Comparison between GSTAR and GSTAR-Kalman Filter models on inflation rate forecasting in East Java

Jessica Rahma Prillantika\textsuperscript{1}, Erna Apriliani\textsuperscript{2}, Nuri Wahyuningsih\textsuperscript{3}

\textsuperscript{1} Student at Mathematics Department, Sepuluh Nopember Institute of Technology, Indonesia
\textsuperscript{2,3} Lectures at Mathematics Department, Sepuluh Nopember Institute of Technology, Indonesia

Abstract—Up to now, we often find data which have correlation between time and location. This data also known as spatial data. Inflation rate is one type of spatial data because it is not only related to the events of the previous time, but also has relevance to the other location or elsewhere. In this research, we do comparison between GSTAR model and GSTAR-Kalman Filter to get prediction which have small error rate. Kalman Filter is one estimator that estimates state changes due to noise from white noise. The final result shows that Kalman Filter is able to improve the GSTAR forecast result. This is shown through simulation results in the form of graphs and clarified with smaller RMSE values.

1. Introduction

Up to now, we often find data which have correlation between time and location. This data also known as spatial data. Inflation rate is one type of spatial data because it is not only related to the events of the previous time, but also has relevance to the other location or elsewhere. Inflation rate can not be destroyed clearly. We only can reduce and control it. Because the controlled inflation can increase economics matters \cite{1}. This fact makes forecast analysis be the important things in inflation rate in Malang, Probolinggo, and Surabaya. Because these locations have an important role to oversee the economics stability in East Java.

One of the methods that can be used to forecast the inflation rate is Generalized Space Time Autoregressive method (GSTAR). This method is used to forecast data which have correlation between time and location. In forecasting method, we need some parameters from GSTAR model. To estimate the value of the GSTAR parameters, we need method that can give the best forecasting. In this research, Ordinary Least Square methods (OLS) and Kalman Filter Estimator are used to estimate the GSTAR parameters \cite{2}.

Filter Kalman method is one of the estimator that estimate toward the conversion of state because of white noise disruption. The advantage of Kalman Filter is in estimation process, recursive form is used to increase Root Mean Square Error (RMSE) value and error \cite{3}. In the past researches, Kalman Filter had done to estimate the ARIMA and VAR parameters. The result showed that estimation using Kalman Filter was better than OLS. It was shown with RMSE value in estimation using Kalman Filter is smaller than OLS \cite{4,5,6}.

In this research, we’ve already comparison the forecasting result in the inflation rate month to monthly in Malang, Probolinggo, and Surabaya using GSTAR method and GSTAR-Kalman Filter method. The best method is the method that have smaller RMSE.

2. GSTAR (Generalized Space-Time Autoregressive) Model

The GSTAR model is specific form of VAR (Vector Autoregressive) model. It reveals linear dependencies of space and time. The main difference is on the spatial dependent, that in GSTAR model, it is expressed by weight matrix. Let \{\textit{Z}(t): t = 0, \pm 1, \pm 2, \ldots\} be a multivariate time series of N components. In matrix notation, the GSTAR model of autoregressive order p and spatial orders \(\lambda_1, \lambda_2, \ldots, \lambda_p\), GSTAR \((\lambda_1, \lambda_2, \ldots, \lambda_p)\) could be written as \cite{7}:
$Z(t) = \sum_{s=1}^{p} \left[ \Phi_{s0} + \sum_{k=1}^{\lambda_p} \Phi_{sk} W^{(k)} \right] Z(t-s) + e(t)$

(1)

where

$\Phi_{s0} = \text{diag}(\phi_{s0}^1, ..., \phi_{s0}^N)$ and $\Phi_{sk} = \text{diag}(\phi_{sk}^1, ..., \phi_{sk}^N)$.

$W$ (weight) are choosen to satisfy $w_{ii}^{(k)} = 0$ and $\sum_{i\neq j} w_{ij}^{(k)} = 1$.

