Vector autoregressive model approach for forecasting outflow cash in Central Java

Abdul hoyyi1, Tarno1, Di Asih I Maruddani1, Rita Rahmawati1
1Department of Statistics, Faculty of Science and Mathematics, Diponegoro University
Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia
E-mail: hoyyistat@live.undip.ac.id

Abstract. Multivariate time series model is more applied in economic and business problems as well as in other fields. Applications in economic problems one of them is the forecasting of outflow cash. This problem can be viewed globally in the sense that there is no spatial effect between regions, so the model used is the Vector Autoregressive (VAR) model. The data used in this research is data on the money supply in Bank Indonesia Semarang, Solo, Purwokerto and Tegal. The model used in this research is VAR (1), VAR (2) and VAR (3) models. Ordinary Least Square (OLS) is used to estimate parameters. The best model selection criteria use the smallest Akaike Information Criterion (AIC). The result of data analysis shows that the AIC value of VAR (1) model is equal to 42.72292, VAR (2) equals 42.69119 and VAR (3) equals 42.87662. The difference in AIC values is not significant. Based on the smallest AIC value criteria, the best model is the VAR (2) model. This model has satisfied the white noise assumption.

Keywords: outflow, Cash, VAR, OLS, AIC

1. Introduction
Cash in Indonesia are broadly known as Rupiah. Bank Indonesia persistently oversees and calculate the value of Rupiah circulating among the societies and banks, literally known as the cash in circulation. Cash serve as the efficient payment method, particularly for the transaction in retail and relatively small value of transaction [1]. The timeliness in appropriately projecting the cash circulating among the societies will help Bank Indonesia, as the holder of authority to print and distribute the cash, in planning the issuance and distribution of notes in Rupiah in Indonesia [12]. This forecast of inflow and outflow is paramount as its relation to the bank liquidity which may affect the monetary policy must be run. Were the value of both aspects high, accordingly the bank liquidity is about to increase, while be it too low, the bank liquidity will also drop down [8]. A study on the cash demand in Indonesia has been conducted by Prayitno, Sanjaya, and Llewelyn [9], to which time the model in use was regression analysis approach with log model to analyse the impact of the government’s expenditure, foreign exchange reserves, and money multiplier against the value of cash circulating in Indonesia for the pre-crisis period, post-crisis period and whole period. Untoro [12] with his time series data approach under ARIMA (Autoregressive Integrated Moving Average) Model and VAR (Vector Autoregressive) Model forecasts the cash demand in Indonesia. His study results in a conclusion that ARIMA model is way better than the VAR model.
VARIMA (Vector Autoregressive Integrated Moving Average) model is a qualitative forecasting approach normally adopted in the multivariate time series data. This model depicts the correlation between an observation done to a certain variable at a particular time and an observation to the other variables done in the previous period of time. Several empirical studies frequently engage multivariate time series data. Meanwhile, this study adopted VAR Model to design the outflow of cash at four Representative Office of Bank Indonesia located in Central Java, namely Semarang, Solo, Purwokerto and Tegal. Those four variables constituted the time series data. The correlation between those four variables brought the resulted model to fit the criteria, that one variable constituted the functions of other variables. The proper multivariate analysis for the data with such characteristic was Vector Autoregressive (VAR).

2. Research Objectives
This research aims to generate the result of estimated parameter of VAR Model in the outflow of cash in Central Java through the Representative Office of Bank Indonesia Solo, Representative Office of Bank Indonesia Purwokerto, and Representative Office of Bank Indonesia Tegal. Apply and reveal the accuracy of VAR Model in the outflow of cash in Central Java through the Representative Office of Bank Indonesia Solo, Representative Office of Bank Indonesia Purwokerto, and Representative Office of Bank Indonesia Tegal.

