Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 16(2) 2023: 215-226
http://ejournal.undip.ac.id/index.php/media_statistika

MODEL COMPARISON OF VECTOR AUTOREGRESSIVE RESHAPED AND SARIMA IN SEASONAL DATA (A CASE STUDY OF TEA PRODUCTION IN PT PERKEBUNAN NUSANTARA VIII INDONESIA)

Dewi Juliah Ratnaningsih[1], Fia Fridayanti Adam[2]
[1] Statistics Study Program, Universitas Terbuka, Jakarta, Indonesia
[2] Vocational Education Program, Universitas Indonesia, Jakarta, Indonesia

e-mail: djuli@ecampus.ut.ac.id

DOI: 10.14710/medstat.16.2.215-226

Article Info:
Received: 15 March 2023
Accepted: 20 February 2024
Available Online: 30 June 2024

Keywords:
Tea Production; Time Series; Seasonal; Spacetime; VAR Reshaped
.

Abstract: PT Perkebunan Nusantara VIII (PTPN VIII) is a State-Owned Enterprise (BUMN). It operates in the plantation sector. The leading commodity is tea. The demand for tea produced by PTPN VIII is increasing. Thus, planning tea production is necessary. One of the production planning efforts is through forecasting based on previous data. Tea production data is time series data. It contains seasonal elements and is dependent on other locations. We will analyze data with these criteria using space-time models, one of which is Vector Autoregressive (VAR). VAR models the relationship between observations on certain variables at one time. It also models the observation of the variable itself at previous times. Additionally, VAR models the relationship between observations and other variables at previous times. This paper explains how to forecast tea production. It uses the reconstituted VAR and Seasonal Autoregressive Moving Average (SARIMA) models. The results showed that the reconstituted VAR model was better than the SARIMA model in predicting tea production. The tea production prediction was at the Sedep and Santosa plantations in Bandung Regency.

1. INTRODUCTION

Tea is one of the significant plantation products in Indonesia’s economy. Indonesia is one of the world's tea exporters. But, the volume and value of its exports have continued to decline. It is the most important export commodity income after oil and gas. Aside from its economic function, tea has a medicinal function compared to other drinks. The production and volume of Indonesian tea exports will decrease every year. Indonesian tea is strong. Yet, it lags behind India, Kenya, Sri Lanka, and Vietnam. Indonesia exports tea to the international market during the export expansion stage (Nursodik et al., 2021).

Tea originated in China. It is now the second most consumed beverage globally, after water. People drink 3 billion cups of tea every day (Voora et al., 2019). 11 provinces in Indonesia have tea plantations. These provinces are North Sumatera, West Sumatera, Jambi, South Sumatera, Bengkulu, West Jawa, Central Java, Yogyakarta, East Java, East Kalimantan, and South Sulawesi. PT Perkebunan Nusantara VIII manages most of the tea plantations in West Java, the largest province out of 11 in Indonesia (Elpawati et al., 2019).

As one of the state-owned plantations, PTPN VIII has the biggest market share, ie. around 70 % of Indonesia's tea production. PTPN VIII is responsible for 24 tea plantations in 25905.3 HA productive lands in West Java. Tea plantations in West Java sit in 6 regencies. They're in Sukabumi, Bogor, Cianjur, Subang, Bandung, West Bandung, and Garut. Sukabumi has 2 areas, Bogor has 2 areas, Cianjur has 3 areas, Subang has 2 areas, Bandung has 1 area, West Bandung has 12 areas, and Garut has 3 areas. Sedep and Santosa plantations in Pengalengan, Bandung Regency, are the biggest tea contributors. This is based on 2014 tea production data. The 1992-2011 tea production data also showed the same results. The average tea production in Santosa and Sedep plantations are 230 734 and 338 922 tonnes respectively. Concerning the increasing demand for the product, estimation of annual tea production planning is important.

Tea production data is a time series that is closely related to a certain location, or other location, and seasonal elements. Location element in this kind of data is known as spatial. Time series data that contains spatial elements can be modeled in spatial or space-time models. The Space-time model developed by Pfeifer & Deutch (1980) is a combination of location and time elements in time series and location data. Space-time autoregressive (STAR) is one of the space-time models--a specification of the Vector Autoregressive (VAR) model. Wei (2006) stated that the VAR model is a time series model that can involve more than one-time series variable. Rosadi (2006) also stated that VAR can also be used to model space-time data. Halim & Chandra (2011) model multivariate time series automatically with the VAR approach.

