ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

PhD THESIS

Selma TOKER

IMPROVED ESTIMATORS IN THE SIMULTANEOUS EQUATIONS MODEL

DEPARTMENT OF STATISTICS

ADANA, 2015

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

IMPROVED ESTIMATORS IN THE SIMULTANEOUS EQUATIONS MODEL

Selma TOKER

PhD THESIS

DEPARTMENT OF STATISTICS

We certify that the thesis titled above was reviewed and approved for the award of degree of the Doctor of Philosophy by the board of jury on 23/06/2015.

.....	…….	….……..…….....
Prof.Dr. Selahattin KAÇIRANLAR SUPERVISOR	Prof.Dr. Aşır GENÇ MEMBER	Assoc.Prof.Dr. M. Revan ÖZKALE MEMBER
……..	……...…….....	
Assoc.Prof.Dr. Mehmet ÖZMEN MEMBER	Assoc.Prof.Dr. Gülin TABAKAN MEMBER	

This PhD Thesis is written at the Statistics Department of Institute of Natural And Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

This thesis is supported by

- **Çukurova University Scientific Research Projects Unit Project Number: FDK-2015-2852**
- **The Scientific and Technological Research Council of Turkey (TUBITAK)-Departments to Support Scientists (BIDEB).**

Note: The usage of the presented specific declerations, tables, figures, and photographs either in this thesis or in any other reference without citiation is subject to "The law of Arts and Intellectual Products" number of 5846 of Turkish Republic

ABSTRACT

PhD THESIS

IMPROVED ESTIMATORS IN SIMULTANEOUS EQUATIONS MODEL

Selma TOKER

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor : Prof. Dr. Selahattin KAÇIRANLAR
Year: 2015, Pages: 101
Jury : Prof. Dr. Selahattin KAÇIRANLAR
: Prof. Dr. Aşır GENÇ
: Assoc. Prof. Dr. M. Revan ÖZKALE
: Assoc. Prof. Dr. Mehmet ÖZMEN
: Assoc. Prof. Dr. Gülin TABAKAN

Two stage least squares is a widely used method of estimating the parameters of a single structural equation in a system of simultaneous equations. The extension of the biased estimation procedures at both stages of the two stage least squares technique in simultaneous equations model is desirable so as to cope with the problem of multicollinearity. Such a two stage technique is applied with the help of ridge regression by Vinod and Ullah (1981). In a similar manner, three different kinds of Liu estimators which are named with regard to their implementation stages are proposed in this study. The mentioned estimators are compared theoretically and the results are supported by a numerical example.

In addition, inequality restrictions are necessary to maintain structural consistency in a system of simultaneous equations. Therefore, the idea of inequality restrictions is applied to the two and three stage ridge regression estimation in the presence of multicollinearity by following Liew (1976a). Inequality constrained two stage and three stage ridge regression estimators are proposed by reducing the primal-dual relation to the fundamental problem of Dantzig-Cottle (1967, 1974) and solving with Lemke (1962) algorithm.

Key Words: Biased estimators, Multicollinearity, Simultaneous equations model, Two stage estimation.

ÖZ

DOKTORA TEZİ

EŞ ANLI DENKLEMLER MODELİNDE İYİLEŞTİRİLMİŞ TAHMİN EDİCİLER

Selma TOKER

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman : Prof. Dr. Selahattin KAÇIRANLAR
Yıl: 2015, Sayfa: 101
Jüri : Prof. Dr. Selahattin KAÇIRANLAR
: Prof. Dr. Aşır GENÇ
: Doç. Dr. M. Revan ÖZKALE
: Doç. Dr. Mehmet ÖZMEN
: Doç. Dr. Gülin TABAKAN

İki aşamalı en küçük kareler, eşanlı denklemler sisteminde bir tek yapısal denklemin katsayılarını tahmin etmede kullanılan yaygın bir metottur. Çoklu iç ilişki problemini gidermek için eşanlı denklemler modelinde iki aşamalı en küçük kareler tekniğinin her iki aşamasında yanlı tahmin yöntemlerinin geliştirilmesine ihtiyaç duyulmaktadır. Böyle bir iki aşamalı teknik ridge regresyon yardımıyla Vinod ve Ullah (1981) tarafından uygulanmıştır. Benzer şekilde, bu çalışmada uygulama aşamalarına göre adlandırılan üç farklı Liu tahmin edici önerilmiştir. Söz konusu tahmin ediciler teorik olarak karşılaştırılmış ve sonuçlar nümerik uygulama ile desteklenmiştir.

Ayrıca, eşitsizlik kısıtı eşanlı denklemler sisteminde yapısal tutarlılığı sağlamak için gereklidir. Bu yüzden, çoklu iç ilişki olması durumunda eşitsizlik kısıtı fikri Liew (1976a) çalışmasını izleyerek iki ve üç aşamalı ridge regresyona uygulanmıştır. Primal-dual problem Dantzig-Cottle (1967, 1974) temel problemine indirgenip Lemke (1962) algoritmasıyla çözülerek iki ve üç aşamalı ridge tahmin edici önerilmiştir.

Anahtar Kelimeler: Yanlı tahmin ediciler, Çoklu iç ilişki, Eşanlı denklemler modeli, İki aşamalı tahmin.

ACKNOWLEDGEMENTS

First and foremost I will always be grateful to my mentor and respected supervisor Prof. Dr. Selahattin KAÇIRANLAR for his never ending patience, encouragement, guidance and support throughout my doctoral study. His idealism, optimism and vigorous search for perfection have illuminated the long and arduous path to the successful completion of this work. He will always be a role model for me.

I would like to express my special thanks to the members of the dissertation commitee. I would like to thank my professors and colleagues in my department whom I benefited much from their experiences, worldviews, attitudes and intellects. Also, I was lucky to have special friends and I'm deeply indebted to them for their close friendship.

At last but not least, I would like to extend my heartfelt thanks to my family for their everlasting love, prayers, tremendous support and encouragement to achieve my goal.

CONTENTS **PAGE**

ABSTRACT.....I
ÖZ.....II
ACKNOWLEDGEMENTS.....III
CONTENTS.....IV
LIST OF FIGURES.....VI
LIST OF ABBREVIATONS.....VIII
1. INTRODUCTION.....1
2. LINEAR REGRESSION MODEL AND SOME BIASED ESTIMATORS.....5
2.1. The Linear Model.....5
2.2. Principal of Ordinary Least Squares and the Need for Alternative Estimators.....7
2.3. Multicollinearity.....10
2.3.1. Identifying the Multicollinearity.....12
2.3.2. Solutions to the Multicollinearity Problem.....14
2.3.2.1.Traditional Solutions to the Multicollinearity Problem.....14
2.3.2.2. Ad Hoc Solutions to the Multicollinearity Problem.....15
2.4. Ridge Estimator.....15
2.4.1. Properties of the Ridge Estimator.....16
2.4.2. Choice of the Ridge Parameter.....17
2.4.2.1. Ridge Trace.....17
2.4.2.2. Estimation of k.....18
2.5. Stein Type Estimator.....21
2.6. Liu Estimator.....23
3. SIMULTANEOUS EQUATIONS MODEL.....27
3.1. Nature of Simultaneous Equations Model.....27
3.2. Problem of Identification.....33
3.2.1. Underidentification.....33
3.2.2. Exact Identification.....34
3.2.3. Overidentification.....35

3.2.4. Identification by Zero Restrictions.....36
3.2.5. Rules for Identification.....39
3.3. Methods of Estimation.....40
3.3.1. Limited Information Estimators.....40
3.3.1.1. Indirect Least Squares Estimator.....42
3.3.1.2. Two Stage Least Squares Estimator.....44
3.3.1.3. Ridge Type Estimators.....49
3.3.1.4. Stein Type Estimators.....51
3.3.1.5. Instrumental Variable Estimators.....52
3.3.2. Full Information Estimators.....54
4. TWO STAGE LIU ESTIMATOR.....57
4.1. Comparisons of the Two Stage Ridge Estimator and the Two Stage Liu Estimator.....58
4.2. Selection of the Biasing Parameter.....63
4.2.1. Biasing Parameter for the Liu Esimator at the Second Stage.....63
4.2.2. Biasing Parameter for the Liu Esimator at the First Stage.....64
4.2.3. Biasing Parameter of the Two Stage Liu Estimator.....66
4.3. Numerical Example.....69
5. INEQUALITY RESTRICTIONS IN SIMULTANEOUS EQUATIONS MODEL.....79
5.1. Inequality Constrained Least Squares Estimator.....79
5.2. Inequality Constrained Ridge Regression Estimator.....82
5.3. Inequality Constrained Two Stage Least Squares Estimator.....84
5.4. Inequality Constrained Two Stage Ridge Regression Estimator.....88
6. CONCLUSIONS.....93
REFERENCES.....95
CURRICULUM VITAE.....101

LIST OF FIGURES **PAGE**

Figure 4.1(a). Estimated mse values of the TSLS estimator and the two stage LE for the first equation.....72

Figure 4.1(b). Estimated mse values of the TSLS estimator and the two stage LE for the first equation.....72

Figure 4.2. Estimated mse values of the TSLS estimator and the two stage LE for the second equation.....73

Figure 4.3. Estimated mse values of the TSLS estimator and the two stage LE for the third equation.....74