3. Material and Methods
3.1. VARMA (Vector Autoregressive Moving Average) Model
Time series data in numbers of study are frequently composed of several variables or known as the multivariate time series data [3]. For the instance, the variables possibly involved in a study on the good sales will be the sale volume, price, and advertisement cost of a product in several close or interconnected marketing area. Given \( z_j(t) \) with \( t \in T, T = \{1, 2, K, T\} \) and \( i = \{1, 2, K, N\} \) which are the parameter index of calculated and limited time and variables (i.e, the different location or different types of product), accordingly VARMA model, in general, may be shown in the below equation [15].

\[
\Phi_p (B) Z(t) = \Theta_q (B) a(t) \tag{1}
\]

where \( Z(t) \) is the multivariate time series vector corrected for its average value, \( \Phi_p (B) \) and \( \Theta_q (B) \) are the autoregressive moving average matrix polynomial of orders \( p \) and \( q \). When \( q=0 \), the process become a vector AR(p) model or VAR(p). The identification of multivariate time series model is similar to the univariate time series model. The identification of stationary on the multivariate time series model may be done by considering Matrix Autocorrelation Function (MACF) plot.

Matrix Autocorrelation Function (MACF)
Given a vector time series of \( T \) observations, namely \( Z_1, Z_2, L, Z_T \), accordingly the correlation matrix equation of its sample will be as follow [15]:

\[
\hat{\rho}(k) = [\hat{\rho}_{ij}(k)]
\]

where \( \hat{\rho}_{ij}(k) \) are the sample cross-correlation for the -ith and -jth component series

\[
\hat{\rho}_{ij}(k) = \frac{\sum_{t=1}^{T-k}(Z_{i,t} - \bar{Z}_i)(Z_{j,t-k} - \bar{Z}_j)}{\left[ \sum_{t=1}^{T}(Z_{i,t} - \bar{Z}_i)^2 \sum_{t=1}^{T}(Z_{j,t} - \bar{Z}_j)^2 \right]^{1/2}} \tag{3}
\]
\( \overline{Z}_i \) and \( \overline{Z}_j \) are the sample means of the corresponding component series.

Bartlett (1966) in [15] has depicted the variance and covariance of the cross-correlation index resulted from the sample. Bartlett suggests this following equation:

\[
\text{Var} [\hat{\rho}_y (k)] \approx \frac{1}{T-k} \left[ 1 + 2 \sum_{\lambda=1}^{\infty} \rho_y (\lambda) \rho_y (\lambda) \right], |k| > q
\]  

Where, \( Z_i \) and \( Z_j \) are white noise series which will further generate this following equation:

\[
\text{Cov} [\hat{\rho}_y (k), \hat{\rho}_y (k + \lambda)] \approx \frac{1}{T-k}
\]  

\[
\text{Var} [\hat{\rho}_y (k)] \approx \frac{1}{T-k}
\]  

For big-sizes sample, \((T-k)\) in (6) equation is frequently replaced with \(T\).

Box and Tiao (1981) in [15] introduce a method summarizing the correlation result of samples. This method use (+), (-), and (.) symbols on the \(i\)-line and \(j\)-column of the sample correlation matrix, where:

1. (+) symbol denotes that \( \hat{\rho}_{ij} (k) \) value greater than 2 times the estimated standard error, meaning that the \((i,j)\) components have positive correlation,
2. (-) symbol denotes that \( \hat{\rho}_{ij} (k) \) value less than than -2 times the estimated standard error, meaning that the \((i,j)\) components have negative correlation, and
3. (.) symbol denotes that \( \hat{\rho}_{ij} (k) \) value within 2 estimated standard errors, meaning that the \((i,j)\) components are uncorrelated.

The Data said to be stationary if MACF plot shows lower (+) and (-), and nearly all signs symbolled with (.) [7]. The multivariate time series only checks the stationarity in mean [11].

3.2. Matrix Partial Autocorrelation Function (MPACF)

Partial autocorrelation function (PACF) is required in the univariate time series to determine the order in AR model. The generalization of the PACF concept into the time series vector form is done by Tiao and Box (1981) in [15]. Tiao and Box (1981) in [15] define the Matrix Partial Autocorrelation Function (MPACF) on the \(k\)-lag denoted by \( \mathcal{P}(k) \). \( \mathcal{P}(k) \) formula will be as follow:

\[
\mathcal{P}(k) = \left\{ \begin{array}{ll}
(\Gamma'(1)[\Gamma(0)]^{-1}, & k = 1 \\
(\Gamma'(k) - c'(k)A(k))^{-1}b(k)\{\Gamma(0) - b'(k)A(k))^{-1}b(k) \}, & k > 1 
\end{array} \right.
\]  

