# An Improved Two-Stage Estimator of Simultaneous Equations Models

Qingli Pan<sup>#1</sup>

<sup>#</sup> The college of Information Science and Technology & Department of Mathematics & Jinan University Guangzhou, China

## Abstract

For an over-identified simultaneous equation model with multiple co-linearity, this paper proposes a twostage k-d type Liu estimate method to estimate such models. Moreover, the properties of the new estimate are verified by the Monte Carlo experiment. Compared with two-stage least square estimate, two-stage Liu estimate, and two-stage ridge estimate, the new estimate performs quite well. The standard deviations of the new estimate are rather small and its practicality is stronger than others.

**Keywords** - Two-stage k-d type Liu estimate, Two-stage least square estimate, Two-stage Liu estimate, Two-stage ridge estimate.

## I. INTRODUCTION

For practical economic problems, the simultaneous equation models (SEM, e.g. [1]) is an equation system that describes the coherence between variables. In the mid-20th century, the simultaneous equation models have received widespread attention from many researchers. Haavelmo[2] has showed the importance of using the SEM to explain economic activity. Anderson and Rubin[3] have proposed a parameter estimation method of a single equation in a complete system of linear stochastic equations. Dr è ze[4] has analysed the Bayesian estimation of SEM under diffuse prior.

The parameter estimation of SEM becomes quite complicated due to the problems regarding the identification. Hausman[5] has presented that the identification of SEM can be reflected by its rank and order conditions. The order condition reflects overall identification while the rank condition reflects local identification. For a general SEM, the ordinary least squares (OLS) estimate is theoretically inappropriate because of its bias and inconsistency. The instrumental variable (IV) method and the indirect least squares (ILS) method are generally only applicable to the estimation of the structural equation that is just identified in the SEM. However, the SEM is generally over-identified in actual economic activities. The two-stage least square (TSLS) have proved to be applicable to the estimation of structural equations for over-recognition (e.g. [6]-[7]). López-Espín et al[8] have studied the two-stage least squares method for SEM using QR-decomposition.

Although the TSLS estimation has many excellent properties, the method does not perform well when there is multiple co-linearity between the explanatory variables of the regression equations. Hoerl and Kennard[9] have proposed the ridge regression to deal with the non-orthogonal problems. William and Ligori[10] have discussed the modified OLS method to overcome the ill-effects of co-linearity of two-variable regression models. Shyti and Valera[11] have learned the properties of class estimators, both asymptotic and small sample properties. Vinod and Ullah[12] have showed the ridge estimator(RE) and Stein-type estimator in SEM with multiple co-linearity. Toker et al[13] have discussed the properties of two-stage Liu estimator(TSLE) for SEM that is just identified.

In this paper, I proposed a new estimation, two-stage k-d type Liu estimation, to Improve the accuracy of estimate for SEM. Ping Hu and Kaiding Li[14] have studied the k-d type Liu estimation for linear regression model and found that the k-d type Liu is better than the standard Liu estimate and the ridge estimate under certain conditions. Moreover, I use the standard deviation (SD) to compare the accuracy of TSLS, two-stage ridge estimate (TSRE), TSLE and new estimate of SEM via Monte Carlo simulation.

### **II. THE SIMULTANEOUS EQUATIONS MODELS**

Consider the following m equations simultaneous equations models

$$Y\Gamma = XB + U , \qquad (1)$$

where Y and X are the  $n \times m$  matrix of observations on m endogenous variables and the  $n \times k$  matrix of observations on k exogenous variables, respectively. Further,  $\Gamma_{m \times m}$  and  $B_{k \times m}$  are the matrices of structural coefficients and  $U_{n \times m}$  denotes the matrix of structural disturbances. The structural disturbances U have zero

mean and they are non-singular. Here we assumed that this system is over-identified and  $u_i \sim N(0, \Sigma)$ , where  $U_{i_0}$  is the *i*-th row of  $U_i$ ,  $i = 1, 2, \dots, n$ .