LIST OF ABBREVIATONS

TSLS : Two stage least squares

OLS : Ordinary least squares

RE : Ridge Estimator

LE : Liu Estimator

ICLS : Inequality constrained least squares

GICLS : Generalized inequality constrained least squares

MSE : Matrix mean square error

mse : Scalar mean square error

BLUE : Best linear unbiased estimator

VIF : Variance inflation factors

ILS : Indirect least squares

3SLS : Three stage least squares

GRE : Generalized ridge estimator

GIV : Generalized instrumental variable

FGIV : Full information generalized instrumental variable

ICRR : Inequality constrained ridge regression

p.d. : Positive definite

$E(\cdot)$: Expected value operator

$Var(\cdot)$: Variance-covariance operator

$Bias(\cdot)$: Bias operator

$\otimes$: Kronocker product

$\hat{\beta}$: Ordinary least squares estimator

$\hat{\beta}_k$: Ridge estimator

$\hat{\beta}_{ST}$: Stein type estimator

$\hat{\beta}_d$: Liu estimator

δ_1^{LS} : Two stage least squares estimator

δ_1^{LE} : Two stage Liu estimator

δ_1^{RE} : Two stage Ridge estimator

δ_1^{ST} : Two stage Stein type estimator

$\hat{\delta}_{3SLS}$: Three stage least squares estimator

β_{IC}^* : Inequality constrained least squares estimator

$\beta_{IC}^*(k)$: Inequality constrained ridge regression estimator

δ_{IC}^* : Inequality constrained two stage least squares estimator

δ_{IC}^{**} : Inequality constrained three stage least squares estimator

$\delta_{IC}^*(k)$: Inequality constrained two stage ridge regression estimator

$\delta_{IC}^{**}(k)$: Inequality constrained three stage ridge regression estimator

1. INTRODUCTION

Single equation regression models are considered to have a single dependent variable that is expressed as a function of one or more explanatory variables. Unidirectional relationship between the dependent and the explanatory variables for the single equation regression model cannot be always maintained. In the case of a bilateral or feedback relationship between the dependent and the explanatory variables, more than one regression equation is required. That kind of regression models, having more than one equation and feedback relationship between variables, are called as simultaneous equations model. There are two approaches to estimate the structural coefficients, namely single-equation methods (limited information methods) such as two stage least squares (TSLS) and system methods (full information methods) such as three stage least squares.

The TSLS has been widely used method of estimating the parameters of a single structural equation since it is easy to apply and frequently preferred over complete system techniques. The TSLS procedure is essentially such an instrument variable technique that the explanatory endogeneous variables are replaced by their instrumental variables in the first stage. These are ordinary least squares (OLS) estimates obtained by using the exogeneous variables as predictors. At the second stage, the OLS estimation is used to obtain the regression coefficients. The TSLS method is proposed more or less independently by Theil (1953a, 1953b, 1954, 1961), Basmann (1957) and Sargan (1958). Additional insights are pointed out by Amemiya (1974), Kadiyala and Nunns (1976), Koutsoyiannis (1977), Intriligator (1978), Anderson et al. (1986), Farebrother (1999) and Anderson (2005).

Multicollinearity relevant to the linear regression model is one of the most widely known problems. The problem of multicollinearity which is native to single equation estimation is addressed here by introducing simultaneous equations model. This problem may still exists in individual equations of the simultaneous model. Due to near linear relationship among predetermined variables and/or predictions of right hand side endogenous variables, the TSLS often fails to provide stable and accurate solutions. Namely, the estimation process turns out with poor precision and inflated

parameter estimates with possible wrong signs. These considerations motivates the development of the biased estimation. Initially, we examine several biased estimators which are proposed to cope with the problem of multicollinearity in the linear regression model in the second section.

In the linear regression model, one of the biased estimators is Stein estimator which is started with seminal paper of Stein (1956) on estimating the multivariate normal mean with a result that the usual estimator, the sample mean, can be improved with respect to squared error loss. A shrinkage estimator that has uniquely smaller quadratic risk than the sample mean is derived by the famous paper of James and Stein (1961). Stein's result has been extended to a wide class of distributions and loss functions and estimators which dominate standard estimators are often called Stein estimators, Stein-rule estimators or Stein-type estimators to appreciate the performance of Stein. Ridge estimator (RE) of Hoerl and Kennard (1970a) is motivated from the fact that the matrix $X'X$ is stabilized by adding small, positive quantities to each of its main diagonal elements. The RE has been shown to give more precise estimates of the regression parameters because its variance and its mean square error are generally smaller than that of the OLS estimators. Although the RE is practical, the choice of its biasing parameter is difficult since the RE has a complicated function of this parameter. Besides the advantage of being linear function of its biasing parameter, Stein type estimator is not efficient in practise due to the same shrinkage of each element. Kejian (1993) introduced a new estimator which combines the advantages of the RE and the Stein type estimator to handle multicollinearity. This estimator, as a linear function of its biasing parameter, yields convenient selection opportunity of the biasing parameter. Subsequently, Akdeniz and Kaçıranlar (1995) entitled this new estimator as Liu estimator (LE). The matrix mean square error comparisons between the OLS estimator and the LE were given in Sakallıoğlu et al. (1996). Sakallıoğlu et al. (2001) compared the RE and the LE with the same criterion. There are lots of similar works performed about the LE in linear regression model, afterwards.

Various kinds of improved estimators to overcome the collinearity issue have recieved attention in the literature of simultaneous equations model and cited in the

third section. Classical and Bayesian interpretations are presented in relation to the ridge-like, shrinkage and some Bayesian estimators by Zellner and Vandaele (1975), Mehta and Swamyy (1978), Maasoumi (1978), Maasoumi (1980), Srivastava et al. (1978), Vinod and Ullah (1981) and Ullah and Srivastava (1988). Alternative simultaneous equation estimators are examined by a Monte Carlo approach by Wagner (1958), Nagar (1960), Raj (1980), Capps and Grubbs (1991), Agunbiade and Iyaniwura (2010) and Agunbiade (2011). Vinod and Ullah (1981) presented the Stein type and the ridge type estimators at the first and/or second stage of the TSLS in a simultaneous equations model. For the improved estimation of coefficients in a structural equation, we consider three different kinds of the LEs' which are named with regard to their implementation stages in the fourth section. Selection methods for the biasing parameter are analyzed. The conditions for superiority of the proposed estimator over conventional estimators like the TSLS estimator and the RE are obtained under the mean square error criterion. The mentioned estimators are compared theoretically and the results are supported by a numerical example and a Monte Carlo simulation study.

To meet the demand to use prior and sample information for parameter estimation of a regression model, inequality restrictions are utilized. In order to maintain structural consistency, many empirical studies entails inequality restrictions on parameters of a regression equation. That's the reason for the inequality constrained least squares (ICLS) estimation to be proposed. Liew (1976b) introduced ICLS estimator by reducing primal dual relation to the to the fundamental problem of Dantzig-Cottle (1967, 1974) and solving with Lemke (1962) algorithm. Toker and Kaçıranlar (2013) applied inequality restrictions to the ridge regression estimation to deal with multicollinearity and proposed inequality constrained ridge regression (ICRR) estimator. Such inequality restrictions are also required to sustain structural consistency in a system of simultaneous equations. Liew (1976a) suggested a generalized inequality constrained least squares (GICLS) estimation model and enlarged the model to a two and three stage least squares estimation when the parameters of the structural equations are constrained by inequalities. In the fifth section, we based on this idea. To circumvent the problem of multicollinearity, we

offer inequality constrained two stage and three stage ridge regression estimators.

2. LINEAR REGRESSION MODEL AND SOME BIASED ESTIMATORS

2.1. The Linear Model

Let y and $x_1, x_2, \ldots, x_p$ denote the dependent and explanatory variables, respectively. The relationship between the variables can be shown by a function f as

$$y = f\left(x_1, x_2, \ldots, x_p\right) + \varepsilon .$$

The multiple linear regression model can be written as follows for the linear function of f:

$$y_t = x_{1t}\beta_1 + x_{2t}\beta_2 + \ldots + x_{pt}\beta_p + \varepsilon_t \qquad t = 1, 2, \ldots, T \tag{2.1.}$$

There are T sets of observations on y, $\left(x_1, x_2, \ldots, x_p\right)$ and ε. The parameters $\beta_1, \beta_2, \ldots, \beta_p$ are the regression coefficients. The equation (2.1) can be written in the matrix form:

$$y = X\beta + \varepsilon \tag{2.2.}$$

where X is a $T \times p$ design matrix, y is a $T \times 1$ vector of observable random variables, β is a $p \times 1$ vector of unknown regression coefficients and ε is a $T \times 1$ vector of disturbances. The following conventional assumptions are often used:

- X is a non-stochastic matrix with $p < T$. If $T < p$, there will exist fewer observations, $t = 1, 2, \ldots, T$, than unknowns $\beta_1, \beta_2, \ldots, \beta_p$ that makes it hard to think any reasonable estimation process. The assumption of a nonstochastic nature of the elements in X is often emphasized by writing

$E(y|X) = X\beta$ instead of $E(y) = X\beta$. This assumption is often weakened to the assumption that the source of data in X is unrelated to the source of ε in econometrics. However, if there is a relation between X and ε, classical regression methods are supposed to yield inadequate results that requires to consider alternative methods.

- The rank of X is p (full column rank). If the values of explanatory variables are measured qualitatively, this asspumption is often satisfied and this prevents the exact linear dependencies between the columns of X. The full rank assumption is not disturbed by a time trend and/or certain dummy variables.
- The elements of ε are non-observable random variables such that $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 I_T$ with $\sigma^2 > 0$.The variance σ^2 is unknown, but it is the same for each observation. Also, $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, namely the covariances between the two random errors corresponding to two different observations are zero.

Let $\Lambda = diag(\lambda_1, \dots, \lambda_p)$ be a diagonal matrix whose diagonal elements are eigenvalues of $X'X$. Let T be a $p \times p$ matrix whose columns are the eigenvectors of $X'X$ and $TT' = T'T = I_p$. Then,

$$y = Z\alpha + \varepsilon,\ \varepsilon \sim (0, \sigma^2 I_p) \qquad (2.3.)$$

is called canonical model where $Z = XT$, $\alpha = T'\beta$ and $Z'Z = T'X'XT = \Lambda$.