For \( k \geq 2 \), the value of \( A(k) \), \( b(k) \), and \( c(k) \) will be as follow:

\[
A(k) = \begin{bmatrix}
\Gamma'(0) & \Gamma'(1) & \cdots & \Gamma'(k-2) \\
\Gamma(0) & \Gamma(1) & \cdots & \Gamma'(k-3) \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma'(k-2) & \Gamma(k-3) & \cdots & \Gamma'(0) \\
\end{bmatrix}
\]  

\[
b(k) = \begin{bmatrix}
\Gamma'(k-1) \\
\vdots \\
\Gamma'(1) \\
\end{bmatrix}, \quad c(k) = \begin{bmatrix}
\Gamma(1) \\
\Gamma(2) \\
\vdots \\
\Gamma(k-1) \\
\end{bmatrix}
\]

where \( \Gamma(k) \) is \( k\)-lag covariance matrix.

The estimated sample of \( \mathcal{P}(k) \) may be calculated by replacing the unrevealed \( \Gamma(k) \) with sample covariance matrix \( \hat{\Gamma}(k) \):

\[
\hat{\mathcal{P}}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (\overline{Z}_t - \overline{Z})(\overline{Z}_{t+k} - \overline{Z})', k = 1, 2, ... \]
where $\bar{Z}$ is the sample average vector.

The identification of data based on MPACF values is also denoted in (+) and (-) and (.) forms as in MACF. Similar to partial autocorrelation equations in univariate cases, $\mathcal{P}(k)$ autocorrelation partial matrix equation also has a cut-off property for the AR process vector.

### 3.3 Parameter Estimation

Parameter estimation of VAR model by using Ordinary Least Square (OLS).

Parameter estimator of VAR model under the OLS method is as follow:

$$\hat{\phi} = [I_M \otimes (X'X)^{-1} X']Z$$

The covariance matrix of $\hat{\phi}$ will be:

$$\sum_{\phi} = \sum_{\phi} \otimes (X'X)^{-1}$$

Diagnostic examination of whether the residual is white-noise, can be seen from the plot of the autocorrelation matrix whose elements do not show a particular pattern and not significant. Selection of the best model using Akaike Information Criteria (AIC). For an m-dimensional VAR(p) model,

$$AIC(p) = \ln(|S(p)|) + \frac{2pm^2}{n}$$

where $S(p)$ is the residual sum of square and cross products, and one select a model that gives the smallest AIC value [15].

### 3.4 Methods

This research engaged the secondary data in form of the outflow of cash in Central Java circulated by the Representative Office of Bank Indonesia Semarang, Representative Office of Bank Indonesia Solo, Representative Office of Bank Indonesia Purwokerto, and Representative Office of Bank Indonesia Tegal for the observation period of January 2010 to March 2015, done under 63 observations. While the variables in use in this research were the data of cash outflow in Central Java, namely: the outflow of cash circulated by the Representative Office of Bank Indonesia Semarang ($Z_1$), Representative Office of Bank Indonesia Solo ($Z_2$), Representative Office of Bank Indonesia Purwokerto ($Z_3$), and Representative Office of Bank Indonesia Tegal ($Z_4$). The procedure to modeling Vector Autoregressive is as follows:

a. exploration of data to find out whether the data has the same relative fluctuations
b. identification of stationary condition on multivariate time series model using MACF plot
c. identification of VAR model order using MACF plot
d. perform the VAR model
e. perform forecasting using the best model

### 4. Results and Discussion

#### 4.1 Data Exploration

The exploration of data from each variable was aimed to reveal the data pattern in general. The exploration of data conducted was performed by arranging the time series plot for those 4 variables. The result was presented below:
The result of time series plot on Figure 1 shows that those 4 variables were fluctuating to the relatively same pattern. It could be defined that the increased or decreased outflow of cash in one region would be coupled with the increased and decreased outflow of cash in the other three regions. The correlation between those variables indicated that the established model was one variable constituting the function of the other variables, therefore the appropriate model in this data was VAR model.

