Int. J. Advance Soft Compu. Appl, Vol. 9, No. 2, July 2017
ISSN 2074-8523

Top-down Forecasting for High Dimensional Currency Circulation Data of Bank Indonesia

I Gede Surya Adi Prayoga, Suhartono and Santi Puteri Rahayu

Department of Statistics,
Institut Teknologi Sepuluh Nopember, Surabaya 60111
e-mail: prayoga10@mhs.statistika.its.ac.id, suhartono@statistika.its.ac.id, santi_pr@statistika.its.ac.id

Abstract

This paper provides a solution for forecasting high dimensional time series, in the case of currency circulation in Indonesia. Currency circulation data are divided to currency inflow and outflow. Each of them are treated as hierarchical time series separately. The top-down method is applied based on historical proportion, thus only the total series of inflow and outflow need to be modeled. We have compared the implementation of some time series models in top-down forecasting, including Naïve, decomposition, Winters', ARIMA, and two levels ARIMAX with Eid al-Fitr effect. Each model was specified with varying type of proportion and historical period for calculating the proportions. The results showed that the best method is top-down method with historical proportions type 2 that use the forecast of Naïve method. The proportions are best calculated by using historical data from the last 12 months.

Keywords: *currency circulation, hierarchical time series, top-down, historical proportion.*

1 Introduction

Currency circulation data in Indonesia are recorded by 40 branch offices of Bank Indonesia. The data can be considered as high dimensional multivariate time series because they are correlated among series. The use of multivariate method such as vector autoregressive (VAR) and space-time autoregressive (STAR) are difficult to be implemented on high dimensional data because there will be many parameters to be estimated. However, using univariate method for forecasting each series will not be able to capture the effect of corresponding series, moreover it needs hard efforts to specify each model individually. Therefore, in this paper, we provides efficient solution by considering the data as a hierarchical time series.

There are some forecasting methods for hierarchical time series, such as bottom-up, top-down and optimal combination. Previous studies have been conducted to evaluate the performance of those methods [1, 2, 3]. The most efficient hierarchical forecasting method is top-down method based on historical proportion. This method involves forecasting the completely aggregated series, and then disaggregating the forecast based on the proportions. Top-down method will be very useful, especially when the aggregated series is made up of components that have similar patterns of variation [4].

However, top-down method will give varying results according to the type of proportion and length of historical series used for proportion calculation. This study will find the best combination of those properties. For forecasting the total series, we implemented some time series models, such as Naïve, decomposition, Winters', ARIMA, and two levels ARIMAX with Eid al-Fitr effect. The performance of methods are evaluated based on the out-of-sample root mean squared error (RMSE).

2 Data

The data used in this study are secondary data obtained from Bank Indonesia. The data are monthly currency inflow and outflow, which are analyzed separately. All series are divided into in-sample data for modeling and out-of-sample data for model selection and performance evaluation. The in-sample data are from January 2003 until December 2013, and the out-of-sample data are from January until December 2014. Both data have the same hierarchical structure with two levels (K=2), and m_K = 40 is the number of series at the bottom level. On Fig. 1, the location of branch offices are pointed by the red star symbol, and corresponding islands are marked by different colors.

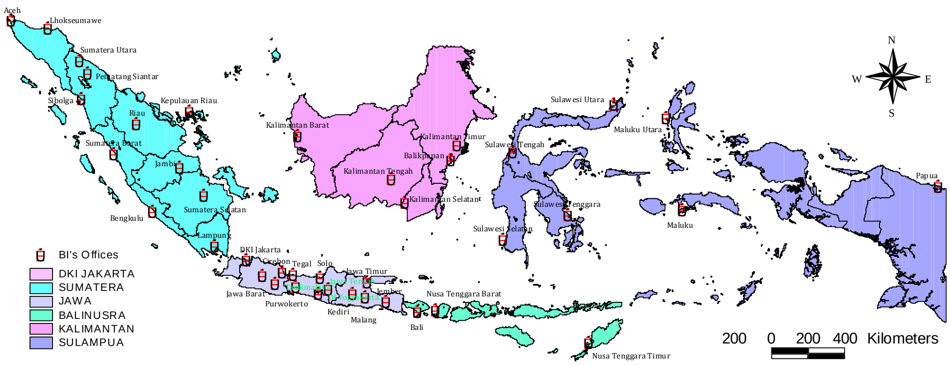


Fig. 1: Bank Indonesia's office location map.