The statistician's task is now to estimate the unknown vector β of the regression parameters in the model (2.2) on the basis of observations and assumptions already stated. Linear estimators for the $p \times 1$ vector of β are of the form

$$\tilde{\beta} = Cy + d$$

where C is an $p \times T$ non-stochastic matrix and d is a $p \times 1$ non-stochastic vector. If d is equal to the $p \times 1$ zero vector, then the linear estimator is called homogeneous estimator, otherwise the linear estimator is called heterogeneous (non-homogeneous) estimator. A linear estimator $\tilde{\beta}$ is called unbiased for β if $E(\tilde{\beta}) = \beta$.

Comparisons of the performances of the different estimators can be made on the basis of their mean square errors. The matrix mean square error (MSE) of an estimator $\tilde{\beta}$ is

$$\begin{aligned} MSE(\tilde{\beta}) &= E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right] \\ &= Var(\tilde{\beta}) + [Bias(\tilde{\beta})][Bias(\tilde{\beta})]' \end{aligned}$$

where the variance matrix of an estimator $\tilde{\beta}$ is

$$Var(\tilde{\beta}) = E\left[\left(\tilde{\beta} - E(\tilde{\beta})\right)\left(\tilde{\beta} - E(\tilde{\beta})\right)'\right]$$

and the bias is

$$Bias(\tilde{\beta}) = E(\tilde{\beta}) - \beta.$$

The scalar mean square error (mse) of an estimator $\tilde{\beta}$ is defined as

$$\begin{aligned} mse(\tilde{\beta}) &= E\left[(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)\right] \\ &= Tr[Var(\tilde{\beta})] + [Bias(\tilde{\beta})]'[Bias(\tilde{\beta})]. \end{aligned}$$

2.2. Principle of Ordinary Least Squares and the Need for Alternative Estimators

Statisticians often use the method of least squares to estimate the parameters of a linear regression model. The OLS estimator is obtained by minimizing the sum

of squares of the error terms for the model (2.2). Thus,

$$F(\beta) = (y - X\beta)'(y - X\beta)$$

is differentiated with respect to β and equated to zero as

$$\frac{\partial F(\beta)}{\partial \beta} = X'X\beta - X'y = 0.$$

The normal equations are yielded as

$$X'X\beta = X'y. \tag{2.4.}$$

If X is full rank p, then $X'X$ is nonsingular and the unique solution of (2.4.), namely the OLS estimator is:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{2.5.}$$

If X is not of full rank, the equation (2.4) has a set of solutions

$$\hat{\beta} = (X'X)^{-} X'y + \left(I - (X'X)^{-} X'X\right)w,$$

where $(X'X)^{-}$ is a generalized inverse of $X'X$ and w is an arbitrary vector.

The most common estimator for the unknown σ^2 is

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - p}.$$

which is based on the residual vector of the least squares residuals $\hat{\varepsilon} = y - X\hat{\beta}$ and it is unbiased for σ^2.

One of the properties of the OLS estimator is

$$E\left(\hat{\beta}\right)=\beta$$

which shows the unbiasedness of $\hat{\beta}$ that results in the $Var\left(\hat{\beta}\right)=MSE\left(\hat{\beta}\right)$ as

$$MSE\left(\hat{\beta}\right)=\sigma^{2}\left(X'X\right)^{-1}. \tag{2.6.}$$

Gauss - Markov theorem establishes that $\hat{\beta}$ is uniformly not worse than any estimator $\tilde{\beta}$ from the set of linear unbiased estimators for β. Namely, the OLS estimator of β is best linear unbiased estimator (BLUE) having the smallest variance among the class of all unbiased estimators. However this theorem does not ensures that this variance will be small. Multicollinearity results in large variances for the OLS estimators of the regression coefficients implying that confidence intervals on β will be wide and the point estimate $\hat{\beta}$ will be very unstable. We can find a biased estimator $\tilde{\beta}$ which has a smaller variance than the unbiased estimator $\hat{\beta}$ to ease this problem. Decrease in variance will be more than the increase in bias of $\tilde{\beta}$ so the MSE of $\tilde{\beta}$ will be less than the variance of $\hat{\beta}$ implying that $\tilde{\beta}$ is a more stable estimator of β. Besides this, confidence intervals on β will be much narrower by using the biased estimator.

The optimality of the OLS estimator within the sets of linear unbiased estimators does not mean that it is necessarily a good estimator as a whole. Although $\hat{\beta}$ is unbiased for β, its expected squared lenght is greater than the squared lenght of β resulting with the unreliable results. The squared distance of any other linear unbiased estimators can be even greater than the squared lenght of $\hat{\beta}$, but nevertheless estimation with least squares in such situations can take the risk of an individual estimate $\hat{\beta}$ being far away from the true parameter β. The squared distance from $\hat{\beta}$ to the true parameter vector β is

$$L_1{}^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

and the expected squared distance, $E(L_1{}^2)$, is

$$E(L_1{}^2) = \sigma^2 Tr(X'X)^{-1}$$
$$= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \qquad (2.7.)$$

where $\lambda_j > 0$, $j = 1, \dots, p$ are the eigenvalues of $X'X$. In the presence of multicollinearity, at least one of the λ_j will be small and (2.7) implies that the distance from the least squares estimate $\hat{\beta}$ to the true parameter β may be large. Namely the vector $\hat{\beta}$ is generally longer than the vector β. This implies that the method of least squares produces estimated regression coefficients that are too large in absolute value. Alternative biased estimators $\tilde{\beta}$ turn out to be not uniformly better than the OLS estimator but nevertheless shorten $\hat{\beta}$.

2.3. Multicollinearity

The numbers of observations are sometimes limited, variables do not vary over a very wide range and thus the samples often do not provide enough information. This lack of sufficient information and the corresponding loss of precision of statistical results based on such data leads to what is commonly called the multicollinearity problem. Multicollinearity means linear dependence among the columns of X. A distinction is often made between exact multicollinearity and near multicollinearity defined as follows:

Exact multicollinearity exists when the rank of X is less than p, namely the columns of X denoted by $x_1, x_2, \dots, x_p$ are linearly dependent. In other words, there exist nonzero constants a_j, $j = 1 \dots . p$ such that

$$\sum_{j=1}^{p} a_j x_j = 0.$$

Near multicollinearity exists when there are nonzero constants a_j such that

$$\sum_{j=1}^{p} a_j x_j \approx 0$$

where ≈ denotes approximate equality.

Exact multicollinearity implies that $X'X$ is singular, i.e. the smallest eigenvalue is $\lambda_p = 0$. Near multicollinearity means that $X'X$ is nearly singular and the smallest eigenvalue is close to zero.

When the condition of exact multicollinearity exists, the $X'X$ matrix is singular and the OLS estimator $\hat{\beta}$ is not defined. Thus we can not obtain estimates of β using the least squares estimation rule. When the variables exhibit near linear dependencies or collinearity, the following consequences are:

- The variance of the OLS estimator is $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ so this means that the sampling variances, standard errors and covariances of the OLS estimator may be large. Large standard errors for the OLS estimators imply that sampling variability is high, interval estimates are wide and the information provided by the sample data about the unknown parameters is relatively imprecise.
- To explain high variance of $\hat{\beta}$ clearly, the trace of variance matrix (2.6) of the OLS estimator, the mse is:

$$mse(\hat{\beta}) = \sigma^2 Tr(X'X)^{-1}. \quad (2.8.)$$

- In terms of the eigenvalues of $X'X$, the equation (2.8) is defined to be

$$MSE(\hat{\beta}) = \sigma^2 \sum_{j=1}^{p} {\lambda_j}^{-1} > \sigma^2 {\lambda_p}^{-1}. \quad (2.9.)$$

For $\lambda_p \to 0$ and (2.9) implies that the $MSE \to \infty$ which means that $\hat{\beta}$ is subject to very large variance. High variance of regression coefficients would

generally mean that the null hypothesis $H_0: \beta_i = 0$ will be more likely to be accepted. Namely, the usual t-tests will lead to the conclusion that parameter values are not significantly different from zero. This outcome occurs despite possibly high R^2 or F-values indicating significant explanatory power of the model for this combination of variables. The problem is that collinear variables do not provide enough information to estimate their separate effects. Another practical difficulty with the estimated students' t values based on multicollinear data is that these values are highly unstable and often change their signs and relative magnitudes with minor perturbations in data.

- Multicollinearity can result the estimated regression coefficients appearing to have the wrong sign (Farrar and Glauber (1967), Mullet (1976)) i.e. opposite to be a prior expectations of the researcher. Excessive caution is required to trust on the appearance of wrong signs because wrong sign may not even be wrong.Visco (1978) and Leamer (1975) have shown by algebraic methods that omitting a variable with relatively low t value cannot correct the wrong sign of a regressor having a higher t value. However, replacing a less significant regressor by another can indeed remove the wrong sign problem in certain cases. Also omission of two collinear regressors at the same time can change the signs of significant regressors and even make them insignificant.

2.3.1. Identifying the Multicollinearity

The methods that have been proposed for detecting the presence and nature of multicollinearity are discussed.

Examination of Correlation Matrix: Simple correlations r_{ij} between pairs of regessors can indicate linear associations between them. However, there is no assurance that any of the pairwise correlations will be large for collinear relationships involving three or more variables.