3. Modified GSTAR Model

The first step to get the GSTAR model is identifying model that include stationary test and spatial and autoregressive order selection. The Box-Cox transformation and Matrix Autocorrelation Function (MACF) plot are employed for testing stationary conditions before GSTAR model is proceeded. The Box-Cox is employed to identify the stationary in varians. If rounded value $\neq 1$, then data is not stationer in varians and transformation must be employed. Whereas MACF plot is employed to identify the stationary in mean. If the data is not stationar, then differencing must be employed [7].

After the data is stationer in mean and varians, then GSTAR order is selected using VAR model. VAR model can be identified from Matrix Partial Autocorrelation Function (MPACF) plot. In this research, 72 in-sample data of the inflation rate is used to modify the GSTAR Model Autoregressive order can be selected by Schwarz Bayesian Criterion (SBC) and spatial order is generally restricted on order 1, since the higher order is difficult to be interpreted [7]. Then estimating the parameters of the model that include determination of space weight and autoregressive parameter. In this research, determination of space weight by using inverse distance [8]. Whereas, OLS and Kalman Filter can be used in estimating parameters. The next step after finding significant model is evaluating the residuals to verify normality and white noise of the error. The last step is comparing the RMSE result of the GSTAR and GSTAR-Kalman Filter [9].

MPACF plot from inflation rate in Malang, Probolinggo, and Surabaya is presented in Figure 1. Figure 1 indicates cut off in lag 1, 2, 4, and 10. It means that there are several possibility order based on these lag. Then we use SBC as the criterion, we select GSTAR model from the lowest value of SBC.

\[ \text{SBC value is presented in Table 1. Table 1 shows that the smallest SBC value for MA 0 is in the lag 2. It indicates that order AR 2 or } p = 2. \text{ Therefore, we obtain GSTAR } (2, 1) \text{ for inflation rate in 3 locations [4].} \]
Table 1. SBC value in any autoregressive order of inflation rate in 3 locations

| Lag | MA 0  |
|-----|-------|
| AR 0 | -17.76274 |
| AR 1 | -17.85476 |
| AR 2 | -18.31683 |
| AR 3 | -18.25295 |
| AR 4 | -18.25391 |
| AR 5 | -17.97148 |
| AR 6 | -17.82173 |
| AR 7 | -17.44618 |
| AR 8 | -17.1605 |
| AR 9 | -16.85751 |
| AR 11| -16.65356 |

4. Parameters Estimation
To estimate the parameters, we must determinate the space weight. In this research, we use the distance inverse space weight ($W$). Then the parameters are estimated using OLS and Kalman Filter [7].

4.1 Determination of Space Weight
The distance inverse space weight ($W$) in GSTAR($2_t$) are generally taken as follow:

$$W = \left[ w_{ij} \right] = \begin{bmatrix}
0 & w_{12} & \ldots & w_{1j} \\
wx1 & 0 & \ldots & w_{2j} \\
\vdots & \vdots & \ddots & \vdots \\
w_{t1} & w_{t2} & \ldots & w_{ij}
\end{bmatrix}$$

(2)

where $w_{ij}$ is distance between $i$ - $th$ location with $j$ - $th$ location. And $w_{ij}$ can be declared as:

$$w_{ij} = \begin{cases}
\frac{1}{b_{ij}}, & i \neq j \\
\frac{1}{\sum_{j=1}^{N} b_{ij}}, & i = j
\end{cases}$$

(3)

Based on Equation (2) and (3), we can calculate the distance inverse space weight ($W$) in GSTAR($2_t$). The calculation result is expressed as follows:

$$W = \begin{bmatrix}
0 & 0.5148 & 0.4852 \\
0.5353 & 0 & 0.4647 \\
0.5202 & 0.4795 & 0
\end{bmatrix}$$

(4)

