International Journal of Statistics and Applications 2017, 7(1): 18-25
DOI: 10.5923/j.statistics.20170701.03

MM and MM Ridge Estimators for SUR Model

Tarek Mahmoud Omara

Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Kafrelsheikh University, Kafrelsheikh, Egypt

Abstract In this paper, we propose the MM and MM ridge estimators for SUR model to deal with outliers. The MM estimators are the type of robust regression with high breakdown point and have more efficient than other robust estimators. Since, outliers, frequently, appear with multicollinearity problem, then we propose MM ridge estimators for SUR mode. In these estimators, the shrink parameter was chosen by minimize robust Cross Validation Criteria (CV_{MM}) which depend on MM estimators. This choice achieves high breakdown point for given shrink parameter. Therefore, the MM ridge estimator has strong robust features. In addition, the asymptotical properties for the MM and MM ridge estimators ware also investigated. The median ASE (average squared error) was used to compeer the efficiency for estimator and to compute the estimators we designed two algorithm. Furthermore, the Simulations study was executed to test the performance of GL S, S, MM and MM ridge estimators for SUR model.

Keywords Seemingly unrelated regression (SUR), GLS estimator, S-estimator, MM-estimator, MM ridge estimators, robustness properties, robust Cross Validation Criteria (CV_{MM})

1. Introduction

The seemingly unrelated regression (SUR) Model proposed by (zellner, 1962) which it depends on general least squares estimator (GLS) and assumes data without outliers but in some cases this cannot be achieved. The robust methods considered the one important approach to deal with outliers which allow the unequal weight for observations. (Koenker etal, 1990) introduce M-estimators method, as a robust methods, to estimate SUR model when the data within outliers, yet the asymptotic efficiency for M-estimators depend on initial estimate and have breakdown point (bp) equal 0. Therefore, (Bilodeau etal., 2000) suggested S-estimator for SUR model which interest same asymptotic properties for M-estimators and have high (bp) reach to 50% of the observations. In the same side, (Garcia etal., 2006) and (Roelandt etal., 2009) proposed **τ**-estimates for multivariate regression which have excellent robustness and high efficiency under normality of error. (Roelandt etal., 2009) extract generalized S estimates for multivariate regression which consider high breakdown estimation when analysis the independent component. The MM estimators for linear regression model was introduced by (Yohai, 1987)), which begin at the first with a highly initial robust regression estimator like S estimators which depend on a loss function $\rho_0$and then used this initial estimator to obtain M estimator with other loss function ρ_1 . This estimator has two advantage, the high asymptotic efficiency for normality error and high breakdown point reach to 50%. The study of (Berrendero etal., 2007) show that, the asymptotic bias for mm estimator is lower than the τ estimators when we contaminations error lower than 0.20 and the MM estimator has best lower maximum bias curve than other robust estimators.

(Salibian etal., 2006) show that, the breakdown point of MM-estimator in finite-sample is equal to or greater than initial S-estimator. The MM estimators begin at the first with a highly initial robust regression estimator like S estimators which depend on a loss function ρ_0 and then used this initial estimator to obtain M estimator with other loss function ρ_1 .(Kudraszow, 2011) suggested MM estimator for multivariate linear model and study the consistency and asymptotic normality with elliptical distribution for error. On other hand, frequently, outliers appear with degree of multicollinearity. The ridge method solved multicollinearity problem has been discussed by (Hoerl etal., 1970). This method is a way of proceeding in solving the problem by adding specific information to remove the ill-condition. The SUR model possibly is under the influence of multicollinarity. (Srivastava etal., 1987) suggested general ridge estimator to remove the ill-condition in SUR model (Alkhamisi. etal, 2006) developed ridge estimators of the SUR model, when the data are transformed in a canonical form. The critical point in ridge estimate is how to choose the shrink parameter. So, (Firinguetti, 1997), (Kibria etal., 2003) and (Alkhamisi. etal, 2006) suggested non robust Criteria to choice the shrink parameter in ridge estimator for SUR model. These formulas assume choosing ridge parameter when data without outliers. (Jung, 2009) proposed robust cross validation in ridge regression when data are within outliers. (El-hosany etal., 2011) used robust Cross

* Corresponding author:
Tarek_em@yahoo.com (Tarek Mahmoud Omara)
Published online at http://journal.sapub.org/statistics

Validation Criteria to choice shrink parameter in SUR model.

In fact, frequently, outliers appear with degree of multicollinearity. Therefore, (El-hosany etal., 2011) mingle between ridge estimator and S- estimator to get robust ridge estimator. (Maronna, 2011), (Moawad etal., 2011) and (Mariam. etal., 2012) have developed MM-estimators to deal with leverage point and absorption problem of multicollinearity. This estimators builded by combined between ridge regression and mm-type to reach to mm ridge estimators.

There is no study use the MM estimator in the SUR model. So, we suggests, MM and MM ridge estimator for SUR model to deal with bad high leverage points and the multicollinearity problem associated with it. In MM ridge estimator, we develop robust Cross Validation Criteria, to choice shrink parameter, depended on MM estimator. This paper organized as follows: The SUR model and estimators defined in Section two. In section three, we study the asymptotic properties of MM and MM ridge estimators in SUR model while section four will be devoted to make the simulation study. In section five, we develop algorithm to compute MM and MM ridge estimator for SUR model.

