The Optimal Determination of Space Weight in GSTAR Model by using Cross-correlation Inference

SUHARTONO[1], SUBANAR[2]

[1]Statistics Department, Institut Teknologi Sepuluh Nopember, Indonesia
PhD Student, Mathematics Department, Gadjah Mada University, Indonesia
suhartono@statistika.its.ac.id

[2]Mathematics Department
Gadjah Mada University, Indonesia
subanar@yahoo.com

Abstract. The aim of this paper is to discuss and develop the optimal determination of space weight in GSTAR (Generalized Space-Time Autoregressive) model by applying statistical inference of cross-correlation between locations (spaces) at the appropriate time lag. Our previous research showed that the directly used of cross-correlation normalization as space weight give improper coefficient between locations in GSTAR model; i.e. these coefficients tend to be significant even though the true condition is insignificant. In this paper, we propose a statistical test to validate the cross-correlation between locations that used as basic of space weight determination in GSTAR model. We focus on the GSTAR(1_1) model and use three kinds relationship between locations as case studies. The results show that statistical inference process to validate cross-correlation between locations yields valid (unbiased) space weight estimates in GSTAR(1_1) model. In general, we can conclude that determination of space weight by using normalization of statistical inference to the cross-correlation between locations at the appropriate time lag is the optimal procedure in GSTAR modeling.

2000 Mathematics Subject Classification: 62M45, Secondary 62M02

Key words and phrases: GSTAR(1_1), space weight, statistical inference, normalization, cross-correlation

1. Introduction

In daily life, we frequently deal with the data that depend not only on time (with past observations) but also depend on site or space, called spatial data. Space-time model is a model that combines time and space dependence which is happened in a certain multivariate time series data. This model firstly proposed by Pfeifer and Deutsch (see [**5, 6**]).

GSTAR model is a tool that usually used for modeling and forecasting space-time series data. This model is an extension of STAR model proposed by Pfeifer and Deutsch. In practical problems, GSTAR model is frequently applied to geology and ecology [**4**]. The other model that also can be used for modeling space-time series data is VAR (Vector Autoregressive) model [**7, 8**].

Determination of space weight is one of the main problems in GSTAR model. This paper discusses the used of space weight based on the statistical inference to the cross-correlation between location at the appropriate time lag.

2. GSTAR (*Generalized Space-Time Autoregressive*) Model

GSTAR model is a more flexible model as a result of STAR model generalization. Mathematically, the notation of GSTAR(p_1) model is the same as STAR(p_1) model. The main difference is the parameters of GSTAR(p_1) model at the same space must not equal. In matrix notation, GSTAR(p_1) model could be written as (see [**1**])

$$Z(t) = \sum_{k=1}^{p} [\Phi_{k0} + \Phi_{k1} W] Z(t-k) + e(t) \tag{2.1}$$

where

- $\Phi_{k0} = \text{diag}(\phi_{k0}^{1}, \ldots, \phi_{k0}^{N})$ and $\Phi_{k1} = \text{diag}(\phi_{k1}^{1}, \ldots, \phi_{k1}^{N})$,
- weights are choosen to satisfy $w_{ii} = 0$ and $\sum_{i \neq j} w_{ij} = 1$.

For instance, GSTAR(1_1) model represent oil production at three locations can be written as

$$Z(t) = [\Phi_{10} + \Phi_{11} W] Z(t-1) + e(t) \tag{2.2}$$

where

$$Z(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix}, \ \Phi_{10} = \begin{pmatrix} \phi_{10} & 0 & 0 \\ 0 & \phi_{20} & 0 \\ 0 & 0 & \phi_{30} \end{pmatrix}, \ \Phi_{11} = \begin{pmatrix} \phi_{11} & 0 & 0 \\ 0 & \phi_{21} & 0 \\ 0 & 0 & \phi_{31} \end{pmatrix},$$

$$W = \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix}, \ Z(t-1) = \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix}, \text{ and } e(t) = \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}.$$

Parameter estimation of GSTAR model can be done by using Least Square Method. The theory and methodology about parameter estimation of GSTAR model can be read extensively in [**1**] and [**3**].