The "level 2" series are the inflow and outflow at each branch offices of Bank Indonesia. The "level 1" series are the aggregation of "level 2" series based on the corresponding islands as seen on Fig. 1, and the "level 0" series are the most aggregated data, which are the national inflow and outflow. The structure of the hierarchy is shown on Fig. 2.

Table 1: Hierarchical structure of currency inflow and outflow data

Level	i	Series	Notation	Number of series
0	1	Indonesia	$Y_{1,t}$	1
1	2	Jakarta	$Y_{2,t}$	6
	3	Sumatera	$Y_{3,t}$	
	⋮	⋮	⋮	
	7	Sulampua	$Y_{7,t}$	
2	8	Jakarta	$Y_{8,t}$	40
	9	Aceh	$Y_{9,t}$	
	⋮	⋮	⋮	
	47	Papua	$Y_{47,t}$	

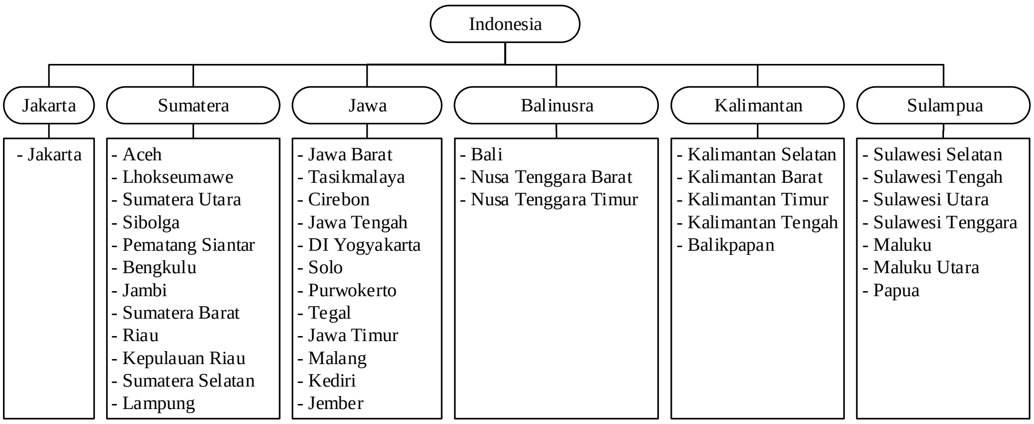


Fig. 2: Hierarchical structure of currency inflow and outflow data.

The hierarchical structure of currency inflow or outflow data can be written in matrix notation as:

$$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{Y}_{K,t} \tag{1}$$

where:

$$\boldsymbol{Y}_t = [\boldsymbol{Y}_{1,t} \quad \boldsymbol{Y}_{2,t} \quad \cdots \quad \boldsymbol{Y}_{47,t}]'$$

$$\boldsymbol{Y}_K = [\boldsymbol{Y}_{8,t} \quad \boldsymbol{Y}_{9,t} \quad \cdots \quad \boldsymbol{Y}_{47,t}]'$$

$$\boldsymbol{Y}_{i,t} = [Y_{i,1} \quad Y_{i,2} \quad \cdots \quad Y_{i,n}]$$

and $\boldsymbol{S}$ is the summing matrix that satisfies Equation 1, which is:

$$\boldsymbol{S} = \begin{bmatrix} 1111111111111111111111111111111111111111 \\ 1000000000000000000000000000000000000000 \\ 0111111111111000000000000000000000000000 \\ 0000000000000111111111111000000000000000 \\ 0000000000000000000000000111000000000000 \\ 0000000000000000000000000000111110000000 \\ 0000000000000000000000000000000001111111 \\ \\ \boldsymbol{I}_{(40\times 40)} \end{bmatrix}.$$

3 Top-down Method

The final forecast of hierarchical forecasting is obtained by using the following equation [1]

$$\tilde{\boldsymbol{Y}}_n(l) = \boldsymbol{SP}\hat{\boldsymbol{Y}}_n(l) \tag{2}$$

where $\hat{\boldsymbol{Y}}_n(l)$ is the individual l step ahead forecast at all hierarchy level, and $\boldsymbol{P}$ is the proportion matrix based on the hierarchical method used. For top-down method, the proportion matrix is:

$$\boldsymbol{P} = [\,\boldsymbol{p}\,|\,\boldsymbol{0}_{m_K \times (m-1)}] \tag{3}$$

where $\boldsymbol{p} = [\,p_1, p_2, ..., p_{m_k}\,]'$ is a set of proportions of the bottom level series [1]. This $\boldsymbol{P}$ matrix will disaggregate the top level forecast to forecasts the bottom level series.