2. GLS, S, MM and Weight MM Estimator for SUR Model

Conceder the SUR model

$$Y = X\beta + \mu \quad (1)$$

where $Y = [Y_1', \dots, Y_q']'$ is a nq × 1 vector of dependant variable in q equation, $X = \mathrm{diag}[X_1, \dots, X_q]$ is a nq×kq matrix of independent variable and $\mu = [\mu_1', \dots, \mu_q']'$ is a nq × 1 vector of random error in q equation. Suppose $E(\mu_i)=0$, $\mathrm{var}(\mu_i) = \sigma_{ii} I_n$ for all i=1,2,…q and $\mathrm{Cov}(\mu_i, \mu_j) = \sigma_{ij} I_n$ for all i,j=1,2,…q. We can right the SUR model in the multivariate form

$$\bar{\bar{Y}} = \bar{\bar{X}}\bar{\bar{\beta}} + \bar{\bar{\mu}} \quad (2)$$

where $\bar{\bar{Y}} = [Y_1, \dots, Y_q]$ is a matrix of n × q observations with Y_1 a n × 1 vector, $\bar{\bar{X}} = [X_1, \dots, X_q]$ is a n × kq matrix, $\bar{\bar{\beta}} = \mathrm{diag}[\beta_1, \dots, \beta_q]$ is a kq×q matrix of coefficients. and $\bar{\bar{\mu}} = [\bar{\bar{\mu}}_1, \bar{\bar{\mu}}_2, \dots, \bar{\bar{\mu}}_n]$ is a n×q matrix of residual with $\bar{\bar{\mu}}_i$ a q × 1 vector. Let $\bar{\bar{\mu}} = A_i\mu, i = 1,2,\dots,n$ where $A_i = \mathrm{diag}(a_i, a_i, \dots, a_i)$ is a nq× q matrix and a_i is vector contains one at position i and zero elsewhere. (Srivastava and Giles, 1987) suggested the GLS estimator for SUR model

$$\beta_{GLS} = [X'(\Sigma^{-1} \otimes I_n)X]^{-1}X'(\Sigma^{-1} \otimes I_n)Y \quad (3)$$

where $\sum$ is q×q variance covariance matrix for error between equations. The GLS estimator calculate by use $\hat{\sum} = \frac{1}{n}\hat{\mu}'\hat{\mu}$ as consistent estimator for $\sum$. (Bilodeau etal, 2000) introduce S estimator to deal with outliers for SUR model. This estimator used the Huber function at the form

$$\rho(T) = \begin{cases} \frac{T^2}{2} - \frac{T^4}{2C^2} + \frac{T^6}{6C^4} |T| \le C \\ \frac{C^2}{6} |T| \ge C \end{cases} \quad (4)$$

where ρ is symmetric, continuous, differentiable and for $C > 0$ it is strictly increasing on[0; c], constant on [c,∞] and ρ(0)= 0. The percentage of breakdown point depend on C. (Ruppert, 1992) show that, when choice C=1.5476 the of breakdown point reach to 50%. The S estimator for SUR model minimize

$$L = \log|\Sigma| - \lambda_s \left[\frac{1}{n}\sum_{i=1}^{n} \rho_0 \left[\{\mu_i' A_i' \Sigma^{-1} A_i \mu_i\}^{\frac{1}{2}}\right] - b\right]$$

where b=EΦ(ρ(μ)) and Φ is the standard normal distribution. The S estimator satisfy the following equation

$$\beta_S = \left(X'(\Sigma^{-1} \otimes W_\mu)X\right)^{-1} X'(\Sigma^{-1} \otimes W_\mu)Y \quad (5)$$

$$\Sigma = q(\bar{\bar{Y}} - \bar{\bar{X}}\bar{\bar{\beta}})' W_\varepsilon (\bar{\bar{Y}} - \bar{\bar{X}}\bar{\bar{\beta}}) / \textstyle\sum_{i=1}^{n} \upsilon(w_i) \quad (6)$$

Where

$w_i^2 = \mu' A_i' \Sigma^{-1} A_i \mu$, i=1,2,……,n , $W_\varepsilon = \mathrm{diag}\left(\frac{\rho_0'(w_i)}{w_i}\right)$,

$\sum_{i=1}^{n} \frac{\rho_0'(w_i)}{w_i} A_i' \Sigma^{-1} A_i' = (\Sigma^{-1} \otimes W_\varepsilon)$ and $\upsilon(w) = \rho_0'(w)\, w - \rho_0'(w) + b$.

Lemma (1):

Let the ρ_1 and ρ_1 have symmetric, continuous, differentiable, for $C > 0$ it is strictly increasing on [0; c], constant on [c,∞] and ρ(0)= 0. In adding, the function ρ_o and ρ_1 achieve $\rho_1(\mu) \le \rho_o(\mu)$ and $\mathrm{Sup}\, \rho_1(\mu) = \mathrm{Sup}\, \rho_0(\mu)$. We use Huber function at the form (4). In the SUR model in (1). We can extracted the MM-estimator for SUR model (new) by

$$\min_{(\beta,\Sigma)} \left[\frac{1}{n}\textstyle\sum_{i=1}^{n} \rho_1\left(\left(\bar{\bar{\mu}}_i' \Sigma_s^{-1} \bar{\bar{\mu}}_i\right)^{1/2}\right)\right] \quad (7)$$

Where, Σ_s estimated by

$$\min_{(\beta,\Sigma)} \left[\frac{1}{n}\textstyle\sum_{i=1}^{n} \rho_0\left[\left(\bar{\bar{\mu}}_i(\beta_0)' \Sigma_0^{-1} \bar{\bar{\mu}}_i(\beta_0)\right)^{1/2}\right] = b\right] \quad (8)$$

Where β_0 and Σ_0 are initial estimators.

Then the MM-estimator for SUR model satisfy the following equations

$$\beta_{mm\ SUR} = \left(X'(\Sigma_s^{-1} \otimes W_{mm})X\right)^{-1} X'(\Sigma_s^{-1} \otimes W_{mm})Y \quad (9)$$

$$\Sigma_s = q(\bar{\bar{Y}} - \bar{\bar{X}}\bar{\bar{\beta}})' W_{mm} (\bar{\bar{Y}} - \bar{\bar{X}}\bar{\bar{\beta}}) / \textstyle\sum_{i=1}^{n} \upsilon(w_{mm\,i}) \quad (10)$$

Where

${w_{mm}}_{i}^{2} = \mu' A_i' \Sigma_s^{-1} A_i \mu$, i=1,2,……,n,

$W_{mm} = \mathrm{diag}\{\varepsilon(w_{mm\,i})\}$, $\varepsilon(W_{mm}) = \rho 1'(W_{mm}) / W_{mm}$,

$\upsilon(W_{mm}) = \rho 1'(W_{mm})d - \rho_1(W_{mm}) + b$ and $E_\Phi\left(\rho_0(\mu)\right) = b$.

