Hybrid model for forecasting time series with trend, seasonal and calendar variation patterns

Suhartono, Rahayu, S.P., Prastyo, D.D., Wijayanti, D.G.P., Juliyanto  
Department of Statistics, Institut Teknologi Sepuluh Nopember, Indonesia  
E-mail:suhartono@statistika.its.ac.id

Abstract. Most of the monthly time series data in economics and business in Indonesia and other Moslem countries not only contain trend and seasonal, but also affected by two types of calendar variation effects, i.e. the effect of the number of working days or trading and holiday effects. The purpose of this research is to develop a hybrid model or a combination of several forecasting models to predict time series that contain trend, seasonal and calendar variation patterns. This hybrid model is a combination of classical models (namely time series regression and ARIMA model) and/or modern methods (artificial intelligence method, i.e. Artificial Neural Networks). A simulation study was used to show that the proposed procedure for building the hybrid model could work well for forecasting time series with trend, seasonal and calendar variation patterns. Furthermore, the proposed hybrid model is applied for forecasting real data, i.e. monthly data about inflow and outflow of currency at Bank Indonesia. The results show that the hybrid model tend to provide more accurate forecasts than individual forecasting models. Moreover, this result is also in line with the third results of the M3 competition, i.e. the hybrid model on average provides a more accurate forecast than the individual model.

1. Introduction
Since the publication from Makridakis and Hibon [1] about “M3 competition: results, conclusions, and implications”, particularly the third result that stated “the accuracy when various methods are being combined outperforms, on average, the individual methods being combined and does very well in comparison to other methods”, many researches about hybrid method in forecasting have been done in spectacular way. Zhang [2] is one of the researchers who firstly proposed hybrid model by combining ARIMA as a linear model and Neural Networks as a nonlinear model for time series forecasting. Zhang’s work has influenced many forecasting researchers to develop hybrid method for solving forecasting problems.

Most of the monthly time series data in economics and business in Indonesia and other Moslem countries not only contain trend and seasonal pattern, but also affected by calendar variation pattern. There are two types of calendar variation effects, i.e. the effect of the number of working days or trading and holiday effects. Monthly data about fashion sales in retail company [3], inflation both national and in certain city [4], and inflow-outflow of currency at central bank [5] are some examples about economics and business data that have trend, seasonal, and calendar variation patterns.

There are some methods that usually used for forecasting time series with trend, seasonal, and calendar variation patterns. Liu [6] and Cleveland and Devlin [7] were among the initial researchers who studied the calendar variation effect on time series. Currently, Suhartono et al. [3] proposed two level ARIMAX and regression models for modeling time series data with calendar variation effects due to Eid ul-Fitr, trend and seasonal patterns. In general, these models that were developed for forecasting trend, seasonal, and calendar variation patterns only focused on individual model, such as
time series regression models, ARIMAX models, and models based on artificial intelligent methods such as Neural Networks.

This paper focuses on developing a hybrid model by combining classical linear time series models (such as time series regression and ARIMAX modes) and modern nonlinear models based on artificial intelligent methods particularly Neural Networks for forecasting time series with trend, seasonal, and calendar variation patterns. The proposed models are two level models that are developed based on classical linear time series model in the first level and neural networks in the second level. This model is applied to two empirical data, i.e. simulation and real data about currency inflow and outflow in Bank Indonesia.

2. Method
In this section, three methods for handling trend, seasonal, and calendar variation patterns are presented, i.e. ARIMAX as classical linear models, Neural Networks as modern nonlinear model, and hybrid model that combining classical linear and modern nonlinear models.

2.1. ARIMAX
Cryer and Chan [8] stated that ARIMAX model is ARIMA model with additional variable or also known as exogenous variables. The general ARIMAX model for forecasting data with trend, seasonal, and calendar variation pattern is as follows [3]

\[ y_t = \beta_1 t + \sum_{i=1}^{L} \alpha_i S_{i,t} + \sum_{j=1}^{M} V_{j,t} + \frac{\theta_q(B^L) \theta_q(B^L)}{\phi_p(B) \phi_p(B^L)(1-B)^d(1-B^L)^D} \epsilon_t . \]  

(1)

where

- \( t = 1,2, ..., n \)
- \( S_{i,t} \) is dummy variables for seasonal pattern
- \( V_{j,t} \) is dummy variable for \( j \)-th calendar variation effect
- \( \epsilon_t \) is a white noise process
- \( \phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \)
- \( \Phi_p(B^L) = 1 - \Phi_1 B^L - \Phi_2 B^{2L} - \cdots - \Phi_p B^{pL} \)
- \( \theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \)
- \( \Theta_q(B^L) = 1 - \Theta_1 B^L - \Theta_2 B^{2L} - \cdots - \Theta_q B^{qL} \),

and \( L \) is the seasonal period, \( B \) is the backshift operator, and \( \epsilon_t \) is a sequence of white noise with zero mean and constant variance.