Variance Inflation Factors: Diagonals of $(X'X)^{-1}$ are called variance inflation factors (VIF) by Marquardt (1970). For the standardized data, $X'X$ is the correlation

matrix and the (i,j)th element of the inverse $(X'X)^{-1}$ may be denoted as r^{ij}. Farrar and Glauber (1967) were the first to suggest looking at the values of r^{jj} to diagnose multicollinearity. Marquardt (1970) suggests a rule of thumb that $VIF_j = r^{jj} > 5$ indicates harmful multicollinearity. Theil (1971) shows that

$$r^{jj} = \frac{1}{(1-{R_j}^2)\|x_j\|^2}$$

where $\|x_j\|^2 = x_j'x_j$ and where ${R_j}^2$ represents the squared multiple correlation coefficient when x_j (j-th column of X) is regressed on the remaining $p-1$ regressors. When multicollinearity is present, ${R_j}^2$ will be large (close to 1) and the denominator of r^{jj} will be close to zero, hence r^{jj} will be large.

Eigensystem Analysis of $X'X$: The analysis of the characteristic roots of $X'X$ can reveal the presence and perhaps the nature of multicollinearity. One or more small eigenvalues imply that there are near linear dependence among the columns of X. The condition number of $X'X$ is

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}}.$$

There is no serious problem with multicollinearity if the condition number is less than 100. However condition numbers between 100 and 1000 imply moderate to strong multicollinearity and if κ exceeds 1000, severe multicollinearity is indicated. The condition indices of the $X'X$ matrix are

$$\kappa_{j=}\frac{\lambda_{max}}{\lambda_j}, j = 1,2, \dots p.$$

Clearly the largest condition index is the condition number. The number of condition indices that are large ($say \geq 1000$) is a useful measure of the number of near-linear dependence in $X'X$.

2.3.2. Solutions to the Multicollinearity Problem

Several techniques have been proposed for dealing with the problems caused by multicollinearity.

2.3.2.1. Traditional Solutions to the Multicollinearity Problem

Additional Sample Information: The data do not include sufficient information about the individual effects of control variables to enable us to estimate the parameters of the statistical model precisely, in the presence of multicollinearity. One solution is to obtain more information and include it in the analysis. Unfortunately collecting data is not always possible because of economic constraints or because the process being studied is no longer available for sampling. Also, the new observations may suffer the same collinear relationships.

An alternative to new sample information is to add structure by introducing non-sample information in the form of restrictions on the parameters.

Exact linear Constraints: One effect of using exact parameter restrictions is to reduce the sampling variability of the estimators. The imposition of binding constraints, even if incorrect restrictions produce biased parameter estimators. Exact linear restrictions may be employed in the case of both extreme and near-extreme multicollinearity.

Stochastic Linear Restrictions: Stochastic linear restrictions arise from prior statistical information, usually in the form of previous estimates of parameters that are also included in a current model. These linear stochastic restrictions act just like additional observations, if the statistical information is unbiased, and thus they may serve to aid in both situations of extreme and near extreme multicollinearity.

Linear Inequality Restrictions: Their use can improve the precision of estimation when near-extreme multicollinearity is present and make it possible to estimate a function that cannot be estimated because inexact restrictions may not effect the dimensionality of the parameter space.

2.3.2.2. Ad Hoc Solutions to the Multicollinearity Problem

It is noted that conventional procedures for combining sample and non sample information can be used in an attempt to reduce the ill effects of multicollinearity. Some of the ad hoc solutions are the biased estimators such as the RE, LE and Stein type estimator, because the information it introduce is specific to sample in hand.

2.4. Ridge Estimator

The RE

$$\hat{\beta}_k = \left(X'X + kI_p\right)^{-1}X'y\,, \quad k \geq 0 \tag{2.10.}$$

is a biased linear estimator introduced by Hoerl and Kennard (1970a). In the presence of collinearity, this estimator can yield better estimates than the OLS estimator $\hat{\beta}$. The use of this estimator is motivated from the fact that the matrix $X'X$ is stabilized when a positive number k is added to each of its main diagonal elements. By increasing k, the ridge of matrix $X'X + kI$ is raised so that $X'X + kI$ resembles more and more a diagonal matrix, which has an optimal condition. The ultimate goal is to find some k which is large enough to reduce the variance compared to the OLS estimator, but which is small enough to produce some acceptable low bias.

Originally Hoerl and Kennard (1970a) obtained the RE by finding the point on the ellipsoid centered at the OLS estimator $\hat{\beta}$ that is closest to the origin. This consist of minimizing the distance $\beta'\beta$ of a vector of parameters β from the origin subject to the side condition:

$$\left(\beta - \hat{\beta}\right)'X'X\left(\beta - \hat{\beta}\right) = \Phi_0.$$

The optimization problem is solved using Lagrange multipliers. Differentiate

$$L = \beta'\beta + \frac{1}{k}\left[\left(\beta - \hat{\beta}\right)'X'X\left(\beta - \hat{\beta}\right) - \Phi_0\right]$$

to obtain

$$\beta + \frac{1}{k}X'X\left(\beta - \hat{\beta}\right) = 0. \qquad (2.11.)$$

The RE in the equation (2.10) is obtained by solving the equation (2.11) for β.

When the squared length of the regression vector β is fixed at c^2, in order to make the sum of squares minimum, the Lagrange multipliers is used to solve the problem:

$$L = (y - X\beta)'(y - X\beta) + \frac{1}{k}(\beta'\beta - c^2).$$

2.4.1. Properties of the Ridge Estimator

The RE can be written in the form

$$\hat{\beta}_k = (X'X + kI)^{-1}X'y = WX'y, \quad k \geq 0.$$

Since

$$E\left(\hat{\beta}_k\right) = WX'X\beta = \beta - kW\beta,$$
$$Bias\left(\hat{\beta}_k\right) = -kW\beta,$$
$$Var\left(\hat{\beta}_k\right) = \sigma^2 WX'XW,$$

the MSE of the RE is

$$MSE(\hat{\beta}_k) = W(\sigma^2 X'X + \mathrm{k}^2\beta'\beta)W. \quad (2.12.)$$

The mse of the $\hat{\beta}_k$ is obtained from the (2.12) as

$$mse(\hat{\beta}_k) = Tr\left(MSE(\hat{\beta}_k)\right)$$
$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j+k)^2} + k^2\beta'(X'X+kI)^{-2}\beta \quad (2.13.)$$

where $\lambda_j > 0, j = 1, \dots, p$ are the eigenvalues of $X'X$.

The total variance decreases as k increases, while the squared bias increases with k. There are values of k for which the mean square error is less for $\hat{\beta}_k$ than it is for the usual solution $\hat{\beta}$. These can be summarized in the following results:

- The total variance is a continuous, monotonically decreasing function of k.
- The squared bias is a continuous, monotonically increasing function of k.
- Existence Theorem (Hoerl and Kennard, 1970a): There always exists a $k > 0$ such that $mse(\hat{\alpha}_k) < mse(\hat{\alpha})$ and

$$k < \frac{\sigma^2}{\alpha_{max}^2} \quad (2.14.)$$

where $\hat{\alpha}$, $\hat{\alpha}_k$ and α_{max}^2 are the OLS estimator, the RE and the maximum coefficient from the canonical model (2.3), respectively.

2.4.2. Choice of the Ridge Parameter

There exist several ways to determine k. For determining an appropriate k, two methodical approaches are subjective and objective methods.

2.4.2.1. Ridge Trace

Hoerl and Kennard (1970a) suggested ridge trace method (subjective method)

which considers the p elements of the RE as functions of k, all of them plotted into one figure. Theil (1963) seems to have independently suggested a similar plot. When data is collinear the regression coefficients appear to change very rapidly as k varies for small values. At some value of k they stabilize; the rate of change slows down to almost zero. The user then chooses the value of k at which the functions begin to stabilize. This methods of choosing k is subjective since the judgment of the ridge trace depends on the user. Also, the impression of stability will be quite different when the ridge trace is considered for different range of k values.

Hoerl and Kennard (1970b) recommended standardizing the data before plotting on a ridge trace, so as to retain numerical comparability of regression coefficients. Vinod and Ullah (1981) points out that the appearance of a ridge trace that does not plot standardized regression coefficients may be dramatically changed by a simple translation of the origin and scale transformation of the variables.

2.4.2.2. Estimation of k

An estimator $\hat{k}$, as a choice of k, can be seen as an objective method in the sense that the choice of k only depends on y.

For any given parameter values ($\beta \neq 0,\ \sigma^2$) there exists some optimal k_{opt} for which $MSE\left(\hat{\beta}_{k_{opt}}\right) \leq MSE(\hat{\beta}_k)$for any $k \geq 0$ and

$$k_{opt} = \frac{p\sigma^2}{\beta'\beta}. \tag{2.15.}$$

Based on this idea, some ad-hoc possibilities for determining an estimator for k is considered.

Estimator from Hoerl and Kennard (1970a): Hoerl and Kennard (1970a) suggested an estimator for k in the equation (2.14). By replacing the OLS estimates of the σ^2 and α_{max}^2, estimator of k is determined as

$$k < \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2}.$$

Estimator from Theobald (1974): Theobald (1974) extended the result of Hoerl and Kennard (1970a) above. Weighted mean square error $E\left(\tilde{\beta}-\beta\right)'B\left(\tilde{\beta}-\beta\right)$, where B is non negative definite matrix for different estimators $\tilde{\beta}$, is used to compare the OLS estimator and the RE. Thus, Theobald (1974) suggested an estimator for k

$$\hat{k}_T = \frac{2\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

where $\hat{\sigma}^2$ and $\hat{\beta}$ are the OLS estimates.

Estimator from Hoerl, Kennard and Baldwin (1975): Hoerl, Kennard and Baldwin (1975) used the two results in improving an algorithm for selecting k. Firstly, k in the equation (2.15) is considered. Secondly, the general form of optimal values for $k_j = \frac{\sigma^2}{\alpha_j^2}$ is evaluated. For combining these individual k_i to obtain a single value of k, Hoerl, Kennard and Baldwin (1975) proposed the use of harmonic mean of the k_j values. So, an estimator of k for automatic selection of choice is

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}.$$

Estimator from Hoerl and Kennard (1976): The estimator $\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ which is proposed by Hoerl, Kennard and Baldwin (1975) often produces small value of k. An iteration of estimation process is proposed to remedy this problem by considering a sequence of estimates of β and k:

$$\hat{\beta} : k_0 = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

$$\hat{\beta}_k(k_0) : \; k_1 = \frac{p\hat{\sigma}^2}{\hat{\beta}_k'(k_0)\hat{\beta}_k(k_0)}$$

$\hat{\beta}_k(k_1)$: $k_2 = \frac{p\hat{\sigma}^2}{\hat{\beta}_k'(k_1)\hat{\beta}_k(k_1)}$

.