Proof of lemma (1)

If we differentiate (7) with respect to β and equalize the result to zero then

$$-\textstyle\sum_{i=1}^{n} \rho_1' \left\{\left[\mu_i(\beta)' A_i' \Sigma_s^{-1} A_i \mu_i(\beta)\right]^{\frac{1}{2}}\right\} \left\{\left[\mu_i(\beta)' A_i' \Sigma_s^{-1} A_i \mu_i(\beta)\right]^{-\frac{1}{2}}\right\}$$

$\left(X'^{A_i}{}'\Sigma_s{}^{-1}A_iX\beta_i - X'A_i'\Sigma_s{}^{-1}A_iY\right) = 0$

$\sum_{i=1}^{n}\frac{\rho_1'(w_{mm\,i})}{w_{mm\,i}}X'A_i'\Sigma_s{}^{-1}A_iY - \sum_{i=1}^{n}\frac{\rho_1'(w_{mm\,i})}{w_{mm\,i}}$

$X'A_i'\Sigma_s{}^{-1}A_iX\beta = 0$

Where $w_{mm\,i}^{\;2} = \mu'A_i{}'\Sigma_s{}^{-1}A_i\mu$, $i = 1,2,\ldots\ldots,n$

Then $X'(\Sigma_s{}^{-1}\otimes W_{mm})Y = X'(\Sigma_s{}^{-1}\otimes W_{mm})X\beta$

Where $(\Sigma_s{}^{-1}\otimes W_{mm}) = \sum_{i=1}^{n}\frac{\rho_1'(w_{mm\,i})}{w_{mm\,i}}A_i'\ \Sigma_s{}^{-1}\ A_I$

and W_{mm}=diag$\left[\frac{\rho_1'(w_{mm\,1})}{w_{mm\,1}},\ldots,\frac{\rho_1'(w_{mm\,n})}{w_{mm\,n}}\right]$

Then complete the proof.

Lemma (2)

If ρ_1 and ρ_1 satisfy the condition in lemma (1), then we extract MM ridge estimator for SUR model (new) by

$$\min_{(\beta,\Sigma)}\left[\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\left(\bar{\bar{\mu}}_i'\Sigma_s{}^{-1}\bar{\bar{\mu}}_i\right)^{\frac{1}{2}}\right) + \lambda\sum_{i=1}^{q}\sum_{j=1}^{k}\beta_{ij}{}^2\right] \quad (11)$$

Where, Σ_s estimated by (8) and λ is ridge parameter. The MM ridge estimator for SUR model (new) in (1) can be written as

$$\beta_{mm\ ridge\ SUR} = \left(X'(\Sigma_s{}^{-1}\otimes W_{mm})X + \lambda I_{kq}\right)^{-1}X'(\Sigma_s{}^{-1}\otimes W_{mm})Y \quad (12)$$

Proof of lemma (2)

If we differentiate (11) with respect to β and equalize the result to zero then:

$$-\sum_{i=1}^{n}\rho_1{}'\left[\{A_i'\mu_i(\beta)'\Sigma^{-1}A_i\mu_i(\beta)\}^{\frac{1}{2}}\right]\left[\left(A_i'\mu_i(\beta)'\Sigma^{-1}A_i\mu_i(\beta)\right)^{-\frac{1}{2}}\right]$$

$$\left[X'A_i{}'\Sigma^{-1}A_iX\beta_i - X'A_i'\Sigma^{-1}A_iY\right] + \lambda\sum_{i=1}^{q}\sum_{j=1}^{K}\beta_{ij} = 0$$

$$\sum_{i=1}^{n}\frac{\rho_1'(w_{mm\,i})}{w_{mm\,i}}X'A_i'\Sigma_s{}^{-1}A_iY - \sum_{i=1}^{n}\frac{\rho_1'(w_{mm\,i})}{w_{mm\,i}}X'A_i'\Sigma_s{}^{-1}A_iX\beta + \lambda\beta = 0$$

Then complete the proof.

3. Asymptotic Properties for MM and MM Ridge Estimators for SUR Model

In this section, we study the asymptotic properties MM and MM ridge estimators.

Lemma (3)

Consider the SUR model with Z observe, $Z = (y_{ij}, x_{ij})$, $1 \le i \le n$, $1 \le j \le q$ which independent random vectors distributed as pq-variate normal distribution $\mathcal{H}_0(Z) = \mathcal{M}_0(X)_0(\mu)$ with mean μ= Zβ and variance ∑s, where $\mathcal{M}_0(X)$ is distribution of X and $G_0(\mu)$ is distribution of μ.

Then

$$\sqrt{n}\left(\beta_{mm\ SUR} - \beta\right)\overset{d}{\rightarrow}$$
$$N\left[0, (E(\Psi^2(|\Sigma_s{}^{-1}\bar{\bar{\mu}}|))/q\beta^2)E(Z'\Sigma_s{}^{-1}Z)^{-1}\right]$$

Where

$\Psi(\mu) = \partial\rho(\mu)/\partial\mu$,

$\beta = E\left\{\left(1-\frac{1}{q}\right)u_{mm}\left(|\Sigma_s{}^{-1/2}\bar{\bar{\mu}}|\right) + \Psi'\left(|\Sigma_s{}^{-1/2}\bar{\bar{\mu}}|\right)\right\}$

Proof of lemma (3):

We use the proof for the Theorem (4-1) in (Yohai, 1987) and the asymptotical properties for S-estimator for SUR model in (Bilodeau etal, 2000) to access the prove. The MM-estimator of β,∑, in SUR model, can be defined as a solution of M-type equation, then the anther form for equations (8) and (9) are

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta,\Sigma) = 0,\ \Psi = (\Psi_1', \Psi_2')' \quad (14)$$

Where:

$\Psi_1(Z,t,\beta,\Sigma) = u_{MM}(d)\left(Z\,\Sigma_s{}^{-1}(t - Z\beta)\right)$,

$\Psi_2(Z,t,\beta,\Sigma) = vec\{qu_{MM}(d)(t - Z\beta)(t - Z\beta)' - v_{MM}(d)\Sigma_s\}$

By using Mean Value Theorem (MVT),

$\exists\beta_n^* \in (\hat{\beta}_n, \beta_0)\ S.th$

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\hat{\beta}_n,\Sigma_s) = \frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_0,\Sigma_s) + \frac{1}{n}\sum_{i=1}^{n}\Psi_{\beta'}(Z,t,\beta_n^*,\Sigma_s)\,(\beta_n^* - \beta_0)$$

For (14)

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_0,\Sigma_s) + \frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_n^*,\Sigma_s)\,(\beta_n^* - \beta_0) = 0$$

Then

$$\sqrt{\boldsymbol{n}}\left((\beta_n^* - \beta_0)\right) = -\left(\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_n^*,\Sigma_s)\right)^{-\mathbf{1}}\frac{\mathbf{1}}{\sqrt{\boldsymbol{n}}}\sum_{i=1}^{n}\Psi(Z,t,\beta_0,\Sigma_s)$$

For uniform laws of large numbers (ULLN)

$$\operatorname*{plim}_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_n^*,\Sigma_s) - E[\Psi(Z,t,\beta_n^*,\Sigma_s)]\right| = 0$$

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_n^*,\Sigma_s)$$
$$= \operatorname*{plim}_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta_n^*,\Sigma_s) - E[\Psi(Z,t,\beta_n^*,\Sigma_s)]\right|$$
$$+\operatorname*{plim}_{n\to\infty}\left|E[\Psi(Z,t,\beta_n^*,\Sigma_s)] - E[\Psi(Z,t,\beta_0,\Sigma_s)]\right|$$
$$+E[\Psi(Z,t,\beta_0,\Sigma_s)]$$

By central limit theorem (CLT), we get

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_1(Z,t,\beta_n^*,\sum_s)\xrightarrow{d}$$

$$N(0,[\mathrm{Var}(\Psi_1(Z,t,\beta,\sum)),\mathrm{Var}(\Psi_2(Z,t,\beta,\sum))])$$

And then

$$\sqrt{\boldsymbol{n}}\left((\beta_n^*-\beta_0)\right)\xrightarrow{d}N$$

$$\left(0,\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum)]^{-1}\mathrm{Var}\left(\Psi_1(Z,t,\beta_0,\sum_s)\right)\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum)]^{'-1}\right)$$

Under the assumption u elliptical. We study the asymptotic variance for β_{mm}.

Hence

$$\mathrm{Cov}(\Psi_1,\Psi_1)=\mathrm{E}\left[qu_{MM}(d)v_{MM}(d)Z'{\sum_s}^{-1}(t-Z\beta)\right]$$

$$-E\left[u_{MM}(d)Z'{\sum_s}^{-1}(t-Z\beta)\right]\mathrm{Vec}\left((t-Z\beta)'(t-Z\beta)\right)=0$$

Let:

$$\mathrm{E}[\Psi(Z,t,\beta_0,\sum)]^{-1}\mathrm{Var}\left(\Psi(Z,t,\beta_0,\sum_s)\right)\mathrm{E}[\Psi(Z,t,\beta_0,\sum)]^{'-1}$$

$$=\begin{bmatrix}\mathrm{E}[\Psi_1(Z,t,\beta,\sum_s)]^{-1} & 0\\ 0 & \mathrm{E}[\Psi_2(Z,t,\beta,\sum_s)]^{-1}\end{bmatrix}$$

$$\begin{bmatrix}\mathrm{Var}(\Psi_1(Z,t,\beta,\sum_s)) & 0\\ 0 & \mathrm{Var}(\Psi_2(Z,t,\beta,\sum_s))\end{bmatrix}$$

$$\begin{bmatrix}\mathrm{E}[\Psi_1(Z,t,\beta,\sum)]^{-1} & 0\\ 0 & \mathrm{E}[\Psi_2(Z,t,\beta,\sum)]^{-1}\end{bmatrix}'$$

while $u_{MM}(d){\Sigma_s}^{-1/2}(t-Z\beta)$ is spherical then

$$\mathrm{Var}\left({\Sigma_s}^{-1/2}(t-Z\beta)\right)=\alpha I$$

$$\propto=\frac{1}{q}\mathrm{E}\left[\Psi^2\left(\left|{\sum_s}^{-1/2}(t-Z\beta)\right|\right)\right]$$

$$-\mathrm{E}\left[\Psi\left(\left|{\sum_s}^{-1/2}(t-Z\beta)\right|\right)\right]$$

$$=\frac{1}{q}\mathrm{E}\left[\Psi^2\left(\left|{\sum_s}^{-1/2}(t-Z\beta)\right|\right)\right]$$

and then $\mathrm{Var}(\Psi_1(Z,t,\beta,\sum_s))=\alpha\mathrm{E}(Z'{\sum_s}^{-1}Z)$.

Extension for Lemma (5-1) (Lopuhaa, 1989). We get

$$\mathrm{E}[\Psi_1(Z,t,\beta,\sum)]^{-1}=-\frac{1}{\beta}\left[\mathrm{E}(Z'{\sum_s}^{-1}Z)\right]^{-1}$$

Where

$$\beta=\mathrm{E}\left\{\left(1-\frac{1}{q}\right)u_{mm}\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)+\Psi'\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)\right\}$$

