SEEMINGLY UNRELATED REGRESSION (SUR) ESTIMATION OF GSTARX(1,1) MODEL USING R

Soekardi Hadi Prabowo[1]
Lisana Sumarah Pratignyo[2]
[1,2]Mathematics Department, Faculty of Science and Technology
University of Islamic As-Syafi'iyah
Email:sh.prabowo59@gmail.com

ABSTRACT

In this paper we studied SUR estimation parameter of GSTARX(1,1) Model, which used overcome the weakness Ordinary Least Square (OLS) ignoring the information Errors between the equations of each location are correlated so that there is a problem Contemporaneous correlation, which willresult in inefficient estimators SUR. We built the script using R software to estimate the parameter using seemingly unrelated regression Method. This case study to apply the GSTARX model to Contaminant Load of BOD With Exogenous Variable of COD in Watershed Citarum River is significance for the result test hypothesis simultaneous model.

Keyword : *OlS, GSTARX(1,1), Contemporaneous, SUR*

1. INTRODUCTION

GSTARX(1,1) is the development from the GSTAR(1,1) model with the addition of one exogenous variable (Astuti, D, 2015)).While the GSTAR(1,1) model is one of the models used to overcome time series data that have interrelationships over time, it was first introduced by Ruchjana, B.N (2002) by reversing the weakness of the STAR model proposed by pfeifer and Deutch (1980), by not requiring space-time parameter And a homogeneous autoregressive, but location parameters that have heterogeneous characteristics. And determination of location weights using uniform weightsThird of the models Studied above of is a models parameter estimation is on a still limited using is an ordinary least square (OLS) method. The method is an assumes that error model certain location has no correlation with an error the other location. But, OLS method can't be used if the occur correlation an error between location, it will produce an estimator in inefficient, in the sense that the resulting variance-covariance matrix of error has no constant.Its parameter estimation method is used overcome the weakness of OLS. SUR method to introduced by Zellner (1962). Its method can accommodate correlations between different of location , for the increasing of estimator parameter. So that SUR is a develop Generalized Least Square (GLS) estimation method, which only involved components of variance error in the model parameter estimation process. While SUR in addition to involving the variance error also adds a matrix of Identity with size T X T through kronecker product. We are used R software to estimate its parameter and apply the model toward real phenomena of data, such as a Biological Oxygen demand (BOD) by involving one exogenous variable of chemical oxygen demand (COD), on three location point observe in the watershed of Citarum. ie Inlet Jatiluhur, outlet Jatiluhur and RengasDengklok.

2. TOERETICAL STUDY

In this section we describe the three steps of Box-Jenkins method (1994), that is used for statistical estimation.

2.1 Identification

Otherwise of data the assumption is fulfilled with plot Autocorrelation function (ACF) and formal test Augmented dickey fuller (ADF) is significance. Because This research focuses the study develop of a model GSTARX with time and spatial lag one for three location and can be used estimate of SUR method, hereinafter written by GSTARX(1,1)-SUR.The model we write

$$Z(t) = \Phi_{10} + \Phi_{11}W^{(l)}]\, Z(t-1) + \gamma X(t) + e(t) \quad (1)$$

Where Z(t), Z(t-1), X(t) and e(t) the vectors by size $((3\times(T-1))\times 1)$, while Φ, , W and γ thr matrix by size (3× 3) (Dewi, A, 2015). Thus equation (1) can be written in matrix form as follows

$$\begin{bmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \end{bmatrix} = \left\{ \begin{matrix} \Phi_{10} & 0 & 0 \\ 0 & \Phi_{20} & 0 \\ 0 & 0 & \Phi_{30} \end{matrix} \right\} \begin{bmatrix} \Phi_{11} & 0 & 0 \\ 0 & \Phi_{12} & 0 \\ 0 & 0 & \Phi_{13} \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t=1) \\ Z_3(t-1) \end{bmatrix} + \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_3(t) \\ e_3(t) \end{bmatrix} \quad (2)$$