Selection or determination of space weight is one of the main problems at GSTAR modeling. Some methods for determining space weight have been proposed to the application of GSTAR model, i.e. (see [**1, 3, 9**])

(i) Uniform weight, i.e. $w_{ij} = \frac{1}{n_i}$, where n_i number of spaces or locations where are located near to location i,
(ii) Binary weight, i.e. $w_{ij} = 0$ or 1, depends on certain constraint,
(iii) Inverse of distance,
(iv) Weight based on semi-variogram or covariogram of variable between locations, and
(v) Weight based on the normalization of cross-correlation between locations at the appropriate time lag. Method (iv) and (v) give negative value possibility to space weight.

3. The used of statistical inference to the cross-correlation for determining space weight GSTAR(1_1) model

Determination of space weight by using the normalization result of cross-correlation between locations at the appropriate time lag is firstly proposed by Suhartono and Atok (see [**9**]). In general, cross-correlation between two variables or location i and j at the time lag k, $\text{corr}[Z_i(t), Z_j(t-k)]$, defined as (see [**2, 10**])

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{\sigma_i \sigma_j}, k = 0, \pm 1, \pm 2, \ldots \tag{3.1}$$

where $\gamma_{ij}(k)$ is cross-covariance between observation in location i and j at the time lag k, σ_i and σ_j is standard deviation of observation in location i and j. The estimated of cross-correlation in sample data is

$$r_{ij}(k) = \frac{\sum_{t=k+1}^{n}[Z_i(t) - \bar{Z}_i][Z_j(t-k) - \bar{Z}_j]}{\sqrt{(\sum_{t=1}^{n}[Z_i(t) - \bar{Z}_i])^2 (\sum_{t=1}^{n}[Z_j(t) - \bar{Z}_j])^2}}. \tag{3.2}$$

Bartlett (1955) has derived variance and covariance of cross-correlation estimated from sample data (see [**10**]). Under hypothesis that two time series data Z_i and Z_j are uncorrelated, Bartlett showed that

$$Variance[r_{ij}(k)] \cong \frac{1}{n-k}[1 + 2\sum_{s=1}^{\infty} \rho_{ii}(s)\rho_{jj}(s)]. \tag{3.3}$$

Hence, for Z_i and Z_j are white noise series, we have

$$Variance[r_{ij}(k)] \cong \frac{1}{n-k}. \tag{3.4}$$

For large sample size, $(n-k)$ in equation (3.4) frequently replaced by n. Under assumption of normal distribution, the cross-correlation estimated from sample can be tested whether significant different from zero. In this paper, testing hypothesis or statistical inference is done by using interval confidence, i.e.

$$r_{ij}(k) \pm [t_{\alpha/2; df=n-k-2} \frac{1}{\sqrt{n}}]. \tag{3.5}$$

Then, determination of space weight could be done by normalization of the statistical inference to the cross-correlation between locations at the appropriate time lag. This process generally yields space weight for GSTAR(1_1) model, i.e.

$$w_{ij} = \frac{r_{ij}(1)}{\sum_{k \neq i} |r_{ik}(1)|}, \tag{3.6}$$

where $i \neq j$, and satisfies $\sum_{j \neq i} |w_{ij}| = 1$.

Space weights by using the normalization of statistical inference to the cross-correlation between locations at the appropriate time lag give all form possibilities of the relationship between locations. Hence, there is no strict constraint about the weight values, i.e. it must depend on distance between locations. This weight also gives flexibility on the sign and size of the relationship between locations.