Multiplying both sides of (1) by  $\Gamma^{-1}$ , the unrestricted reduced form equations are

$$Y = X\Pi + V , \qquad (2)$$

where the associated reduced form coefficients are  $\Pi = B\Gamma^{-1}$  and  $V = U\Gamma^{-1}$ . The row of V,  $V_{i_n} \sim N(0, \Omega)$ ,  $i = 1, 2, \dots, n$ , where  $\Omega = (\Gamma^{-1})' \Sigma \Gamma^{-1} \Gamma^{-1}$ .

Without loss of generality the estimation of the parameters of the first equation of the system in (1) is considered. Suppose  $\Gamma$  has a diagonal element of 1. This equation is given by

$$y_1 = Y_1 \gamma_1 + X_1 \beta_1 + u_1, \qquad (3)$$

$$Y_{1} = X \Pi_{1} + V_{1}, \qquad (4)$$

where  $Y_1$  is a  $n \times m_1$  matrix of observations on  $m_1$  included endogenous variables,  $X_1$  is a  $n \times k_1$  matrix of observations on  $k_1$  included exogenous variables,  $u_1$  is the first column of U.  $\Pi_1$  is the corresponding coefficient of reduced form and  $V_1$  is the  $n \times m_1$  error matrix. Equation (3) can also be rewritten as follows:

$$y_1 = Z_1 \delta_1 + u_1, (5)$$

where  $Z_1 = [Y_1, X_1]$  and  $\delta_1 = [\gamma_1, \beta_1]'$ .

## **III.PARAMETER ESTIMATION**

The two-stage least squares method is less sensitive to both specification error and multiple co-linearity and is a very practical estimation method. The first stage of the TSLS is to apply the OLS method to the reduced equations to find the estimated parameters we need. And the second phase of the task is to obtain a consistent estimate of the structural parameters through a special form of instrumental variable method.

The TSLS estimator (e.g. [1]) may be written as:

$$\delta_{1}^{tshs} = \left(\hat{Z}_{1}'\hat{Z}_{1}\right)^{-1}\hat{Z}_{1}'y_{1}, \qquad (6)$$

where  $\hat{Z}_{1} = [\hat{Y}_{1}, X_{1}], \hat{Y}_{1} = X \hat{\Pi}_{1}$  and  $\hat{\Pi}_{1} = (X X)^{-1} X Y_{1}$ .

However, when the regression model exhibits multiple co-linearity, the TSLS prediction is not effective. A well-known biased estimation method dedicated to collinear data analysis is ridge estimation. Ridge regression is essentially a method of improving the OLS estimate by abandoning the unbiasedness of the least squares estimate and at the expense of partial information. The two-stage ridge estimator, discussed by Vinod and Ullah[12], is as follows:

$$\delta_{1}^{kre} = \left(\tilde{Z}_{1}'\tilde{Z}_{1} + kI\right)^{-1}\tilde{Z}_{1}'y_{1},$$
(7)

where  $\widetilde{Z}_{_1} = [\widetilde{Y}_{_1}, X_{_1}], \ \widetilde{Y}_{_1} = X \widetilde{\Pi}_{_1}, \ k > 0$ , I is the unit matrix and

$$\widetilde{\Pi}_{1} = \left(X'X + kI\right)^{-1}X'Y_{1}.$$
(8)

Toker et al.[13] have proposed the two-stage Liu estimator of the SEM and show that the TSLE perform better than TSLS and TSRE when the model is just identified. The TSLE is presented as follows:

$$\delta_1^{isle} = \left(\overline{Z}_1'\overline{Z}_1 + I\right)^{-1} \left(\overline{Z}_1'y_1 + d\overline{\delta}_1\right), \tag{9}$$