4.2 Parameter Estimation of GSTAR($2_t$) using OLS
Based on Equation (1) with the distance inverse space weight ($W$) from Equation 4, GSTAR($2_t$) model is represented as follow:
\[
\begin{bmatrix}
z_i(t) \\
z_j(t) \\
z_k(t)
\end{bmatrix} = \begin{bmatrix}
\Phi_{10}^t & 0 & 0 \\
0 & \Phi_{10}^t & 0 \\
0 & 0 & \Phi_{10}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-1) \\
z_j(t-1) \\
z_k(t-1)
\end{bmatrix} + \begin{bmatrix}
\Phi_{11}^t & 0 & 0 \\
0 & \Phi_{11}^t & 0 \\
0 & 0 & \Phi_{11}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-1) \\
z_j(t-1) \\
z_k(t-1)
\end{bmatrix} + \begin{bmatrix}
\Phi_{20}^t & 0 & 0 \\
0 & \Phi_{20}^t & 0 \\
0 & 0 & \Phi_{20}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-2) \\
z_j(t-2) \\
z_k(t-2)
\end{bmatrix} + \begin{bmatrix}
\Phi_{21}^t & 0 & 0 \\
0 & \Phi_{21}^t & 0 \\
0 & 0 & \Phi_{21}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-2) \\
z_j(t-2) \\
z_k(t-2)
\end{bmatrix} + \begin{bmatrix}
e_1(t) \\
e_2(t) \\
e_3(t)
\end{bmatrix}
\]

(5)

Based on equation (5), the result of parameter value using OLS method can be seen in Table 2. Then these parameters are eliminated using a significance level of 0.05. To eliminate these parameters, we use t-student test with value of t-table is 1.97. The parameter is catagorized as not significant parameter if \(|t-value| < t-table\). And the parameter is categorized as significant parameter if \(|t-value| > t-table\).

### Table 2. Estimated parameters of GSTAR(2,4) model for inflation rate data

| Coef   | SE Coef | T  | Conclusion     |
|--------|---------|----|----------------|
| \Phi_{10}^t | -0.1696 | 0.2431 | -0.70 | Not significant |
| \Phi_{11}^t | 1.2088  | 0.2718 | 4.45  | Significant     |
| \Phi_{20}^t | -0.9206 | 0.2372 | -3.88 | Significant     |
| \Phi_{21}^t | 0.6460  | 0.2527 | 2.56  | Significant     |
| \Phi_{30}^t | 1.0855  | 0.2062 | 5.26  | Significant     |
| \Phi_{31}^t | -0.4064 | 0.2086 | -1.95 | Not significant |
| \Phi_{40}^t | 0.7975  | 0.2346 | 3.40  | Significant     |
| \Phi_{41}^t | -0.6977 | 0.2183 | -3.20 | Significant     |
| \Phi_{50}^t | -0.5816 | 0.2504 | -2.35 | Significant     |
| \Phi_{51}^t | 1.4074  | 0.2211 | 6.36  | Significant     |
| \Phi_{60}^t | -0.4240 | 0.2514 | -1.69 | Not significant |
| \Phi_{61}^t | 0.3754  | 0.2735 | 1.37  | Not significant |

Using significant parameters, GSTAR model can be represented as follow

\[
\begin{bmatrix}
z_i(t) \\
z_j(t) \\
z_k(t)
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 \\
0 & \Phi_{10}^t & 0 \\
0 & 0 & \Phi_{10}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-1) \\
z_j(t-1) \\
z_k(t-1)
\end{bmatrix} + \begin{bmatrix}
\Phi_{11}^t & 0 & 0 \\
0 & \Phi_{11}^t & 0 \\
0 & 0 & \Phi_{11}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-1) \\
z_j(t-1) \\
z_k(t-1)
\end{bmatrix} + \begin{bmatrix}
\Phi_{20}^t & 0 & 0 \\
0 & \Phi_{20}^t & 0 \\
0 & 0 & \Phi_{20}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-2) \\
z_j(t-2) \\
z_k(t-2)
\end{bmatrix} + \begin{bmatrix}
\Phi_{21}^t & 0 & 0 \\
0 & \Phi_{21}^t & 0 \\
0 & 0 & \Phi_{21}^t
\end{bmatrix} \begin{bmatrix}
z_i(t-2) \\
z_j(t-2) \\
z_k(t-2)
\end{bmatrix} + \begin{bmatrix}
e_1(t) \\
e_2(t) \\
e_3(t)
\end{bmatrix}
\]