4. Implementation of space weight determination based on the normalization of the statistical inference to the cross-correlation for GSTAR(1_1) model

This section gives the results of simulation study of the statistical inference application to the cross-correlation between locations for determining

TABLE 1. The result of cross-correlation between locations and their confidence interval for simulation data at case 1

Parameter	Coefficient estimated	95 percent Lower bound	95 percent Upper bound	Conclusion
$r_{12}(1)$	0.245912	0.132562	0.359262	Valid and
$r_{13}(1)$	0.245017	0.131667	0.358367	concurrent
$r_{21}(1)$	0.249190	0.135840	0.362540	Valid and
$r_{23}(1)$	0.176879	0.063529	0.290229	concurrent
$r_{31}(1)$	0.179549	0.066199	0.292899	Valid and
$r_{32}(1)$	0.270282	0.156932	0.383632	concurrent

space weight at GSTAR(1_1) model. As in Suhartono and Atok [**9**], there are three cases that relate to the size and sign of relationship coefficient; i.e. (1) same, (2) different size, but the same sign, and (3) different signs. In this simulation study, the GSTAR(1_1) is generated as follows

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} \phi^*_{11} & \phi^*_{12} & \phi^*_{13} \\ \phi^*_{21} & \phi^*_{22} & \phi^*_{23} \\ \phi^*_{31} & \phi^*_{32} & \phi^*_{33} \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}, \tag{4.1}$$

where $\phi^*_{ii} = \phi_{i0}$, and $\phi^*_{ij} = w_{ij}\phi_{i1}$ for $i \neq j$.

4.1. Case 1. In this section, we give an example of GSTAR(1_1) model with coefficient parameters between locations are equal, i.e.

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.25 & 0.2 & 0.2 \\ 0.15 & 0.2 & 0.15 \\ 0.15 & 0.15 & 0.2 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}, \tag{4.2}$$

where $e_i(t)$ is white noise vector with mean 0 and variance 0.25. The simulation is done for sample size 300.

The result of cross-correlation between locations at the time lag 1, $r_{ij}(1)$ where $i \neq j$, and their 95 percent confidence interval can be seen in Table 1. This statistical inference result shows that cross-correlation between locations are valid and concurrent. It means the magnitude of correlation between location 2, 3 at time $(t-1)$ and location 1 at time t are equal. Its condition also happened to cross-correlation between other locations. Thus, we can use uniform weight, i.e.

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}. \tag{4.3}$$

This result explains that space weight based on statistical inference is valid. It's caused the result of space weight is the same as the postulated weight. By using this weight, we yield the parameter estimates of GSTAR(1_1) model as shown in Table 2.

From table 2, we can see clearly that all parameter estimates of GSTAR(1_1) model are significant different from zero. By applying matrix operation, i.e adding all coefficients at GSTAR(1_1) model, we have

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.2455 & 0.1776 & 0.1776 \\ 0.1744 & 0.2082 & 0.1744 \\ 0.1702 & 0.1702 & 0.2003 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}. \tag{4.4}$$

TABLE 2. The result of parameter estimates GSTAR(1_1) model by using space weight of cross-correlation inference normalization at case 1

Parameter	Coefficient estimated	Standard Error	t-value	p-value
ϕ_{10}	0.24545	0.05568	4.41	0.000
ϕ_{20}	0.20823	0.05458	3.82	0.000
ϕ_{30}	0.20028	0.05401	3.71	0.000
ϕ_{11}	0.35515	0.06991	5.08	0.000
ϕ_{21}	0.34485	0.07814	4.41	0.000
ϕ_{31}	0.34045	0.07028	4.84	0.000

TABLE 3. The result of cross-correlation between locations and their confidence interval for simulation data at case 2

Parameter	Coefficient estimated	95 percent Lower bound	95 percent Upper bound	Conclusion
$r_{12}(1)$	0.222863	0.109513	0.336213	Valid
$r_{13}(1)$	0.016784	-0.096566	0.130134	Invalid
$r_{21}(1)$	0.196791	0.083441	0.310141	Valid
$r_{23}(1)$	0.351704	0.238354	0.465054	Valid
$r_{31}(1)$	0.312338	0.198988	0.425688	Valid
$r_{32}(1)$	0.026139	-0.087211	0.139489	Invalid

This final model has relatively equal parameter coefficients to the model in equation (4.2), both size and sign.