Where 
$$\overline{\delta_1} = (\overline{Z_1'Z_1})^{-1} \overline{Z_1'y_1}, \overline{Z_1} = [\overline{Y_1}, X_1], \overline{Y_1} = X\overline{\Pi_1} \text{ and } \overline{\Pi_1} = (XX + I)^{-1} (XY_1 + d\widehat{\Pi_1}).$$

## **IV.NEW ESTIMATION**

Although the standard two-stage Liu estimate improves the estimation accuracy of the simultaneous equations models to a certain extent, it only applies to the case where the regression model is just identified. However, most regression models are over-identified in actual economic activities. In order to solve the above problems, we proposed a new estimation, two-stage k-d type Liu estimation, as follows:

$$\delta_{1}^{new} = \left( \dot{Z}_{1}' \dot{Z}_{1} + I \right)^{-1} \left( \dot{Z}_{1}' y_{1} + d \dot{\delta}_{1} \right), \tag{10}$$

where  $\dot{S}_{1} = (\dot{Z}_{1}\dot{Z}_{1} + kI)^{-1}\dot{Z}_{1}\dot{Y}_{1}$ ,  $\dot{Z}_{1} = [\dot{Y}_{1}, X_{1}]$ ,  $\dot{Y}_{1} = X\dot{\Pi}_{1}$ , and  $\dot{\Pi}_{1} = (X\dot{X} + I)^{-1}(X\dot{Y}_{1} + d\tilde{\Pi}_{1})$ ,  $\tilde{\Pi}_{1}$  is defined by (8).

## V. SIMULATION

In this section, we apply our method to Monte Carlo experiments. I use standard deviation to compare the properties of these four estimators for SEM. I generate the data using the following structural model

$$Y\Gamma = XB + U , \qquad (11)$$

where Y is a  $n \times 3$  matrix of endogenous variables, X is random drawing from a multivariate normal distribution with zero mean and covariance matrix

$$\Sigma_{x} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

The row of structural error U,  $U_{i_{u}}(i = 1, 2, \dots, n)$ , are independently drawn from a multivariate normal distribution with zero mean and covariance matrix

$$\Sigma = \left( \begin{array}{cccc} 0.\ 07 & 0.\ 05 & 0.\ 04 \\ 0.\ 05 & 0.\ 045 & 0.\ 035 \\ 0.\ 04 & 0.\ 035 & 0.\ 03 \end{array} \right).$$

As for coefficient matrix

$$\Gamma = \begin{pmatrix} 1 & 0 & \gamma_{1,3} \\ -\gamma_{2,1} & 1 & 0 \\ 0 & \gamma_{3,2} & 1 \end{pmatrix}, \ B = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & 0 \\ 0 & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & 0 & \beta_{3,3} \end{pmatrix},$$

we set  $\gamma_{1,3} = 0.3$ ,  $\gamma_{2,1} = 0.5$ ,  $\gamma_{3,2} = -0.1$ ,  $\beta_{1,1} = \beta_{2,2} = \beta_{3,1} = 0.1$ ,  $\beta_{1,2} = 0.5$ ,  $\beta_{2,3} = 0.2$  and  $\beta_{3,3} = 0.6$ . Let n = 25, k = 0.02, d = 0.96 and  $\rho = 0.8$ . I also repeated the same Monte Carlo experiments 10000 times. What's more, we studied the effect of  $\rho$  by using various values of  $\rho = 0.5:001:0.99$ .

Table 1: The mean value and standard deviations (sds) for four estimations

		TSLS		TSLE		TSRE		New	
	TV	Mean	SDs	Mean	SDs	Mean	SDs	Mean	SDs
$\gamma_{_{2,1}}$	0.5	-0.2478	45.4685	0.4205	32.6772	0.4845	0.5821	0.4943	0.5796
${oldsymbol{eta}}_{1,1}$	0.1	0.4897	23.9968	0.1427	18.1700	0.1086	0.3268	0.1040	0.3262
${oldsymbol{eta}}_{{}_{3,1}}$	0.1	0.1778	6.2094	0.1296	4.2208	0.1017	0.1248	0.0991	0.1227

As we can see in table 1, for the first equation, the mean value of TSLS and TSLE are very different from the true value and the value of SDs are both quite large. I can also find that our proposed estimation method performs very well, its estimate is very close to the actual value and the standard errors are very small.