(6)

Based on equation (6), in the same way, the value of significant parameters are estimated. Estimated significant parameters are represented in Table 3.
### Table 3. Estimated significant parameters of GSTAR($2_1$) model for inflation rate data

| Parameter | Coef  | SE Coef | $t$-value | Conclusion |
|-----------|-------|---------|-----------|------------|
| $\phi_{10}$ | 1.0400 | 0.1233 | 8.43      | Significant |
| $\phi_{20}$ | -0.9213 | 0.2363 | -3.90     | Significant |
| $\phi_{11}$ | 0.6644 | 0.2504 | 2.65      | Significant |
| $\phi_{01}$ | 0.7501 | 0.1158 | 6.48      | Significant |
| $\phi_{21}$ | -0.7875 | 0.2178 | -3.62     | Significant |
| $\phi_{12}$ | -0.5572 | 0.2430 | -2.29     | Significant |
| $\phi_{02}$ | 1.3849 | 0.2169 | 6.38      | Significant |

### 4.3 Diagnostic Check

After finding the significant model of GSTAR, the last step is checking the white noise and normality of the error [10]. MACF plot can be used to identify the assumption of white noise. MACF plot of the error is represented in Figure 2.

**Figure 2.** MACF plot of error inflation rate in 3 locations

Based on Figure 2, error of the GSTAR($2_1$) model using distance inverse space weight is suitable for white noise assumption. It is represented from the sign (.) that is more than sign (+) and (-) [10]. Whereas, we can use q-q plot to check the normality of the error. This plot is represented in Figure 3. Figure 3 shows that the plot tends to form into a straight line. We can conclude that error of the GSTAR($2_1$) model is normal multivariate distribution. Because of these error, white noise and normal multivariate distribution test are fulfilled, then GSTAR($2_1$) with distance inverse space weight is the best model for inflation rate in 3 locations.
4.4 Parameter Estimation of GSTAR(21) using Kalman Filter

After building the best GSTAR model, Kalman Filter is applied to estimate the GSTAR parameters. In this research, parameters estimated using Kalman Filter are GSTAR parameters that have been estimated using OLS and significant, i.e. \( \phi_{10}^1, \phi_{20}^1, \phi_{21}^1, \phi_{10}^2, \phi_{20}^2, \phi_{21}^2, \phi_{11}^1 \), and \( \phi_{21}^3 \). Equation (6) can be modified in to sistem model as follows:

\[
\begin{bmatrix}
  z_1(t) \\
  \Phi_1^1 \\
  \Phi_2^1 \\
  \Phi_3^1 \\
  \vdots \\
  z_n(t)
\end{bmatrix}_{t+1} =
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
  \vdots \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  \vdots 
\end{bmatrix}
\begin{bmatrix}
  z_1(t) \\
  \Phi_1^1 \\
  \Phi_2^1 \\
  \Phi_3^1 \\
  \vdots \\
  z_n(t)
\end{bmatrix}_t
\]

where

\[
\begin{align*}
  a_{12} &= 0.5148 z_2(t-1) + 0.4852 z_3(t-1) \\
  a_{13} &= z_3(t-2) \\
  a_{14} &= 0.5148 z_2(t-2) + 0.4852 z_3(t-2) \\
  a_{56} &= z_2(t-1) \\
  a_{57} &= z_3(t-2) \\
  a_{58} &= 0.5353 z_2(t-2) + 0.4647 z_3(t-2) \\
  a_{410} &= z_3(t-1) \\
  a_{911} &= 0.5205 z_1(t-1) + 0.4795 z_2(t-1)
\end{align*}
\]

Then using Measurement model of Kalman Filter that is declared as: \( e_t = H x_t + v_t \), It can be written as follows.
We can determine matrix $H$ in Equation (7) by considering the data that we have. Because we only have the data of $z_1(t), z_2(t),$ and $z_3(t)$, then matrix $H$ can be written as in Equation (7).

