CESIS

Centre of Excellence for Science and Innovation Studies

CESIS Electronic Working Paper Series

Paper No. 258

Median Regression for SUR Models with the Same Explanatory Variables in Each Equation

Ghazi Shukur

Zangin Zeebari

October 2011

The Royal Institute of technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

Median Regression for SUR Models with the Same Explanatory Variables in Each Equation

By

Ghazi Shukur and Zangin Zeebari

Department of Economics, Finance and Statistics,
Jönköping International Business School,

P.O. Box 1026, SE-55111 Jönköping, Sweden

Ghazi.shukur@jibs.hj.se   Zangin.zeebari@jibs.hj.se

Abstract

In this paper we introduce an interesting feature of the Generalized Least Absolute Deviations (GLAD) method for Seemingly Unrelated Regression Equations (SURE) models. Contrary to the collapse of Generalized Least Squares (GLS) parameter estimations of SURE models to the Ordinary Least Squares (OLS) estimations of the individual equations when the same regressors are common between all equations, the estimations of the proposed methodology are not identical to the Least Absolute Deviations (LAD) estimations of the individual equations. This is important since contrary to the least squares methods, one can take advantage of efficiency gain due to cross-equation correlations even if the system includes the same regressors in each equation. This kind of methodology is useful say when estimating the factors that affect firms' innovation investments across European countries.

Key-words: Median Regression, Robustness, Efficiency, SURE Models, Innovation Investment

JEL Classification: C30, C31

1. Introduction

The Generalized least squares (GLS) method of estimation for Seemingly Unrelated Regression Equations (SURE) models proposed by Zellner (1962) is considered as one of the most successful and efficient methods for estimating seemingly unrelated regressions. The proposed SURE model has stimulated a countless theoretical and empirical results in econometrics and other areas, (see Zellner, 1962; Rao, (1975), Brown and Payne, 1975; Srivastava and Giles, 1987; Chib and Greenberg, 1995). For example, the methodology is applicable to political behavior such as voting, biometric problems, allocation models, demand functions for a number of commodities (i.e., Almost Ideal Demand Systems, AIDS), investment functions for a number of firms, income or consumption functions for subsets of populations or different regions, to mention some. The efficiency of the GLS estimation over the Ordinary Least Squares (OLS) estimation of SURE models increases when the correlations between the error terms from the different equations included in the system also increase. However, in situations when the system contains the same regressors in each equation, the GLS estimation of SURE models collapses to the OLS estimation of SURE models which in turn is equivalent to the OLS estimation of individual equations. In this case there will not be any gain of efficiency from applying the GLS estimation even when the cross-equation errors are highly correlated.

Shukur and Zeebari (2009) conducted a Generalized Least Absolute Deviations (GLAD) estimation method for SURE models which is more robust and more efficient than the usual GLS estimation method when the distributions of the error terms are not symmetric. The authors also showed that the properties of the GLAD estimations will not deteriorate compared to the GLS estimations of the SURE models when the distributions of the error terms are symmetric. Moreover, they found that even with no correlations between the equations, using GLAD estimations for SURE models instead of Least Absolute Deviations (LAD) estimations of single equation median regressions will not damage the estimations of the parameters.

The most interesting feature of the GLAD estimators is that, contrary to the least-squares SURE formulation, if the system contains the same regressors in each equation the GLAD estimations do not collapse to the LAD estimations of individual equations. This will be of great importance when estimating conventional multivariate regression models with errors of the equations contemporaneously correlated to each other.

The aim of this paper is to further investigate and prove this issue by analytical results and Monte Carlo simulations. Moreover, we conduct an empirical example to illustrate our proposed methodology.

The rest of the paper is organized as follows: In Section 2 we describe the methodology we used in this paper; Section 3 presents the Monte Carlo simulations and results; In Section 4 we give description of the data and the model we used in our empirical example with the results from this part; Section 5, finally, gives conclusions and a summary of findings.

2. Methodology

Except the ease of mathematical tractability of the properties of conditional mean function estimators, one may not prefer the conditional mean function to the conditional median function. If the conditional mean function suffices to estimate the regression parameters when the error terms are symmetrically distributed, such as with the Gaussian distribution, with asymmetrically distributed error terms, other measures of central tendency, like median, could be taken as more suitable alternatives to the mean for location behavior of the errors.

With a sample of n observations for the conditional median function of Y given $\mathbf{x}$ in regression analysis, it is well known that we want a minimizer $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(Y, \mathbf{X})$ that minimizes the sum of absolute deviations,

$$\min_{\boldsymbol{\beta} \in \mathrm{R}^p} \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i' \boldsymbol{\beta} \right|. \tag{2.1}$$

Despite the huge available literature on univariate median regression there is relatively much less work on the multivariate context. Much more work has been done on the least squares estimation methods for the conventional multivariate linear regression analysis when there is a system of linear regression equations. In the multivariate context, the problem is not only due to the computational difficulties with the LAD method of estimation, but also to the definition of multivariate median.

The OLS method of estimating multivariate linear regression deals with each regression equation of the system separately, i.e., it gives the OLS estimations of the parameters of individual regression equations. However, there is a need of special care when the equations are related to each other, since the OLS method does not take into account the correlations between the equations in the multivariate linear regression analysis. A special case is the existence of correlations between contemporaneous cross-equation error terms when the endogenous variable in each regression equation is only a function of the exogenous variables and the error term. Zellner's introduced SURE model is of such case with a change in the structure of the design matrices.

There are some special cases where the GLS estimation of the SURE models collapses to the OLS estimation of SURE models, which is in turn equivalent to the OLS estimation of individual regression equations (Srivastava & Giles 1987, page17). A widely used case in which the GLS estimates of SURE models are equivalent to the OLS estimates of the individual regression equations is the possession of the same values of regressors in each regression equation. Consequently, the correlations between cross-equation errors are not beneficial to the gain of efficiency in GLS estimations.