There are two types of historical proportions. The first one is named as top-down historical proportions 1 (TDHP-1) that considers:

$$p_i = \frac{1}{n}\sum_{t=1}^{n}\frac{Y_{i,t}}{Y_{\text{Total},t}} \tag{4}$$

whereas the second one is named as top-down historical proportion 2 (TDHP-2) that considers:

$$p_i = \sum_{t=1}^{n}\frac{Y_{i,t}}{n}\bigg/\sum_{t=1}^{n}\frac{Y_{\text{Total},t}}{n}\,. \tag{5}$$

4 Forecasting Methods for Total Series

This section explain the methods used for forecasting currency inflow and outflow of Indonesia. The methods are including Naïve, Decomposition, Winters', ARIMA, and two levels ARIMAX with Eid al-Fitr effect.

4.1 Naïve Model

Naïve model assumes that the recent periods are the best predictors of the future. Naïve model for stationary data is:

$$\hat{Y}_t = Y_{t-1}\,. \tag{6}$$

Naïve model for trend data is:

$$\begin{aligned} \hat{Y}_t &= Y_{t-1} + (Y_{t-1} - Y_{t-2}),\ \text{or} \\ \hat{Y}_t &= Y_{t-1}\left(Y_{t-1}/Y_{t-2}\right). \end{aligned} \tag{7}$$

Naïve model for seasonal data is:

$$\hat{Y}_t = Y_{t-s}\,. \tag{8}$$

Naïve model for data with trend and seasonality is:

$$\hat{Y}_t = Y_{t-s} + (Y_{t-s} - Y_{t-2s}) \tag{9}$$

where Y_t the time series value at period t, $\hat{Y}_t$ is the forecast for 1 period ahead and s is the length of seasonality.

4.2 Decomposition Method

The general mathematical representation of decomposition approach is [5]:

$$Y_t = f(S_t, T_t, E_t) \tag{10}$$

where:

Y_t the time series value at period t,

S_t is the seasonal component at period t,

T_t is the trend component at period t, and

E_t is the irregular component at period t.

Decomposition methods can assume an additive or multiplicative model and can be in varying forms. For example, the decomposition method of simple averages assumes the additive model

$$Y_t = S_t + T_t + E_t \tag{11}$$

whereas the ratio-to-trend method uses a multiplicative model

$$Y_t = S_t \times T_t \times E_t\,. \tag{12}$$

4.3 Winters' Method

Winters' method can assume an additive or multiplicative model. The basic equation for Winters' additive method is [5]:

$$\hat{Y}_{t+m} = L_t + b_t m + S_{t-s+m} \tag{13}$$

where:

$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$

$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$

$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}$

$\hat{Y}_{t+m}$ is the forecast for m periods ahead, s is the length of seasonality, L_t represents the level of the series, b_t denotes the trend, and S_t is the seasonal component.

The basic equation for Winters' multiplicative method is:

$$\hat{Y}_{t+m} = (L_t + b_t m) S_{t-s+m} \tag{14}$$

where:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-s}.$$

4.4 ARIMA Model

ARIMA model is a flexible time series model that can capture the effect of autoregressive (AR) and moving average (MA). This model can be applied both for non-seasonal or seasonal data and for stationary or non-stationary data. The general form of ARIMA $(p,d,q)(P,D,Q)^s$ model is:

$$Y_t = \frac{\theta_q(B)\Theta_Q(B^s)}{\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D} a_t \tag{15}$$

where:

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - .. - \phi_p B^p)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)$$

$$\Phi_P(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps})$$

$$\Theta_Q(B^s) = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - ... - \Theta_Q B^{Qs})$$

B is the backshift operator, s is the seasonal period, a_t is a white noise process with zero mean and constant variance, t = 1,2,...,n, and n is the number of observation [6].