Fig. 1 show the variation of the standard deviations with the value of  $\rho$ . Here only two estimators that perform well may be considered. Regardless of the value of  $\rho$ , the SDs of new estimated we propose are smaller than that of TSRE. And except for  $\gamma_{2,1}$ , the SDs of other parameters increase as  $\rho$  increases. The new estimate and TSRE are minimally affected by  $\rho$  and the new estimate performs well.

## VI.CONCLUSIONS

In this paper, according to the insensitivity of k-d type Liu estimation to multiple co-linearity and the characteristics of two-stage least squares, two-stage k-d type Liu estimation is proposed. By comparing with TSLS, TSLE and TSRE, it is found that the new estimate is suitable for the case where the regression model is over-identified and there is multiple co-linearity. Moreover, the new estimation performs very well in the Monte Carlo simulation. Of course, this new estimation method can also be applied to other models in my future research.

#### ACKNOWLEDGMENT

I would like to express our appreciation for the valuable comments made by editors and unknown referees so that we can make further improvements to this paper.

## REFERENCES

- [1] William H. Greene, Econometric Analysis, 7th ed. Prentice Hall, 2012.
- [2] T. Haavelmo, The statistical implications of a system of simultaneous equations, Econometrica, 1943, vol. 11, pp. 1-12.
- [3] T. W. Anderson and H. Rubin, Estimation of the parameters of a single equation in a complete set of stochastic equations, Annals of Mathematical Statistics, 1949, vol. 21, pp. 570-582.
- [4] J. H. Drèze, Bayesian limited information analysis of the simultaneous equation model, Econometrica, 1976, vol. 44, pp. 1045-1075.
- [5] J. A. Hausman, Specification and estimation of simultaneous equations systems, In Z. Griliches M.D. Intrilligator (eds.), Handbook of Econometrics, Amsterdam: North-Holland, 1983, vol. 1.
- [6] H. Theil, Repeated least squares applied to complete equation systems, The Hague: Central planning bureau, 1953
- [7] R. L. Basmann, A generalized classical method of linear estimation of cofficients in a structural equation, Econometrica, 1957, vol. 25, pp. 77-83.
- [8] Jose J. López-Espín, Antonio M. Vidal, and Domingo Giménez, Two-stage least squares and indirect least squares algorithms for simultaneous equations models, Journal of Computational and Applied Mathematics, 2012, vol. 235, pp. 3676-3684.
- [9] A. E. Hoerl and R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 1970, vol. 12, pp. 55-67.
- [10] Martin L. William and L.Maria Alphonse Ligori, A Modified Least-Squares Approach to Mitigate the Effect of Collinearity in TwoVariable Regression Models, International Journal of Mathematics Trends and Technology, 2016, vol. 30(1), pp. 48-52.
- [11] Bederiana Shyti and Dhurata Valera, The Regression Model for the Statistical Analysis of Albanian Economy, International Journal of Mathematics Trends and Technology, 2018, vol. 62(2), pp. 90-96.
- [12] H. D. Vinod and A. Ullah, Recent advances in regression methods, New York: Marcel Dekker, 1981.
- [13] Selma Toker, Selahattin Kaçıranlar, and Hüseyin Güler, Two-stage Liu estimator in a simultaneous equations model, Journal of Statistical Computation and Simulation, 2018, vol. 88, pp. 2066-2088.
- [14] Ping Hu and Kaiding Li, New Research on liu Estimation of Linear Regression Model, Mathematica Applicata, 2007, vol. 20, pp.37-39.