2.2. Neural Networks
Neural networks (NN) is one of machine learning techniques that has been developed as a generalization of the mathematical model of the biological nervous system. The neural networks model that mostly used in time series forecasting is feed forward neural networks (FFNN) or multilayer perceptron (MLP) [9]. The accuracy of neural networks model is determined by three components, i.e the network architecture, training methods or algorithms, and activation functions.

The model of FFNN in Figure 1 can be written as follows:

\[ f(\mathbf{x}, \mathbf{v}, \mathbf{w}) = g_2 \left\{ \sum_{j=0}^{L} v_j g_1 \left[ \sum_{i=0}^{K} w_{ij} x_i \right] \right\}, \]

(2)

where \( \mathbf{w} \) is the weights that connect the input layer to the hidden layer, \( \mathbf{v} \) is the weights that connect the hidden layer to the output layer, \( g_1(\cdot) \) and \( g_2(\cdot) \) is the activation function, while \( w_{ij} \) and \( v_j \) are the weights. The widely used activation function are logistic sigmoid and tangent hyperbolic.
2.3. Hybrid Model

Hybrid model is a combination between linear and nonlinear models that usually be used for increasing the forecast accuracy. In general, the mathematical form of combination between linear and nonlinear models is as follows [2]:

$$ Y_t = L_t + N_t + e_t $$  \hspace{1cm} (3)

where $L_t$ is a linear component and $N_t$ is a nonlinear component of the model. In this paper, NN is used for modeling the nonlinear component as proposed by Zhang [2].

Estimation of this hybrid model is done in two steps. The first is modelling the linear component to get the residual and then applying a nonlinear model to this residual for handling the nonlinear component. In this paper, ARIMAX model is used for handling the linear component. Assume $a_t$ is residual at period $t$ from the first linear model or ARIMAX model, i.e.

$$ a_t = Y_t - \hat{L}_t $$  \hspace{1cm} (4)

where $\hat{L}_t$ is the forecast of linear model at period $t$. Then, NN is applied for modelling $a_t$ as follows:

$$ a_t = f(a_{t-1}, a_{t-2}, \ldots, a_{t-p}) + e_t = \hat{N}_t + e_t $$  \hspace{1cm} (5)

where $f(.)$ is a nonlinear function from the NN model and $e_t$ is the residual of this NN model. Hence, the forecast value of the hybrid ARIMAX-NN model is as follows

$$ \hat{Y}_t = \hat{L}_t + \hat{N}_t $$  \hspace{1cm} (6)

3. Results

In this section, the results from both simulation studies and real case studies are presented. Simulation studies is done for evaluating the performance of each method for forecasting data with trend, seasonal, and calendar variation patterns. Then, the proposed hybrid method applies to real monthly data about outflow and inflow currency at Bank Indonesia.

3.1. The results of simulation study

The simulation data that contains trend, seasonal, and calendar variation patterns are generated by using the model as follows:

$$ Y_t = T_t + S_t + V_t + N_t $$  \hspace{1cm} (7)

where $T_t$ is trend component, $S_t$ is seasonal component (homogenous and heterogenous), $V_t$ is calendar variation component, $N_t$ is noise component that follows linear and nonlinear model. Therefore, four types of simulation data are used in this research, i.e.

- Scenario 1: trend, seasonal homogenous, calendar variation, and linear noise patterns.
- Scenario 2: trend, seasonal homogenous, calendar variation, and nonlinear noise patterns.
- Scenario 3: trend, seasonal heterogenous, calendar variation, and linear noise patterns.
- Scenario 4: trend, seasonal heterogenous, calendar variation, and nonlinear noise patterns.
Three forecasting methods are applied to these four scenarios of simulation study and the forecast accuracy are compared.

The comparison results of forecast accuracy between ARIMAX, NN, and Hybrid models in testing data based on RMSE are shown in Table 1. The results show that the Hybrid model yield more accurate forecast than ARIMAX and NN in each scenario. Hence, the Hybrid model is the best model for forecasting these four scenarios data, i.e. data with trend, seasonal both homogenous and heterogenous, calendar variation, and both linear and nonlinear noise patterns. This result is also in line with the third results of the M3 competition, i.e. the hybrid model on average provides a more accurate forecast than the individual model.