Then

$$\mathrm{E}[\Psi(Z,t,\beta_0,\sum)]^{-1}\mathrm{Var}(\Psi(Z,t,\beta_0,\sum_s))\mathrm{E}[\Psi(Z,t,\beta_0,\sum)]^{-T}$$

$$=\begin{bmatrix}-\frac{1}{\beta}\left[\mathrm{E}(Z'{\sum_s}^{-1}Z)\right]^{-1} & 0\\ 0 & \mathrm{E}[\Psi_2(Z,t,\beta,\sum_s)]^{-1}\end{bmatrix}$$

$$\begin{bmatrix}\alpha\mathrm{E}(Z'{\sum_s}^{-1}Z) & 0\\ 0 & \mathrm{Var}(\Psi_2(Z,t,\beta,\sum_s))\end{bmatrix}$$

$$\begin{bmatrix}-\frac{1}{\beta}\left[\mathrm{E}(Z'{\sum_s}^{-1}Z)\right]^{-1} & 0\\ 0 & \mathrm{E}[\Psi_2(Z,t,\beta,\sum_s)]^{-1}\end{bmatrix}'$$

Then

$$\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum)]^{-1}\mathrm{Var}\left(\Psi_1(Z,t,\beta_0,\sum_s)\right)\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum)]^{-T}$$

$$=(\alpha/\beta^2)\left[\mathrm{E}(Z'{\sum_s}^{-1}Z)\right]^{-1}$$

Then completes the proof.

Lemma (4):

Consider the same condition in proposition (3-1) and $n^{-1/2}\lambda=\lambda_0$ then

$$\sqrt{n}\left(\beta_{\text{mm ridge SUR}}-\beta\right)\xrightarrow{d}N$$

$$\left[0,(\mathrm{E}(\Psi^2(\left|{\sum_s}^{-1}\bar{\bar{\mu}}\right|+2\lambda_0\beta)/q\beta^2)\mathrm{E}(Z'{\sum_s}^{-1}Z)^{-1}\right] \quad (15)$$

Where $\Psi(\mu)=\partial\rho(\mu)/\partial\mu$,

$$\beta=\mathrm{E}\left\{\left(1-\frac{1}{q}\right)u\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)+\Psi'\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)\right\}$$

and then $\mathrm{Var}[\Psi_1(Z,t,\beta,\sum_s)]=\alpha\mathrm{E}(Z'{\sum_s}^{-1}Z)$. Following by the proof for lemma (3) then completes the proof.

$$=\frac{1}{q}E\left[\Psi^2\left(\left|{\Sigma_s}^{-1/2}(t-Z\beta)\right|+2\lambda_0\beta\right)\right]$$

Where

$$\Psi(\mu)=\partial\rho(\mu)/\partial\mu,$$

$$\beta=\mathrm{E}\left\{\left(1-\frac{1}{q}\right)u_{mm}\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)+\Psi'\left(\left|{\sum_s}^{-1/2}\bar{\bar{\mu}}\right|\right)\right\}$$

and $u(\mu)=[\partial\rho(\mu)/\partial\mu]/\mu$

Proof of lemma (4):

The MM ridge estimator of $\beta,\sum$, in SUR model, can be defined as a solution of M-type equation, then the anther form for equations (9) and (12) are

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(Z,t,\beta,\Sigma_s,\lambda)=0,\ \Psi=\left(\Psi_1',\Psi_2'\right)' \quad (16)$$

Where

$$\Psi_1(Z,t,\beta,\Sigma_s,\lambda)=u_{MM}(d)Z'{\Sigma_s}^{-1}(t-Z\beta)+2\lambda\beta,$$

$$\Psi_2(Z,t,\beta,\sum)=\mathrm{vec}\{qu_{MM}(d)(t-Z\beta)(t-Z\beta)'-v_{MM}(d)\sum_s\}$$

By using (MVT, (ULLN) and (CLT)) theorem in proofs for proportion (3-2). In adding, consider the condition $n^{-1/2}\lambda=\lambda_0$ then

$$\sqrt{n}\left((\beta_n^*-\beta_0)\right)\xrightarrow{d}N\left(0,\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum_s,\lambda_0)]^{-1}\right)$$

$$\mathrm{Var}[\Psi_1(Z,t,\beta_0,\sum_s,\lambda_0)]\mathrm{E}[\Psi_1(Z,t,\beta_0,\sum_s,\lambda_0)]'^{-1}$$

Under the assumption u elliptical. We study the asymptotic variance for $\hat{\beta}$.

Hence

$$\operatorname*{plim}_{n\to\infty}\left[\frac{1}{n}\mathrm{Cov}(\Psi_1,\Psi_1)\right]$$

$$=\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathrm{E}\left[qu_{MM}(d)Z'{\sum_s}^{-1}(t-Z\beta)+2\lambda_0\beta\right]$$

Let:

$$\mathrm{E}[\Psi(Z,t,\beta_0,\sum)]^{-1}\mathrm{Var}[\Psi(Z,t,\beta_0,\sum_s)]\mathrm{E}[\Psi(Z,t,\beta_0,\sum_s)]^{'-1}$$

$$=\begin{bmatrix}E[\Psi_1(Z,t,\beta,\Sigma_s,\lambda)]^{-1} & 0\\ 0 & E[\Psi_2(Z,t,\beta,\sum_s)]^{-1}\end{bmatrix}$$

$$\begin{bmatrix} Var(\Psi_1(Z,t,\beta,\Sigma_s,\lambda)) & 0 \\ 0 & Var(\Psi_2(Z,t,\beta,\Sigma_s)) \end{bmatrix}$$

$$\begin{bmatrix} E[\Psi_1(Z,t,\beta,\Sigma_s,\lambda)]^{-1} & 0 \\ 0 & E[\Psi_2(Z,t,\beta,\Sigma_s,\lambda)]^{-1} \end{bmatrix}'$$

While $u_{MM}(d){\Sigma_s}^{-1/2}(t-Z\beta)+2\lambda_0\,\beta$ is spherical then $Var[\,u_{MM}(d){\Sigma_s}^{-1/2}(t-Z\beta)+2\lambda_0\,\beta\,]=\alpha I$

Where

$$\propto=\frac{1}{q}E[\Psi^2(|{\Sigma_s}^{-1/2}(t-Z\beta)+2\lambda_0\,\beta|)]$$
$$-\frac{1}{q}E[\Psi(|{\Sigma_s}^{-1/2}(t-Z\beta)+2\lambda_0\,\beta|)]$$
$$=\frac{1}{q}E(\Psi^2(|{\Sigma_s}^{-1/2}(t-Z\beta)|+2\lambda_0\,\beta))$$

and then $Var(\Psi_1(Z,t,\beta,\Sigma_s))=\alpha E(Z'{\Sigma_s}^{-1}Z)$.