.

$\hat{\beta}_k(k_t)$: $k_{t+1} = \frac{p\hat{\sigma}^2}{\hat{\beta}_k'(k_t)\hat{\beta}_k(k_t)}$.

The sequence is continued if $\frac{\hat{k}_{j+1} - \hat{k}_j}{\hat{k}_j} > 20T^{-1.3}$ where $T = \frac{1}{p} tr\left(X'X\right)^{-1}$, otherwise it is terminated and $\hat{\beta}_k\left(k_j\right)$ is used. Since T increases with the spread in the eigenvalues of $X'X$ allowing further shrinkage as the degree of ill conditioning in the data increases, this choice of termination criterion is selected.

Estimator from McDonald and Galarneau (1975): McDonald and Galarneau (1975) offer choosing k so that

$$\hat{\beta}_k'\hat{\beta}_k = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \sum_{j=1}^{p} \lambda_j{}^{-1}. \qquad (2.16.)$$

If the right hand side of (2.16) is negative, k is taken zero (OLS) or k is taken ∞ $\left(\hat{\beta}_k = 0\right)$. Both methods were not better than OLS in all cases.

Estimator from Lawless and Wang (1976): Lawless and Wang (1976) give a different, Bayesian oriented motivation for the corresponding estimator $\hat{k}_{HKB}$ as follows:

$$\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\hat{\beta}'X'X\hat{\beta}} .$$

It is also indicated that the usage of $\hat{k}_{LW}$ increases the performance of the RE in comparisons with the OLS estimator with a Monte Carlo simulation.

2.5. Stein Type Estimator

Stein (1956) pointed out that the standart estimator of the normal mean can be improved under certain conditions. At first, this result was confusing for the statisticians. Then, the result interpretated in many papers for the regression analysis. A family of biased estimators, called Stein-type or Stein-rule estimators, with smaller mean square error are offerred as an alternative to the OLS estimator.

The Stein estimator for β is

$$\hat{\beta}_S = \left(1 - \frac{c\sigma^2}{\hat{\beta}'X'X\hat{\beta}}\right)\hat{\beta} \qquad (2.17.)$$

where $c > 0$ is any constant and $\hat{\beta}$is the OLS estimator. When σ^2 has an unknown value in (2.17), James and Stein (1961) consider the following estimator

$$\hat{\beta}_{JS} = \left[1 - \left(\frac{c}{T-p}\right)\frac{\hat{u}'\hat{u}}{\hat{\beta}'X'X\hat{\beta}}\right]\hat{\beta}$$

where c is an arbitrary constant and σ^2 has been substituted by its unbiased estimator

$$s^2 = \frac{\hat{u}'\hat{u}}{T-p} = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{T-p}.$$

The Stein type estimator can be written in a form as follows:

$$\hat{\beta}_{ST} = \left(1 - \frac{k_1\hat{u}'\hat{u}}{\hat{\beta}'X'X\hat{\beta}}\right)\hat{\beta} \qquad (2.18.)$$

where $k_1 > 0$ is an arbitrary scalar constant.

Some properties of the Stein type estimator (2.18) are:

$$Bias(\hat{\beta}_{ST}) = -\frac{(T-p)k_1\sigma^2}{\beta'X'X\beta}\beta,$$

$$mse\left(\hat{\beta}_{ST}\right) = \sigma^2 Tr(X'X)^{-1} + \frac{(T-p)k_1\sigma^4}{(\beta'X'X\beta)^2}\left[\beta'\beta\left(4 + k_1(T-p+2)\right) - 2(\beta'X'X\beta)Tr(X'X)^{-1}\right].$$

The Stein type estimator dominates the OLS in the sense of mean square error. For the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ of the matrix $X'X$, $\lambda_p \leq \beta'X'X\beta/\beta'\beta \leq \lambda_1$ is concluded from Rao (1973) and then it is approved that from $mse\left(\hat{\beta}_{ST}\right)$ that $mse\left(\hat{\beta}_{ST}\right) < mse\left(\hat{\beta}\right)$ for

$$0 < k_1 \leq \frac{2}{T-p+2}(d-2)$$

where $d = \lambda_p \sum_{i=1}^{p} \lambda_i^{-1} > 2$ and $\sum_{i=1}^{p} \lambda_i^{-1} = Tr(X'X)^{-1}$.

Mayer and Willke (1973) introduced an estimator called shrinkage estimator or contraction estimator

$$\hat{\beta}(\delta) = \frac{1}{1+\delta}\hat{\beta},\ \delta \geq 0 \qquad (2.19.)$$

as an alternative biased estimator to the OLS estimator.When, the shrinkage parameter δ is fixed, (2.19) is a homogeneous estimator that satisfies

$$E\left(\hat{\beta}(\delta)\right) = \frac{1}{1+\delta}\beta,$$
$$Var\left(\hat{\beta}(\delta)\right) = \frac{\sigma^2}{(1+\delta)^2}(X'X)^{-1}.$$

The OLS estimator is multiplied by the same factor $1/(1+\delta)$ so the sign and the proportion of the individual estimates are resumed. The factor $\gamma = 1/(1+\delta)$ lies between 0 and 1 when $\delta \geq 0$. In the case of $X'X = I_p$, $\hat{\beta}_k = \hat{\beta}(k)$, namely the ridge estimator with parameter k is the same as the shrinkage estimator with $\delta = k$.

2.6. Liu Estimator

The RE and the Stein-type estimator are two fundamental biased estimators which are proposed in the case of multicollinearity. Both of the estimators have some advantages and disadvantages.

Although the RE is practical, we have difficulty in the choice of its biasing parameter since the RE has a complicated function of this parameter. The condition number of $X'X + kI$ decreases for sufficiently greater values of k. However, the values of k is expected to be smaller in practice. Therefore, stability of the ridge regression can not be provided.

Besides the advantage of being linear function of its biasing parameter, Stein type estimator is not efficient in practise due to the same shrinkage of each element.

Liu (1993) introduced a new estimator which combines the advantages of the RE and Stein type estimator to handle multicollinearity:

$$\hat{\beta}_d = \left(X'X + I\right)^{-1}\left(X'y + d\hat{\beta}\right), \quad 0 < d < 1. \tag{2.20.}$$

This estimator (2.20), as a linear function of its biasing parameter, yields convenient selection opportunity of the biasing parameter. Subsequently, Akdeniz and Kaçıranlar (1995) entitled this new estimator as the LE.

The equality $0 = k^{1/2}\beta + \varepsilon'$ is considered with the model (2.2) by using the least squares estimation and the RE is derived. In a similar manner, the LE is obtained by using the equation $d\hat{\beta} = \beta + \varepsilon'$.

Some properties of this estimator are discussed below:

- For $d = 1$, $\hat{\beta}_d = \hat{\beta}$. Besides this, $\hat{\beta}_d$ can be written as

$$\hat{\beta}_d = \left(X'X + I\right)^{-1}\left(X'X + dI\right)\hat{\beta} \tag{2.21.}$$

and

$$E\left(\hat{\beta}_d\right)=\left(X'X+I\right)^{-1}\left(X'X+dI\right)\beta$$
$$Var\left(\hat{\beta}_d\right)=\sigma^2 M_d\left(X'X\right)^{-1}M_d'$$

where $M_d=(X'X+I)^{-1}(X'X+dI)$. There exists $0<d<1$ such that the mse of $\hat{\beta}_d$ outperforms the mse of $\hat{\beta}$. The MSE and the mse of the $\hat{\beta}_d$ (2.21) are:

$$MSE\left(\hat{\beta}_d\right)=\sigma^2 M_d\left(X'X\right)^{-1}M_d'+\left(M_d-I\right)\beta\beta'\left(M_d-I\right),$$
$$mse\left(\hat{\beta}_d\right)=Tr\left(\sigma^2 M_d\left(X'X\right)^{-1}M_d'\right)+\beta'\left(M_d-I\right)\left(M_d-I\right)\beta.$$

- For the canonical form of the model (2.3), the LE is $\hat{\alpha}_d=\left(\Lambda+I\right)^{-1}\left(\Lambda+dI\right)\hat{\alpha}$ and the mse of the $\hat{\alpha}_d$ is

$$mse\left(\hat{\alpha}_d\right)=\sigma^2\sum_{i=1}^{p}\frac{\left(\lambda_i+d\right)^2}{\lambda_i\left(\lambda_i+1\right)^2}+\left(d-1\right)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{\left(\lambda_i+1\right)^2}. \qquad (2.22.)$$

(2.22) is minimized at

$$d=\sum_{i=1}^{p}\frac{\alpha_i^2-\sigma^2}{\left(\lambda_i+1\right)^2}\Bigg/\sum_{i=1}^{p}\frac{\sigma^2+\lambda_i\alpha_i^2}{\lambda_i\left(\lambda_i+1\right)^2}$$
$$=1-\sigma^2\left(\sum_{i=1}^{p}\frac{1}{\lambda_i\left(\lambda_i+1\right)}\right)\Bigg/\left(\sum_{i=1}^{p}\frac{\sigma^2+\lambda_i\alpha_i^2}{\lambda_i\left(\lambda_i+1\right)^2}\right). \qquad (2.23.)$$

α_i^2 and σ^2 are replaced with their unbiased estimate $\hat{\alpha}_i^2-\hat{\sigma}^2/\lambda_i$ and $\hat{\sigma}^2$ in the equation (2.23), thus

$$\hat{d} = 1 - \hat{\sigma}^2 \left(\sum_{i=1}^{r} \frac{1}{\lambda_i (\lambda_i + 1)} \Big/ \sum_{i=1}^{r} \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2} \right). \tag{2.24.}$$

(2.24) is generalized to

$$\hat{d}_{opt} = 1 - h\hat{\sigma}^2 \left(\sum_{i=1}^{r} \frac{1}{\lambda_i (\lambda_i + 1)} \Big/ \sum_{i=1}^{r} \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2} \right) \tag{2.25.}$$

where $h > 0$.