Table 1. The comparison results of forecast accuracy based on RMSE at out-sample (testing data).

| Scenario | Method     | Replication |
|----------|------------|-------------|
| 1        | ARIMAX     | 1.635       |
|          | NN         | 18.208      |
|          | Hybrid     | 1.489       |
|          | ARIMAX     | 4.138       |
|          | NN         | 15.442      |
|          | Hybrid     | 4.120       |
| 3        | ARIMAX     | 2.998       |
|          | NN         | 15.760      |
|          | Hybrid     | 2.976       |
| 4        | ARIMAX     | 5.356       |
|          | NN         | 16.210      |
|          | Hybrid     | 5.340       |

3.2. The results of real data

In this paper, two real data are used as case studies, i.e. monthly currency inflow and outflow at Bank Indonesia in West Java Province for the period 2004:M1 until 2016:M12. The last 12 data in year 2016 is used as testing data. Figure 2 illustrates the training data in a time series plot.

Figure 2. Time series plot of currency inflow (a) and outflow (b) at Bank Indonesia in West Java.

This plot shows the presence of calendar variation due to the celebration of Eid every year. Moreover, Figure 2a illustrates that the largest inflow in every year is at one month before or during month Eid holidays depends on the week that Eid occurred. Similarly, Figure 2b shows that the largest
outflow in every year is at one month after or during month Eid holidays depends on the week that Eid occurred. The results of three methods for forecasting data with trend, seasonal, and calendar variation patterns are presented as follows.

3.2.1. The results of ARIMAX model
The best ARIMAX models for both inflow and outflow data are found by applying identification, estimation, and check diagnostic steps. For illustration, the following model is the best ARIMAX model for outflow data, i.e.

\[
y_t^* = -0.01 t - 2.72 I_{1,t} - 2.57 I_{2,t} + 0.03 t I_{1,t} + 0.03 t I_{2,t} - 0.001 S_{1,t} + 0.27 S_{2,t} + 0.57 S_{3,t} + 0.85 S_{4,t} + 0.65 S_{5,t} + 0.93 S_{6,t} + 0.72 S_{7,t} + 0.38 S_{8,t} + 0.58 S_{9,t} + 0.49 S_{10,t} + 0.614 S_{11,t} + 1.37 S_{12,t} - 0.70 V_{1,t} + 1.50 V_{2,t} + 1.13 V_{3,t} + 1.54 V_{4,t} + 1.63 V_{4,t-1} + 1.08 V_{2,t-1} + 0.29 V_{3,t-1} + 0.47 V_{4,t-1} + \frac{1}{(1-0.20 B-0.24 B^2+0.32 B^3+0.24 B^4)}(1-0.21 B^{12}) \varepsilon_t,
\]  

where \( y_t^* \) is transformation data by Box-Cox transformation with \( \lambda = 0.2 \), \( t = 1, 2, \ldots, n \); \( I_{1,t} \) and \( I_{2,t} \) are the second and third trend respectively, \( S_{j,t} \) dummy variables for seasonal pattern, \( V_{j,t} \) is dummy variable for \( j \)-th calendar variation effect, and \( \varepsilon_t \) is a white noise process.

3.2.2. The results of Neural Networks model
The inputs are determined by using PACF or Partial Autocorrelation Function from the data as proposed by Crone and Kourentzes [10]. For illustration, the best NN model for outflow data is the NN architecture as shown at Figure 3.

![Figure 3](image)

**Figure 3.** The best architecture of Neural Networks for outflow data.

This graph shows that the best inputs for forecasting outflow data are lag 1, 2, 3, 5, 12, 13, and 14 of outflow, and 2 neurons in hidden layer.

3.2.3. The results of Hybrid model
Hybrid ARIMAX-NN model is a combination between ARIMAX as linear model and NN as nonlinear model. Firstly, ARIMAX model is fitted to original data and then the residuals are fitted by NN. Finally, the forecast values are calculated by summing the forecast values of ARIMAX and NN.

As an illustration, the inputs of NN at this hybrid model for outflow data are the lags of residual from the ARIMAX model, i.e. lag 1, 2, 12, 13, 14, 25, and 26. By trying the number of neurons in hidden layer from 1 to 15, the results show that 3 neurons in hidden layer is the best model for forecasting the outflow data. Thus, NN with 7 lags input and 3 neurons in hidden layer is the best NN in this hybrid model with the nonlinear equation as follows:
\[ \hat{N}_t = 0.29 + 0.26 h_{1,t} + 0.28 h_{2,t} - 0.36 h_{3,t} \]  

(9)

where

\[ h_{1,t} = [1 + \exp\left( -(1.27 + 0.51 a_{t-1} + 2.35 a_{t-2} - 2.35 a_{t-12} + \cdots + 9.06 a_{t-26}) \right)]^{-1} \]

\[ h_{2,t} = [1 + \exp\left( -(0.94 + 2.07 a_{t-1} - 0.96 a_{t-2} + 19.79 a_{t-12} + \cdots - 9.28 a_{t-26}) \right)]^{-1} \]

\[ h_{3,t} = [1 + \exp\left( -(3.01 + 1.67 a_{t-1} - 3.83 a_{t-2} - 3.22 a_{t-12} + \cdots + 7.06 a_{t-26}) \right)]^{-1}. \]