4.2. Identification of VAR model

The identification of stationary condition on multivariate time series model may be done by observing the MACF plot, as follow:

| Variable/ | Z1 | Z2 | Z3 | Z4 |
|-----------|----|----|----|----|
| Lag       | ++++ | .... | .... | .... |
|           | .... | .... | .... | .... |
|           | .... | .... | .... | .... |
|           | .... | .... | .... | .... |
|           | .... | .... | .... | .... |
|           | .... | .... | .... | .... |
|           | .... | .... | .... | .... |

+ is > 2*std error, - is < -2*std error, . is between

Figure 2. MACF Data of the Outflow of Cash in Four Regions

On Figure 2, it can be seen that nearly all signs symbolized with (.) after lag (0) showed none of correlation occurrence. As for the sign (+) and (-) there was no lag that came out of ± 2 times the standard error limit simultaneously. This indicated that the data was stationary in the mean that was viewed subjectively. In the multivariate time series only checked the stationary of mean.
The subsequent step was to determine the order of VAR model which may be conducted by identifying the partial cross-correlation matrix scheme among the variables or MPACF and observing its AIC value. All lags on MPACF appearing beyond the standard error limit may be used as the order of temporary estimate model. The results of identification of MPACF may be seen on this following figure:

| Variable/ Lag | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Z1            | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  |
| Z2            | ..-.  | .+..  | ....  | .-..  | ....  | ....  | ....  | ....  | ....  | ....  |
| Z3            | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  | ....  |
| Z4            | ...+  | ....  | ....  | ....  | ....  | .-..  | ....  | ....  | ....  | ....  |

+ is > 2*std error, - is < -2*std error, . is between

Figure 3. MPACF Data of the Outflow of Cash in 4 regions

Figure 3 shows that there were several lags appearing beyond the standard error limits. As the total lags appearing beyond the standard error limits were more than 1, accordingly there will be more than one model to be formed. It is therefore necessary to identify the AIC values in some order. Order with the smallest AIC value can form the corresponding VAR model. Results of identification of AIC values for all models of interim allegations are as follows:

| Lag | MA 0          | MA 1          | MA 2          | MA 3          | MA 4          | MA 5          |
|-----|---------------|---------------|---------------|---------------|---------------|---------------|
| AR 0| 43.193956     | 43.887573     | 43.971770     | 44.087056     | 44.310970     | 45.223479     |
| AR 1| 42.664127     | 42.829534     | 42.931481     | 43.350071     | 43.754947     | 44.562739     |
| AR 2| 42.722990     | 43.214542     | 43.595219     | 44.241583     | 44.855456     | 45.975469     |
| AR 3| 42.975704     | 43.578720     | 44.193073     | 45.178845     | 46.058586     | 47.534216     |
| AR 4| 43.219355     | 43.645792     | 44.466471     | 45.737529     | 47.332450     | 49.568261     |
| AR 5| 43.853172     | 44.501470     | 45.678765     | 47.325763     | 49.614203     | 53.582221     |

Figure 4. AIC value from Several Model Order

Figure 4 shows that the smallest AIC values were in ARMA (1.0) or AR (1) of 42.664127, so at the identification stage the most suitable model had an autoregressive order equal to one (p = 1). In this experiment, p = 2 or VAR (2) and p = 3 or VAR (3).

4.3. Parameter Estimation

Parameter Estimation of VAR model was done by regressing Zt on one of the regions as the response variable, and the predictor variable was Zt in the other three regions and the lags of Zt. Lag time auto regression used was (p = 1) determined based on the identification results on MPACF and the smallest AIC value. The values of the autoregressive parameter of the VAR model in this study were estimated using the least square method.

The AIC values for the three models were obtained as follows:
Table 1. VAR Model and AIC Value

| Model   | AIC Value  |
|---------|------------|
| VAR (1) | 42.72292   |
| VAR (2) | 42.69119   |
| VAR (3) | 42.87662   |

Diagnostic examination of whether the residual is white-noise, can be seen from the plot of the residual autocorrelation matrix. Plot of the residual autocorrelation matrix for the following three models is as follow:

Variable/
Lag 0 1 2 3 4 5
Z1 ++++ ---- .... .... .... ....
Z2 ++++ ---- .... .... .... ....
Z3 ++++ ---- .... .... .... ....
Z4 ++++ --.. .... .... .... ....

