Thailand Statistician
January 2011; 9(1) : 77-92
http://statassoc.or.th
Contributed paper

Generalized Least Squares Transformation with the Second-order Autoregressive Error

Warangkhana Keerativibool

Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phattalung 93110 Thailand, and Centre of Excellence in Mathematics, CHE, Si Ayutthaya Rd., Bangkok 10400, Thailand.
e-mail: walex27@hotmail.com

Received: 19 September 2010
Accepted: 21 December 2010

Abstract

The major problem found in the fitted simultaneous equations model (SEM) maybe the autocorrelation and/or moving average problems. These problems have an affect on the ordinary least squares (OLS) estimators are not efficient. A generalized least squares (GLS) transformation matrix in order to correct the second-order autocorrelation [AR(2)] is proposed in this article by using the Cholesky decomposition and also shown in theoretically and empirically, it is a correction of AR(2) problem. The simulation study shows that the AR(2) problem in the errors makes the values of sum squared error (SSE) higher than they should because the estimated values of the residuals still overestimate.

Keywords: generalized least squares (GLS) transformation matrix, second-order autocorrelation [AR(2)], simultaneous equations model (SEM).

1. Introduction

A common problem maybe occurred in the fitted model is the discovery of autocorrelation and/or moving average problems in the residuals. This problem may occur when we start the plausible structural form of a misspecified model, due to lack of complete the accurate data. The consequence of this problem is that the OLS estimators are still linear unbiased, they are not efficient [1]. A lot of literary attention to this issue, such as Ullah, et al. [2] derived the large sample asymptotic approximation for the covariance matrix of the two stage Prais-Winsten estimator of the regression coefficients and analyzed numerically the efficiency properties of this estimator with respect to OLS and GLS with a known autocorrelation coefficient. Choudhury and Power [3] constructed a new approximate GLS estimator for the linear regression model with ARMA(p,q) disturbances which this estimator consists of the two-step procedure followed by OLS estimation of the transformed model: the first step eliminates the autoregressive component of the disturbance and the second step addresses the moving average component. Galbraith and Zinde-Walsh [4] gave a transformation of the general ARMA error-components in the panel model to yield spherical disturbances. Marazzi and Yohai [5] proposed new estimators to transform the response variables which are based on the minimization of a robust measure of residual autocorrelation. These estimators are robust and consistent even if the assumptions of normality and homoscedasticity do not hold. Hwang, et al. [6] constructed the GLS estimator for explosive AR(1) processes with conditionally heteroscedastic errors. The model under this consideration accommodates diverse conditionally heteroscedastic processes including Engle's ARCH, threshold-ARCH and beta-ARCH processes. Vougas [7] proposed an approximations of the usual GLS transformation matrices for estimation with AR(1) and AR(2) error that remove boundary discontinuities. This method avoids constrained optimization that unnecessarily enforces estimated parameters to be in the interior. Keerativibool [8] and Keerativibool, et al. [9-11] proposed a transformation matrix to correct the AR(2) problem in a SEM by extended the Prais-Winsten transformation and next Keerativibool [12] proposed a transformation matrix to correct the first-order moving average [MA(1)] problem in a regression model.

This article proposes a GLS transformation matrix to correct the AR(2) problem in a SEM as in [8-11] by using the Cholesky decomposition to find a square root of inverse matrix of the error variance instead of extended the Prais-Winsten transformation. We argue that, the proposed transformation would be better than the old transformation because it corrects the AR(2) problem along with the consideration of

contemporaneous correlation while the transformation in [8-11] is not consider this relationship. The remainder of this article is organized as follows. In Section 2, we introduce a SEM with serially and contemporaneously correlated error vector. The methodology is proposed in the 3rd Section. Section 4 shows the simulation study and results. Finally, Section 5 is the conclusion.

2. A Simultaneous Equations Model (SEM)

The structural form of the SEM [13] may be represented as

$$\mathbf{Y\Gamma} + \mathbf{XB} = \mathbf{U}, \tag{1}$$

where $\mathbf{Y}$ is a $\mathrm{T}\times\mathrm{M}$ matrix of observations y, $\mathbf{\Gamma}$ is a $\mathrm{M}\times\mathrm{M}$ nonsingular matrix of coefficients of endogenous variables γ, $\mathbf{X}$ is a $\mathrm{T}\times\mathrm{K}$ design matrix x of full-column rank, $\mathbf{B}$ is a $\mathrm{K}\times\mathrm{M}$ matrix of coefficients of predetermined variables β, and $\mathbf{U}$ is a $\mathrm{T}\times\mathrm{M}$ matrix of serially and contemporaneously correlated errors u. The reduced form model described by the structural form model in (1) can be written as

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}, \tag{2}$$

where $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ is a $\mathrm{K}\times\mathrm{M}$ matrix of unknown parameters π, and $\mathbf{V} = \mathbf{U\Gamma}^{-1}$ is a $\mathrm{T}\times\mathrm{M}$ matrix of serially and contemporaneously correlated errors v. For the t^{th} time of the SEM is

$$\mathbf{y}_t' = \mathbf{x}_t'\mathbf{\Pi} + \mathbf{v}_t'. \tag{3}$$

Combine all T times of the model in (3) to a stack model as follows:

$$\mathbf{y} = \tilde{\mathbf{\Pi}}'\mathbf{x} + \mathbf{v}, \tag{4}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix}, \quad \tilde{\mathbf{\Pi}}' = \begin{bmatrix} \mathbf{\Pi}' & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}' & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Pi}' \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_T \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_T \end{bmatrix},$$