After modifying the model system and measurement model, then do initialization. Initial values for $Z_1, Z_2,$ and $Z_3$ are taken from the first data of the inflation rate that already stationer. Whereas the initial value of the variance of the noise given $Q = 0.3$ dan $R = 0.3$. For initial $\hat{x}_0$ and covariant values are given as follows:

$$\hat{x}_0 = [0.5787 \quad 1.1234 \quad -0.8787 \quad 0.7667 \quad 0.7394 \quad -0.7012 \quad 0.8099 \quad -0.8097 \quad 0.6925 \quad -0.6767 \quad 1.2348]^T.$$

The initial value $\hat{x}_0$ can be determined from the data that we have in the first period in each location. Beside that, we also determine the initial value $\hat{x}_0$ from predicted value of the parameter that we calculate using OLS method.

With $P_0 = I_{11 \times 11}$, and $Q_z = I_{11 \times 11} \times Q$, then for Prediction Stage are given from these equation:

$$\hat{x}_{r+1} = A_r x + B_r u_r$$

$$P_{r+1} = A_r P_r A_r^T + G_r Q G_r^T$$

After do Prediction Stage, we do Measurement update. Measurement update is also called Correction Stage. In this stage, given Kalman Gain as this equation:

$$K_{r+1} = P_{r+1} M_{r+1}^T (M_{r+1} P_{r+1} M_{r+1}^T + R_{r+1})^{-1}$$

Then, $\hat{x}_{r+1}$ value is estimated using $\hat{x}_{r+1}$ value from Prediction Stage, then we get:

$$\hat{x}_{r+1} = \hat{x}_{r+1} + K_{r+1} (Z_{r+1} - M_{r+1} \hat{x}_{r+1})$$

$P_{r+1}$ value can be estimated using $P_{r+1}$ value from Prediction Stage, then it can be written:

$$P_{r+1} = (I - K_{r+1} H_{r+1}) P_{r+1}$$

Parameters estimation using Kalman Filter of the inflation rate in 3 locations are represented in Table 4. Table 4 shows the value of each parameters of GSTAR($2_1$) that is estimated using Kalman Filter. Then these values are substituted in to Equation (6). The equation using parameters value from Table 4 is used to compare the forecasting result with parameters that is estimated using OLS.

5. **Comparison Prediction The Rate Inflation Using GSTAR and GSTAR-Kalman Filter**

Comparison between GSTAR and GSTAR-Kalman Filter model is used to see the efficiency of these models. Directly we can compare and see the most efficiency models from the graphic that we can see in Figure 4. But we can fix it mathematically using RMSE value. These models are compared using RMSE value from out-sample data. The most efficiency model is the model with the smallest RMSE. RMSE can be calculated using this equation [9]:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2}$$

where $y_t$ is the true value and $\hat{y}_t$ is the predicted value.
\[
RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (Y_t - \hat{Y}_t)^2}
\]  
(8)

**Table 4.** Parameter estimation of GSTAR model using Kalman Filter

| Model       | Parameter | Coefficient |
|-------------|-----------|-------------|
| GSTAR(2)    | \(\Phi_{11}^1\) | 0.9229      |
|            | \(\Phi_{20}^1\) | -1.0012     |
|            | \(\Phi_{21}^1\) | 0.7623      |
|            | \(\Phi_{10}^2\) | 0.9072      |
|            | \(\Phi_{20}^2\) | -0.2844     |
|            | \(\Phi_{21}^2\) | -0.2849     |
|            | \(\Phi_{10}^3\) | -0.5984     |
|            | \(\Phi_{11}^3\) | 1.4200      |

From Figure 4, we can see that prediction data using GSTAR-Kalman Filter is closer with factual data than prediction data using GSTAR. This is indicated by the GSTAR-Kalman Filter graph that is more keep up the factual data graph than GSTAR graph. To determine the most efficiency model in forecasting the inflation rate, we use RMSE value. Based on Equation (8), the result of RMSE calculation of this research can be represented in Table 5.