Consider a general system of M linear regression equations given by

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \qquad i = 1, 2, \ldots, M \tag{2.2}$$

where, $\mathbf{Y}_i$ is a $T \times 1$ vector of the dependent variables, $\mathbf{e}_i$ is a $T \times 1$ vector of random errors with $E(\mathbf{e}_i) = \mathbf{0}$, $\operatorname{var}(\mathbf{e}_i) = \sigma_i^2 \mathbf{I}_T$, $\mathbf{X}_i$ is a $T \times k_i$ matrix of observations on k_i independent variables including a constant term and $\boldsymbol{\beta}_i$ is a $k_i \times 1$ vector of coefficients to be estimated. The equations in the system (2.2) can be combined into a more comprehensive model like

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{pmatrix}. \tag{2.3}$$

The above model can be rewritten compactly as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{e} \tag{2.4}$$

where, $\mathbf{Y}$ and $\mathbf{e}$ are of dimension $TM \times 1$, $\mathbf{X}$ is of dimension $TM \times k$, and finally $\mathbf{B}$ is of the dimension $k \times 1$, with $k = \sum_{i=1}^{M} k_i$. We also suppose that $\mathbf{e}$ has a continuous finite mixture distribution with $E(\mathbf{e}) = \mathbf{0}$ and $E(\mathbf{ee}') = \mathbf{\Sigma} \otimes \mathbf{I}_T$, where $\mathbf{\Sigma} = \left[\sigma_{ij}\right]_{M \times M}$ is positive definite and $\otimes$ is the Kronecker product.

The following transformation is conducted due for the GLS estimation of the SURE models:

$$\mathbf{Y}^* = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\,\mathbf{Y}$$

$$\mathbf{X}^* = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\,\mathbf{X}$$

$$\mathbf{e}^* = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\,\mathbf{e},$$

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^* \tag{2.5}$$

$$\operatorname{var}\left(\mathbf{e}^*\right) = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)(\mathbf{\Sigma} \otimes \mathbf{I}_T)(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T) = \mathbf{I}_{TM}\,.$$

The GLS estimation of a SURE model is the OLS estimation of the transformed SURE model. Usually an estimation, $\mathbf{S}$, of the unknown covariance matrix $\mathbf{\Sigma}$ is used in (2.5).

In our method, instead of using the OLS method conducted on the transformed model, which gives us Aitken's GLS estimates, we use the median regression and get the GLAD estimates. In the OLS method, the 2-norm (squared Euclidian distance) is used to minimize the distance between the observations and estimations, which after transformation is transformed to Mahalanobis distance, whereas in our method, the 1-norm (taxicab or city-block) distance is used and after the transformation, the distance is transformed to a new form of distance underlies the GLAD method. This enables us, while estimating, to take into account the information embedded in the correlations between the cross-equation errors.

Shukur and Zeebari (2009) showed that, due to the structure of the SURE models, the GLS-OLS association in some aspects is reflected in the GLAD-LAD association, unless the case when each equation has the same regressors. For instance, they showed that with no correlations between the equations, using GLAD estimation method instead of LAD estimations does not damage the estimations of the parameters, hence the harmless of GLAD method even in the lack of correlations between the equations.

Another fact is that the OLS estimation of SURE models results in OLS estimations of individual equations. Shukur and Zeebari (2009) showed that the same argument holds when applying LAD method on SURE model, i.e., it is the same as applying the LAD method on individual equations. This means that applying OLS or LAD method on SURE models abandons the information imbedded in the correlation matrix of cross-equation errors.

The minimization problem (2.1) for the GLAD estimation method becomes as follows,

$$\min_{\boldsymbol{\beta}\in \mathrm{R}^k} \sum_{i=1}^{TM} \left| y_i^* - \mathbf{x}_i'^* \boldsymbol{\beta} \right|, \tag{2.6}$$

which can be simplified as

$$\min_{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_M)\in \mathrm{R}^{k_1}\times \mathrm{R}^{k_2}\times\ldots\times \mathrm{R}^{k_M}} \sum_{i=1}^{M}\sum_{j=1}^{T}\left| \sum_{m=1}^{M} \delta^{im}\left(y_{mj} - \mathbf{x}_{mj}' \boldsymbol{\beta}_m \right) \right|, \tag{2.7}$$

where δ^{im} is the im^{th} element of $\boldsymbol{\Sigma}^{-1/2}$, and both of y_{mj} and $\mathbf{x}_{mj}$ are respectively the j^{th} observation on the dependent variable and independent variables of the m^{th} equation, for $i, m = 1,\ldots,M$. With the values of the explanatory variables identical in all equations, the objective function of (2.7) can be further simplified as

$$\sum_{i=1}^{M}\sum_{j=1}^{T}\left| \sum_{m=1}^{M} \delta^{im}\left(y_{mj} - \mathbf{x}_{mj}' \boldsymbol{\beta}_m \right) \right| = \sum_{i=1}^{M}\delta^{ii}\sum_{j=1}^{T}\left| (y_{ij} - \mathbf{x}_j' \boldsymbol{\beta}_i) + \sum_{m=1, m\neq i}^{M} \frac{\delta^{im}}{\delta^{ii}} y_{mj} - \mathbf{x}_j' \sum_{m=1, m\neq i}^{M} \frac{\delta^{im}}{\delta^{ii}} \boldsymbol{\beta}_m \right|. \tag{2.8}$$

Then, the minimizer $\hat{\boldsymbol{\beta}}_k$ of the k^{th} equation in (2.8) will be as follows,