2.2 Parameter estimate of GSTARX(1,1)-SUR Model

Parameter estimation which is the second step of box – jenkins method, can be done if the equation (1) has presented like in linear regression model, by using matrix representation, then could be written equation as follows

$$Y = X\beta + \varepsilon \quad (3)$$

Where

$Y'_{((3\times(T-1))\times 1)}$ =Z'(t) $[Z_1(t)Z_2(t)Z_3(t)]$

$X_{it} = [Z_i(t-1)V_i(t-1)X_i(t)]_{((3\times(T-1))\times(3\times 3))}$, for i = 1, 2, 3

$\beta'_{((3\times(T-1))\times 1)} = [\Phi_{i0}\Phi_{i1}\gamma_i]$, I = 1, 2 , 3

$\varepsilon''_{((3\times(T-1))\times 1)} = [\varepsilon_1(t)\varepsilon_2(t)\varepsilon_3(t)]$, vector of error with assuming that $\varepsilon \sim IID\ N(0, \sigma^2 I_T)$

$\varepsilon''_{((3\times(T-1))\times 1)} = [\varepsilon_1(t)\varepsilon_2(t)\varepsilon_3(t)]$Vektor error dengan asumsi$\varepsilon \sim IID\ N(0, \sigma^2 I_T)$

The next could be obtained estimator parameter by OLS method, through the way of minimized sum square of error (ε'ε), so that would be obtained estimating

$$\hat{\beta}^{SOLS} = (X'X)^{-1}X'Y \quad (4)$$

Estimate the parameters β by OLS on the equation (5) will be efficient, if the error between equation is not correlation. But it is called inefficient if the errors for different location are correlated with to display of contemporaneous correlation. These things result implies the variance- covariance matrix error of result OLS estimation becoming inconsistent. SUR is an estimation method which could overcome the problem of correlation between errors in different location.

Therefore, before applicated the SUR model on equation (3), fFirst, the detection of contemporaneous correlation problems, certain for size correlate between errors of N = 3 equation different on the time of the same (Dufour Jean M & Lynda Khalaf, 2000). These correlations are those exploited by zellner's SUR estimation (1962). The contemporaneous correlation can be used with Lagrange multiplier test ((λ_{LM}) by means of formula (Dufour et.al, 2000)

$$\lambda_{LM} = T \sum_{i=}^{N=3} \sum_{j=1}^{i-1} r_{ij}^2 \ , \quad \text{for } T = n-1 \tag{5}$$

where n size of observation,r_{ij}^2 is the ij^{th}error correlation coefficient and criterion of test can be done by means of comparison betweenλ_{LM} value with χ^2_{Tabel} for level of significance α and degree of freedom d, with $d = \frac{N\times(N-1)}{2}$, for N =3. So that d is 6. If value of λ_{LM} then conclusion there is problem of contemporaneous correlation in model.Its this means the condition of validity application SUR method on the estimation parameters model is fulfilled. The estimate parameters of GSTARX(1,1)-SUR model, so that would be obtained estimator

$$\hat{\beta}^{SUR} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y \tag{6}$$

where $\Omega = \sum \otimes J_{(T\times\times T)}$, $\sum$ = variance-covariance matrix of size $(N \times N)$, ie for N = 3,there is

$$\sum = \begin{bmatrix} \sigma_{11} & \alpha_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \tag{7}$$

and $J_{(T\times\times T)}$is identity matrix of size $(T \times T)$, $T = n-1$. While $\otimes$ is notation Kronecker product, with one of the proferties for the inverse of two matrix A and D is define$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$. So for the formula$\Omega^{-1} = \left(\sum \otimes J_{(T\times\times T)}\right)^{-1}$and the result it is

$$\Omega^{-1} = \Sigma^{-1} \otimes J_{(T\times\times T)} \tag{8}$$

By using notation Kronecker product representation for example with matrix identity of size (2×2), then the result

$$\Omega = \begin{bmatrix} \sigma_{11} & \alpha_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}_{(3\times 3)} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{(2\times 2)} = \begin{bmatrix} \sigma_{11}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \alpha_{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sigma_{13}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \sigma_{21}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sigma_{22}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sigma_{23}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \sigma_{31}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sigma_{32}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sigma_{33}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix} \tag{9}$$