Suhartono & Wustqa (2007) researched tea production in the Bandung Regency. They found that VAR gave more specific model formation steps. In determining model order, VAR did not have to be autoregressive compared to STAR. Besides VAR offered better forecast accuracy. Unfortunately, VAR had not been used for seasonal data. Several researchers used the SARIMA (Seasonal ARIMA) model for seasonal univariate data series forecasting. Goswami & Hazarika (2017), and Hussain & Ali (2017) used the SARIMA model to predict Dibrugarh Station in Assam, India, and Pakistan. They did this because they had monthly and seasonal temperature data. The analysis revealed the best seasonal models to describe the data. SARIMA (2,1,1) (0,1,1)12 is the best model for monthly maximum temperature data, while SARIMA (2,1,1) (0,1,1)12 is the best for monthly minimum temperature data.

Suhartono (2011) mentioned using the SARIMA model to forecast multiplicative and additive seasonal time series data. The results stated that there are differences in the pattern of lag due to seasonal and non-seasonal lags. Determining the order of identification of subsets, multiplication or additive sequences needs to be considered to produce the best model. Mohamed et al., (2011), researchers used the SARIMA model to improve short-term load forecasting in energy systems. The data had seasonal patterns, either daily or weekly. The goal was to prevent system failure. Choi et al., (2011), Gijo & Etuk (2013) stated that researchers should analyze the detailed scale. This scale contains the seasonal and stochastic components. They should use SARIMA. In their research, Choi et al., (2011) combined two methods are SARIMA method and wavelet transform for sales forecasting. Mercy & Kihoro (2015) showed that the VAR model offered better forecasting performance than SARIMA for seasonal, univariate unemployment.

Tea production data at several interconnected locations is a time series. Data models like this can be forecasted by using VAR. Suhartono & Wustqa (2007) used VAR. However, they did not analyze seasonal data elements. However, they found seasonal data in tea production data. Mercy & Kihoro (2015) suggested reshaping the data to handle seasonal elements in time series data. This paper will discuss a forecasting model for seasonal data by VAR reshaped. It will also compare its performance with SARIMA. It will forecast tea production data at the Sedep and Santosa tea plantations of PTPN VIII.

2. LITERATURE REVIEW

2.1. Vector Autoregressive (VAR)

Developed by Sims (1986), VAR is the most frequently utilized multivariate time series forcasting model in predicting stationary data. Time series y_t where $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})$ declares variable with the size of $n \times 1$ is a VAR (*p*) model or VAR with *p* order provided that it meets the following equation:

$$\boldsymbol{Y_t} = \boldsymbol{\Pi} + \boldsymbol{\Phi_1 y_{t-1}} + \cdots + \boldsymbol{\Phi_{t-p} y_{t-p}} + \boldsymbol{e_t} \qquad t = 1, 2, 3, \ldots \tag{1}$$

In a matrix form, the equation can be put as follows:

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \Pi_1 \\ \vdots \\ \Pi_n \end{bmatrix} + \begin{bmatrix} \Phi_{11}^1 & \ldots & \Phi_{1n}^1 \\ \vdots & \ldots & \vdots \\ \Phi_{n1}^1 & \ldots & \Phi_{nn}^1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ \vdots \\ y_{nt-1} \end{bmatrix} + \begin{bmatrix} \Phi_{11}^2 & \ldots & \Phi_{1n}^2 \\ \vdots & \ldots & \vdots \\ \Phi_{n1}^2 & \ldots & \Phi_{nn}^2 \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ \vdots \\ y_{nt-2} \end{bmatrix} + \cdots + \begin{bmatrix} \Phi_{11}^p & \ldots & \Phi_{1n}^p \\ \vdots & \ldots & \vdots \\ \Phi_{n1}^p & \ldots & \Phi_{nn}^p \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ \vdots \\ y_{nt-p} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ \vdots \\ e_{nt} \end{bmatrix} \tag{2}$$

Where $\boldsymbol{\Pi}$ vector has a dimension of n, $\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}, \ldots, \boldsymbol{\Phi_p}$ is a parameter matrix $n \times n$, and $\boldsymbol{e_t}$ vector error which are free and have identical distribution. Error $\boldsymbol{e_t}$ is a vector with the size of $n \times 1$ and has a multivariate normal distribution and it fits with the assumption that $E(e_t) = 0$, $E(\boldsymbol{e_t}\ \boldsymbol{e_t'}) = s$, and $E(\boldsymbol{e_t}\ \boldsymbol{e_{t-n}'}) = 0$ for every non-zero *n*.

VAR(*p*) model can also be described as follows:

$$\left(\boldsymbol{I} - \boldsymbol{\Phi_1 B} - \cdots - \boldsymbol{\Phi_p B^p}\right)\boldsymbol{Y_t} = \boldsymbol{\Pi} + \boldsymbol{e_t} \tag{3}$$

With B as backward operator.