$\mathbf{y}_t$ is a $\mathrm{M}\times 1$ vector of observations, $\mathbf{\Pi}'$ is a $\mathrm{M}\times\mathrm{K}$ matrix of unknown parameters, $\mathbf{x}_t$ is a $\mathrm{K}\times 1$ vector of independent variables, and $\mathbf{v}_t$ is a $\mathrm{M}\times 1$ vector of serially and contemporaneously correlated errors which has the form

$$\mathbf{v}_t = \rho_1\mathbf{v}_{t-1} + \rho_2\mathbf{v}_{t-2} + \boldsymbol{\varepsilon}_t \;;\; t = 1, 2, \dots, T, \tag{5}$$

where the l -periods back error $\mathbf{v}_{t-l}$ is called the l^{th} lag of error $\mathbf{v}_t$, the autoregressive parameters ρ_1 and ρ_2 of the model must satisfy the following conditions to ensure the stationary of the error terms [14],

$$\rho_1 + \rho_2 < 1,\ \rho_2 - \rho_1 < 1,\ |\rho_2| < 1, \tag{6}$$

and $\boldsymbol{\varepsilon}_t$ is a multivariate normal given zero means and contemporaneous covariance matrix $\boldsymbol{\Sigma}$, which is assumed nonsingular and positive symmetric definite,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{bmatrix}.$$

That is,

$$\boldsymbol{\varepsilon}_t \sim \mathrm{N}_M(\mathbf{0}, \boldsymbol{\Sigma}). \tag{7}$$

3. Methodology

Theorem A. The expectation of error $\mathbf{v}$ in (4) is equal to zero, so that the variance of error $\mathbf{v}$ is

$$\begin{aligned} \boldsymbol{\Phi} &= \mathrm{Var}(\mathbf{v} \mid \mathbf{x}) = \mathrm{E}(\mathbf{v}\mathbf{v}') = \begin{bmatrix} \boldsymbol{\phi}(0) & \boldsymbol{\phi}(1) & \cdots & \boldsymbol{\phi}(\mathrm{T}-1) \\ \boldsymbol{\phi}(1) & \boldsymbol{\phi}(0) & \cdots & \boldsymbol{\phi}(\mathrm{T}-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\phi}(\mathrm{T}-1) & \boldsymbol{\phi}(\mathrm{T}-2) & \cdots & \boldsymbol{\phi}(0) \end{bmatrix} \\ &= \frac{1}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \begin{bmatrix} \phi'(0) & \phi'(1) & \cdots & \phi'(\mathrm{T}-1) \\ \phi'(1) & \phi'(0) & \cdots & \phi'(\mathrm{T}-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi'(\mathrm{T}-1) & \phi'(\mathrm{T}-2) & \cdots & \phi'(0) \end{bmatrix} \otimes \boldsymbol{\Sigma} \\ &= \boldsymbol{\phi} \otimes \boldsymbol{\Sigma}, \end{aligned} \tag{8}$$

where

$$\boldsymbol{\phi}(|\mathrm{t}-\mathrm{s}|) = \mathrm{E}(\mathbf{v}_t \mathbf{v}_s');\ \mathrm{t, s} = 1, 2, \ldots, \mathrm{T},$$

$$\boldsymbol{\phi}(0) = \frac{1-\rho_2}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \boldsymbol{\Sigma} = \frac{1}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \phi'(0) \boldsymbol{\Sigma},$$

$$\boldsymbol{\phi}(1)=\frac{\rho_1}{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}\boldsymbol{\Sigma}=\frac{1}{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}\phi'(1)\boldsymbol{\Sigma},$$

$$\boldsymbol{\phi}(2)=\frac{\rho_1^2+\rho_2(1-\rho_2)}{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}\boldsymbol{\Sigma}=\frac{1}{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}\phi'(2)\boldsymbol{\Sigma},$$

$$\boldsymbol{\phi}(l)=\rho_1\boldsymbol{\phi}(l-1)+\rho_2\boldsymbol{\phi}(l-2)=\frac{1}{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}\phi'(l)\boldsymbol{\Sigma}\ ;\ l=3,\,4,\,\ldots,\,\mathrm{T}-1.$$

Theorem B. The inverse of the error variance in (8) is

$$\boldsymbol{\Phi}^{-1}=\boldsymbol{\phi}^{-1}\otimes\boldsymbol{\Sigma}^{-1}, \tag{9}$$

where

$$\boldsymbol{\phi}^{-1}=\begin{bmatrix}
1 & -\rho_1 & -\rho_2 & 0 & \cdots & 0 & 0 & 0\\
-\rho_1 & 1+\rho_1^2 & -\rho_1(1-\rho_2) & -\rho_2 & \cdots & 0 & 0 & 0\\
-\rho_2 & -\rho_1(1-\rho_2) & 1+\rho_1^2+\rho_2^2 & -\rho_1(1-\rho_2) & \cdots & 0 & 0 & 0\\
0 & -\rho_2 & -\rho_1(1-\rho_2) & 1+\rho_1^2+\rho_2^2 & \cdots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & \cdots & 1+\rho_1^2+\rho_2^2 & -\rho_1(1-\rho_2) & -\rho_2\\
0 & 0 & 0 & 0 & \cdots & -\rho_1(1-\rho_2) & 1+\rho_1^2 & -\rho_1\\
0 & 0 & 0 & 0 & \cdots & -\rho_2 & -\rho_1 & 1
\end{bmatrix}.$$