$$\Omega = \begin{bmatrix} \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{11} \end{bmatrix} & \begin{bmatrix} \alpha_{12} & 0 \\ 0 & \alpha_{12} \end{bmatrix} & \begin{bmatrix} \sigma_{13} & 0 \\ 0 & \sigma_{13} \end{bmatrix} \\ \begin{bmatrix} \sigma_{21} & 0 \\ 0 & \sigma_{21} \end{bmatrix} & \begin{bmatrix} \sigma_{22} & 0 \\ 0 & \sigma_{22} \end{bmatrix} & \begin{bmatrix} \sigma_{23} & 0 \\ 0 & \sigma_{23} \end{bmatrix} \\ \begin{bmatrix} \sigma_{31} & 0 \\ 0 & \sigma_{31} \end{bmatrix} & \begin{bmatrix} \sigma_{32} & 0 \\ 0 & \sigma_{32} \end{bmatrix} & \begin{bmatrix} \sigma_{32} & 0 \\ 0 & \sigma_{32} \end{bmatrix} \end{bmatrix}_{(6\times 6)} \tag{10}$$

So that, substituting of equation (9), then the form of equation (7) can be changed the form to

$$\hat{\beta}^{SUR} = \left(X'\left(\Sigma^{-1} \otimes J_{(T\times\times T)}\right)X\right)^{-1} X'\left(\Sigma^{-1} \otimes J_{(T\times\times T)}\right)Y \tag{11}$$

when the covariance matrix $\sum$ is unknown, a feasible SUR estimator is defined by replacing a widely used estimate 0f $\sum$ is

$$\Sigma = \left(\hat{\sigma}_{ij}\right) \tag{12}$$

Where $\hat{\sigma}_{ij} = \frac{1}{T} e_t' e_t$ and e_t is the OLS errors of the k^{th} equation (Moon Hyungsik R &Benoit P, 2006)., in this by manss k = 3.

Otherwise the SUR estimator is a three step, The first compute $\hat{\beta}^{OLS}$ in used to obtain error e_t and an estimator of$\sum$, more than can be done detection problem of contemporaneous on the second step, so that the three step compute $\hat{\beta}^{SUR}$ based on the estimator $\sum$ in the first step.

2.3 Check Diagnostic model

On the third step of box-jenkins method can be used check diagnostic, with prepared to known as the estimate model which the result is fulfilled goodness of fit. In this step as in significance parameter model, detection assumption multivariate error white noise and normality multivariate test of error. But in this paper we only can be detection significance parameter model. The first detection, is simultaneous significance test of parameterwith canbeusedstatistic

$$F_{value} = \frac{\frac{sum\ square\ of\ regression}{k}}{\frac{sum\ square\ of\ error}{n-k-1}} = \frac{\frac{sSSR}{k}}{\frac{SSE}{n-k-1}} \tag{13}$$

With the criterion of test, significance simultaneous parameters is significance if F_{value} more than or equal to F_{Table}, for level of significance α and degree of freedom numerator = k, de-numerator = n-k-1, but if its less than is not significance.The second, test of individual significance parameter, with can be used statistic

$$t_i = \frac{\hat{\beta}}{S_e c_{ii}} \tag{14}$$

With $S_e = \frac{e_t' e_t}{n-k}$ and c_{ii} element ofinverse variance-covariance matrix errorof model. It's the criterion test, individual parameters is significance if t_i more than or equal to t_{Table}, for level of significance α and degree of freedom (df) = n-k-1, but if its less than, so that one of individual parameteris not significance.

3. RESEARCH METHODS

This research is research that applies the Generalized Space Time Autoregressive (GSTAR) model involving one exogenous variable X (GSTARX), to obtain parameter estimates using the SUR method. Which was built based on R Software, in secondary data the BOD contamination load involved one exogenous COD variable at three monitoring point locations in the Lower Citarum River for 55 periods during the time interval from January 2017 to July 2021, using uniform weighting.

The study of Model Parameter Estimation as the second stage of the four stages of the Space Time model which adopts the BOX-Jenkins time series model in this paper is the result of further research from the research the author conducted and the author presented the results at the 2018 National Seminar on Mathematics and its Applications with the title Identification GSTARX-(P,q) model with data processing using R Software (BOD Pollution Load Study at three Citarum Watershed Monitoring Point Locations.

