22, No. 1, pp. 397-408.

Prechelt, L. (1997), Investigation of the CasCor Family of Learning Algorithms, *Journal of Neural Networks*, **10**, 885-896.

Suhartono, (2005). Neural Networks, ARIMA and ARIMAX Models for Forecasting Indonesian Inflation. *Jurnal Widya Manajemen & Akuntansi*, Vol. **5**, No. 3, hal. 45-65.

Suhartono and Subanar, (2006). The Effect of Decomposition Method as Data Preprocessing on Neural Networks Model For Forecasting Trend and Seasonal Time Series. *JURNAL TEKNIK INDUSTRI: Jurnal Keilmuan dan Aplikasi Teknik Industri*, Vol. **9**, No. 2, pp. 27-41.

Swanson, N. R. and H. White (1995), A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks, *Journal of Business and Economic Statistics*, **13**, 265–275.

Swanson, N. R. and H. White (1997), A model-selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, *Review of Economic and Statistics*, **79**, 540–550

Tsay, R.S. (2005). *Analysis of Financial Time series*. John Wiley & Sons. New Jersey.

White, H. (1990), Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mapping, *Neural Networks*, **3**, 535-550.

Wei, W.W.S. (1990). *Time series Analysis: Univariate and Multivariate Methods.* Addison-Wesley Publishing Co., USA.