3. SIMULTANEOUS EQUATIONS MODEL

3.1. Nature of Simultaneous Equations Model

The cause and effect relationship in the single equation models ran from explanatory variables (X) to dependent variable (Y). However, a one way relationship like that is not valid for simultaneous equations models. There is a two way, simultaneous relation between X and Y therefore it will be more complicated to determine the dependent and explanatory variables. In such models, if one does not consider the information provided by other equations in the system, parameters of a single equation may not be estimated. If the explanatory variables are stochastic or nonstochastic and are not distributed independently of the stochastic disturbance term the assumption of the OLS estimators are inconsistent. Consider the following system of equations

$$Y_{1i} = \beta_{10} + \beta_{12}Y_{2i} + \gamma_{11}X_{1i} + u_{1i} \quad (3.1.)$$

$$Y_{2i} = \beta_{20} + \beta_{21}Y_{1i} + \gamma_{21}X_{1i} + u_{2i} \quad (3.2.)$$

where Y_1 and Y_2 are mutually dependent variables; X_1 is an explanatory variable and u_1 and u_2 are the stochastic disturbance terms. Y_1 and Y_2 are both stochastic so, if the explanatory variable Y_2 in the equation (3.1) and Y_1 in the equation (3.2) are not distributed independently with u_1 and u_2, respectively, the OLS estimation of these equations will be inconsistent. For the simple Keynesian model of income determination is as follows

$$C_t = \beta_0 + \beta_1 Y_t + u_t,\ 0 < \beta_1 < 1 \quad (3.3.)$$

$$Y_t = C_t + I_t. \quad (3.4.)$$

The variables C, Y, I, t and u are defined as consumption expenditure,

income, investment, time and stochastic disturbance term, respectively. β_0 and β_1 are the parameters. The first equation is the consumption function and the second one is the national income identity. From economic theory, β_1 is expected to be between 0 and 1. It is clear that C and Y are interdependent and Y_t in equation (3.3) is not expected to be independent of the disturbance term. Since the change in u_t affects C_t, Y_t is also affected because of bilateral relationship. Thus, the OLS method is inapplicable to equation (3.3). While estimating the parameters of the consumption function, the assumptions of the classical linear regression model are $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_{t+j}) = 0$ (for $j \neq 0$) and $Cov(I_t, u_t) = 0$. Under these assumptions, it can be shown that Y_t and u_t are correlated and $\hat{\beta}_1$ is inconsistent for β_1.

Replacing C_t in the equation (3.3) to the equation (3.4), Y_t can be written as:

$$Y_t = \beta_0 + \beta_1 Y_t + u_t + I_t$$
$$= \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t.$$

Covariance between Y_t and u_t is

$$Cov(Y_t, u_t) = E[Y_t - E(Y_t)][u_t - E(u_t)]$$
$$= \frac{\sigma^2}{1-\beta_1} \qquad (3.5.)$$

and it is different from zero since σ^2 is positive. To show the inconsistency of the OLS estimator $\hat{\beta}_1$, the probability limit is checked whether it is equal to its true value. $\hat{\beta}_1$ can be written as

$$\hat{\beta}_1 = \frac{\sum(C_t - \bar{C})(Y_t - \bar{Y})}{\sum(Y_t - \bar{Y})^2}$$
$$= \beta_1 + \frac{\sum y_t u_t}{\sum y_t^2}$$

where the lower case letter shows deviations from the mean value. Its probability limit is as follows

$$plim(\hat{\beta}_1) = plim\,(\beta_1) + plim\,\left(\frac{\sum y_t u_t/n}{\sum y_t^2/n}\right)$$
$$= \beta_1 + \frac{plim(\sum y_t u_t/n)}{plim(\sum y_t^2/n)}.$$

As the sample size n increases, the sample covariances between Y and u will approach to population covariance (3.5) and similarly the sample variance of Y will be close to its population variance σ_Y^2. Finally,

$$plim(\hat{\beta}_1) = \beta_1 + \frac{\sigma^2/(1-\beta_1)}{\sigma_Y^2}.$$

Since, $0 < \beta_1 < 1$ and σ^2 and σ_Y^2 are both positive, $plim(\hat{\beta}_1)$ is always greater than β_1.

For the standard notation of simultaneous equations model, consider M structural equations that present the relationships among M jointly determined endogenous variables, N predetermined variables and M random errors. The model can be stated as

$$\begin{aligned}
&\gamma_{11}y_1 + \gamma_{21}y_2 \ldots + \gamma_{M1}y_M + \beta_{11}x_1 + \beta_{21}x_2 \ldots + \beta_{N1}x_N = u_1\\
&\gamma_{12}y_1 + \gamma_{22}y_2 \ldots + \gamma_{M2}y_M + \beta_{12}x_1 + \beta_{22}x_2 \ldots + \beta_{N2}x_N = u_2\\
&\qquad\qquad\vdots\\
&\gamma_{1M}y_1 + \gamma_{2M}y_2 \ldots + \gamma_{MM}y_M + \beta_{1M}x_1 + \beta_{2M}x_2 \ldots + \beta_{NM}x_N = u_M
\end{aligned} \qquad (3.6.)$$

and represents T observations on M endogenous variables by $(T \times 1)$ vectors $y_1, \ldots, y_M$; N predetermined variables by $(T \times 1)$ vectors $x_1, \ldots, x_N$ and M random error variables by $(T \times 1)$ vectors $u_1, \ldots, u_M$. The equations are known as structural or behavioral equations which describes the structure of an economic model. The $\gamma's$ and $\beta's$ are structural parameters or coefficients of the endogeneous and

predetermined variables, respectively. Endogenous variables are determined within the model and stochastic. Predetermined variables are determined outside the model and nonstochastic. Predetermined variables consist of exogenous-lagged or not- and lagged endogenous.

In the matrix notation, the model (3.6) can be written as

$$\mathrm{Y\Gamma + XB = U} \tag{3.7.}$$

where

$$Y = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{M1} \\ y_{22} & y_{22} & \cdots & y_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{MT} \end{bmatrix}$$

and

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{22} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{NT} \end{bmatrix}$$

are the matrices of observations on M jointly dependent and N predetermined variables. Further, in (3.7)

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{M1} \\ \gamma_{22} & \gamma_{22} & \cdots & \gamma_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{1T} & \gamma_{2T} & \cdots & \gamma_{MT} \end{bmatrix}$$

and

$$B = \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{N1} \\ \beta_{22} & \beta_{22} & \cdots & \beta_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1T} & \beta_{2T} & \cdots & \beta_{NT} \end{bmatrix}$$

are the matrices of structural coefficients and

$$U = \begin{bmatrix} u_{11} & u_{21} & \dots & u_{M1} \\ u_{22} & u_{22} & \dots & u_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1T} & u_{2T} & \dots & u_{MT} \end{bmatrix}$$

denotes the matrix of structural disturbances. The columns of Γ and B are the coefficients of different equations in (3.6), the first column of Γ and B are the coefficients of first structural equation and last column of Γ and B are the coefficients of M^{th} equation.

The assumptions for the model are:

1. M structural equations are set interdependently in the simultaneous equations of (3.6), by M jointly dependent and N predetermined variables. $y_t' = (y_{1t}, \dots, y_{Mt})$ and $x_t' = (x_{1t}, \dots, x_{Nt})$ are T row vectors of observations on all jointly dependent and predetermined variables. Γ in (3.7) is a nonsingular matrix.
2. The elements of each row of the matrix X are independently distributed of the structural disturbances in the corresponding row of U. Equivalently, it can be stated that the elements of X are nonstochastic and fixed.
3. The rank of X is $\mathrm{N} \leq \mathrm{T}$. This implies that the columns of X are linearly independent and $X'X$ is nonsingular.
4. For $t = 1, \dots, T$, $u_t' = (u_{1t}, \dots, u_{Mt})$ are the row vectors of the structural disturbances matrix U. The T row vectors are independent random variables with mean $E(u_{it}) = 0$, variance $Var(u_{it}) = \sigma_{ii}$ and covariances $Cov(u_{it}, u_{jt'}) = \sigma_{ij}\ (t = t')$ and $Cov(u_{it}, u_{jt'}) = 0\ (t \neq t')$ where $i, j = 1, \dots, M$. Thus the structural disturbances have zero mean and they are homoscedastic.

The reduced form equations and the associated reduced form coefficients can be obtained from the structural equations. The reduced form of the equations can be derived by solving the equations for jointly dependent variables in terms of the

predetermined variables and stochastic disturbances. Matrix notation (3.7) is postmultiplied by Γ^{-1}

$$Y\Gamma\Gamma^{-1} + XB\Gamma^{-1} = U\Gamma^{-1}$$

and the reduced form is yielded as follows:

$$Y = X\Pi + V. \tag{3.8.}$$

Coefficients matrix of the reduced form is

$$\Pi = -B\Gamma^{-1} = \begin{bmatrix} \pi_{11} \cdots & \pi_{1M} \\ \vdots & \vdots \\ \pi_{N1} \cdots & \pi_{NM} \end{bmatrix} \tag{3.9.}$$

and disturbances matrix of the reduced form is

$$V = U\Gamma^{-1} = \begin{bmatrix} v_{11} \cdots & v_{M1} \\ \vdots & \vdots \\ v_{1T} \cdots & v_{MT} \end{bmatrix}. \tag{3.10.}$$

The OLS methods can be used for estimating the reduced form coefficients since predetermined variables and stochastic disturbances seem on the right side equations, solely and predetermined variables are not correlated with disturbances.