4.5 Two Levels ARIMAX Model

ARIMAX model is an ARIMA model with the addition of exogenous variables [7]. In case of modelling the effect of calendar variation by Eid al-Fitr, the inputs of ARIMAX can be dummy variables representing the presence of Eid al-Fitr in certain month. The general form of the first level model is:

$$\begin{aligned} Y_t = \mu_0 t + \sum_{m=1}^{s}\delta_m M_{m,t} + \sum_{j=0}^{30}\alpha_j H_{j,t+1} + \sum_{j=0}^{30}\beta_j H_{j,t} + \sum_{j=0}^{30}\gamma_j H_{j,t-1} \\ + \sum_T \frac{1}{(1-B)}\upsilon_T S_t^{(T)} + \sum_T \omega_T A_t^{(T)} + \frac{\theta_q(B)\Theta_Q(B^s)}{\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D} a_t \end{aligned} \tag{16}$$

where $M_{m,t}$ is the dummy variable representing each month, $H_{j,t}$ is the dummy variable representing the presence of Eid al-Fitr at period *t*, *j* is the number of days before Eid al-Fitr in certain month, $S_t^{(T)}$ is the dummy variable for level shift (LS) outlier, $A_t^{(T)}$ is the dummy variable for additive outlier (AO), and *B* is the backshift operator. The role of $\mu_0 t$ or $(1-B)^d$ is to make the data stationary in mean, whereas the effect of seasonality can be captured by $\sum_{m=1}^{s}\delta_m M_{m,t}$ or $\Theta_Q(B^s)/\Phi_P(B^s)$.

In many cases, the length of time series are limited, thus the parameters of H_j cannot be estimated for all *j* in the first level model. This becomes problem because the forecasting requires the values of the unknown parameters. Therefore, the second level models are needed to predict the parameters for every possibility number of days before Eid al-Fitr. As the second level model, a linear function can be applied as follows:

$$\hat{\alpha}_j = \xi_0 + \xi_1 j \tag{17}$$

$$\hat{\beta}_j = \psi_0 + \psi_1 j \tag{18}$$

$$\hat{\gamma}_j = \zeta_0 + \zeta_1 j \tag{19}$$

where *j* is the number of days before Eid holidays in corresponding month. In this model, the response variables are the estimated parameters from the first level modeling [8].

5 Results

This section explains the forecasting results of the total series and the comparison of top-down forecast performance from each method with specific type of proportions and length of historical series for calculating the proportions.

5.1 Forecast of Total Series

The currency inflow and outflow patterns are displayed on Fig. 3. The time series plots show that both inflow and outflow data have trend and seasonal patterns. However, the inflow and outflow data have different patterns in term of calendar variation effect. If the Eid al-Fitr happens in the end of the month, the inflow increase in the next month. Otherwise, the inflow increase in the month containing Eid al-Fitr. If the Eid al-Fitr happens in the End of the month, the outflow increase in that month. Otherwise, the outflow increase in the previous month.

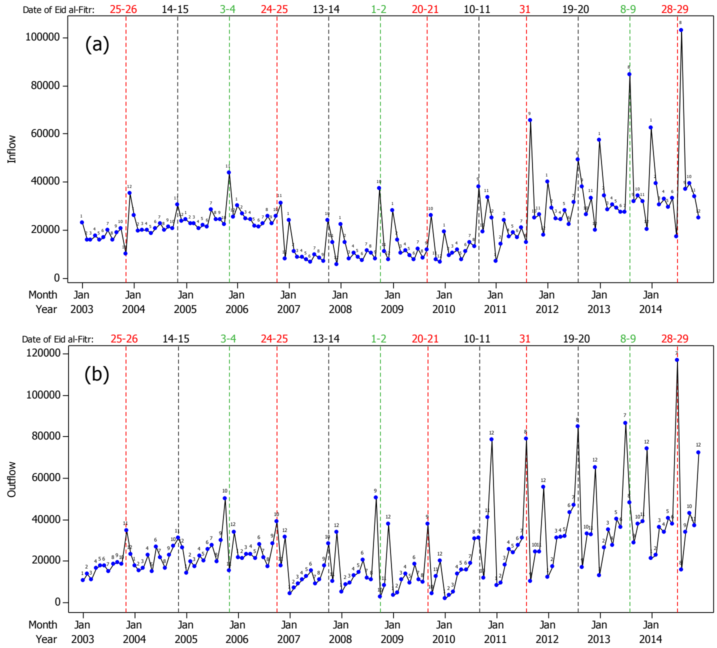


Fig. 3: Time series plot of national currency (a) inflow and (b) outflow.