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}(u_k, \mathbf{X}) - \hat{\boldsymbol{\gamma}}_k, \tag{2.9}$$

where

$$u_k = \sum_{m=1}^{M} \frac{\delta^{km}}{\delta^{kk}} y_m \tag{2.10}$$

and

$$\hat{\boldsymbol{\gamma}}_k = \sum_{m=1, m\neq k}^{M} \frac{\delta^{km}}{\delta^{kk}} \hat{\boldsymbol{\beta}}_m . \tag{2.11}$$

The formula (2.9) shows that generally the LAD estimation of the k^{th} individual regression equation $\tilde{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}(Y_k, \mathbf{X})$ is different from the GLAD estimation of that equation $\hat{\boldsymbol{\beta}}_k$. For calculating the LAD estimation of the k^{th} individual regression equation only the dependent variable of that equation is involved, whereas the GLAD estimation of the k^{th} equation is a function of other equations' dependent variables and a linear function of the GLAD estimations of other equations. A rare special case in which the two estimates could be equal is when $u_k = Y_k + X\hat{\boldsymbol{\gamma}}_k$, for $k = 1, \ldots, M$, or when there is no correlation between cross-equation errors. In Section 4, we show the inequality of GLAD and LAD estimation of SURE models through an example.

Here, some questions may arise. Does the GLAD method of estimation result in any gain of efficiency when the values of regressors of all equations are the same? Is the GLAD method preferable to the OLS method even if the cross-equation errors have a symmetric distribution, like a multivariate normal? In the next section we perform a Monte Carlo simulation to answer those questions.

3. Monte Carlo experiment

In our Monte Carlo simulation, we change some factors that intuitively may change the efficiency of the estimators. These factors are: the number of equations (M), the sample size (T), the skewness of intra-equation errors, the correlations between cross-equation errors and the number of regressors (k). Fairly medium-sized systems with five equations and the smallest possible systems (with 2 equations) are included in our simulation (see Table 3.1). Samples of size 30, 100 and 1000 observations that cover small, modest and relatively large samples are generated for each combination of number of equations, number of regressors, skewness level and correlation level.

Different levels of correlations between the equations have been imposed. Without loss of generality but just for simplicity, we suppose the same level of correlation between all pairs of equations. When generating the data, we also impose the same level of skewness of the errors for each equation. A proper choice of degrees of freedom, parameter of skewness and parameter of correlation of the multivariate skew t distributions enables us to generate data sets with the desired distributional properties. First, we generate the design matrix for each equation as shown in (3.1). Next, with the fixed design matrices in each of the 2000 replications of the experiment, we generate a set of errors distributed with a proper multivariate skew t distribution with the desired properties. The data for the dependent variable of each equation is computed by the formula (3.2).

The relative efficiency of the GLAD estimator to the efficiency of GLS estimator, which is in turn reduced to OLS estimator, is computed through the generalized sample variance, total sample variance and the MSE of the estimators (see Table 3.3 to Table 3.4).

The design matrix is generated as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_T & \mathbf{X}^* \end{bmatrix}_{T\times k}, \qquad (3.1)$$

where $\mathbf{X}^*$ is a $T\times(k-1)$ matrix, whose rows, $\mathbf{x}^*_j$, are vectors with the multivariate normal distribution $N_{k-1}\left(\boldsymbol{\mu}, \mathbf{I}_{k-1}\right)$, and $j = 1, 2, \ldots, T$. For simplicity, we let $\boldsymbol{\mu} = (1, \ldots, 1)' = \mathbf{1}_{k-1}$, with $k = 3, 6$. Also, we let

$$y_{ij} = \left(\mathbf{x}'_j \boldsymbol{\beta}_i + 1\right)\varepsilon_{ij}, \qquad (3.2)$$

where the constant vector $\boldsymbol{\beta}_i = (1, \ldots, 1)' = \mathbf{1}_k$, and y_{ij} and ε_{ij} are the j^{th} elements of the vectors $\boldsymbol{\varepsilon}_i$ and $\mathbf{Y}_i$, respectively. From (3.2), we can say that

$$\mathbf{Y}_i(\tau) = \mathbf{X}_i\, \boldsymbol{\beta}_i(\tau) + \boldsymbol{\varepsilon}_i(\tau), \qquad (3.3)$$

if we define

$$\boldsymbol{\beta}_i(\tau) = \boldsymbol{\beta}_i\, \boldsymbol{\varepsilon}_i(\tau), \qquad (3.4)$$

where, $\mathbf{Y}_i(\tau)$ is the value of $\mathbf{Y}_i$ corresponding to the τ^{th} quantile of the error term $\boldsymbol{\varepsilon}_i$, for $0 < \tau < 1$, and $\boldsymbol{\varepsilon}_i$'s are together distributed with multivariate skew t distribution, for $i = 1, 2, \ldots, M$.

Table 3.1 Values of factors that vary in different models

Factor	Symbol	Value
No. of Equations in the Model	M	2, 5
Sample Size	T	30, 100, 1000
Level of the Correlations	ρ	Low (0), Medium (0. 5), High (0.9)
Level of Skewness	γ	Low (0), Medium (0.75), High (1.5)
No. of Explanatory Variables (with intercept)	k	3, 6

Let the two independent variables $\mathbf{Z} \sim SN_M\left(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha}\right)$ and $V \sim \chi^2_{(v)}$ have multivariate skew normal and chi-square distributions, respectively. The vector of contemporaneous cross-equation errors, defined as $\boldsymbol{\varepsilon} = \mathbf{Z} / \sqrt{V / v} + \boldsymbol{\xi}$, has as a multivariate skew t distribution $\boldsymbol{\varepsilon} \sim St_M\left(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, v\right)$. The pair of parameters $\boldsymbol{\Omega}$ and $\boldsymbol{\alpha}$ of the given multivariate skew t distribution can be expressed through another pair of parameters $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ (see Appendix).