Theorem C. The square root of $\boldsymbol{\Phi}^{-1}$ in (9) is equal to $\mathbf{P}\otimes\boldsymbol{\Sigma}^{-\frac{1}{2}}$ where the $\mathrm{T}\times\mathrm{T}$ upper triangular matrix $\mathbf{P}$ by using the Cholesky decomposition is defined to be

$$\mathbf{P}=\begin{bmatrix}
1 & -\rho_1 & -\rho_2 & 0 & \cdots & 0 & 0 & 0\\
0 & 1 & -\rho_1 & -\rho_2 & \cdots & 0 & 0 & 0\\
0 & 0 & 1 & -\rho_1 & \cdots & 0 & 0 & 0\\
0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & \cdots & 1 & -\rho_1 & -\rho_2\\
0 & 0 & 0 & 0 & \cdots & 0 & \sqrt{1-\rho_2^2} & -\rho_1\sqrt{\frac{1+\rho_2}{1-\rho_2}}\\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & \sqrt{\frac{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}{1-\rho_2}}
\end{bmatrix}, \tag{10}$$

and

$\boldsymbol{\Sigma}^{\frac{1}{2}}$ is the square root of the positive symmetric definite matrix $\boldsymbol{\Sigma}$.

Theorem D. The $\mathrm{TM}\times\mathrm{TM}$ transformation matrix $\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}$ is defined by

$$\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}=\begin{bmatrix}
\mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{M}} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{\mathrm{M}} & -\rho_1\mathbf{I}_{\mathrm{M}} & -\rho_2\mathbf{I}_{\mathrm{M}} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \sqrt{1-\rho_2^2}\,\mathbf{I}_{\mathrm{M}} & -\rho_1\sqrt{\frac{1+\rho_2}{1-\rho_2}}\,\mathbf{I}_{\mathrm{M}} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \sqrt{\frac{(1+\rho_2)\left[(1-\rho_2)^2-\rho_1^2\right]}{1-\rho_2}}\,\mathbf{I}_{\mathrm{M}}
\end{bmatrix}. \tag{11}$$

The transformation matrix $\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}$ in (11) is used to transform the reduced form model in (4), such that the AR(2) in the errors is eliminated, to give the model

$$\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathbf{y}=\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\tilde{\mathbf{\Pi}}'\mathbf{x}+\boldsymbol{\varepsilon}, \tag{12}$$

where

$$\boldsymbol{\varepsilon}=\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathbf{v},\ \mathrm{E}\left(\boldsymbol{\varepsilon}\right)=\mathbf{0},$$

and

$$\mathrm{Var}\left(\boldsymbol{\varepsilon}\right)=\mathrm{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\left[\mathbf{I}_{\mathrm{T}}\otimes\boldsymbol{\Sigma}\right]=\begin{bmatrix}
\boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}
\end{bmatrix}.$$

After using the GLS transformation matrix $\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}$ to eliminate the AR(2) problem in the SEM, the usual assumptions of the model will satisfy, i.e., the errors of the model are independent. Therefore, either the estimation or testing hypothesis of the parameters in the SEM can be done.

4. Simulation Study and Results

In this simulation study, we use the SAS programming version 9.1 to generate one hundred iterations of the systems of three simultaneous equations with one hundred observations per each equation and four independent explanatory variables as the following steps.

1. Generate 10,000 vectors of the 3x1 multivariate normal $\boldsymbol{\varepsilon}_t$ in (5) by the IML procedure, given zero means and contemporaneous covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\varepsilon}_t \sim \mathrm{N}_3\left(\mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 0.49 & 0.392 & 0.504 \\ 0.392 & 0.64 & 0.648 \\ 0.504 & 0.648 & 0.81 \end{bmatrix}\right); \ t = 1, 2, \ldots, 10,000.$$

2. Construct the 3x1 serially and contemporaneously correlated error vectors $\mathbf{v}_t$ in (5) by the following form where the first-two vectors are arbitrarily given as

$$\mathbf{v}_{-1} = \begin{bmatrix} 3 & 5 & 7 \end{bmatrix}, \ \mathbf{v}_0 = \begin{bmatrix} 4 & 6 & 8 \end{bmatrix},$$

and the first-two autoregressive parameters are arbitrarily given as $\rho_1 = 0.6$, $\rho_2 = -0.5$,

$$\mathbf{v}_t = 0.6\mathbf{v}_{t-1} - 0.5\mathbf{v}_{t-2} + \boldsymbol{\varepsilon}_t \ ; \ t = 1, 2, \ldots, 10,000.$$

Split the series of errors $\mathbf{v}_t$ in sequence to preserve the autocorrelation relationship into 100 samples, each of which consists of 100 vectors. Then estimate the autoregressive parameters, test the AR(2) properties and test the multivariate normality for the residuals by the MODEL and ARIMA procedure.