Simple methods such as Naïve, Decomposition and Winters' can be easily implemented on those data. However, the ARIMA approach may be not appropriate because the calendar variation effect of Eid al-Fitr may cause outliers that make the violation of the normality assumption.

By using Box-Jenkins procedure, the ARIMA model for national currency inflow data is:

$$Y_t = \frac{(1-0.82B)(1+0.59B^{12})}{(1-B)} a_t \tag{20}$$

whereas the ARIMA model for national currency outflow data is:

$$Y_t = \frac{(1-0.87)}{(1-B)(1-0.70B^{12})} a_t \,. \tag{21}$$

Those ARIMA models have meet the assumption of independent residual and all parameters are significant. However, as expected, the residuals are not normally distributed.

Considering this problem, we applied more sophisticated model, i.e. two levels ARIMAX model. The first level model for national currency inflow data is:

$$Y_t = 306.4t + \sum_{j=0}^{30} \beta_j H_{j,t} + \sum_{j=0}^{30} \gamma_j H_{j,t-1} - \frac{1}{(1-B)} 12109.6 S_t^{(48)} + 15453.8 A_t^{(95)}$$
$$+11413.7 A_t^{(96)} - 25265.1 A_t^{(97)} - 7048.5 A_t^{(98)} + 3631.5 A_t^{(121)} \quad (22)$$
$$+ \frac{1}{(1-0.46B-0.41B^2)(1-0.85B^{12})} a_t$$

and the second level models are:

$$\hat{\beta}_j = 36396 + 1387 j \quad (23)$$

$$\hat{\gamma}_j = -3651 + 1088 j\,. \quad (24)$$

For national currency outflow data, the first level model is:

$$Y_t = 366.9t + \sum_{j=0}^{30} \alpha_j H_{j,t+1} + \sum_{j=0}^{30} \beta_j H_{j,t} - \frac{1}{(1-B)} 9661.4 S_t^{(49)} + 36081.9 A_t^{(96)}$$
$$+ \frac{1}{(1-0.29B-0.50B^3)(1-0.95B^{12})} a_t \quad (25)$$

and the second level models are:

$$\hat{\alpha}_j = 33547 - 1278 j \quad (26)$$

$$\hat{\beta}_j = 860 + 1666 j\,. \quad (27)$$

By including the effect of Eid al-Fitr and some outliers, the two levels ARIMAX models have satisfied the assumption of significant parameters, independent and normally distributed residual. The performance of two levels ARIMAX compared to the other methods are shown on Table 2.

Table 2: RMSE comparison on total series

Method	Currency Inflow		Currency Outflow	
	In-sample	Out-of-sample	In-sample	Out-of-sample
Naïve	13643.1	**7463.2**	18613.2	**8237.2**
Decomposition	11664.6	24925.2	13426.1	25730.7
Winters'	11361.3	20046.4	13395.1	23095.8
ARIMA	9022.6	15525.6	11302.9	16014.0
Two levels ARIMAX	**3449.8**	15626.3	**5575.0**	12280.7

According to Table 2, the two levels ARIMAX model clearly outperformed the other methods for in-sample data, whereas the best forecast for out-of-sample data are generated by Naïve method. The forecasts of each method are visualized on Fig. 4 and 5. The time series plots show that the forecasts of decomposition, Winters' and ARIMA method are far from the actual data, especially in the months that contain Eid al-Fitr. Whereas, the two levels ARIMAX method are able to capture the effect of Eid al-Fitr, thus the forecasts are nicely fits the actual data. However, the Naïve method has the worst performance on in-sample data, but surprisingly forecast the out-of-sample data almost perfectly and outperforms the other methods.