Following by the proof for lemma (3) then completes the proof.

4. The Simulation Study

In this section, we provide a simulation study to illustrate the performance of four estimators, the (GLS), S, MM and MM ridge estimator for SUR model. This simulation process is executed to generate data for the following equation $y_{li}=\sum_{j=1}^{k}x_{lij}\,\beta_{ij}+\mu_{li}\,, l=1,2,\dots,n\,, i=1,2,\dots,q$. Where $x_{li1}=1$. In this simulation, we set the initial value for β= [1,2,3] for k=3. The explanatory variables are generated by multivariate normal distribution MNNk=3 (0,∑x) where diag(∑x)=1, off-diag(∑x)= ρX= 0.15 for low interdependency and ρx= 0.70 for high interdependency. Where ρx is correlation between explanatory variables. We chose two sample size 25 for small sample and 100 for large sample. The specific error in equations μi, i=1,2,…..,n, we generated by MVNk=3 (0, ∑ε), ∑ε the variance covariance matrix of errors, diag(∑ε)= 1, off-diag(∑ε)= ρε= 0.15. To investigate the robustness of the estimators against outliers, we chosen different percentages of outliers (20%, 45%). We choose shrink parameter in (12) by minimize the new robust Cross Validation (CV_{MM}) criterion which avoided outliers. This criterion depend on MM estimators and takes the form $CV_{MM}(\lambda_d)=\left[\frac{\mu_{MM}{}'\mu_{MM}}{\left(1-\frac{tr(H(\lambda))}{nq}\right)^2}\right]$. Where μ_{MM} is a residual depend on MM estimators and

$$H(\lambda)=X\left({\Sigma_s}^{-\frac{1}{2}}\otimes W_{mm}\right)$$
$$\left(X'({\Sigma_s}^{-1}\otimes W_{mm})X+\lambda I_{qk}\right)^{-1}\left({\Sigma_s}^{-\frac{1}{2}}\otimes W_{mm}\right)X'$$

We measure the goodness of fit of estimated model for β by use the median average squared error [Median(ASE_j)] where ASE_j is defined by

$$ASE_j=\left[\frac{1}{nq}\left(X_j\beta_0-(X_j\breve{\beta}_j)\right)'\left(X_j\beta_{j,0}-(X_j\breve{\beta}_j)\right)\right]$$

where j=1,2,……N. Where β̆ is MM estimator for β and N is a number of iteration. We run N=1,2,…, 500 replications by using software MATHCAD.

5. Algorithm

In this paper, we use two algorithm to compute the MM and MM ridge estimators. β and Σ_s is initial estimate for β and Σ_s respectively, then iterative solve for (9) is

$$\beta_{j+1}\left(X'\ddot{W}_{mm\,j}X\right)^{-1}=X'\ddot{W}_{mm\,j}Y \quad (16)$$

And the then iterative solve for (12) is

$$\beta^*{}_{j+1}\left(X'\ddot{W}_{mm\,j}X+\lambda I_{kq}\right)^{-1}=X'\ddot{W}_{mm\,j}Y\,, j=1,\dots,N \quad (17)$$

Where N is a number of iteration,

$${W_{mm}}^2_{i\,j}=\mu'{A_i}'{\Sigma_{0_s}}^{-1}A_i\mu\ ,$$

$$W_{mm\,j}=diag\left(\frac{\rho_1'(W_{mm\,ij})}{W_{mm\,ij}}\right), i=1,2,\dots,n$$

$$and\ \ddot{W}_{mm\,j}=\left({\Sigma_s}^{-1}\otimes W_{mm\,j}\right).$$

The equation (17) is a weight version for ridge estimator. This equation is rewrite as normal equations

$$\beta_{\,j+1}\left(X_{j,\lambda}'X_{j,\lambda}\right)^{-1}=X_{j,\lambda}'Y_j\,, j=1,\dots,N \quad (18)$$

Where N is a number of iteration,

$${W_{mm}}^2_{i\,j}=\mu'A_i'{\Sigma_{0_s}}^{-1}A_i\mu,$$

$$W_{mm\,j}=diag\left(\frac{\rho_1'(W_{mm\,ij})}{W_{mm\,ij}}\right), i=1,2,\dots\dots,n$$

and $\ddot{W}_{mm\,j}=\left({\Sigma_s}^{-1}\otimes W_{mm\,j}\right)$

The equation (14) is a weight version for ridge estimator. This equation is rewrite as normal equations

$$\beta_{\,j+1}\left(X_{j,\lambda}'X_{j,\lambda}\right)^{-1}=X_{j,\lambda}'Y_j\,, j=1,\dots,N \quad (15)$$

Where $X_{j,\lambda}=\begin{bmatrix}{\ddot{W}_{mm\,j}}^{1/2}X\\ \sqrt{\lambda}I_{kq}\end{bmatrix}$ and $Y_j=\begin{bmatrix}{\ddot{W}_{mm\,j}}^{1/2}Y\\ 0_{kq\times 1}\end{bmatrix}$

Algorithm (1)

In this algorithm, we compute MM estimators for SUR model $(\breve{\beta},\breve{\Sigma})$. by developed the algorithms for (Tharmaratnam et al., 2008) and (Al-hosany et al., 2011) We drawing the algorithm (1) by the following steps:

Step (1): Let $\tilde{\beta}_i^{(0)}, j=1,.,J$ is initial candidates estimate for β.