4.2. Case 2. In this section, we give a brief result of GSTAR(1_1) model with coefficient parameters between locations are different size but the same sign, i.e.

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.25 & 0.2 & 0 \\ 0.15 & 0.2 & 0.3 \\ 0.25 & 0 & 0.25 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}, \tag{4.5}$$

where $e_i(t)$ is a white noise vector as in case 1.

The cross-correlation between locations at the time lag 1 and their 95 percent confidence interval can be seen in Table 3. We can see clearly that cross-correlations between location 2 and 1, 1 and 2, 3 and 2, also location 1 and 3, are statistically significant. This condition is the same as the postulated model in equation (4.5).

Based on this result, we can use space weights between location 2 and 1, 3 and 1, are respectively 1 and 0, as binary weight. The space weights between location 1 and 2, 3 and 2, are respectively 1/3 and 2/3, and between location 1 and 3, 2 and 3, are respectively 1 and 0. Thus, the completely given space weights are

$$W = \begin{pmatrix} 0 & 1 & 0 \\ 0.33 & 0 & 0.67 \\ 1 & 0 & 0 \end{pmatrix}. \tag{4.6}$$

TABLE 4. The result of parameter estimates GSTAR(1_1) model by using space weight of cross-correlation inference normalization at case 2

Parameter	Coefficient estimated	Standard Error	t-value	p-value
ϕ_{10}	0.25133	0.05310	4.73	0.000
ϕ_{20}	0.17003	0.05428	3.13	0.000
ϕ_{30}	0.23893	0.05359	4.46	0.000
ϕ_{11}	0.21116	0.05364	3.94	0.000
ϕ_{21}	0.50468	0.06896	7.32	0.000
ϕ_{31}	0.29430	0.05309	5.54	0.000

TABLE 5. The result of cross-correlation between locations and their confidence interval for simulation data at case 3

Parameter	Coefficient estimated	95 percent Lower bound	95 percent Upper bound	Conclusion
$r_{12}(1)$	0.141557	0.028207	0.254907	Valid and
$r_{13}(1)$	-0.207770	-0.321120	-0.094420	different sign
$r_{21}(1)$	-0.220560	-0.333910	-0.107210	Valid and
$r_{23}(1)$	0.120653	0.007303	0.234003	different sign
$r_{31}(1)$	0.224607	0.111257	0.337957	Valid and
$r_{32}(1)$	-0.251830	-0.365180	-0.138480	different sign

This result shows that space weight based on statistical inference is valid, because it equal to the postulated weight. Then, we use this wight and yield the parameter estimates of GSTAR(1_1) model as shown in Table 4.

Table 4 shows that all parameter estimates of GSTAR(1_1) model are significant different from zero. By adding all coefficients at GSTAR(1_1) model, we have

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.251 & 0.211 & 0 \\ 0.168 & 0.170 & 0.336 \\ 0.294 & 0 & 0.239 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}. \tag{4.7}$$

This final model has equal sign and relatively similar size of parameter coefficients with the model in equation (4.5).

4.3. Case 3. In this section, we provide a brief result of GSTAR(1_1) model with coefficient parameters between locations are the same size but different sign, i.e.

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.25 & 0.2 & -0.2 \\ -0.15 & 0.2 & 0.15 \\ 0.15 & -0.15 & 0.25 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}, \tag{4.8}$$

where $e_i(t)$ is a white noise vector as in case 1.

Table 5 illustrate the result of cross-correlation between locations at the time lag 1 and their confidence interval. We can observe clearly that all cross-correlations between locations are statistically significant. Again, this condition is the same as the postulated model in equation (4.8).