The type of data used in this research is secondary data on BOD and COD pollution loads from three locations along the lower part of the Citarum River. The location weights used in this research are weights

4. RESULT AND DISCUSSION

The data used to estimate the GSTARX model is secondary data from www.jasatirta2.go.id, at BOD contaminant load (Z(t)) with involving exogenous variable of COD (X(t)), for 55 monthly in three location point observe in watershed of CitarumRiver. But can be used for the modeling CSTAX(1,1)-SUR is only 44 monthly which mention in sample data, part of our programming is following command in R software for estimate step parameter for spatial weight matrix of uniform

$$W = \begin{bmatrix} 0 & 0,5 & 0.5 \\ 0,5 & 0 & 0,5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

```
# Preparation of input data into Script of  R Software#
    BOD.i^th location<-c(z_i(1),z_i,(2) ⋯ ,z_i(n))
    COD.i^th location<-c(x_i(1),x_i,(2) ⋯ ,x_i(n)), I = 1, 2, 3
    # Definition of variable #
    Z_i.1<-BOD.i^th location[-1]
    Z_i43<-BOD.i^th location[-44]
    V_i.43<-o.5(Z_j43 + Z_k43)
    X_i.1<-COD.i^th location[-1]
    N.43<-rep(0,43)
    X<-
    matrix(c(Z_1.43,N.43,N.43,N.43,Z_2.43,N.43,N.43,N.43,Z_3.43,V_1.43,N.43,N.43,N.43,V_1.43,N
    .43,N.43,N.43,. V_1.43, X_1.1,N.43,N.43,N.43, X_2.1,N.43,N.43,N.43,X_3.1),ncol=9)
    Y<-matrix(c(Z_i.1,Z_2.1,Z_3.1),ncol=1)
    # Estimation parameter by OLS #
    Kp.1<-t(X)%*%X
    Kp2<-t(x)%*%Y
    Estpar.OLS<-solve(Kp1)%*%Kp2
    # contemporaneous correlation test by Lagrange multiplier (LM) #
    Zi.EST<-estimatintunctioni^th of HSTARX(1,!)-OLS
   eit<-Zi.1-Zi.EST, # I =1, 2, 3 #
   et<-cbind(e1t,e2t,e3t)
   cor(et)
   n<-length(Zi.1)
   T<-n-1
   Lambda.LM<-T*(cor(et)[1,2]^2+cor(et)[1,3]^2+cor(et)[2,3]^2)
        # Estimation parameter by SUR #
   sg<-var(et)
   Isg<-solve(sg)
        I43<-rep(0,43)
        Ivkij<-Isg[1,j]*I43 for i, j = 1, 2, 3
```

```
      #operation of kronecker product ⊗#
IOMG<-
      matrix(c(ivk11,ivk12,ivk13,ivk21,ivk22,ivk23,ivk31,ivk32,ivk33),nrow=129,ncol=129)
hkmtr<-t(X)%*%IOMG
komp1<-hkmtr%*%X
komp2<-hkmtr%*%Y
EST.SUR<-solve(komp1)%*% komp2
EST.SUR
      # simultaneous of Fit parameter #
k<-9
SST<-n*(t(Y)%*%Y)
SSE<-t(et)%*%et
SSREG<-SST-SSE
MSREG<-SSREG/(k+1)
MSE<-SSE/(n-k-1)
F.c.st<-MSREG/MSE
F.Table<-qf(0.95,10,33)
      # individual of Fit parameter #
ni<-length(eit.lb)
Z<-matrix(c(Z1.43,Z2.43,Z3.43),ncol=3)
KZ<-t(Z)%*%Z
IKZ<-solve(KZ)
SSEZ<-t(eit.lb)%*%eit.lb
SEZ<-sqrt(SSEZ/(ni-3))
SB1<-SEZ*sqrt(IKZ[i,i])
EST.PARP<-matrix(c(0.156755,2.511082,0.209790),ncol=3)
ti<-EST.PARP[i,j]/SB1
t.Table<-qt(0.975,40)
```

and then we got the test contemporaneous correlation of error the result of OLS method in the GSTARX(1,1) model with using Lambda multiplier (λ_{LM}) is following Table.1