Finally, the forecast values for hybrid model are calculated by summing the forecast from linear model in equation (8) or \( \hat{L}_t \), and nonlinear model in equation (9) or \( \hat{N}_t \), i.e.

\[ \hat{Y}_t = \hat{L}_t + \hat{N}_t. \]

3.2.4. The evaluation of forecast accuracy

The root mean squared error (RMSE) in testing data is used as an evaluation index for evaluating the performance of three forecasting models that be applied for forecasting inflow and outflow data. The results of RMSE in testing data obtained using these models are listed in Table 2. Moreover, the graph between forecast and actual values from each method at testing data are illustrated in Figure 4.

| Data   | RMSE | Ratio RMSE to ARIMAX |
|--------|------|-----------------------|
|        | ARIMAX | NN | Hybrid | ARIMAX | NN | Hybrid |
| Inflow | 1.85  | 3.03 | 1.94  | 1       | 1.64 | 1.05   |
| Outflow| 1.28  | 3.76 | 1.24  | 1       | 1.79 | 0.97   |

![Figure 4. Time series plot between actual and forecast values of inflow (a) and outflow (b) data.](image)

The results at Table 2 and Figure 4 show that the hybrid model generate more accurate forecasted values than ARIMAX and NN for outflow data. It supports the third results of M3 competition, i.e. the hybrid model on average provides a more accurate forecast than the individual model. Otherwise, ARIMAX yield more accurate forecast than NN and Hybrid for inflow data. This result is also in line with the first results of M3 competition, i.e. the complex methods do not necessary yield better forecast than the simpler one.

4. Conclusion and future work

In general, forecasting of time series data that have trend, seasonal and calendar variation patterns need special treatment. This paper showed that ARIMAX, NN, and Hybrid model could be used for forecasting these kind of time series. Moreover, the most important part in applying these models is how to determine or choose the appropriate inputs for each model, particularly the inputs for tackling the calendar variation effects.
The results from simulation study showed that the hybrid model yield better forecast than ARIMAX and NN models. Otherwise, the results at real data showed that the hybrid model only yielded better forecast in one of two case studies, i.e. for forecasting outflow data. These results of simulation and outflow data are in line with the conclusion of many previous researches, particularly the third result of M3 competition [1] and Prayoga et al. [5] that stated hybrid model in average will give better forecast than individual model. In addition, the results at the second case study, i.e. inflow data, showed that simpler model yielded better forecast than the complex models. This result also in line with the first result of M3 competition [1] and Suhartono et al. [3] that concluded the complex model do not necessary give better forecast than the simpler one. Moreover, further research is needed to validate these results and to compare the forecast accuracy with other more intelligent methodologies.

Acknowledgment

This research was supported by DRPM-DIKTI under scheme of “Penelitian Berbasis Kompetensi”, project No. 532/PKS/ITS/2017. The authors thank to the General Director of DIKTI for funding and to anonymous referees for their useful suggestions.

References

[1] Makridakis, S. and Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting, 16*, 451-476.

[2] Zhang, G.P. (2003). Time Series Forecasting using a Hybrid ARIMA and Neural Network Model. *Neurocomputing, 50*, 159-175.

[3] Suhartono, Lee, M.H., & Prastyo, D.D. (2015). Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects. *AIP Conference Proceedings, 1691*, no. 050026. http://scitation.aip.org/content/aip/proceeding/aipcp.

[4] Eksiandayani, S., Suhartono, & Prastyo, D.D. (2015). Hybrid Arimax-NN Model for Forecasting Inflation. *Proceeding International Conference on Science, Technology and Humanity, 181*-187.

[5] Prayoga, I.G.S.A., Suhartono, Rahayu, S.P. (2017). Forecasting currency circulation data of Bank Indonesia by using hybrid ARIMAX-ANN model. *Proceedings the 3rd ISM International Statistical Conference 2016: Bringing Professionalism and Prestige in Statistics, ISM 2016; Institute of Mathematical Sciences, University of Malaya, Malaysia.*

[6] Liu, L.M. (1980). Analysis of Time Series with Calendar Effects. *Management Science, 2*, 106-112.

[7] Cleveland, W.S. and Devlin, S.J. (1980). Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods. *Journal of the American Statistical Association, 75*(371), 487-496.

[8] Cryer, J.D., and Chan, K.S. (2008). *Time Series Analysis: with Application in R*. 2nd edition, New York: Springer-Verlag.

[9] Zhang, G.P. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research, 160*, pp. 501-514.

[10] Crone, S.F. and Kourentzes, N. (2009). Input-variable Specification for Neural Networks - an Analysis of Forecasting low and high Time Series Frequency. *International Joint Conference on Neural Networks*, pp. 14-19.