In Table 3.2, with 5 degrees of freedom, $v = 5$, and $\boldsymbol{\xi} = \mathbf{1}_M$ a vector of ones, we give the values of the parameters $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ to obtain the desired skewness level of intra-equation errors and correlation level of cross-equation errors. Note that as we said before just for simplicity we suppose the same skewness parameter λ and the same correlation parameter ψ for each pair of the components, as long as $-(M-1)^{-1} < \psi < 1$.

Table 3.2: Parameters of skew t distribution

			Correlation (ρ) 0	0.5	0.9	Mean	Median
Skewness (γ)	0	λ	0	0	0	1	1
		ψ	0	0.5	0.9		
	0.75	λ	0.6812335	0.68123357	0.6812336	1.53430	1.46394
		ψ	-	0.39334988	0.8786699		
	1.5	λ	1.5080071	1.5080071	1.5080071	1.79092	1.66100
		ψ	-	-	0.7954784		
Maximum magnitude of possible λ		M=2	1.4750284	2.5548241	6.4294996		
		M=5	0.7375141	1.8065334	5.0020646		
Maximum magnitude of attainable Skewness γ		M=2	1.4764728	2.0081796	2.4400950		
		M=5	0.8099441	1.6889051	2.3746932		

From Table 3.2, we see that for $\rho = 0$ and $\gamma = 1.5$ the correlation matrix will not be positive definite. The problem arises from the fact that the unrestricted skewness parameter λ exceeds the upper bound of the allowable skewness parameter as defined in Appendix (A.19). To solve this problem for $\rho = 0$, we can generate components of the multivariate skew t distribution independently of each other through univariate skew t distributions with the desired skewness parameter.

Table 3.3: Relative efficiency of the GLS estimation to the GLAD estimation of SURE Model parameters with 2 Equations

Regressors	Sample Size	Ratio	Correlation Low			Correlation Medium			Correlation High		
			Skewness			Skewness			Skewness		
			Low	Medium	High	Low	Medium	High	Low	Medium	High
2 Regressors	30	MSE	0.02060	0.01422	0.01221	0.01460	0.00954	0.00742	0.01043	0.00668	0.00629
		Determinant	7.1E-10	2.6E-10	2.2E-10	8.4E-10	3.4E-10	1.7E-10	2.2E-09	4.5E-10	4.9E-10
		Trace	0.0304	0.0297	0.0294	0.0241	0.0232	0.0174	0.0161	0.0147	0.0171
	100	MSE	0.00412	0.00217	0.00161	0.00280	0.00129	0.00105	0.00205	0.00100	0.00082
		Determinant	1.8E-11	1.7E-11	8.1E-12	4.9E-11	9.9E-12	1.7E-11	1.8E-10	7.8E-11	9.8E-11
		Trace	0.0149	0.0164	0.0132	0.0143	0.0102	0.0101	0.0095	0.0088	0.0091
	1000	MSE	0.00035	0.00016	0.00012	0.00021	0.00010	0.00008	0.00016	0.00007	0.00006
		Determinant	3.5E-12	1.5E-12	1.1E-12	4.0E-12	1.6E-12	1.6E-12	4.3E-11	1.7E-11	1.8E-11
		Trace	0.0114	0.0102	0.0080	0.0087	0.0070	0.0056	0.0063	0.0057	0.0047
5 Regressors	Sample Size	Ratio	Correlation Low			Correlation Medium			Correlation High		
			Skewness			Skewness			Skewness		
			Low	Medium	High	Low	Medium	High	Low	Medium	High
	30	MSE	0.00881	0.00671	0.00714	0.00698	0.00470	0.00401	0.00467	0.00441	0.00404
		Determinant	8.9E-25	5.6E-25	2.4E-24	3.1E-24	2.6E-25	7.1E-25	1.6E-23	1.2E-22	2.1E-22
		Trace	0.0106	0.0097	0.0128	0.0088	0.0071	0.0064	0.0056	0.0073	0.0079
	100	MSE	0.00507	0.00348	0.00284	0.00371	0.00242	0.00186	0.00260	0.00177	0.00149
		Determinant	1.9E-25	1.9E-25	3.1E-25	4.2E-25	5.6E-25	7.7E-25	1.2E-23	1.6E-23	3.5E-22
		Trace	0.0087	0.0097	0.0101	0.0069	0.0079	0.0075	0.0045	0.0053	0.0064
	1000	MSE	0.00066	0.00030	0.00023	0.00043	0.00020	0.00015	0.00034	0.00015	0.00011
		Determinant	9.2E-27	1.1E-26	3.0E-26	2.4E-26	3.6E-26	1.5E-25	1.6E-23	1.1E-23	6.4E-23
		Trace	0.0064	0.0070	0.0077	0.0050	0.0055	0.0059	0.0035	0.0039	0.0044

Table 3.4: Relative efficiency of the GLS estimation to the GLAD estimation of SURE Model parameters with 5 Equations