3. Generate 10,000 observations of the series of independent explanatory variables x_{t2}, x_{t3}, and x_{t4} by the UNIFORM function and the first element of independent explanatory variables x_{t1} is given as a constant 1,

$$x_{t2} \sim \mathrm{U}(1,7), \ x_{t3} \sim \mathrm{U}(2,9), \text{ and } x_{t4} \sim \mathrm{U}(3,6).$$

Again, split the series of independent variables in sequence into 100 samples, each of which consists of 100 observations. Then test the multicollinearity problem for the series of independent variables.

4. Construct 100 samples of the 3x1 dependent vectors $\mathbf{y}_t$ in (3), using the independent variables obtained in Step 3 and the autocorrelated errors obtained in Step 2, as the following form where the parameters of the model are arbitrarily given,

$$\mathbf{y}_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ y_{t3} \end{bmatrix} = \begin{bmatrix} 20 & 5 & 7 & 12 \\ 12 & 4 & 11 & 20 \\ 15 & 6 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 \\ x_{t2} \\ x_{t3} \\ x_{t4} \end{bmatrix} + \begin{bmatrix} v_{t1} \\ v_{t2} \\ v_{t3} \end{bmatrix}; \; t = 1, 2, \ldots, 100.$$

For each sample, we combine all one hundred times of $\mathbf{y}_t$, x_t's and $\mathbf{v}_t$ in Step 4, 3, and 2, respectively, as a stack model in (4).

5. Construct the estimate of proposed transformation matrix $\mathbf{P} \otimes \mathbf{I}_M$ in (11) for each sample by using the estimated values of autoregressive parameters which obtained in Step 2 and use it to eliminate the AR(2) problem of the model in Step 4, to give the model in (12).

6. Estimate the parameters π of the model before and after transformation by the MODEL procedure. For the model after transformation, examine the AR(2) problem in the residuals and the normality of the residuals by the ARIMA and the MODEL procedures, respectively. The tests confirm that the residuals of all 100 transformed samples are independent. Therefore, the proposed transformation matrix in (11) has a transformation percentage of 100%.

7. Use the residuals of the models before and after transformation to calculate the SSE. Compare the SSE by the TTEST procedure. The result shows that, when ignoring the AR(2) problem in the errors, the SSE is greater than the one of the transformed model at the level of significance 0.05 which this result corresponding to Theorem A-5 in [8]. The SSE values of all 100 samples from the model before and after transformation are shown in Figure 1. The summary statistics of the SSE before and after transformation are shown in Table 1.

5. Conclusions

This article has proposed a new transformation matrix in order to correct the AR(2) problem by using the Cholesky decomposition and also shown theoretically that it can be a correction of this problem without any condition. The results of simulation study confirm that the transformed errors of the SEM are independent and that before removing the autocorrelation in the errors of the model, the SSE is greater than the one after the autocorrelation problem has been corrected at the level of significance 0.05 because the estimated values of the residuals are overabundant.

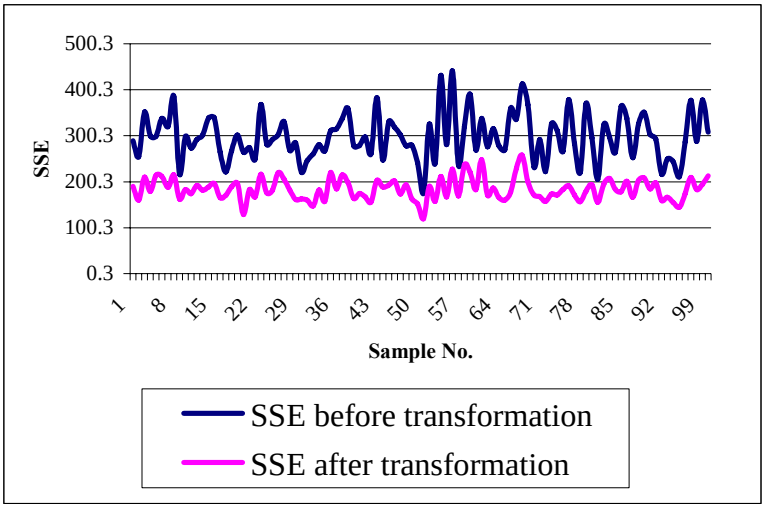


Figure 1. The SSE of 100 samples from the model before and after transformation.

Table 1. Summary statistics of the SSE in 100 samples of 100 observations.

Statistics	SSE	
	Before Transformation	**After Transformation**
Mean	298.2052	184.4247
S.D.	51.4663	24.2265
Min	176.4408	119.5157
Max	441.6474	258.0641
t-test	20.00	
p-value	< 0.001	

Acknowledgements

This article is partially supported by the Centre of Excellence in Mathematics, the Commission of Higher Education, Thailand. The author is thankful for the help and support of several people in my organization, Thaksin University.