Step (2): Design the variable (X_l,μ_l), l=1,2,…,q. For each $\tilde{\beta}_j^{(0)}$:

a. Compute $\hat{W}_\varepsilon(\hat{\beta}_j^{(0)})$, $\hat{\Sigma}(\hat{\beta}_j^{(0)})$

b. Compute S estimators as in (8) for some ρ-function ρ0 by set m=0, and get the following steps:

i. Let

$$\hat{\beta}_{S\,j}^{(m+1)} = \left(X'\left(\hat{\Sigma}^{-1}(\hat{\beta}_j^{(0)}) \otimes \hat{W}_{\varepsilon}(\hat{\beta}_j^{(0)})\right)X\right)^{-1} X'\left(\hat{\Sigma}^{-1}(\hat{\beta}_j^{(0)}) \otimes \hat{W}_{\varepsilon}(\hat{\beta}_j^{(0)})\right)Y$$

ii. If either m equal the maximum number of iterations or $\left\|\hat{\beta}_{S\,j}^{(m)} - \hat{\beta}_{S\,j}^{(m+1)}\right\| < ℃ \left\|\hat{\beta}_{S\,j}^{(m)}\right\|$ where $℃ > 0$ is affixed small

iii. constant, then $\hat{\beta}_{s_j}^{f} = \hat{\beta}_{S\,j}^{(m)}$ go to step (2)

iv. Else Compute $\hat{W}_{\varepsilon}\left(\hat{\beta}_{S\,j}^{(m+1)}\right)$, $\hat{\Sigma}\left(\hat{\beta}_{S\,j}^{(m+1)}\right)$ and put m←m+1

Step (3): Select $\hat{\beta}_S$ which active

$$\min_{1 \le j \le J} \left[\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\left(\bar{\bar{\mu}}_i{}'\hat{\Sigma}^{-1}\left(\hat{\beta}_{s_j}^{f}\right)\bar{\bar{\mu}}_i\right)^{1/2}\right)\right]$$

Step(4): Compute MM estimator as in (7) for some ρ-function ρ1: Let g=0 and compute $\hat{W}_{mm}\left(\tilde{\beta}_j^{(0)}\right)$ and let $\Sigma_s{}^{-1} = \hat{\Sigma}^{-1}\left(\hat{\beta}_{s_j}^{f}\right)$

$$\beta_{j+1}{}^{(g+1)} = \left[X'\hat{W}_{mm}\left(\hat{\beta}_j^{(g)}\right)X\right]^{-1} X'\ddot{W}_{mm\,j}{}^{(g)}Y$$

Step (5): If either m equal the maximum number of iterations

or $\left\|\hat{\beta}_{mm\,j}^{(g)} - \hat{\beta}_{mm\,j}^{(g+1)}\right\| < ℃ \left\|\hat{\beta}_{mm\,j}^{(g)}\right\|$ where $℃ > 0$ is affixed small constant, then

$\hat{\beta}_{mm\,j}^{f} = \hat{\beta}_{mm\,j}^{(g)}$ break

Else Compute $\hat{W}_{\varepsilon}\left(\hat{\beta}_{mm\,j}^{(m+1)}\right)$ and put g←g+1

Step (6): Select $\hat{\beta}_{mm}$ which active

$$\min_{1 \le j \le J} \left[\frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\left(\bar{\bar{\mu}}_i{}'\Sigma_s{}^{-1}\bar{\bar{\mu}}_i\right)^{1/2}\right)\right]$$

The J random subsample for initial candidates $\tilde{\beta}_i^{(0)}, j = 1, \ldots, J$ ware chosen as set for OLS estimators. To avoid ill condition, we take J by large size.

Algorithm (2)

In this algorithm, we compute MM ridge estimators for SUR model $(\breve{\beta}, \breve{\Sigma})$ by develop (Maronna R,2011), (Tharmaratnam et al., 2008) and (Al-hosany et al., 2011) algorithms.

We drawing the algorithm (2) by the following steps:

Step (1): We use the steps (1,2,3) in algorithm (1).

Step (2): Compute MM ridge estimator as in (11) for some ρ-function ρ_1. Let g = 0 and compute $\hat{W}_{mm}\left(\hat{\beta}_j^{(0)}\right)$, λ (2), $\Sigma_s{}^{-1} = \hat{\Sigma}^{-1}\left(\hat{\beta}_{s_j}^{f}\right)$ and let

$$\beta_{\,j+1} = \left(X_{j,\lambda}{}^{(0)\prime}X_{j,\lambda}{}^{(0)}\right)^{-1} X_{j,\lambda}{}'Y_jY$$

Step (3): If either m equal the maximum number of iterations

Or

$$\left\|\hat{\beta}_{mmridge\;j}^{(g)} - \hat{\beta}_{mmridge\;j}^{(g+1)}\right\| < ℃ \left\|\hat{\beta}_{mmridge\;j}^{(g)}\right\|$$

where $℃ > 0$ is affixed small constant, then $\hat{\beta}_{mmridge\;j}^{f} = \hat{\beta}_{mmridge\;j}^{(g)}$ break

Else Compute $\hat{W}_{\varepsilon}\left(\hat{\beta}_{mm\,j}^{(g+1)}\right)$ and put g←g+1

Step (4): Select $\hat{\beta}_{mm}$ active

$$\left[\frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\left(\bar{\bar{\mu}}_i{}'\Sigma_s{}^{-1}\bar{\bar{\mu}}_i\right)^{1/2}\right) + \lambda \sum_{i=1}^{q}\sum_{j=1}^{k} \hat{\beta}_{mmridge\;ij}^{f\;2}\right]$$

6. Result for Simulation

Table (1). Median ASE for GLS, S, MM and MM ridge estimators for SUR model, (percentages of outliers(ν) = 20%

Estimators	q	ρx	n	Median (ASE)
GLS S MM MM ridge	10	0.15	25	0.204 0.215 0.211 0.201
GLS S MM MM ridge			100	0.194 0.199 0.196 0.196
GLS S MM MM ridge		0.70	25	0.114 0.119 0.115 0.115
GLS S MM MM ridge			100	0.098 0.101 0.098 0.098
GLS S MM MM ridge	2	0.15	25	0.201 0.205 0.202 0.200
GLS S MM MM ridge			100	0.190 0.195 0.192 0.191
GLS S MM MM ridge		0.70	25	0.122 0.129 0.126 0.125
GLS S MM MM ridge			100	0.086 0.091 0.089 0.088