Regressors	Sample Size	Ratio	Correlation								
			Low			Medium			High		
			Skewness			Skewness			Skewness		
			Low	Medium	High	Low	Medium	High	Low	Medium	High
2 Regressors	30	MSE	0.01873	0.01406	0.00373	0.00811	0.00497	0.00414	0.00500	0.00297	0.00268
		Determinant	2.6E-24	2.6E-25	4.1E-27	1.3E-25	2.5E-26	3.8E-26	2.3E-23	2.1E-25	2.4E-25
		Trace	0.0260	0.0270	0.0045	0.0141	0.0131	0.0115	0.0088	0.0078	0.0084
	100	MSE	0.00431	0.00214	0.00173	0.00136	0.00063	0.00049	0.00079	0.00038	0.00031
		Determinant	1.7E-27	1.2E-28	2.8E-28	6.0E-28	4.0E-29	9.7E-29	5.1E-26	4.3E-27	3.7E-27
		Trace	0.0165	0.0136	0.0130	0.0076	0.0061	0.0045	0.0040	0.0036	0.0034
	1000	MSE	0.00035	0.00016	0.00012	0.00010	0.00004	0.00004	0.00006	0.00003	0.00003
		Determinant	8.3E-30	1.2E-30	8.4E-31	7.1E-30	3.0E-31	6.4E-31	1.2E-25	6.6E-27	1.0E-26
		Trace	0.0108	0.0093	0.0077	0.0050	0.0037	0.0025	0.0028	0.0025	0.0021

Regressors	Sample Size	Ratio	Correlation								
			Low			Medium			High		
			Skewness			Skewness			Skewness		
			Low	Medium	High	Low	Medium	High	Low	Medium	High
5 Regressors	30	MSE	0.00848	0.00782	0.00736	0.00271	0.00347	0.00262	0.00216	0.00199	0.00182
		Determinant	1.3E-61	2.2E-61	3.2E-60	2.0E-63	1.5E-60	1.1E-60	3.1E-59	2.7E-58	1.3E-56
		Trace	0.0100	0.0116	0.0124	0.0032	0.0063	0.0047	0.0027	0.0034	0.0036
	100	MSE	0.00494	0.00337	0.00287	0.00197	0.00118	0.00098	0.00110	0.00076	0.00060
		Determinant	9.8E-64	1.5E-63	3.5E-62	1.5E-62	1.4E-62	2.8E-60	2.4E-56	3.3E-56	7.6E-55
		Trace	0.0080	0.0087	0.0100	0.0040	0.0043	0.0048	0.0020	0.0024	0.0027
	1000	MSE	0.00065	0.00030	0.00023	0.00021	0.00009	0.00007	0.00013	0.00006	0.00005
		Determinant	2.6E-66	4.1E-66	3.1E-64	4.2E-64	2.0E-64	1.3E-62	3.9E-53	4.4E-53	4.4E-51
		Trace	0.0062	0.0068	0.0078	0.0031	0.0031	0.0031	0.0015	0.0016	0.0020

From Table 3.3 and Table 3.4 we can see that the relative efficiency of GLAD estimations to OLS estimations, which are in turn equivalent to GLS estimations, increases with any increase in the level of skewness of intra-equation errors and/or the level of correlation between the cross-equation errors. The same argument holds with an increase in the sample size. It may not look reasonable to compare models of different sizes; like in our simulation, models with 2 equations and 5 equations or equations with 2 regressors and 5 regressors, for relative efficiency of GLAD method to the OLS method. For instance, the relative efficiency based on the generalized variance is a power of the dimension of the covariance matrix (number of equations and number of regressors). Nevertheless, under the same circumstances (level of skewness, level of correlation and sample size), the relative efficiency of GLAD method to the OLS method increases as the dimensions of the model become larger.

4. Empirical example

To illustrate our methodology, we take the example 7.25 from Johnson and Wichern (1998) page 454. The dependents variables and regressors are the followings.

Y_1 : Total TCAD plasma level (TOT)

Y_2 : Amount of amitriptyline present in TCAD plasma level (AMI)

X_1 : Gender, 1=female, 0=male (GEN)

X_2 : Amount of antidepressants taken at time of overdose (AMT)

X_3 : PR wave measurement (PR)

X_4 : Diastolic blood pressure (DIAP)

X_5 : QRS wave measurement (QRS)

First, we model the example as a conventional multivariate linear regression model, like,

$$[\mathbf{Y}_1 \quad \mathbf{Y}_2] = \mathbf{X}[\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2] + [\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2] \tag{4.1}$$

and find the OLS estimates of the parameters (see Table 4.1).

Table 4.1: OLS Estimates

Dependent/ Independent Variable		Intercept	GEN	AMT	PR	DIAP	QRS
TOT	Coefficient	-2879.478	675.651	0.285	10.272	7.251	7.598
(1st Equation)	(Standard Error)	(893.260)	(162.056)	(0.061)	(4.255)	(3.225)	(3.849)
AMI	Coefficient	-2728.708	763.030	0.306	8.896	7.206	4.987
(2nd Equation)	(Standard Error)	(928.847)	(168.512)	(0.063)	(4.424)	(3.354)	(4.002)

When further investigating the data, we realize that the two dependent variables of the model have a very high correlation coefficient about 0.976 and very high skewness indices about 2.416 and 2.478. These two reasons are enough to strongly recommend someone to express the example in the form of Zellner's SURE model, like,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \qquad (4.2)$$

and calculate the GLAD estimations of the parameters (see Table 4.2).