+ is > 2*std error, - is < -2*std error, . is between

Figure 5. Plot of the residual autocorrelation matrix of VAR(1) Model

On Figure 5 it can be seen in lag 1 that many of cross-correlation values indicated significantly negative results reflecting the non-fulfilment of white noise assumption.

Variable/
Lag 0 1 2 3 4 5
Z1 ++++ .... .... .... .... ....
Z2 ++++ .... .... .... .... ....
Z3 ++++ .... .... .... .... ....
Z4 ++++ ---- .... .... .... ....

+ is > 2*std error, - is < -2*std error, . is between

Figure 6. Plot of the residual autocorrelation matrix of VAR(2) Model

Figure 6 shows that after lag 0, the residual autocorrelation symbolized with (.) (nonsignificant) indicated the fulfilment of white noise assumption.

Variable/
Lag 0 1 2 3 4 5
Z1 ++++ ---- .... .... .... ....
Z2 ++++ ---- .... .... .... ....
Z3 ++++ ---- .... .... .... ....
Z4 ++++ ---- .... .... .... ....

+ is > 2*std error, - is < -2*std error, . is between

Figure 7. Plot of the residual autocorrelation matrix of VAR(3) Model

Figure 7 shows that after lag 0, the residual autocorrelation symbolized with (+) (significant) indicated the fulfilment of white noise assumption.
Based on the AIC value and the assumption of white noise, the best model was the VAR(2) Model with the AIC value = 42.69119 and the residual autocorrelation plot indicated the fulfillment of white noise assumption. The VAR(2) Model in the form of the equation is written as follow:

\[
\begin{bmatrix}
Z_{1,t} \\
Z_{2,t} \\
Z_{3,t} \\
Z_{4,t}
\end{bmatrix}
= \begin{bmatrix}
1.76741 & -3.55138 & -0.62162 & 2.06277 \\
0.81619 & -1.85411 & -0.13680 & 1.05192 \\
0.93777 & -2.03478 & -0.23358 & 1.10131 \\
0.047422 & -1.23284 & -0.16436 & 0.98972
\end{bmatrix}
\begin{bmatrix}
Z_{1,t-1} \\
Z_{2,t-1} \\
Z_{3,t-1} \\
Z_{4,t-1}
\end{bmatrix}
+ \begin{bmatrix}
0.89739 & 0.81971 & -0.27939 & -2.34222 \\
0.44460 & 0.63058 & -0.39880 & -1.11153 \\
0.39574 & 0.52354 & -0.01874 & -1.44647 \\
0.22852 & 0.38192 & -0.04124 & -0.83401
\end{bmatrix}
\begin{bmatrix}
da_{1,t} \\
da_{2,t} \\
da_{3,t} \\
da_{4,t}
\end{bmatrix}
\]

4.4. Forecast

Based on the best model of VAR(2) Model, it can be used to obtain forecast values for each region. Data were forecasted for the next 5-year periods. The results of the forecast using the VAR(2) Model for currency outflow data in four regions are as follows:

**Table 2.** Forecasting result with VAR(2) Model

| Year | Month | Semarang | Solo | Purwokerto | Tegal |
|------|-------|----------|------|------------|------|
| 2015 | April | 1738.69  | 871.79 | 839.14     | 510.44|
|      | May   | 1067.71  | 506.28 | 448.52     | 268.24|
|      | June  | 1208.47  | 574.30 | 552.15     | 343.89|
|      | July  | 1082.03  | 524.68 | 505.64     | 309.81|
|      | August| 969.30   | 464.04 | 441.30     | 275.71|

Plots of the forecasting result are presented below:

**Figure 8.** Data Plots of the Forecasting Results for the subsequent 5-year periods

In Figure 8, the data plot for each region had the same relative data pattern. In this plot, shows the predicted value for Z1 (Outflow of cash in the Representative Office of Bank Indonesia Semarang)
was higher compared to 3 other regions. While the forecast value for Z4 (Outflow of cash in the Representative Office of Bank Indonesia Tegal) had the lowest forecast value compared to the other 3 regions.