2.2. Seasonal Time Series Model

Seasonal time series denotes a seasonal phenomena that is repetitive in a certain period of time. The shortest time period of the occurence is called seasonal period. Seasonal time data series like that can be done through SARIMA (*Seasonal* ARIMA) model as follows:

$$A_p(\boldsymbol{Z})A_P(\boldsymbol{Z^s})\boldsymbol{x_t} = B_q(\boldsymbol{Z})B_Q(\boldsymbol{Z^s})\boldsymbol{\varepsilon_t} \tag{4}$$

Where $A_p(\boldsymbol{Z}), A_P(\boldsymbol{Z^s}), B_q(\boldsymbol{Z})$, and $B_Q(\boldsymbol{Z^s})$ are in a polynomial sequence order p, P, q, Q, Z are backward operator, and s is seasonal period of time series. A_p is the process of AR has p order, A_P is the process of AR with seasonal component order, B_q is the process of MA order q, and B_Q is the process of MA with seasonal component order. It is also found that $\boldsymbol{x_t} = \Delta^d \boldsymbol{y_t}$.

However, SARIMA model has several limitations. One of which is the high-cost construction because modelling is done through p, q, and s orders that take longer time to complete. Mercy & Kihoro (2015) overcame the shortcoming by reshaping seasonal time series model and completed it with VAR. In quarter data, for example, the seasonal period was 4 ($s = 4$). Reshaping means dividing y_1, y_2, y_3, y_4 data into 4 new datasets with data division, as inTable 1.

Table 1. Time Series Data Reshaping Concept

	Y_1	Y_2	Y_3	Y_4
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
	Q_1	Q_2	Q_3	Q_4
Y_1	x_1	x_2	x_3	x_4
Y_2	x_5	x_6	x_7	x_8
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_T	x_{n-3}	x_{n-2}	x_{n-1}	x_n
Y_{T+1} real value	$\hat{x}_{n+1}$	$\hat{x}_{n+2}$	$\hat{x}_{n+3}$	$\hat{x}_{n+4}$

Thus, one data set of Y_t will be changed into Q_1, Q_2, Q_3, Q_4. In other words, one univariate variable becomes multivariate variables. The 4 datasets are then completed by VAR(p) model as follows:

$$\begin{bmatrix} Q_{1t} \\ \vdots \\ Q_{4t} \end{bmatrix} = \begin{bmatrix} \Pi_1 \\ \vdots \\ \Pi_4 \end{bmatrix} + \begin{bmatrix} \Phi_{11}^1 & \dots & \Phi_{14}^1 \\ \vdots & \dots & \vdots \\ \Phi_{41}^1 & \dots & \Phi_{44}^1 \end{bmatrix} \begin{bmatrix} Q_{1t-1} \\ \vdots \\ Q_{4t-1} \end{bmatrix} + \begin{bmatrix} \Phi_{11}^2 & \dots & \Phi_{14}^2 \\ \vdots & \dots & \vdots \\ \Phi_{41}^2 & \dots & \Phi_{44}^2 \end{bmatrix} \begin{bmatrix} Q_{1t-2} \\ \vdots \\ Q_{4t-2} \end{bmatrix}$$
$$+ \cdots + \begin{bmatrix} \Phi_{11}^p & \dots & \Phi_{14}^p \\ \vdots & \dots & \vdots \\ \Phi_{41}^p & \dots & \Phi_{44}^p \end{bmatrix} \begin{bmatrix} Q_{1t-p} \\ \vdots \\ Q_{4t-p} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ \vdots \\ e_{4t} \end{bmatrix} \quad (5)$$

3. MATERIAL AND METHOD

This study used secondary data of PTPN VIII tea production. The data came from two locations: Sedep Tea Plantation and Santosa Tea Plantation. We selected sites that are close to each other based on their locations. We collected data from monthly tea production from January 1992 to October 2011. We have 238 data points. In the analysis process, we divided the data into training data and testing. We used softwares for this data analysis are SAS 9.4 and Minitab 16.

Tea production data was forecast in two ways: VAR reshaped models and SARIMA. The model formation steps included four steps. These were: model identification, model estimation, model residual testing, and forecasting. The following was forecasting model on VAR reshaped and SARIMA.

3.1. VAR Reshaped

VAR reshapes the data by dividing it based on seasonal periods. We reshaped the data twice, assuming a seasonal pattern with two periods for tea production $(s = 4)$. We do this to meet the assumption of the number of data for model estimation. Thus, the VAR reshaped model requires the analysis of two new data sets. The first step in VAR analysis is the identification model phase. This step aims to determine data stationary and model order. If the absolute value of the AR Characteristic Polynomial Roots is less than 1, the data is stationary. Meanwhile, we can determine the model order from the MPACF plot. We can also look at the Akaike's Information Criterion (AIC) value, which should be the smallest.

a. Phase Estimation

   It is a model estimation of the datasets resulting from reshaping by using iterative methods of OLS.

b. Model Residual Testing Phase
This step aims to check whether the estimated residual models already meet the assumption of white noise and normal multivariate. The testing is done by using distribution test by using Minitab macros.

c. Forecasting Step
At this step, testing data is used. Data model generated from the process of reshaping is used to forecast with a treatment step to s where $\hat{x}_s$ is equivalent to the first step of forecasting $\hat{Y}_{t+1}$.