Table.1 The Result Statsitic Test of Lagrange Multipler

Test Statistic of Lagrange Multipler	Value
λ_{LM}	90,98
χ^2_{Table}	7,81

Source : The Result of Software R

Based on for table.1 we got the value Lammbda Lagrange Multiplier (λ_{LM}) which more than chi-square Table (χ^2_{Table}), so that the conclusion is the problem of Errors between the equations of each location are correlated so that there is a problem contemporaneous correlation, which will result in inefficient estimators of OLS. SUR so the alternative estimation parameter method in GSTARX(1,1) can be using SUR. SUR method the accommodate matrix of variance covariance error. From result of the data processing with R software, we got inverse variance covariance error of OLS method , there are from the result of multiplication as between

$$\Sigma^{-1} = \begin{bmatrix} 0{,}10283 & 0{,}0006 & -0{,}10778 \\ 0{,}0006 & -0{,}0027 & -0{,}00268 \\ -0{,}10778 & -0{,}00268 & 0{,}13175 \end{bmatrix},$$

and Identity matrix $I_{(43\times 43))}$ with using notation Kronecker Product, ang the result Ω^{-1} which basically process of estimation parameter model with using SUR method.
So we got the seemingly Unrelated regression of GSTARX(1,1) as following Table.2

Table.2 Seemingly unrelated Regression of GSTARX(1,1)

Location	y Φ_0	Reg F Φ_1	of GSTARX Co	Statistic	
				Count	Table
Inlet Jatiluhur	0.1567	0.96444	0.9975		
Outltl Jatiluhur	2,5111	-3,448	-1,91898	55,9147	2,1325
Rengasdengklok	0.20979	0,9974	0,26258		

From above apply the model GSTARX(1,1) toward real phenomena data, such as Contaminant Load Of BOD With Exogenous Variable of COD in Watershed Citarum River, we can explain that the seemingly unrelated regression estimator, for the test simultaneous hypothesis of model, based on table.2 we got is significance.

5. CONCLUSION

In the paper we studied the seemingly unrelated regression of GSTARX model. The estimator has Properties of efficient. We built the script using R software to estimate the parameter using seemingly unrelated regression Method. If should be done, because not yet space time software which can be used to estimate the parameters.

For case study to apply the GSTARX model to Contaminant Load of BOD With Exogenous Variable of COD in Watershed Citarum River is significance for the result test hypothesis simultaneous model.

6. REFERENSI

Baum Christopher F & Boston C, 2001, “Residual Diagnostics for Cross-Section Time Series Regression Models”, The Stats Journal I, number 1. pp. 101-104

Dufour Jean, M and Lynda Khalaf, 2000, ” Exact test for Contemporaneous Correlatio n of Disturbance in Seemingly Unrelated Regression” , Scientific Series, Cirano, Montreal Montreal. pp 59-64. Download August 2022

Henningsen Arne & Jeff d Hamann, 2008, “ A Package for Estimating Systems of Simulation Equation in R”, The Stats Journal I, number II. Pp. 202-205. Download Dec 2021

Kurnia JD, Setiawan, 2015, “The Simulation for Generalized Space Time Autoregressive-X (GSTARX) Model, Proceeding of The ConSSE FSM SMCU, e-Jurnal. Download August 2022

Prabowo Hadi Soekardi, 2018, Identification of the GSTARX-(p,q) Data Processing Model using Software R, Paper National Seminar on Mathematics and its Applications e-ISSN: 2550-0392, Unsoed Purwokerto. Downloaded Dec 2022

Ruchjana, B.N, Svetlana A, Borovkova and P Lopuha, 2012, "Least Square Estimation of Generalized Space Time Autoregressive (GSTAR) Model and Properties ", The 5th International Conference on Research & Education in Mathematics AIP.conf, American Institute of Physics. Downloaded August 2023

Ooms Marius, 2007, " Panel Data, SUR and GLS", Ekonometrics II, Timbergen Institute Amsterdam

Zellner A, 1962, " As an Efficient Methods of Estimation SUR and Test for Agregation Bias", Journal of American Statistical Assosiation. 57. pp 348-368