References

[1] Gujarati, D.N., *Essentials of Econometrics,* 3rd ed., Singapore, McGraw-Hill, 2006.

[2] Ullah, A., Srivastava, V.K., Magee, L., and Srivastava, A., Estimation of Linear Regression Model with Autocorrelated Disturbances, *Journal of Time Series Analysis*, 1983; 4: 127-135.

[3] Choudhury A.H., and Power S., A New Approximate GLS Estimator for the Linear Regression Model with ARMA(p,q) Disturbances, *Economics Letters*, 1995; 48: 119-127

[4] Galbraith, J.W., and Zinde-Walsh, V., Transforming the Error-Components Model for Estimation with General ARMA Disturbances, *Journal of Econometrics*, 1995; 66: 349-355.

[5] Marazzi, A., and Yohai, V.J., Robust Box–Cox Transformations Based on Minimum Residual Autocorrelation, *Computational Statistics & Data Analysis*, 2006; 50: 2752-2768.

[6] Hwang, S.Y., Kim., S., Lee, S.D., and Basawa, I.V., Generalized Least Squares Estimation for Explosive AR(1) Processes with Conditionally Heteroscedastic Errors, *Statistics & Probability Letters*, 2007; 77: 1439-1448.

[7] Vougas, D.V., Generalized Least Squares Transformation and Estimation with Autoregressive Error, *Statistics & Probability Letters*, 2008; 78: 402-404.

[8] Keerativibool, W., *Selection of a System of Simultaneous Equations Models*, Doctoral Dissertation, School of Applied Statistics, National Institute of Development Administration, 2009.

[9] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Autocorrelation Correction in a Regression Model, *Proceedings: Applied Statistics Conference 2009 (ASCONF2009)*, Pattaya, Chon Buri, Thailand, 2009: 519-533.

[10] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Autocorrelation Correction in a Simultaneous Equations Model, *Proceedings: the 5th International Conference on Mathematics, Statistics and their Applications (ICMSA2009)*, West Sumatra, Indonesia, 2009: 192-197.

[11] Keerativibool, W., Jitthavech, J., and Lorchirachoonkul, V., Model Selection in a System of Simultaneous Equations Model, *Communications in Statistics-Theory and Methods*, 2011; 40: 373-393.

[12] Keerativibool, W., Moving Average Correction in a Regression Model, *Thailand Statistician: Journal of Thai Statistical Association*, 2010; 8: 63-80.

[13] Greene, W., *Econometric Analysis,* 6th ed., Upper Saddle River, New Jersey, Prentice Hall, 2008.

[14] Box, G.E.P., Jenkins, G.M., and Reinsel, G.C., *Time Series Analysis : Forecasting and Control,* 3rd ed., Englewood Cliffs, New Jersey, Prentice Hall, 1994.

Appendix

Proof of Theorem A.

The expectation of $\mathbf{v}_t$ in (5) is

$$\mathrm{E}(\mathbf{v}_t) = \rho_1 \mathrm{E}(\mathbf{v}_{t-1}) + \rho_2 \mathrm{E}(\mathbf{v}_{t-2}) + \mathrm{E}(\boldsymbol{\varepsilon}_t).$$

Using the stationary conditions in (6) and the assumption $\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ in (7), we have

$$\mathrm{E}(\mathbf{v}_t) = \rho_1 \mathrm{E}(\mathbf{v}_t) + \rho_2 \mathrm{E}(\mathbf{v}_t)$$

$$(1-\rho_1-\rho_2)\mathrm{E}(\mathbf{v}_t) = \mathbf{0}. \qquad \text{(A.1)}$$

Under the stationary conditions in (6), the value of $1-\rho_1-\rho_2$ in (A.1) is always greater than zero. Therefore, $\mathrm{E}(\mathbf{v}_t) = \mathbf{0}$ and also $\mathrm{E}(\mathbf{v}) = \mathbf{0}$. So that, the variance of error $\mathbf{v}$ in (4) can be written as

$$\boldsymbol{\Phi} = \mathrm{Var}(\mathbf{v} \mid \mathbf{x}) = \mathrm{E}(\mathbf{v}\mathbf{v}') = \mathrm{E}\left(\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' & \mathbf{v}_2' & \cdots & \mathbf{v}_T' \end{bmatrix} \right)$$