3.2. SARIMA

SARIMA model building steps are similar to formation of VAR reshaped. The data used in this modeling is intact on each set of training and testing data.

After forecasting model is obtained through the VAR reshaped and SARIMA, the next analysis step is to determine model performance to determine which model is better in tea production data forecasting in both locations. The model performance criteria is seen from the MSE, RMSE, and MAPE from each model. The third formulation of the model performance criteria is as follows:

$$MSE(\hat{Y}_t) = E(\hat{Y}_t - Y_t)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \tag{7}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|\hat{Y}_t - Y_t|}{Y_t} \times 100 \tag{8}$$

4. RESULTS AND DISCUSSION

Systematic exposure of the results of the analysis in the paper is divided into 3 sub-sections, namely: (a) determining VAR Reshape Model, (b) determining SARIMA model, and (3) comparing the performance of VAR Reshape Model and SARIMA Models in tea production data forecasting at PTPN VIII's Sedep and Santosa Tea Plantations.

4.1. VAR Reshaped Model

There are 4 steps in VAR reshaped model forecasting: (a) model identification which covers stationary identification and VAR order, (b) model parameter estimation, (c) model residual testing covering white noise and normal multivariate tests, and (d) forecasting to be the used in determining MSE, RMSE, and MAPE model values. Each step will be described in the following discussion.

Model Identification

This step is aimed at finding our the stationary data and model order. Data stationary can be seen in Roots of AR Characteristic Polynomial value (Table 2). Table 2 shows that the absolute value of Roots of AR Characteristic Polynomial is smaller than 1, indicating that the data is stationary.

Meanwhile model order can be observed in MPACF plot and Akaike's Information Criterion (AIC) value from data which are already stationary. MPACF plot of tea production data in two locations are presented in Figure 1 while AIC value is presented in Table 3.

Figure 1 shows that significant lags happen until the 3rd lag, and again in the 5th., and 7th. lag. In the meantime, the smallest AIC (Table 3) identifies that the possible model candidates are VAR(1), VAR(2), or VAR(3). From the two criteria, the model is still dynamic, therefore residual assumption has to be checked to see which fulfils the white noise assumption. Wei (2006) stated that the model that met the white noise assumption was the chosen model.

Table 2. Roots of AR Characteristic Polynomial Value of Tea Production in Two Locations

Index	Real	Imaginary	Modulus	ATAN(I/R)	Degree
1	0.67968	0.00000	0.69790	0.00000	0.00000
2	0.60323	0.00000	0.60320	0.00000	0.00000
3	0.28911	0.61775	0.68210	1.13310	64.92030
4	0.28911	-0.61775	0.68210	-1.13310	-64.92030
5	0.19049	0.00000	0.19050	0.00000	0.00000
6	-0.11892	0.56435	0.57670	1.77850	101.89900
7	-0.11892	-0.56435	0.57670	-1.77850	-101.89900
8	-0.20139	0.46349	0.50530	1.98070	113.48530
9	-0.20139	-0.46349	0.50530	-1.98070	-113.48530
10	-0.32243	0.35400	0.47880	2.30960	132.32750
11	-0.32243	-0.35400	0.47880	-2.30960	-132.32750
12	-0.66098	0.00000	0.66100	3.14160	180.00000

Schematic Representation of Partial Cross Correlations

Variable/ Lag	1	2	3	4	5	6	7	8	9	10	11	12
Q1_SED	.+..	....	-...	....	.+..	....	....	....	....	....	.+..	....
Q2_SED	+...	+...	....	....	....	....	....	....	....	....	....	....
Q1_SAN	..++	...+	-...	....	....	....	...+	....	....	....	....	....
Q2_SAN	..++	-..-	+...	....	....	....	....	-...	....	....	....	....

Figure 1. MPACF Plot of Tea Production Data in Two Locations

Table 3. Information Criterion Values of The Tea Production in Two Locations

Lag	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	86.468429	86.431833	86.440272	86.583329	86.718209	86.87199
AR 1	86.084834	86.340397	86.349694	86.467179	86.629834	86.824391
AR 2	86.088486	86.350832	86.527838	86.52788	86.754403	87.004434
AR 3	86.080931	86.389439	86.445177	86.712447	86.974354	87.366589
AR 4	86.390303	86.653224	86.719533	86.999679	87.266046	87.758740
AR 5	86.182055	86.354685	86.702559	87.172354	87.615655	88.169301
AR 6	86.452073	86.687054	87.099731	87.615677	88.062351	88.463483
AR 7	86.548218	87.094809	87.550266	88.126751	88.670589	89.171246
AR 8	87.013361	87.628745	88.314460	88.986941	89.644773	90.277277
AR 9	87.559807	88.258636	89.043474	89.931242	90.712371	91.509826
AR 10	88.118404	88.919874	89.828208	90.064105	92.064105	93.105887