Table 4.2: GLAD Estimates

Dependent/ Independent Variable		Intercept	GEN	AMT	PR	DIAP	QRS
TOT	Coefficient	-3356.086	811.902	0.261	11.789	6.409	9.854
(1st Equation)	(Standard Error)	(779.796)	(141.471)	(0.053)	(3.714)	(2.815)	(3.360)
AMI	Coefficient	-3257.014	1009.126	0.298	10.581	6.117	5.930
(2nd Equation)	(Standard Error)	(810.861)	(147.107)	(0.055)	(3.862)	(2.928)	(3.494)

In (2.9), we showed that generally the GLAD estimations are different from the LAD estimations. Here, we are going to show it numerically through this example. The square root of the inverse of the covariance matrix, $\mathbf{S}^{-1/2}$, is the following.

$$\mathbf{S}^{-1/2} = \begin{bmatrix} 0.005879 & -0.00316 \\ -0.00316 & 0.005587 \end{bmatrix}$$

The estimations of (2.11) are calculated as below.

$$\hat{\boldsymbol{\gamma}}_1 = \frac{\delta^{12}}{\delta^{11}} \hat{\boldsymbol{\beta}}_2 = (1753.0713, -543.15681, -0.16066845, -5.6953225, -3.2925598, -3.1918801)'$$

$$\hat{\boldsymbol{\gamma}}_2 = \frac{\delta^{12}}{\delta^{22}} \hat{\boldsymbol{\beta}}_1 = (1900.8699, -459.85687, -0.14801917, -6.6773844, -3.6302638, -5.5811185)'$$

$$\hat{\boldsymbol{\beta}}(u_1, X) = (-1603.0153, 268.74486, 0.10066724, 6.0939536, 3.1168634, 6.6618804)'$$

$$\hat{\boldsymbol{\beta}}(u_2, X) = (-1356.1444, 549.26897, 0.15048521, 3.9039008, 2.4869515, 0.34904501)'$$

Then, the functional relationship (2.9) can be easily verified. However, for the purpose of comparison, we present the LAD estimations of the individual regression equations (see Table 4.3). Another noticeable issue is the relative efficiency of the GLAD estimations to each of the LAD and the OLS estimations.

Table 4.3: LAD Estimates

Dependent/ Independent Variable		Intercept	GEN	AMT	PR	DIAP	QRS
TOT	Coefficient	-3044.362	569.990	0.285	8.174	7.153	13.783
(1st Equation)	(Standard Error)	(1329.592)	(241.215)	(0.091)	(6.333)	(4.800)	(5.729)
AMI	Coefficient	-3879.222	904.787	0.337	12.925	10.421	5.359
(2nd Equation)	(Standard Error)	(1222.319)	(221.754)	(0.083)	(5.822)	(4.413)	(5.267)

Conclusions

Contrary to the collapse of GLS estimations of SURE models to the OLS estimations of the individual equations with identical regressors, the GLAD estimations are not identical to the LAD estimations of the individual equations. Therefore, the problem of dealing with conventional multivariate linear regression models in which the equations, in some extent, are correlated, and when the OLS estimation of the parameters is aimed for, can be solved through converting the model into a SURE model and finding the GLAD estimations instead. The reason to do so is due to the gain of efficiency through the information embedded in the correlations between the equations. One can get even further efficiency at the presence of a slight skewness in the error terms.

References

[1] Azzalini, A. and Capitanio, A. (2003). “Distributions generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution”, *Royal Statistical Society. Series B*, Vol **56**, No. **2**.

[2] Azzalini, A. and Dalla Valle, A. (1996). “The Multivariate Skew-Normal Distribution”, *Biometrica*, **83**, 4, pp. 715-726.

[3] Chib, S., and Greenberg, E. (1995). “Hierarchical Analysis of SUR model with Extensions to Correlated Serial Errors and Time Varying Parameter Models”, *Journal of Econometrics*, **68**, 339-360.

[4] Johnson, Richard A. and Wichern, Dean W. (1998). “*Applied Multivariate Statistical Analysis*”, 4th edition, USA, Prentice-Hall, Inc. page (454-456).

[5] Rao, C. R. (1975). “Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems”, *Biometrics*, vol. 3, No. 2, 545-554.

[6] Shukur, G., and Zeebari, Z. (2009), "Developing Median Regression for SURE Models, with Application to 3-Generation Immigrants’ Data in Sweden", *Royal Institute of Technology, CESIS Electronic Working Paper Series in Economics and Institutions of Innovation*, No. **138**.

[7] Srivastava, V., and Giles, D. (1987). “*Seemingly Unrelated Regression Equations Models*”, New York: Marcel Dekker.

[8] Zellner, A. (1962). “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”, *American Statistical Association*, Vol **57**, No. **298**, pp. 348-68.

Appendix

We mention a definition of multivariate skew t distributions (Azzalini and Capitanio 2003) which is derived from the multivariate skew normal distribution. Then, we discuss how we control some properties of the distribution through the proper selection of its parameters.

Azzalini and Dalla Valle (1966) introduced two methods of generating data for multivariate skew normal distributions; the conditioning method through a pair of parameters $(\boldsymbol{\alpha}, \boldsymbol{\Omega})$, and the transformation method through another pair of parameters $(\boldsymbol{\lambda}, \boldsymbol{\Psi})$. Moreover, they showed that the pair $(\boldsymbol{\alpha}, \boldsymbol{\Omega})$ is a function of the pair $(\boldsymbol{\lambda}, \boldsymbol{\Psi})$ and that both pairs (methods) result in the same class of densities. We discuss how to control the process of generating the data through the transformation method. Suppose that

$$\begin{pmatrix} V_o \\ \mathbf{V} \end{pmatrix} \sim N_{m+1}\left(\mathbf{0}, \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Psi} \end{bmatrix}\right), \qquad \text{(A.1)}$$

where $\boldsymbol{\Psi}$ is the full rank correlation matrix and m is the dimension of the vector $\mathbf{V}$. The vector of skewness parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)' \in R^m$ is define through

$$\boldsymbol{\Delta} = \mathrm{diag}\left(\sqrt{1\text{-}\delta_1^2}, \ldots, \sqrt{1-\delta_m^2}\right), \text{ and}$$

$$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)',$$

where

$$\delta_i = \lambda_i \Big/ \sqrt{1+\lambda_i^2}, \qquad \text{(A.2)}$$

for $i = 1, \ldots, m$. Then, the new random vector $\mathbf{Z} = |V_o|.\boldsymbol{\delta} + \boldsymbol{\Delta}\mathbf{V}$ has the multivariate skew normal distribution, denoted as $\mathbf{Z} \sim SN_m(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, with the probability density function

$$f_{\mathbf{Z}}(\mathbf{z}) = 2\phi_m(\mathbf{z}; \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\mathbf{z}), \ \mathbf{z}, \boldsymbol{\alpha} \in R^m, \qquad \text{(A.3)}$$

where

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\Delta}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}}},$$

$$\boldsymbol{\Omega} = \boldsymbol{\Delta}(\boldsymbol{\Psi} + \boldsymbol{\lambda}\boldsymbol{\lambda}')\boldsymbol{\Delta}. \tag{A.4}$$