Table (2). Median ASE for (GLS), S, MM and MM ridge estimator for SUR model, (percentages of outliers(ν) = 45%

Estimators	q	ρ_x	n	Median (ASE)
GLS	10	0.15	25	3.022
S				2.024
MM				2.001
MM ridge				1.987
GLS			100	3.018
S				1.802
MM				1.710
MM ridge				1.662
GLS		0.70	25	2.966
S				1.901
MM				1.792
MM ridge				1.624
GLS			100	2.714
S				1.901
MM				1.882
MM ridge				1.251
GLS	2	0.15	25	2.831
S				1.812
MM				1.692
MM ridge				1.620
GLS			100	1.842
S				1.605
MM				1.591
MM ridge				1.312
GLS		0.70	25	2.741
S				1.901
MM				1.555
MM ridge				1.412
GLS			100	1.692
S				1.512
MM				1.405
MM ridge				1.001

We summarize the simulation results in tables (1, 2). These tables shows the median ASE for GLS,S, MM and MM ridge estimators for SUR model under the study factors. When analysis the result of simulation study, we can say that, the GLS estimators, in all case, works better than anther estimators when the percentages of outliers close to zero percentage. Although S estimator has clear advantage when percentages of outliers reach to 20% and 45%, but the MM estimators have more efficiency. When the number of observations were increasing, we note improvement in the work of all the estimators. On other hand, when the number of equations increase, all estimators going to be less efficient. works good when the percentage of outliers and correlation between explanatory variables increases. The MM estimators and MM ridge estimators works good when the percentage of outliers and correlation between explanatory variables increases. Nevertheless, MM ridge estimators get more efficient than MM estimators.

7. Conclusions

To reach the best estimator for the SUR model which have advantages of the robust parameter estimation, we compare between the robust and nonrobust estimators. To attain the robust high level break down point, we used S, MM and MM ridge estimator for SUR model. The MM ridge estimators avoids the multicollinearity problem, which accompany the outliers. The median average squared error [Median (ASE)] was used for trade-off between estimators. The result of simulation show that, when the percentage of outlier and correlation between explanatory are increasing, the MM and MM ridge estimators are the best estimators between the robust and nonrobust estimators under the all factors of study.

REFERENCES

[1] Berrendero J., Mendez B. and Tyler D. (2007), “On the maximum bias functions of MM-estimates and constrained M-estimates of regression”, The Annals of Statistics, Vol.35, No. 1.

[2] Bilodeau M. and Duchesne P. (2000), "Robust estimation of the SUR model", The Canadian Journal of Statistics, Vol. 28.

[3] Copt S. and Heritier S. (2006), “Robust MM-estimation and inference in mixed linear models”, Biometrics, Vo.1.

[4] El- Houssainy A., Sayed M., Alaa A., Naglaa A. and Tarek m. (2011), "Robust cross validation in SUR ridge etimators and SUR robust ridge estimators", Journal of statistic theory and application, Vol. 10, No.1.

[5] Garcia M., Martinez E. and Yohai V. (2006), “Robust estimation for the multivariate linear model based on a τ –scale”, Journal of Multivariate Analysis, Vo. 97.

[6] Hendrik P. and Lopuha Äa, H. P. (1992), “Highly efficient estimators of multivariate location with high breakdown Point’, The Annals of Statistics, Vol. 20, No. 1.

[7] Jose R., Beatriz V. and David E. (2007), “On the maximum bias functions of MM-estimates and constrained M-estimates of regression”, The Annals of statistics, Vo. 35.

[8] Koenker, R. and Portnoy, S. (1990), "M-estimation of multivariate regressions". Journal of American Statistics Association, Vol. 63.

[9] Kudraszow N. and Moronna R. (2011), Estimates of MM type for the multivariate linear model, Journal of Multivariate Analysis, Vol. 102, No. 9.

[10] Lopuha Äa H. (1989), "On the relation between S-estimators and M-estimators of multivariate location and covariance" The Annals of Statistics, Vol.17.

[11] LopuhaÄa H. (1992), “Highly efficient estimators of multivariate location with high breakdown point", Annals of Statistics, Vo,20.

[12] Maronna R. and Zamar R. (2002), “Robust estimates of location and dispersion for high-dimensional datasets”, Technometrics, Vo.44, No.4.

[13] Maronna R. (2011) “Robust ridge regression for high-dimensional Data”, Technometrics, Vo.53, No.1.

[14] Meriam S., Said M., Iqbal M. and Ismail B. (2012),” Weight ridge MM-estimator in robust ridge regression with Multicollinearity”, Mathmatical Models and Methods in Modem Science Conference, 14th.

[15] Moawad E. and Abd E. (2011), ”Estimation methods for multicollinearity problem combined with high leverage data points”, Journal of Mathematics and Statistics, Vo.7, No. 2.

[16] Roelandt E., Van S. and Croux C. (2009). Multivariate Generalized S estimators, Journal of Multivariate Analysis, Vo.100, No.5.

[17] Rousseeuw, P. and Leroy. A (1987), “Robust regression and outlier detection”, John Wiley, New York.

[18] Ruppert D. (1992). "Computing S estimators for regression and multivariate location/dispersion", Journal of computational and graphical statistics, Vol. 32.

[19] Salibian B. and Yohai V. (2008) "High breakdown point robust regression with censored data", The Annals of Statistics, vol. 36.

[20] Srivastava V. and Giles D. (1987), “Seemingly unrelated regression equations Models: Estimation and Inference”, New York: Marcel Dekker.

[21] Tharmaratnam, K. and Claeskens G. (2008). "S-Estimation for Penalized Regression Splines", Comput. Statist., Vol.30.

[22] Yohai V. (1987), "High breakdown point and high efficiency robust estimates for regression," Annals of Statistics, Vo. 15.

[23] Zellner A. (1962), “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias”, Journal of American Statistics Association, Vol. 57.