Model Estimation Step

Model parameter estimation in this paper is done through iterative OLS. Based on the model analysis result of VAR Reshape in two locations, the one that met the white noise residual assumption is as follows:

$$\begin{aligned} Q_{1\_SED(t)} &= 249937 + 0.221 Q_{2_{SED(t-1)}} + 0.426 Q_{2_{SAN(t-1)}} + 0.248 Q_{2_{SED(t-2)}} \\ &\quad -0.347 Q_{2_{SAN(t-2)}} - 0.252 Q_{1_{SED(t-3)}} + 0.176 Q_{2\_SAN(t-3)} + e_{1\_SED(t)} \end{aligned}$$

$$Q_{2\_SED(t)} = 97443 + 0.548 Q_{1\_SED(t-1)} - 0.381 Q_{1\_SAN(t-1)} - 0.195 Q_{2\_SED(t-2)}$$

$$
\begin{aligned}
&+0.322Q_{1\_SAN(t-2)} + 0.189Q_{1\_SAN(t-3)} + 0.582Q_{1\_SED(t-5)} \\
&-0.324Q_{2\_SED(t-3)} + e_{2\_SED(t)}
\end{aligned}
$$

$$
Q_{1\_SAN(t)} = 173647 + 0.413Q_{2\_SAN(t-1)} - 0.127Q_{1\_SED(t-3)} + e_{1\_SAN(t)}
$$

$$
\begin{aligned}
Q_{2\_SAN(t)} = {} & 97040 + 0.17Q_{1\_SED(t-1)} + 0.288Q_{1\_SAN(t-2)} - 0.168Q_{2\_SAN(t-2)} \\
& +0.211Q_{1_{SED(t-5)}} - 0.194Q_{2_{SED(t-5)}} + 0.137Q_{1_{SAN(t-5)}} + e_{2\_SAN(t)}
\end{aligned}
$$