Consequently, from the formulae (A.4) we have

$$\omega_{ij} = \delta_i \delta_j + \psi_{ij} \sqrt{(1-\delta_i^2)(1-\delta_j^2)}\text{, for } i, j = 1, \ldots, m. \tag{A.5}$$

A.1. Multivariate Skew t Distribution

Let $\mathbf{Z} \sim SN_m(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ be independent of $V \sim \chi^2_{(v)}$. Then the vector $\mathbf{Y} = \mathbf{Z}\big/\sqrt{V/v} + \boldsymbol{\xi}$ has a multivariate skew t distribution, denoted as $\mathbf{Y} \sim St_m(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, v)$, with the pdf

$$f_{\mathbf{Y}}(\mathbf{y}) = 2t_m(\mathbf{y}; v) T_1\left(\boldsymbol{\alpha}'(\mathbf{y}-\boldsymbol{\xi})\sqrt{\frac{v+m}{(\mathbf{y}-\boldsymbol{\xi})'\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})+v}}; v+m\right), \tag{A.6}$$

where $t_m(\mathbf{y}; v)$ is the multivariate t probability density function

$$t_m(\mathbf{y}; v) = \frac{\Gamma(\frac{v+m}{2})}{\Gamma(\frac{v}{2})(\pi v)^{m/2}\left|\boldsymbol{\Omega}\right|^{1/2}}\left(1 + \frac{(\mathbf{y}-\boldsymbol{\xi})'\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})}{v}\right)^{-(v+m)/2} \tag{A.7}$$

and $T_1(\ .\ ; v+m)$ denotes the univariate t distribution with $v+m$ degrees of freedom.

The moments of $\mathbf{Y}$ are mathematically tractable since $\mathbf{Z}$ and V are assumed to be independent of each other. For the sake of simplicity in notation, let's define

$$\kappa = \sqrt{\frac{2}{\pi}}\,E\left((V/2)^{-1/2}\right) = \frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\sqrt{\frac{v}{\pi}}. \tag{A.8}$$

Then,

$$E(\mathbf{Y}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\frac{\boldsymbol{\alpha}'\boldsymbol{\Omega}}{\sqrt{1+\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}}}.\frac{(v/2)^{1/2}\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} = \boldsymbol{\xi} + \kappa\boldsymbol{\delta},\quad v > 1 \tag{A.9}$$

and

$$\operatorname{var}(\mathbf{Y}) = \boldsymbol{\Sigma} = \frac{v}{v-2}\boldsymbol{\Omega} - \kappa^2\,\boldsymbol{\delta}\boldsymbol{\delta}', \qquad v > 2. \tag{A.10}$$

From (A.10), we can see that

$$\sigma_{ij} = \frac{v}{v-2}\omega_{ij} - \kappa^2\delta_i\,\delta_j \quad , \quad v > 2. \tag{A.11}$$

A.2. Parameter Selection

Some properties of multivariate skew t distribution, like the covariance matrix and the marginal skewness indices of the components, in special cases can be controlled if we choose proper parameters. If we denote the correlation matrix of $\mathbf{Y}$ by $\boldsymbol{\rho} = \left[\rho_{ij}\right]$, from (A.11) and by substituting ω_{ij} with ψ_{ij} as in (A.5), we get

$$\rho_{ij} = \frac{(\frac{v}{v-2} - \kappa^2)\,\delta_i\,\delta_j + \frac{v}{v-2}\psi_{ij}\sqrt{1-\delta_i^2}\,\sqrt{1-\delta_j^2}}{\sqrt{\frac{v}{v-2} - \kappa^2\delta_i^2}\,\sqrt{\frac{v}{v-2} - \kappa^2\delta_j^2}} \quad , \quad v > 2. \tag{A.12}$$

The above formula means that we can have control over the correlation matrix $\boldsymbol{\rho}$, through suitable parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\lambda}$ (since $\boldsymbol{\lambda}$ is a function of $\boldsymbol{\delta}$, as in A.2). But we should note that some combinations of $\boldsymbol{\Psi}$ and $\boldsymbol{\lambda}$ might result in negative definite matrices of $\boldsymbol{\rho}$. For simplicity, we try to get the same correlation coefficient ρ between all mutual components of $\mathbf{Y}$. Then, the correlation matrix $\boldsymbol{\rho}$ is always positive definite as long as $(1-m)^{-1} < \rho < 1$, where $m > 1$ is the number of rows (or columns). To get the above correlation matrix we choose the same correlation coefficient ψ between all mutual components of $\mathbf{Z}$ and the same skewness parameter λ for all components of $\mathbf{Z}$ defined in (A.1) and (A.2).