As a summary, for the t^{th} row of the variables, the models in (3.7) and (3.8) are written as

$$y_t'\Gamma + x_t'B = u_t',$$
$$y_t' = x_t'\pi + v_t'$$

and

$$Y\gamma_{.1} + X\beta_{.1} = u_1$$

is one of the specific equation (say, the first) in the model.

3.2. Problem of Identification

If the structural parameters can be determined from the estimates of reduced form parameters, that equation is regarded as identified. In other case, it is unidentified or underidentified. Under consideration of identification, two different cases are exactly identified case where unique values of structural parameters can be obtained and overidentified case where more than one value of structural parameters can be acquired.

3.2.1. Underidentification

Consider the demand and supply function, respectively

$$Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t},\ \alpha_1 < 0 \quad (3.11.)$$

$$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t},\ \beta_1 > 0 \quad (3.12.)$$

and the equilibrium condition is $Q_t^d = Q_t^s$. P is the price, Q^d is quantity demand, Q^s is quantity supply, t is time, α, β are the parameters and u_1, u_2 are stochastic disturbances. By solving the equilibrium condition, equilibrium price is obtained as

$$P_t = \Pi_0 + v_t \quad (3.13.)$$

where

$$\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \tag{3.14.}$$

and $v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$. Replacing P_t in the equation (3.13) into the equation (3.11) or (3.12):

$$Q_t = \Pi_1 + w_t \tag{3.15.}$$

where

$$\Pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \tag{3.16.}$$

and $w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$. (3.13) and (3.15) are reduced form equations and reduced form coefficients are given in equations (3.14) and (3.16). Four structural coefficients $\alpha_0,\ \alpha_1,\ \beta_0$ and β_1 can not be estimated uniquely since there exists only two reduced form coefficients which contain all four structural coefficients.

3.2.2. Exact Identification

The former demand and supply function could not be identified since P and Q exists in both functions and there is no more information. Suppose that demand and supply functions are as follows:

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t},\ \ \alpha_1 < 0,\ \alpha_2 > 0 \tag{3.17.}$$

$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t},\ \beta_1 > 0,\ \beta_2 > 0. \tag{3.18.}$$

Demand function includes an additional variable, income of consumer *I* which is an exogenous variable. There is a price lagged one period predetermined variable

additionally for the supply function. By equating demand to supply the equilibrium price is solved as

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 P_{t-1} + v_t \tag{3.19.}$$

where $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\Pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}$, $\Pi_2 = \frac{\beta_2}{\alpha_1 - \beta_1}$ and $v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$.

Substituting P_t into (3.17) or (3.18), Q_t is obtained as

$$Q_t = \Pi_3 + \Pi_4 I_t + \Pi_5 P_{t-1} + w_t \tag{3.20.}$$

where $\Pi_3 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$, $\Pi_4 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}$, $\Pi_5 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$ and $w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$.

There are six structural coefficients α_0, α_1, α_2, β_0, β_1 and β_2 in the equations (3.17-3.18) and six reduced form coefficients Π_0, Π_1, Π_2, Π_3, Π_4 and Π_5 to estimate in the equations (3.19-3.20). The structural coefficients can be estimated uniquely, namely the parameters of both demand and supply equations can be identified since there exists six equations in six unknowns.

3.2.3. Overidentification

Suppose that the supply function remains the same as before and a wealth variable R is added to the demand function.

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}, \ \alpha_1 < 0, \alpha_2 > 0 \tag{3.21.}$$

$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}, \ \beta_1 > 0, \beta_2 > 0. \tag{3.22.}$$

Using market clearing mechanism the equilibrium price and quantity are as follows:

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + \Pi_3 P_{t-1} + v_t \quad (3.23.)$$

$$Q_t = \Pi_4 + \Pi_5 I_t + \Pi_6 R_t + \Pi_7 P_{t-1} + w_t \quad (3.24.)$$

where $\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\Pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}$, $\Pi_2 = -\frac{\alpha_3}{\alpha_1 - \beta_1}$, $\Pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1}$, $\Pi_4 = -\frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$, $\Pi_5 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}$, $\Pi_6 = -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1}$, $\Pi_7 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$, $w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$ and $v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$.

Seven structural coefficients are comprised by the model (3.21-3.22) while the number of reduced form coefficients in the equation (3.23-3.24) is eight. Unique estimation of all parameters is not possible since the number of equations is greater than the number of unknowns.

Although the supply function remains the same as equation (3.18) it is not identified in the system (3.21) and (3.22) because of over sufficiency of information. The exclusion of the income variable from the supply function in the model (3.21) and (3.22) is enough for identification. However in the system (3.21) and (3.22), both of the income and wealth variable is excluded which means too many restrictions on supply function and more variables are excluded than necessary.

3.2.4. Identification by Zero Restrictions

If an equation in a model excludes one or more variables that are present elsewhere in the model, identification of this equation is possible. Since the coefficients of variables do not appear in an equation it is known as zero restrictions criterion. Suppose that $m_1 + 1 \leq M$ joinly dependent and $N_1 \leq N$ exogenous variables enter any one structural equation of the system (say, the first) and to identify one endogeneous variable as the left hand side variable let its associated coefficient equal minus one. With some rearrangements, variables can be written as:

$$Y = [y_1 \quad Y_1 \quad Y_1^*]$$
$$X = [X_1 \quad X_1^*],$$

where Y_1 and Y_1^* are matrices of T observations on m_1 included and $m_1^* = M - m_1 - 1$ excluded jointly dependent variables, respectively. Similarly X_1 and X_1^* are matrices of T observations on N_1 included and $N_1^* = N - N_1$ excluded exogenous variables.

$$\gamma_{.1} = \begin{bmatrix} 1 \\ -\gamma_1 \\ 0 \end{bmatrix}$$

and

$$\beta_{.1} = \begin{bmatrix} -\beta_1 \\ 0 \end{bmatrix}$$

are the first column of Γ and B, respectively with order of $m_1 \times 1$ and $N_1 \times 1$. The order of zero vector in $\gamma_{.1}$ is $m_1^* \times 1$ and the order of zero vector in $\beta_{.1}$ is $N_1^* \times 1$. Thus,

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + u_1 \tag{3.25.}$$

is the first equation of the system where u_1 is the first column of U.

$$[y_1 \quad Y_1 \quad Y_1^*] = [X_1 \quad X_1^*] \begin{bmatrix} \pi_{11} & \Pi_{11} & \Pi_{11}^* \\ \pi_{21} & \Pi_{21} & \Pi_{21}^* \end{bmatrix} + [v_1 \quad V_1 \quad V_1^*] \tag{3.26.}$$

is the partition of the reduced form equation where π_{11}, π_{21} are $N_1 \times 1$ and $N_1^* \times 1$ vectors, Π_{11}, Π_{21} are $N_1 \times m_1$ and $N_1^* \times m_1$ matrices, respectively. v_1, V_1 and V_1^* are $T \times 1$, $T \times m_1$ and $T \times m_1^*$ disturbances corresponding to y_1, Y_1 and Y_1^*. To obtain the relationship between the parameters of the equation (3.26) with the reduced form parameter such a relation from (3.9)-(3.10) is

$$\Pi\Gamma + B = 0 \tag{3.27.}$$

$$V\Gamma = U. \tag{3.28.}$$

For the first column of Γ and B in the equation (3.27)

$$\Pi\gamma_{.1} + \beta_{.1} = 0$$

$$\begin{bmatrix} \pi_{11} & \Pi_{11} & \Pi_{11}^{*} \\ \pi_{21} & \Pi_{21} & \Pi_{21}^{*} \end{bmatrix} \begin{bmatrix} 1 \\ -\gamma_1 \\ 0 \end{bmatrix} + \begin{bmatrix} -\beta_1 \\ 0 \end{bmatrix} = 0$$

is obtained. Then, the identifiability relationships between the structural parameters and the reduced form parameters for the first equation is as follows:

$$\pi_{11} = \Pi_{11}\gamma_1 + \beta_1 \tag{3.29.}$$

$$\pi_{21} = \Pi_{21}\gamma_1. \tag{3.30.}$$

For the first column of Γ in the equation (3.28)

$$V\gamma_{.1} = u_1$$

$$\begin{bmatrix} v_1 & V_1 & V_1^{*} \end{bmatrix} \begin{bmatrix} 1 \\ -\gamma_1 \\ 0 \end{bmatrix} = u_1$$

provides the relation as follows:

$$v_1 = V_1\gamma_1 + u_1. \tag{3.31.}$$

These relationships have been obtained after using a prior restrictions on the coefficients.

3.2.5. Rules for Identification

To determine the identification of an equation in a system of simultaneous equations there are two conditions which are order and rank conditions.

A necessary condition for identification is the order condition and can be stated as:

- If $M-1$ endogenous or predetermined variables are excluded from one equation of the model of M simultaneous equations, it is just identified. It is overidentified under the case of excluding more than $M-1$ variables.
- In a model with M simultaneous equations, the condition for identification for one of the equation is $N-\mathrm{n} \geq m-1$ (equality means just identified) where N is the number of predetermined variables including intercept, n is the number of predetermined variables in a given equation and m is the number of endogenous variables in a given equation.