$$= \begin{bmatrix} \mathrm{E}(\mathbf{v}_1\mathbf{v}_1') & \mathrm{E}(\mathbf{v}_1\mathbf{v}_2') & \cdots & \mathrm{E}(\mathbf{v}_1\mathbf{v}_T') \\ \mathrm{E}(\mathbf{v}_2\mathbf{v}_1') & \mathrm{E}(\mathbf{v}_2\mathbf{v}_2') & \cdots & \mathrm{E}(\mathbf{v}_2\mathbf{v}_T') \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(\mathbf{v}_T\mathbf{v}_1') & \mathrm{E}(\mathbf{v}_T\mathbf{v}_2') & \cdots & \mathrm{E}(\mathbf{v}_T\mathbf{v}_T') \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\phi}(0) & \boldsymbol{\phi}(1) & \dots & \boldsymbol{\phi}(T-1) \\ \boldsymbol{\phi}(1) & \boldsymbol{\phi}(0) & \dots & \boldsymbol{\phi}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\phi}(T-1) & \boldsymbol{\phi}(T-2) & \dots & \boldsymbol{\phi}(0) \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{\phi}(0) = \mathrm{E}(\mathbf{v}_t \mathbf{v}_t') &= \mathrm{E}\left[\mathbf{v}_t (\rho_1 \mathbf{v}_{t-1}' + \rho_2 \mathbf{v}_{t-2}' + \boldsymbol{\varepsilon}_t')\right] \\ &= \rho_1 \mathrm{E}(\mathbf{v}_t \mathbf{v}_{t-1}') + \rho_2 \mathrm{E}(\mathbf{v}_t \mathbf{v}_{t-2}') + \mathrm{E}\left[(\rho_1 \mathbf{v}_{t-1} + \rho_2 \mathbf{v}_{t-2} + \boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}_t'\right] \\ &= \rho_1 \mathrm{E}(\mathbf{v}_t \mathbf{v}_{t-1}') + \rho_2 \mathrm{E}(\mathbf{v}_t \mathbf{v}_{t-2}') + \rho_1 \mathrm{E}(\mathbf{v}_{t-1} \boldsymbol{\varepsilon}_t') + \rho_2 \mathrm{E}(\mathbf{v}_{t-2} \boldsymbol{\varepsilon}_t') + \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \\ \boldsymbol{\phi}(0) &= \rho_1 \boldsymbol{\phi}(1) + \rho_2 \boldsymbol{\phi}(2) + \boldsymbol{\Sigma}, \end{aligned} \tag{A.2}$$

$$\begin{aligned} \boldsymbol{\phi}(1) = \mathrm{E}(\mathbf{v}_{t-1} \mathbf{v}_t') &= \mathrm{E}\left[\mathbf{v}_{t-1} (\rho_1 \mathbf{v}_{t-1}' + \rho_2 \mathbf{v}_{t-2}' + \boldsymbol{\varepsilon}_t')\right] \\ &= \rho_1 \boldsymbol{\phi}(0) + \rho_2 \boldsymbol{\phi}(1) \\ \boldsymbol{\phi}(1) &= \frac{\rho_1}{1-\rho_2} \boldsymbol{\phi}(0), \end{aligned} \tag{A.3}$$

$$\begin{aligned} \boldsymbol{\phi}(2) = \mathrm{E}(\mathbf{v}_{t-2} \mathbf{v}_t') &= \mathrm{E}\left[\mathbf{v}_{t-2} (\rho_1 \mathbf{v}_{t-1}' + \rho_2 \mathbf{v}_{t-2}' + \boldsymbol{\varepsilon}_t')\right] \\ &= \rho_1 \boldsymbol{\phi}(1) + \rho_2 \boldsymbol{\phi}(0). \end{aligned} \tag{A.4}$$

Replacing (A.3) into (A.4),

$$\boldsymbol{\phi}(2) = \frac{\rho_1^2 + \rho_2 (1-\rho_2)}{1-\rho_2} \boldsymbol{\phi}(0). \tag{A.5}$$

Substituting (A.3) and (A.5) into (A.2) gives

$$\boldsymbol{\phi}(0) = \frac{1-\rho_2}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \boldsymbol{\Sigma}. \tag{A.6}$$

Replacing (A.6) into (A.3) and (A.5) yields

$$\boldsymbol{\phi}(1) = \frac{\rho_1}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \boldsymbol{\Sigma},$$

and

$$\boldsymbol{\phi}(2) = \frac{\rho_1^2 + \rho_2 (1-\rho_2)}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \boldsymbol{\Sigma}.$$

For $l = 3, 4, \ldots, T-1$,

$$\begin{aligned}\boldsymbol{\phi}(l) = \mathrm{E}\left(\mathbf{v}_{t-l}\mathbf{v}_t'\right) &= \mathrm{E}\left[\mathbf{v}_{t-l}\left(\rho_1\mathbf{v}_{t-1}' + \rho_2\mathbf{v}_{t-2}' + \boldsymbol{\varepsilon}_t'\right)\right] \\ &= \rho_1\mathrm{E}\left(\mathbf{v}_{t-l}\mathbf{v}_{t-1}'\right) + \rho_2\mathrm{E}\left(\mathbf{v}_{t-l}\mathbf{v}_{t-2}'\right) + \mathrm{E}\left(\mathbf{v}_{t-l}\boldsymbol{\varepsilon}_t'\right) \\ &= \rho_1\boldsymbol{\phi}(l-1) + \rho_2\boldsymbol{\phi}(l-2).\end{aligned}$$

In general,

$$\boldsymbol{\phi}(|t-s|) = \mathrm{E}\left(\mathbf{v}_t\mathbf{v}_s'\right);\ t, s = 1, 2, \ldots, T.$$

Q.E.D.

Proof of Theorem B.