The component-wise skewness of multivariate skew t distribution as given in Azzalini and Capitanio (2003) is as follows.

$$\gamma = \kappa\delta\left(\frac{v(3-\delta^2)}{v-3} - \frac{3v}{v-2} + 2\kappa^2\delta^2\right)\left(\frac{v}{v-2} - \kappa^2\delta^2\right)^{-3/2}, \; v > 3. \tag{A.14}$$

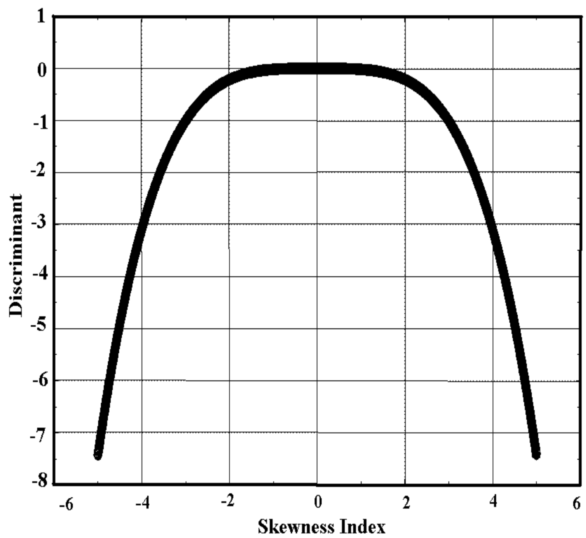
Substituting δ by its function of λ, as in (A.2), gives us a relation between the skewness parameter λ and the skewness coefficient γ, as shown below.

$$\gamma = \frac{\frac{\kappa\lambda\sqrt{v-2}}{v-3}\left(3v+[2\kappa^2(v-2)(v-3)-v(v-5)]\lambda^2\right)}{\left(v+[v-\kappa^2(v-2)]\lambda^2\right)^{3/2}}, \quad v>3. \qquad \text{(A.15)}$$

From (A.15), it appears that we have control over the coefficient of skewness through the parameter of skewness. To get a specific coefficient of skewness γ, we should look for a suitable value of λ. For this purpose, we try to define λ as a function of γ. If we square both sides of the relation in (A.15), it could be expressed in terms of an even function of λ in the form of a sextic equation (polynomial of degree six), which could be reduced into a cubic equation of λ^2, of the form $a(\lambda^2)^3 + b(\lambda^2)^2 + c(\lambda^2) + d = 0$. Also, it is well known that the discriminant of the cubic function is $\Delta = -4b^3d + b^2c^2 - 4ac^3 + 18abcd - 27a^2d^2$.

For $v > 3$, one can verify numerically or graphically in the (v, γ, Δ) coordinate system that the discriminant is always negative. At least for $v = 5$, it can be verified easily (see Figure A.1). This means that there is only one real root, say r, for the cubic equation above. Consequently, there is a one-to-one correspondence between γ^2 and λ^2. By glance at (A.15), one can easily notice that the sings of λ and γ are always the same, hence,

Figur A.1: Discriminant for different Skewness Indeces with v=5.



$$\lambda = \text{sgn}(\gamma)\sqrt{r}\,. \qquad \text{(A.16)}$$

where r is the real root of the cubic polynomial of λ^2.

For instance, taking the same level of skewness for all components of a random vector distributed as multivariate skew t distribution with degrees of freedom $v = 5$ to get skewness coefficients $\gamma = 0$, $\gamma = 0.75$ and $\gamma = 1.5$ at different times, we choose the skewness parameter $\boldsymbol{\lambda} = \mathbf{0}_m$, $\boldsymbol{\lambda} \approx (0.68123357)\mathbf{1}_m$ and $\boldsymbol{\lambda} \approx (1.5080071)\mathbf{1}_m$, respectively.

The chosen skewness parameter obtained from (A.16) then is used with the dependence parameter to get a specific correlation matrix, as given in (A.12). Now, Let us specify the correlation matrix $\boldsymbol{\psi}$ for that purpose. From (A.12) we can get

$$\psi_{ij} = \frac{\rho_{ij}\sqrt{1-\frac{\kappa^2(v-2)}{v}\delta_i^2}\sqrt{1-\frac{\kappa^2(v-2)}{v}\delta_j^2} - \left(1-\frac{\kappa^2(v-2)}{v}\right)\delta_i\delta_j}{\sqrt{1-\delta_i^2}\sqrt{1-\delta_j^2}} . \qquad \text{(A.17)}$$

Substituting each δ_i by λ_i, as in (A.1), and assuming the same skewness parameter λ corresponding to each component and the same coefficient of correlation ρ between mutual components, we can simplify (A.17) for each component ψ to the following relation

$$\psi = \rho - (1-\rho)[1 - \frac{\kappa^2(v-2)}{v}]\lambda^2 . \qquad \text{(A.18)}$$

From (A.3.10), it appears that the skewness parameter has effect on the choice of ψ. Some combinations of λ and ψ may give us a correlation matrix $\boldsymbol{\Psi}$ which is not positive definite. Since $-\frac{1}{m-1} < \psi < 1$ in order to have a positive definite $\boldsymbol{\Psi}$, and since $\kappa < \frac{2\sqrt{3}}{\pi}$ for $v > 3$, using (A.18), the bounds of λ could be obtained from the following inequality,

$$\lambda^2 < \frac{v[\rho(m-1)+1]}{(m-1)(1-\rho)[v-\kappa^2(v-2)]} , \qquad v > 3 \qquad \text{(A.19)}$$

where m is the dimension of the correlation matrix.

We can find the vector of the medians of the component of $\mathbf{Y} \sim St_m(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, v)$ by looking at each component as a univariate skew t distribution and finding numerically the root of the equation $\int_{-\infty}^{x} f_Y(y)\,dy - \frac{1}{2} = 0$, for $f_Y(y)$ defined in (A.6) with $m = 1$.