Model Residual Test Step

After the significant parameter and model were retrieved, an assumption test followed to find out whether the residual met the white noise assumption and normal multivariate. White noise residual can be seen in the residual MACF plot in Figure 2.

```
                    The VARMAX Procedure

      Schematic Representation of Residual Cross Correlations

Variable/
Lag         0     1     2     3     4     5     6     7     8     9     10    11    12

Q1_SED     ++++  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....
Q2_SED     ++++  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....
Q1_SAN     ++++  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....
Q2_SAN     ++++  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....  ....

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

Figure 2. Residual MACF Plot of Tea Production in Two Locations

From Figure 2 we can see that there is no significant lag in the residual MACF plot, meaning the assumption of white noise residual is met. Meanwhile, normal multivariate residual assumption is tested between residual, test data and multinormal with qq plot. The multinormal test result using Minitab macro indicates that the data has multinormal distribution ($t = 0.681$). Based on qq plot and multinormal test result against the residual, it can be stated that the acquired VAR Reshape Model has met multinormal distribution assumption. The qq plot result where the residual has normal multivariate distribution.

Forecasting Step

In this step, information presented is the performance of forecasting result, covering MSE, RMSE, and MAPE values. VAR reshaped model of forcasting performance analysis is presented in Table 4.

Table 4. VAR Reshape Model Performance From The Two Plantations

Model	Variable	Results of Model Performance		
		MSE	RMSE	MAPE
VAR Reshape	Sedep	36.323.39	7.565.69	0.203
	Santosa	27.678.76	5.237.44	0.219

4.2. SARIMA Model

SARIMA Model presented in this section is based on location, i.e. SARIMA for Sedep Plantation and SARIMA for Santosa Plantation. The step in each location is described as follows:

SARIMA Sedep

Model Identification

Like VAR reshaped model, identification of SARIMA Model covers: model stationary and order. You can see data stationary in the time series plot below. Figure 3 shows that tea production data at Sedep Plantation is stationary. We use ACF and PACF to decide the model order. Figures 4 and 5 present both plots. We analyze the following from ACF and PACF plots. First, the data is stationary because of dies-down decreasing lag in ACF and the cut off PACF. Secondly, there is 12th seasonal phenomenon. The results indicate the appropriate model is SARIMA(1,0,0)(1,0,0)^{12}.

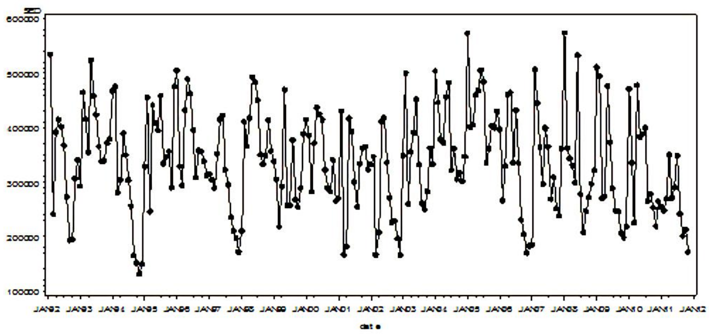

Figure 3. Data Series Plot Of Tea Production At Sedep Plantation

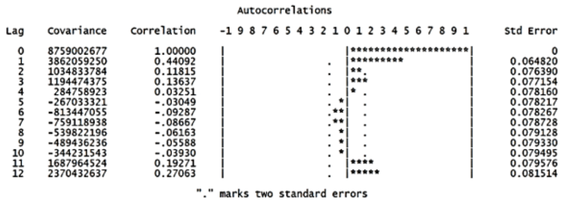

Figure 4. ACF Plot of Tea Production Data at Sedep Plantation

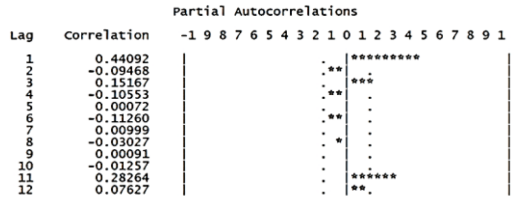

Figure 5. PACF Plot of Tea Production Data at Sedep Plantation

Model Estimation

The 1st. step identifies that the possible for Sedep is SARIMA(1,0,0)(1,0,0)^{12}. Analysis results analysis for that model is presented in Table 5.

Table 5. Parameter Estimation Value on SARIMA Model (1,0,0)(1,0,0)^{12}

Parameter	Estimate	Standard Error	t value	Approx Pr > |*t*|	Lag
MU	337223.3	13417,9	25.13	< 0.0001	0
AR1,1	0.45411	0.05867	7.74	< 0.0001	1
AR2,1	0.30363	0.06336	4.79	< 0.0001	12

The equation model of Table 5 can be derived as follows

$$(1-0.45411B)(1-0.3036B^{12})y_t = 337223.3 + e_t$$

$$y_t = 337223.3 + 0.45411y_{t-1} + 0.3036y_{t-12} - 0.137881y_{t-13} + e_t$$
$$+0.211Q_{1_{SED(t-5)}} - 0.194Q_{2_{SED(t-5)}} + 0.137Q_{1_{SAN(t-5)}} + e_{2\_SAN(t)}$$

Model Residual Test

The assumption test of white noise and normal multivariate on SARIMA Model (1,0,0)(1,0,0)^{12} at Sedep location is presented in below. White noise test result is in Table 6, showing that white noise assumption is met with p-value > 0.05 in all lags.

Next step is residual distribution test by using Anderson Darling and Kolmogorov Smirvov. The results show that the residual met the normal distribution assumption with p-value > 0.05. Normality test results are presented in Table 7.