From Theorem A and suppose T = 4, the autocovariance of $\mathbf{v}_t$ at lag 3 can be written as

$$\boldsymbol{\phi}(3) = \mathrm{E}\left(\mathbf{v}_{t-3}\mathbf{v}_t'\right) = \rho_1\boldsymbol{\phi}(2) + \rho_2\boldsymbol{\phi}(1) = \frac{\rho_1\left[\rho_1^2 + \rho_2(2-\rho_2)\right]}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]}\boldsymbol{\Sigma}.$$

Therefore, the variance of error $\mathbf{v}$ in (8) when given time T equal to 4 is of the form

$$\boldsymbol{\Phi} = \frac{1}{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]} \times \begin{bmatrix} 1-\rho_2 & \rho_1 & \rho_1^2 + \rho_2(1-\rho_2) & \rho_1\left[\rho_1^2 + \rho_2(2-\rho_2)\right] \\ \rho_1 & 1-\rho_2 & \rho_1 & \rho_1^2 + \rho_2(1-\rho_2) \\ \rho_1^2 + \rho_2(1-\rho_2) & \rho_1 & 1-\rho_2 & \rho_1 \\ \rho_1\left[\rho_1^2 + \rho_2(2-\rho_2)\right] & \rho_1^2 + \rho_2(1-\rho_2) & \rho_1 & 1-\rho_2 \end{bmatrix} \otimes \boldsymbol{\Sigma}. \quad \text{(B.1)}$$

The inverse of the matrix $\boldsymbol{\Phi}$ in (B.1) is defined by

$$\boldsymbol{\Phi}^{-1} = \begin{bmatrix} 1 & -\rho_1 & -\rho_2 & 0 \\ -\rho_1 & 1+\rho_1^2 & -\rho_1(1-\rho_2) & -\rho_2 \\ -\rho_2 & -\rho_1(1-\rho_2) & 1+\rho_1^2 & -\rho_1 \\ 0 & -\rho_2 & -\rho_1 & 1 \end{bmatrix} \otimes \boldsymbol{\Sigma}^{-1}. \quad \text{(B.2)}$$

From Theorem A and suppose T = 5, the autocovariance of $\mathbf{v_t}$ at lag 4 can be written as

$$\boldsymbol{\phi}(4) = \mathrm{E}\left(\mathbf{v_{t-4}}\mathbf{v_t'}\right) = \rho_1\boldsymbol{\phi}(3) + \rho_2\boldsymbol{\phi}(2) = \frac{\rho_1^4 + \rho_1^2\rho_2\left(3-\rho_2\right) + \rho_2^2\left(1-\rho_2\right)}{\left(1+\rho_2\right)\left[\left(1-\rho_2\right)^2 - \rho_1^2\right]}\boldsymbol{\Sigma}.$$

Therefore, the variance of error $\mathbf{v}$ in (8) when given time T equal to 5 is of the form

$$\boldsymbol{\Phi} = \frac{1}{\left(1+\rho_2\right)\left[\left(1-\rho_2\right)^2 - \rho_1^2\right]}$$

$$\times\begin{bmatrix} 1-\rho_2 & \rho_1 & \rho_1^2+\rho_2(1-\rho_2) & \rho_1\left[\rho_1^2+\rho_2(2-\rho_2)\right] & \rho_1^4+\rho_1^2\rho_2(3-\rho_2)+\rho_2^2(1-\rho_2) \\ \rho_1 & 1-\rho_2 & \rho_1 & \rho_1^2+\rho_2(1-\rho_2) & \rho_1\left[\rho_1^2+\rho_2(2-\rho_2)\right] \\ \rho_1^2+\rho_2(1-\rho_2) & \rho_1 & 1-\rho_2 & \rho_1 & \rho_1^2+\rho_2(1-\rho_2) \\ \rho_1\left[\rho_1^2+\rho_2(2-\rho_2)\right] & \rho_1^2+\rho_2(1-\rho_2) & \rho_1 & 1-\rho_2 & \rho_1 \\ \rho_1^4+\rho_1^2\rho_2(3-\rho_2)+\rho_2^2(1-\rho_2) & \rho_1\left[\rho_1^2+\rho_2(2-\rho_2)\right] & \rho_1^2+\rho_2(1-\rho_2) & \rho_1 & 1-\rho_2 \end{bmatrix}\otimes\boldsymbol{\Sigma}. \tag{B.3}$$

The inverse of the matrix $\boldsymbol{\Phi}$ in (B.3) is defined by

$$\boldsymbol{\Phi}^{-1} = \begin{bmatrix} 1 & -\rho_1 & -\rho_2 & 0 & 0 \\ -\rho_1 & 1+\rho_1^2 & -\rho_1\left(1-\rho_2\right) & -\rho_2 & 0 \\ -\rho_2 & -\rho_1\left(1-\rho_2\right) & 1+\rho_1^2+\rho_2^2 & -\rho_1\left(1-\rho_2\right) & -\rho_2 \\ 0 & -\rho_2 & -\rho_1\left(1-\rho_2\right) & 1+\rho_1^2 & -\rho_1 \\ 0 & 0 & -\rho_2 & -\rho_1 & 1 \end{bmatrix}\otimes\boldsymbol{\Sigma}^{-1}. \tag{B.4}$$