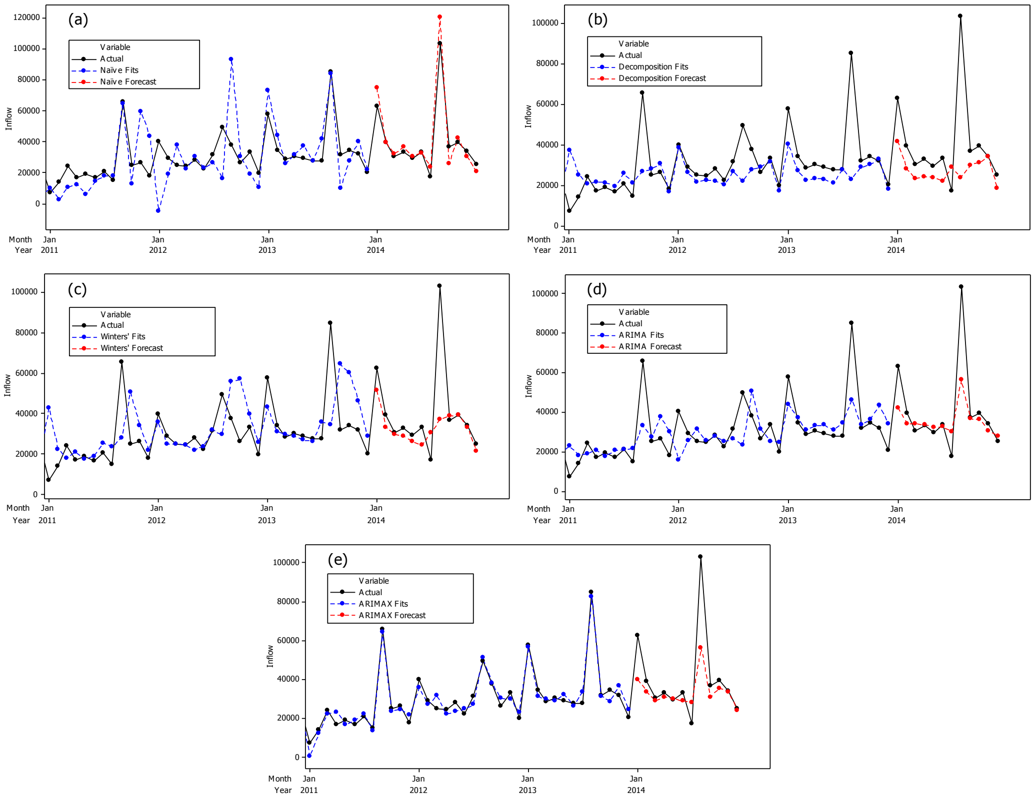
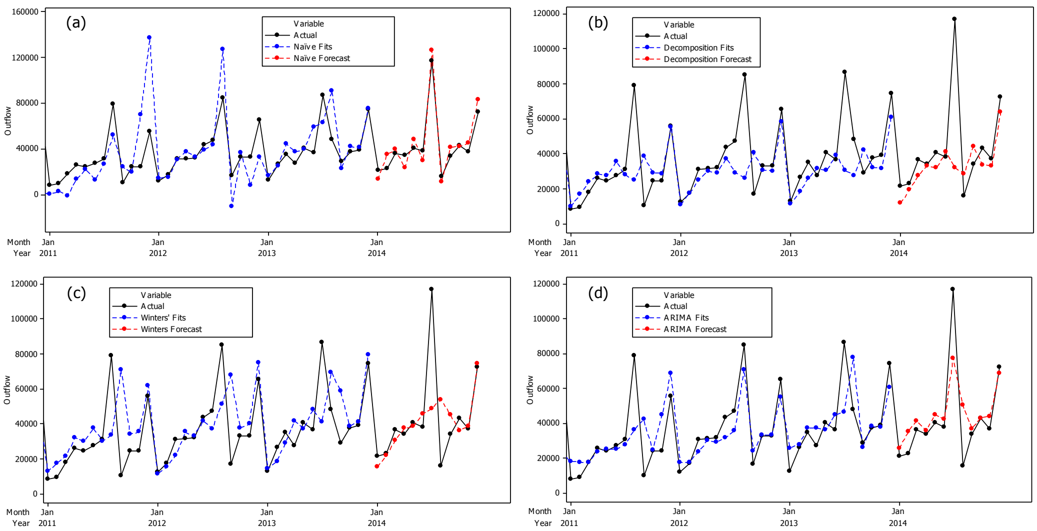


Fig. 4: Forecast of national currency inflow by (a) Naïve, (b) Decomposition, (c) Winters', (d) ARIMA, and (e) two levels ARIMAX model.