**Table 5.** Nilai RMSE *out-sample* di setiap lokasi

| Lokasi     | GSTAR-OLS | GSTAR-Kalman Filter |
|------------|-----------|---------------------|
| Malang     | 0.0661    | 0.0132              |
| Probolinggo| 0.0351    | 0.0132              |
| Surabaya   | 0.0597    | 0.0132              |

From Table 5, RMSE of GSTAR-Kalman Filter have same value in each location. And also, RMSE of GSTAR-Kalman Filter is smaller than GSTAR in each locations. Then we can conclude that GSTAR-Kalman Filter is more efficiency in forecasting the rate inflation than GSTAR in each locations.
Figure 4. Comparison between data actual and prediction data of inflation rate in 3 locations.
6. Conclusion

Based on the results of this research, can be conclude that the best model of inflation rate in Malang, Probolinggo, and Surabaya is GSTAR\( (2) \). The results show that RMSE value in each location of GSTAR-Kalman Filter is smaller than GSTAR that was estimated using OLS. It means that parameter estimation of GSTAR model using distance inverse space weight can be done optimally by using Kalman Filter.

References

[1] Artikel Ekonomi. (2015). “Artikelsiana, Artikel Belajar dan Bermanfaat”. Dalam http://www.artikelsiana.com/2015/02/pengertian-inflasi-jenis-dampak-penyebab.html?m=1.

[2] Kurniawati. (2016). Perbandingan Penerapan Model GSTAR dengan Pembobot Inverse Jarak dan Normalisasi Korelasi Silang pada Laju Inflasi Kota Surakarta, Yogyakarta, dan Surabaya. Tugas Akhir. Universitas Sebelas Maret, Surakarta.

[3] G. Welch, & G.Bishop. (2006). An Introduction to the Filter Kalman. Chapel Hill: Department of Computer Science, University of North Carolina.

[4] I.F. Hamsyah. (2015). “Perbandingan GSTAR dan ARIMA Filter Kalman dalam Perbaikan Hasil Prediksi Debit Air Sungai Brantas”. Tugas Akhir. Departemen Matematika. Institut Teknologi Sepuluh Nopember, Surabaya.

[5] M. H. Pamungkas. (2015). “Estimasi Parameter Model ARIMA Menggunakan Kalman Filter untuk Peramalan Permintaan Darah (Studi Kasus: UTD PMI Surabaya)”. Tugas Akhir. Departemen Matematika. Institut Teknologi Sepuluh Nopember, Surabaya.

[6] P. Febritasari. (2016). “Estimasi Inflasi Wilayah Kerja KPwBI Malang Menggunakan ARIMA-Filter Kalman dan VAR-Filter Kalman”. Tugas Akhir. Departemen Matematika. Institut Teknologi Sepuluh Nopember, Surabaya.

[7] D. U. Wutsqa, Suhartono, & B. Sutijo. (2012). “Aplikasi Model Generalized Space Time Autoregressive Pada Data Pencemaran Udara di Kota Surabaya”. Jounal Pythagoras, Vol.7, No.2.

[8] Mansoer, A. Shaliha. (2016). "Pemodelan Seasonal Generalized Space Time Autoregressive (SGSTAR) (Studi Kasus: produksi Padi di Kabupaten Demak, Kabupaten Boyolali, dan Kabupaten Grobogan)". Jurnal Gaussian, 593-6-2.

[9] Makridakis, McGee, dan W. Wheelright. (1999). Metode dan Aplikasi Peramalan. Edisi kedua. Terj. Andriyanto, U.S. Jakarta: Bina Rupa Aksara.

[10] B.J. Cromwell, et al. (1994). Multivariate Test For Time Series Models. United State of America: Sage Publication, Inc