In general, the inverse of the error variance in (8) can be obtained by extended (B.2) and (B.4) as follows:

$$\boldsymbol{\Phi}^{-1} = \begin{bmatrix} 1 & -\rho_1 & -\rho_2 & 0 & \cdots & 0 & 0 & 0 \\ -\rho_1 & 1+\rho_1^2 & -\rho_1\left(1-\rho_2\right) & -\rho_2 & \cdots & 0 & 0 & 0 \\ -\rho_2 & -\rho_1\left(1-\rho_2\right) & 1+\rho_1^2+\rho_2^2 & -\rho_1\left(1-\rho_2\right) & \cdots & 0 & 0 & 0 \\ 0 & -\rho_2 & -\rho_1\left(1-\rho_2\right) & 1+\rho_1^2+\rho_2^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho_1^2+\rho_2^2 & -\rho_1\left(1-\rho_2\right) & -\rho_2 \\ 0 & 0 & 0 & 0 & \cdots & -\rho_1\left(1-\rho_2\right) & 1+\rho_1^2 & -\rho_1 \\ 0 & 0 & 0 & 0 & \cdots & -\rho_2 & -\rho_1 & 1 \end{bmatrix}\otimes\boldsymbol{\Sigma}^{-1}$$

$$= \boldsymbol{\phi}^{-1}\otimes\boldsymbol{\Sigma}^{-1}. \qquad \textbf{Q.E.D.}$$

Proof of Theorem C.

For easy to proof, we rename the matrix $\boldsymbol{\phi}^{-1}$ in Theorem B to be the matrix $\mathbf{B} = \left[b_{ij} \right]_{T \times T}$. By Cholesky decomposition, there exists the upper triangular matrix $\mathbf{P} = \left[p_{ij} \right]_{T \times T}$ such that $\mathbf{B} = \mathbf{P}'\mathbf{P}$ where each elements of the matrix $\mathbf{P}$ can be found by the following formulas,

$$p_{11} = \sqrt{b_{11}},$$

$$p_{1t} = \frac{b_{1t}}{p_{11}} ; \; t = 2, 3, \ldots, T,$$

$$p_{tt} = \sqrt{b_{tt} - \sum_{s=1}^{t-1} p_{st}^2} ; \; t = 2, 3, \ldots, T,$$

$$p_{st} = \frac{b_{st} - \sum_{u=1}^{s-1} p_{us} p_{ut}}{p_{ss}} ; \; s, t = 2, 3, \ldots, T.$$

Therefore, the $T \times T$ upper triangular matrix $\mathbf{P}$**,** decompose from the matrix $\mathbf{B}$ (or $\boldsymbol{\phi}^{-1}$), can be written as

$$\mathbf{P} = \begin{bmatrix} 1 & -\rho_1 & -\rho_2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -\rho_1 & -\rho_2 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & -\rho_1 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -\rho_1 & -\rho_2 \\ 0 & 0 & 0 & 0 & \ldots & 0 & \sqrt{1-\rho_2^2} & -\rho_1\sqrt{\frac{1+\rho_2}{1-\rho_2}} \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & \sqrt{\frac{(1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right]}{1-\rho_2}} \end{bmatrix}.$$

Q.E.D.

Proof of Theorem D

The expectation of the error vector $\boldsymbol{\varepsilon}$ in (12) is equal to

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathrm{E}\left\{\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathbf{v}\right\} = \left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathrm{E}(\mathbf{v}). \tag{D.1}$$

From Theorem A we have $\mathrm{E}(\mathbf{v}) = \mathbf{0}$, so that the expectation of the error vector $\boldsymbol{\varepsilon}$ in (D.1) is also equal to zero. The variance of the error vector $\boldsymbol{\varepsilon}$ in (12) is equal to

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\varepsilon}) &= \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathrm{E}\left(\left\{\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathbf{v}\right\}\left\{\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathbf{v}\right\}'\right)\\ &= \left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\mathrm{E}(\mathbf{v}\mathbf{v}')\left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]' = \left[\mathbf{P}\otimes\mathbf{I}_{\mathrm{M}}\right]\left[\boldsymbol{\phi}\otimes\boldsymbol{\Sigma}\right]\left[\mathbf{P}'\otimes\mathbf{I}_{\mathrm{M}}\right]\\ &= \left[\mathbf{P}\boldsymbol{\phi}\otimes\boldsymbol{\Sigma}\right]\left[\mathbf{P}'\otimes\mathbf{I}_{\mathrm{M}}\right] = \left[\mathbf{P}\boldsymbol{\phi}\mathbf{P}'\otimes\boldsymbol{\Sigma}\right].\end{aligned}$$

Since, the matrix $\mathbf{P}$ is the square root of the matrix $\boldsymbol{\phi}^{-1}$, we have

$$\boldsymbol{\phi}^{-1} = \mathbf{P}'\mathbf{P}.$$

Thus,

$$\boldsymbol{\phi} = \mathbf{P}^{-1}\left(\mathbf{P}'\right)^{-1},$$

and

$$\mathbf{P}\boldsymbol{\phi}\mathbf{P}' = \mathbf{P}\,\mathbf{P}^{-1}\left(\mathbf{P}'\right)^{-1}\mathbf{P}' = \mathbf{I}_{\mathrm{T}}.$$

Therefore,

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \left[\mathbf{I}_{\mathrm{T}}\otimes\boldsymbol{\Sigma}\right].$$

Q.E.D.