5. Conclusion
According to the results and discussion, the conclusions obtained are as follow:
   a. The best fit VAR Model for those 4 regions is VAR(2) Model which has the smallest AIC Value and satisfy the white noise assumption.
   b. The Forecasting results show that the value is fluctuated to the relatively similar data pattern.

6. Suggestions
It would be better to do further research, as in addition to the time correlation, it is also region correlation from one region to the others. The next research can be continued with the GSTAR-SUR model approach.

References
[1] Bank Indonesia. 2006. *Perhitungan Uang Kartal yang Diedarkan. Bagian Penelitian dan Pengembangan Pengedaran Uang – DPU*. [www.bi.go.id, 12 April 2017].
[2] Borovkova, S.A., Lopuhaa, H.P., and Ruchjana, B.N. 2008. Consistency and asymptotic normality of least square estimators in generalized STAR models. *Statistica Neerlandica*, Vol. 62, No.4, pp. 482-508.
[3] Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control*, 3rd edition, Englewood Cliffs: Prentice Hall.
[4] Faizah, L.A. and Setiawan. 2013. Pemodelan Inflasi di Kota Semarang, Yogyakarta dan Surakarta dengan Pendekatan GSTAR. *Jurnal Sains dan Seni Pomits*. Vol.2. No.2. pp. 2337-3520.
[5] Gerai Info. 2011. *Pengedaran Uang: “Yang Penting Pas Agar Ekonomi Nggak Kolaps”*. Sixteenth edition. Jakarta: Bank Indonesia.
[6] Lütkepohl, H. 2007. *Econometrics Analysis with Vector Autoregressive Models*. EUI Working Papers ECO. 1725-6704.
[7] Mubarak, R. 2015. *Generalized Space-Time Autoregressive with Exogenous Variables Model for Forecasting Cashflow in Bank Indonesia East Java Region*. Surabaya: Sepuluh November Institute of Technology.
[8] Pramono, B et al. 2006. *Dampak Pembayaran Non Tunai Terhadap Perekonomian dan Kebijakan Moneter*. Working Paper Nomor 11. Not published. Bank Indonesia.
[9] Prayitno, L., Sandjaya, H. dan Llewelyn, R. 2002. Faktor-Faktor Yang Berpengaruh Terhadap Jumlah Uang Beredar di Indonesia Sebelum dan Sesudah Krisis: Sebuah Analisis Ekonometrika. *Jurnal Manajemen & Kwirirauahaan*, Vol. 4, No.1, pp. 46–55.
[10] Suhartono dan Atok, R.M. 2006. Pemilihan Bobot Lokasi yang Optimal pada model GSTAR. *Presented at National Mathematics Conference XIII*. Semarang: State University of Semarang.
[11] Suhartono dan Subanar. 2006. The Optimal Determination of Space Weight in GSTAR Model by Using Cross-correlation Inference. *Journal of Quantitative Methods: Journal Devoted the Mathematical and Statistical Application in Various Field*, Vol.2. No.2. pp.45-53.
[12] Untoro. 2007. Mengkaji Efektivitas Penggunaan ARIMA dan VAR Dalam Melakukan Proyeksi Permintaan Uang Kartal di Indonesia. *Buletin Ekonomi Moneter dan Perbankan. Buletin Ekonomi Moneter dan Perbankan*, Vol.10. No.1, pp. 51-83.
[13] Wutsqa, D.U., Suhartono dan Sutijo, B. (2010). *Generalized Space-Time Autoregressive Modeling*. *Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010)*, Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia, pp. 752-761.
[14] Wutsqa, D. U. dan Suhartono. 2010. Seasonal Multivariate Time Series Forecasting on Tourism Data by Using Var-Gstar Model. *Jurnal ILMU DASAR*, Vol. 11, No. 1, pp. 101-109.

[15] Wei, W.W.S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. United State of America: Addison-Wesley Publishing Co., USA.