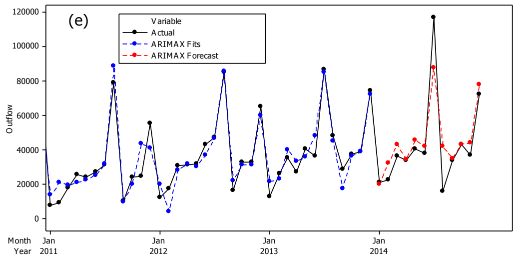


Fig. 5: Forecast of national currency outflow by (a) Naïve, (b) Decomposition, (c) Winters', (d) ARIMA, and (e) two levels ARIMAX model.

5.2 Top-down Forecast Evaluation

The forecast of total series by each method are disaggregated to level 1 and 2 by using top-down method with varying specification on the type of proportion and the number of in-sample for calculating the proportions ($n*$). The performance comparison on inflow and outflow data are displayed on Fig. 6 and 7.

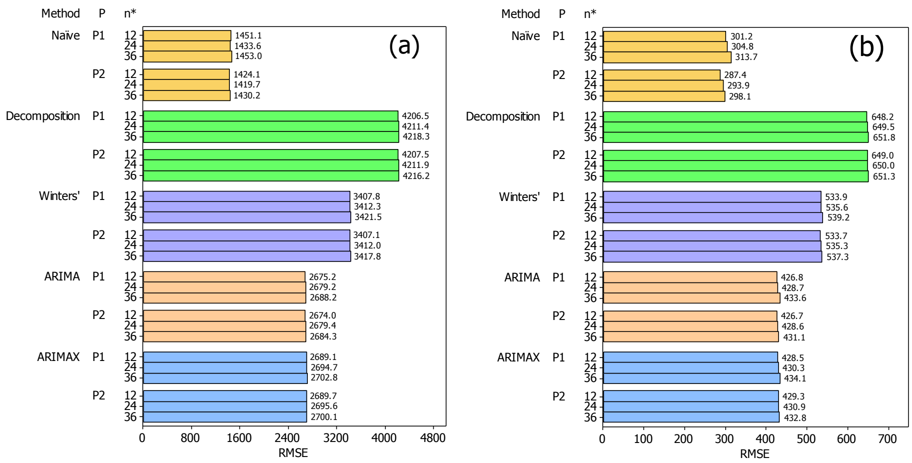


Fig. 6: RMSE comparison on currency inflow data at (a) level 1 and (b) level 2.

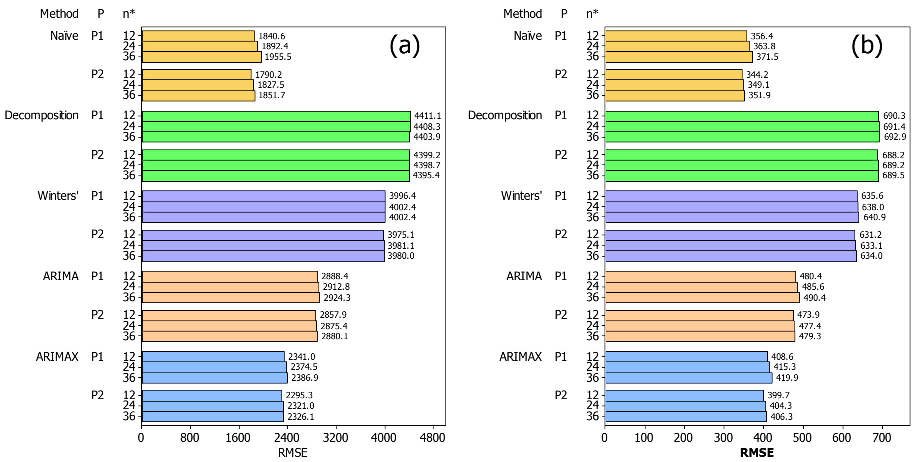


Fig. 7: RMSE comparison on currency outflow data at (a) level 1 and (b) level 2.