TABLE 6. The result of parameter estimates GSTAR(1_1) model by using space weight of cross-correlation inference normalization at case 3

Parameter	Coefficient estimated	Standard Error	t-value	p-value
ϕ_{10}	0.29061	0.05240	5.55	0.000
ϕ_{20}	0.19837	0.05537	3.58	0.000
ϕ_{30}	0.22049	0.05483	4.02	0.000
ϕ_{11}	0.35136	0.07736	4.54	0.000
ϕ_{21}	0.30067	0.07307	4.11	0.000
ϕ_{31}	0.44502	0.07313	6.09	0.000

Based on the result in Table 5, we can use uniform space weights with different sign, i.e.

$$W = \begin{pmatrix} 0 & 0.5 & -0.5 \\ -0.5 & 0 & 0.5 \\ 0.5 & -0.5 & 0 \end{pmatrix}. \tag{4.9}$$

This space weight based on statistical inference is valid, because it equal to the postulated weight. We implement this weight and yield the parameter estimates of GSTAR(1_1) model as seen at Table 6.

By applying matrix operation to all coefficients at GSTAR(1_1) model, we get

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \begin{pmatrix} 0.29 & 0.18 & -0.18 \\ -0.15 & 0.20 & 0.15 \\ 0.22 & -0.22 & 0.22 \end{pmatrix} \begin{pmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \end{pmatrix} + \begin{pmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{pmatrix}. \tag{4.10}$$

This final model has equal sign and relatively similar size of parameter coefficients with the model in equation (4.8). This result shows that the final model is an unbias model estimate.

5. Conclusion

Based on the results at the previous section, it can be concluded that space weight determination at GSTAR model can be done optimally by using normalization of statistical inference to the cross-correlation between locations at the appropriate time lag. Additionally, the results also show that space weight determination by using this method covers uniform and binary space weights.

For further research, it is important to study further about the relationship between statistical inference at the parameters GSTAR model and the statistical inference on the space weights.

References

[1] S. A. Borovkova, H. P. Lopuhaa and B. N. Ruchjana, Generalized STAR model with experimental weights. In M. Stasinopoulos and G. Touloumi (Eds.), *Proceedings of the 17th International Workshop on Statistical Modeling,* Chania, (2002), pp. 139-147.

[2] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd edition, Englewood Cliffs: Prentice Hall.

[3] B. N. Ruchjana, Pemodelan Kurva Produksi Minyak Bumi Menggunakan Model Generalisasi S-TAR, *Forum Statistika dan Komputasi,* IPB, Bogor, 2002.

[4] B. N. Ruchjana, The Stationary Conditions of The Generalized Space-Time Autoregressive Model, *Proceeding of the SEAMS-GMU Conference,* Gadjah Mada University, Yogyakarta, 2003.

[5] P. E. Pfeifer and S. J. Deutsch, A Three Stage Iterative Procedure for Space-Time Modeling, *Technometrics*, Vol. **22**, No. 1 (1980a), 35–47.

[6] P. E. Pfeifer and S. J. Deutsch, Identification and Interpretation of First Order Space-Time ARMA Models, *Technometrics*, Vol. **22**, No. 1 (1980b), 397–408.

[7] Suhartono, Evaluasi pembentukan model VARIMA dan STAR untuk peramalan data deret waktu dan lokasi, Presented at *Workshop and National Seminar on Space Time Models and Its Application*, UNPAD, Bandung, 2005.

[8] Suhartono, Perbandingan antara model VARIMA dan GSTAR untuk peramalan data deret waktu dan lokasi, *Prosiding Seminar Nasional Statistika*, ITS, Surabaya, 2006.

[9] Suhartono dan R. M. Atok, Pemilihan bobot lokasi yang optimal pada model GSTAR, Presented at *National Mathematics Conference XIII,* Universitas Negeri Semarang, 2006.

[10] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods,* Addison-Wesley Publishing Co., USA, 1990.