Table 6. Results of Residual White Noise Test Analysis on SARIMA Model $(1,0,0)(1,0,0)^{12}$

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.000	4	0.0611	0.009	-0.102	0.155	0.008	0.019	-0.044
12	16.86	10	0.0776	-0.025	-0.011	-0.059	-0.073	0.133	-0.065
18	21.06	16	0.1760	0.087	-0.060	-0.027	-0.005	-0.055	-0.039
24	28.37	22	0.1639	-0.024	0.024	-0.008	-0.060	0.116	0.096
30	31.11	28	0.3121	0.032	0.038	0.013	-0.065	-0.035	0.045
36	36.63	34	0.3475	-0.036	-0.076	-0.019	-0.012	-0.002	0.110
42	38.72	40	0.5277	0.052	0.036	-0.002	-0.032	0.006	-0.470

Table 7. Residual Normality Test on SARIMA Model $(1,0,0)(1,0,0)^{12}$

Test	Statistics	p-Value
Shapiro-Wilk	0.985846	0.0184
Kolmogorov-Smirnov	0.041902	>0.1500
Cramer-von Mises	0.097008	0.1261
Anderson-Darling	0.726110	0.0597

The analysis shows that SARIMA model $(1,0,0)(1,0,0)^{12}$ is appropriate to be used in forecasting tea production data at Sedep Plantation.

Forecasting

From forecasting data, model performance of MSE, RMSE, and MAPE values can be derived. The three values are presented in Table 8.

Table 8. Performance of SARIMA Model $(1,0,0)(1,0,0)^{12}$

MSE	RMSE	MAPE
9186.26	8060.13	0.210

Table 8 showing that MAPE model value is smaller than 1 indicates that the model performance is better.

SARIMA at Santosa Location

Model Identification

Model Identification for Santosa location can be seen from the time series plot and ACF and PACF in Figure 6, Figure 7, and Figure 8. The three figures show that the data is stationary because of AFC's dies-down lag decrease. cut off PACF and 12th seasonal phenomenon. SARIMA$(1.0.0)(1.0.0)^{12}$ because of cut off PACF.

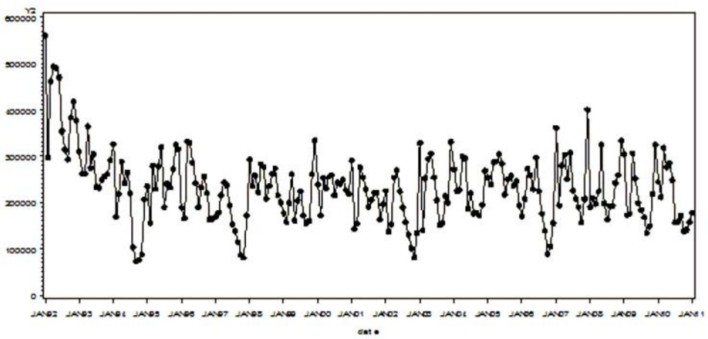

Figure 6. Data Series Plot of Tea Production at Santosa Plantation

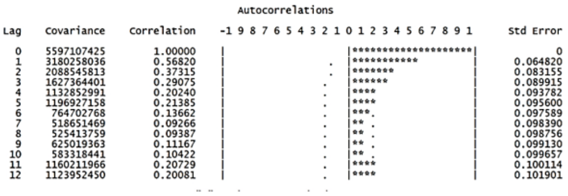

Figure 7. ACF Plot of Tea Production Data at Santosa Plantation

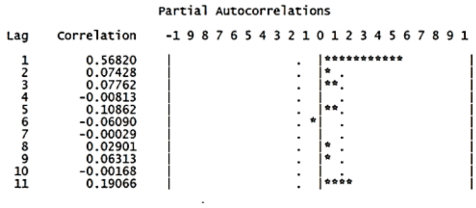

Figure 8. PACF Plot of Tea Production Data at Santosa Plantation

Model Estimation

Analysis of model parameter estimation on tea production data at Santosa location is in Table 9.

Table 9. Parameter Estimation Value on SARIMA Model $(1,0,0)(1,0,0)^{12}$

Parameter	Estimate	Standar Error	t value	Approx Pr > |*t*|	Lag
MU	234508.7	12401.3	18.91	< 0.0001	0
AR1,1	0.60883	0.05195	11.72	< 0.0001	1
AR2,1	0.22357	0.06526	3.43	0.0007	12

Based on the above. SARIMA Model $(1,0,0)(1,0,0)^{12}$ at Santosa location is:

$$(1-0.60883B)(1-0.22357B^{12})y_t = 234508.7 + e_t$$

$$y_t = 234508.7 + 0.60883y_{t-1} + 0.22357y_{t-12} - 0.136116y_{t-13} + e_t$$

Model Residual Test

Model residual white noise at Santosa location is presented in Table 10. showing that p-value>0.05 for 13th lag forward and it means that the residual met the white noise assumption from 13th lag forward. However. 1st. – 6 lags. white noise residual assumption is met on level α=3% and from 7th – 12th lag. it is met on level α=1%.

Table 10. Results of Residual White Noise Test Analysis on SARIMA Model $(1,0,0)(1,0,0)^{12}$

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	10.34	4	0.0350	-0.121	-0.003	0.071	-0.066	0.136	-0.003
12	22.39	10	0.6540	0.051	0.023	0.017	-0.068	0.174	-0.098
18	25.26	16	0.0654	0.038	-0.078	-0.031	0.044	-0.010	-0.025
24	30.87	22	0.0989	-0.019	-0.039	-0.037	-0.024	-0.041	0.125
30	35.65	28	0.1517	-0.064	0.024	0.053	-0.066	-0.063	0.041
36	40.99	34	0.1906	-0.045	-0.021	-0.007	-0.062	0.066	0.091
42	44.08	40	0.3033	0.046	0.041	-0.039	0.007	0.044	-0.058

Table 11 shows that the results of Kolmogorov Smirnov test is that p-value>0.01. meaning the residual met the normal distribution assumption on level α=1%. Both results show that model residual of SARIMA Model $(1,0,0)(1,0,0)^{12}$ at Santosa location met the *white noise* assumption and had normal distribution.

Table 11. Residual Normality Test on SARIMA Model $(1,0,0)(1,0,0)^{12}$

Test	Statistics	p-Value
Shapiro-Wilk	0.951587	<0.0001
Kolmogorov-Smirnov	0.065697	0.0135
Cramer-von Mises	0.309561	<0.0050
Anderson-Darling	1.933587	<0.0050

Forecasting

Performance of SARIMA Model $(1,0,0)(1,0,0)^{12}$ at Santosa location can be seen in Table 12.

Table 12 . Performance of SARIMA Model $(1.0.0)(1.0.0)^{12}$

MSE	RMSE	MAPE
6751.46	6062.77	0.21724

Table 12 shows that model MAPE value is smaller that one. meaning that model performance is better.