Fig. 6 and 7 show that the performance comparisons are consistent over hierarchy level. It means that the forecast accuracy of the bottom series depends on the accuracy of the total series. Accordingly, the method selection for forecasting the total series becomes crucial step in top-down forecasting.

On the same number of in-sample data, proportion type-2 (based on Equation 5) performs better than proportion type-1 (Equation 4), except for decomposition method on inflow data. On the same type of proportion, using more number of historical data tends to yields worse forecast. Except for Naïve method on level 1 inflow and decomposition method on level 1 outflow, the best forecasts are resulted if the proportions are calculated by using in-sample data from the most recent year.

Table 3: RMSE comparison of top-down and individual forecast

Method	Inflow		Outflow	
	Level 1	Level 2	Level 1	Level 2
TDHP-2 based on Naïve forecast	**1424.1**	**287.4**	**1790.2**	**344.2**
Naïve individual forecast	1464.2	308.8	1972.9	422.1
TDHP-2 based on ARIMAX forecast	2689.7	429.3	2295.3	399.7
ARIMAX individual forecast	2602.9	505.0	2198.4	392.3

Table 3 explains that using top-down method based on Naïve forecast is better than forecasting each of the level 1 and 2 series individually. If we concern only on the implementation of ARIMAX, the top-down forecasts are not always better than ARIMAX individual forecast. It is reasonable because each individual series has different pattern in term of the position of outliers. However, the RMSE difference are not too big. Therefore, top-down method still can be an efficient solution because modeling each series individually by using ARIMAX is not easy.

6 Conclusions

Based on the results, we highly recommend using top-down method with historical proportions to forecast currency inflow and outflow data of Bank Indonesia. Besides being more efficient, this method also has shown better performance compared to forecasting each series individually. However, this method needs to be carefully specified. The performance of top-down method are affected by the type of proportion and the number of in-sample data for calculating the proportions. Using proportion type 2 and the most recent historical data tend to yield the best forecast.

The most important step in top-down forecasting is selecting the method for forecasting the total series because different method will significantly give different forecast performance. The results showed that Naïve method gives the best performance consistently for all hierarchy level. However, using Naïve method could be risky because this method showed the worst performance on in-sample data. Two levels ARIMAX model could be a more reliable alternative because it performed well both on in-sample and out-of-sample data.

ACKNOWLWDGEMENTS
The authors acknowledge the Bank Indonesia (BI) for permission to use the monthly currency inflow and outflow data from all branch offices in Indonesia.

References

[1] G. Athanasopoulos, R. A. Ahmed and R. J. Hyndman. 2009. Hierarchical forecasts for Australian domestic tourism, *International Journal of Forecasting*, Vol.25, No.1, 146-166.

[2] R. J. Hyndman, R. A. Ahmed and G. Athanasopoulos. 2011. Optimal combination forecasts for hierarchical time series, *Computational Statistics & Data Analysis*, Vol.55, No.9, 2579-2589.

[3] I. G. S. A. Prayoga, S. P. Rahayu and Suhartono. 2015. Hierarchical forecasting method based on ARIMAX and recurrent neural network for motorcycle sales prediction, *International journal of applied mathematics and statistics*, Vol.53, No.5, 116-124.

[4] L. Lapide. 2006. Top-down & bottom-up forecasting in S&OP, *Journal of business forecasting methods and systems*, Vol.25, No.2, 14.

[5] S. Makridakis, S. C. Wheelwright and R. J. Hyndman. 2008. Forecasting methods and applications, John Wiley & Sons,

[6] B. L. Bowerman and R. T. O'Connell. 1993. Forecasting and time series: An applied approach, 3rd ed., Duxbury Press.

[7] J. D. Cryer and K. S. Chan. 2008. Time series analysis: with applications in R, 2nd ed., Springer.

[8] Suhartono, M. H. Lee and D. D. Prastyo. 2015. Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects, *Proceedings of the 2nd innovation and analytics conference & exhibition*, Vol.1691, No.1, 050026.