4.3. Comparison VAR Reshaped Model and SARIMA

Based on the previous analysis.the following shows comparative performance of the two models.

Table 13. Comparative Performance of VAR *Reshaped* Model and SARIMA

Model	Variable	Performance		
		MSE	RMSE	MAPE
VAR Reshaped	Sedep	**5723.339**	**7565.46**	**0.20324**
	Santosa	**2742.876**	**5237.44**	0.21850
SARIMA	Sedep	6497.259	8060.13	0.21004
	Santosa	3675.464	6062.77	**0.21724**

Table 13 shows that based on MSE and RMSE criteria. VAR Reshaped Model was found to be better than SARIMA Model concerning tea production data at Sedep and Santosa plantations. MAPE values on VAR reshaped model at Sedep location is 0.20342 while SARIMA Model 0.21004. Based on that. VAR Reshaped Model is better than SARIMA Model in terms of tea production forecasting at Sedep location. MAPE VAR Reshaped value of Santosa is higher than SARIMA.

5. CONCLUSION

Forecasting model of data containing seasonal elements is generally analyzed with SARIMA models. However. seasonal data can be analyzed with VAR model. One VAR models used to analyze seasonal data is VAR reshaped. with its main concept of data reshaping which is divided based on the seasonal period.

Based on the MSE and RMSE values of the tea production data at PTPN VIII plantation at Santosa and Sedep. the forecasting model produced by VAR Reshaped is proved to be better than SARIMA while based on MAPE criteria. the VAR Reshaped Model is better than SARIMA models for Sedep location.

.

REFERENCES

Choi, T.M., Yu, Y., & Au, K.F. (2011). A Hybrid SARIMA Wavelet Transform Method for Sales Forecasting. *Decision Support System*. *51*, 130-140.

Elpawati. N.A.A. & Abimanyu. T. (2019). Influence of Product Quality, Price, Promotion, and Quality Service Towards Customer Satisfaction of Walini Tea in Downstream Tea Industry PT. Nusantara Plantations VIII Bandung. *Agribusiness Journal*. *13*(7), 82-95.

Etuk, E. H. (2013). The Fitting of a SARIMA Model to Monthly Naira-Euro Exchange Rates. *Mathematical Theory and Modeling*. *3*(1), 17-26.

Gijo. E. V. (2011). Demand Forecasting of Tea by Seasonal ARIMA Model. *International Journal of Business Excellence*. *4*(1), 111-124.

Halim, S. & Chandra, A. (2011). Automatic Multivariate Time Series. *Journal of Industrial Engineering*. *13*(1), 19-26.

Hussain, M.N. & Ali, A. (2017). Forecasting of Pakistan's Import Prices of Black Tea Using ANN and SARIMA Model. *International Review of Management and Business Research*. *6*(4), 1372-1382.

Mercy, C. & Kihoro, J. (2015). Application of Vector Autoregressive (VAR) Process in Modelling Reshaped Seasonal Univariate Time Series. *Science Journal of Applied Mathematics and Statistics*. *3*(3), 124-135.

Mohamed, N., Ahmad, M.H., Suhartono, S., & Ismail, Z. (2011). Improving Short Term Load Forecasting Using Double Seasonal ARIMA Model. *World Applied Sciences Journal*. *15*(2), 223-231.

Nursodik, H., Santoso, S. I. & Nurfadilah, S. (2021). Competitiveness and Determining Factors of Indonesian Tea Export Volume in The World Market. *Habitat*. *32*(3), 163-172.

Pfeifer, P. E., & Deutsch, S. J. (1980). A Three-Step Iterative Procedure for Space-Time Modeling. *Technometrics*. *22*(1).

Rosadi, D. (2006). *Learning Module: Time series Analysis. An Introduction*. Department of Statistics. Gadjah Mada University. Yogyakarta: Andi Publisher.

Sims, C. A. (1986). Macroeconomics and Reality. *Journal Econometrics*. *48*(1), 38-48.

Suhartono, S. & Wutsqa, D. U. (2007). *Comparative study of VAR Model and STAR on Tea Production Forecast in West Java*. http://staff.uny.ac.id/sites/default-/files/132048772/Suhartono_&_Doriva_2007.pdf access on 25 Juni 2015.

Suhartono, S. (2011). Time Series Forecasting by using Seasonal Autoregressive Integrated Moving Average: Subset. Multiplicative or Additive Model. *Journal of Mathematics and Statistics*. *7*(1), 20-27.

Voora, V., Bermúdez, S., & Larrea, C. (2019). *Global Market Report: Tea. Sustainable Commodities Marketplace Series 2019*. Canada: International Institute For Sustainable Development *(IISD)*. https://www.jstor.org/stable/resrep22027?seq=1#me-tadata_info_tab_contents

Wei, W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education. Boston: Pearson Addison Wesley.