Even the order condition is satisfied, the equation can not be identified since the condition is only a necessary conditon. That is, although $N-n \geq m-1$ is satisfied, the equation can be unidentified since the predetermined variables excluded from this equation but present in the model may not all be independent.Therefore, there may not be one to one correspondence between the structural and reduced form coefficients to estimate the structural coefficients.

Rank condition of identification which is both necessary and sufficient is as follows:

- A necessary and sufficient condition in order to provide the identification of an equation in a model of M equations in M endogeneous variables is to find a nonzero determinant of order $(M-1)(M-1)$ from the coefficients of the variables excluded (endogenous and predetermined) from that equation but included in the other equations of the model.

3.3. Methods of Estimation

It is useful to think of estimators of the equation (3.25) as falling into two main classes. The first principal class of estimators take account of the fact that the equation to be estimated is part of a simultaneous system but uses prior restrictions only on the equation to be estimated, ignoring what may be known about the rest of the system. Such estimators obtained by single equation methods are called limited information estimators:

- Indirect Least squares (ILS)
- Two stage least squares (TSLS)
- Ridge type estimators
- Stein type estimators

The second princial class of estimators uses the fact that the equation to be estimated is part of a simultaneous system; it also uses all the prior restrictions on that system in estimating any equation. The members of full information estimators are*:*

- Full information maximum likelihood estimator
- Three stages least squares (3SLS)

3.3.1. Limited Information Estimators

The first equation of the system (3.25) can be rewritten as

$$\begin{aligned} y_1 &= Y_1\gamma_1 + X_1\beta_1 + u_1 \\ &= Z_1\delta_1 + u_1 \end{aligned} \qquad (3.32.)$$

for

$$Z_1 = [Y_1 \quad X_1]$$

$$\delta_1 = \begin{bmatrix}\gamma_1 \\ \beta_1\end{bmatrix}$$

where y_1 is a $T \times 1$ vector having the left hand jointly dependent variables, Z_1 is a $T \times p_1$ matrix where $p_1 = m_1 + N_1$ whose first m_1 columns contains Y_1 which is our $T \times m_1$ matrix of m_1 right hand jointly dependent variables and the last N_1 columns of Z_1 contains X_1 which is a $T \times N_1$ matrix of included exogenous variables. Also, γ_1 and β_1 are $m_1 \times 1$ and $N_1 \times 1$ coefficient vectors respectively and δ_1 is a $p_1 \times 1$ vector.

From the partitioned form of the reduced form equations (3.26), y_1 and Y_1 can be written as follows:

$$\begin{aligned} y_1 &= [X_1 \quad X_1^*]\begin{bmatrix}\pi_{11} \\ \pi_{21}\end{bmatrix} + v_1 \\ &= X_1\pi_{11} + X_1^*\pi_{21} + v_1 \end{aligned} \tag{3.33.}$$

$$\begin{aligned} Y_1 &= [X_1 \quad X_1^*]\begin{bmatrix}\Pi_{11} \\ \Pi_{21}\end{bmatrix} + V_1 \\ &= X_1\Pi_{11} + X_1^*\Pi_{21} + V_1 \end{aligned} \tag{3.34.}$$

where X_1^* is a $T \times N_1^*$, $N_1^* = N - N_1$, matrix of excluded predetermined variables; π_{11}, π_{21} are $N_1 \times 1$ and $N_1^* \times 1$ vectors, Π_{11}, Π_{21} are $N_1 \times m_1$ and $N_1^* \times m_1$ matrices and v_1, V_1 are $T \times 1$ vector and a $T \times m_1$ matrix, respectively.

$$\begin{bmatrix}\pi_{11} \\ \pi_{21}\end{bmatrix} = \begin{bmatrix}\Pi_{11} & I \\ \Pi_{21} & 0\end{bmatrix}\begin{bmatrix}\gamma_1 \\ \beta_1\end{bmatrix}$$

is obtained from the restrictions (3.29) and (3.30). Let the matrices π_1, J_1 and Π_1 which are order of $N \times 1$, $N \times p_1$ and $p_1 \times 1$, respectively, written as

$$\pi_1 = \begin{bmatrix}\pi_{11} \\ \pi_{21}\end{bmatrix}$$

$$J_1 = \begin{bmatrix} \Pi_1 & I \\ & 0 \end{bmatrix}$$

and

$$\Pi_1 = \begin{bmatrix} \Pi_{11} \\ \Pi_{21} \end{bmatrix}.$$

Π_1, I and 0 in J_1 are $N \times m_1$, $N_1 \times N_1$ and $N_1^* \times N_1$ matrices, respectively. The relationship between the structural parameters of the first equation (3.25) and the reduced form parameters can be written as

$$\pi_1 = J_1 \delta_1. \qquad (3.35.)$$

3.3.1.1. Indirect Least Squares Estimator

The method of ILS obtains the estimates of the structural coefficients from the OLS estimates of the reduced form coefficients. Firstly, the reduced form equations, in which dependent variable is the only endogenous variable and a function of only predetermined variables, are obtained from the structural coefficients. Then, OLS method is applied to the reduced form coefficients individually since the explanatory variables predetermined and uncorrelated with the disturbances. If an equation is just identified due to the one to one correspondence between the structural and reduced form coefficients unique estimates of the structural coefficients are obtained from the estimated reduced form coefficients.

The ILS estimator of γ_1 and β_1 is obtained by using the relation (3.35), $\pi_1 = J_1 \delta_1$. Firstly, π_1 and J_1 are substituted by their OLS estimates from (3.33)-(3.34). From the equation (3.33)

$$\begin{aligned} y_1 &= [X_1 \quad X_1^*] \begin{bmatrix} \pi_{11} \\ \pi_{21} \end{bmatrix} + v_1 \\ &= X\pi_1 + v_1, \end{aligned}$$

the OLS estimate of π_1 is obtained as follows:

$$\hat{\pi}_1 = (X'X)^{-1}X'y_1. \tag{3.36.}$$

Since the equation (3.34)

$$\begin{aligned} Y_1 &= [X_1 \quad X_1^*]\begin{bmatrix}\Pi_{11}\\ \Pi_{21}\end{bmatrix} + V_1 \\ &= X\Pi_1 + V_1 \end{aligned} \tag{3.37.}$$

is solved and $\hat{\Pi}_1$ is yielded as

$$\hat{\Pi}_1 = (X'X)^{-1}X'Y_1. \tag{3.38.}$$

With the help of (3.38), $\hat{J}_1$ is as follows:

$$\hat{J}_1 = \begin{bmatrix}\hat{\Pi}_1 & I\\ & 0\end{bmatrix}. \tag{3.39.}$$

From (3.36) and (3.39), the equation (3.35) becomes

$$\hat{\pi}_1 = \hat{J}_1\delta_1. \tag{3.40.}$$

Secondly, δ_1 is solved from (3.40) to obtain the ILS estimator as

$$\delta_1^* = \hat{J}_1^{-1}\hat{\pi}_1$$

where

$$\gamma_1^* = \hat{\Pi}_{21}^{-1}\hat{\pi}_{21}$$
$$\beta_1^* = \hat{\pi}_{11} - \hat{\Pi}_{11}\hat{\gamma}_1.$$

3.3.1.2. Two Stage Least Squares Estimator

An alternative limited information estimators is the TSLS estimator. The advantages and resilience of the TSLS estimator have been pointed out by numerous analysts; Intriligator (1978), Koutsoyiannis (1977), Namboodiri, Carter and Blalock (1975) and it is the most versatile structural equation estimating techniques. It is consistent, less sensitive than most methods of equation misspecification and performes well in terms of bias and mean square error. As a single equation technique TSLS is easy to apply and is frequently preferred over complete system techniques because it is less sensitive to both specification error and multicollinearity than full information maximum likelihood and three stages least squares methods.

The TSLS procedure is essentially an instrument variable technique, where in the first stage the explanatory endogeneous variables are replaced by their instrumental variable. These are the OLS estimates obtained by using the exogeneous variables as predictors. At the second stage, the OLS estimation is used to obtain the regression coefficients.

For the equation (3.32), the TSLS is another estimator for δ_1 in over-identified case. It can be obtained from (3.40), by premultiplying with X and applying the least squares to solve the equation with

$$S^2 = \left(X\hat{\pi}_1 - X\hat{J}_1\delta_1\right)'\left(X\hat{\pi}_1 - X\hat{J}_1\delta_1\right)$$
$$= \hat{\pi}_1'X'X\hat{\pi}_1 - \hat{\pi}_1'X'X\hat{J}_1\delta_1 - \delta_1'\hat{J}_1'X'X\hat{\pi}_1 + \delta_1'\hat{J}_1'X'X\hat{J}_1\delta_1.$$

By derivating with respect to δ_1 and equating it to zero, it is resulted in

$$-2\hat{J}_1'X'X\hat{\pi}_1 + 2\hat{J}_1'X'X\hat{J}_1\delta_1 = 0$$

and $\hat{\delta}_1$ can be obtained as:

$$\hat{\delta}_1 = \left(\hat{J}_1'X'X\hat{J}_1\right)^{-1}\hat{J}_1'X'X\hat{\pi}_1.$$

Substituting

$$\hat{J}_1 = \begin{bmatrix} \hat{\Pi}_1 & I \\ 0 & \end{bmatrix}$$

and

$$\begin{aligned} X_1 &= [X_1 \quad X_1^*] \begin{bmatrix} I \\ 0 \end{bmatrix} \\ &= X \begin{bmatrix} I \\ 0 \end{bmatrix}, \end{aligned}$$

$\hat{\delta}_1$ is as follows:

$$\begin{aligned} \hat{\delta}_1 &= \left[\begin{bmatrix} \hat{\Pi}_1' \\ I \quad 0 \end{bmatrix} X'X \begin{bmatrix} \hat{\Pi}_1 & I \\ 0 & \end{bmatrix} \right]^{-1} \begin{bmatrix} \hat{\Pi}_1' \\ I \quad 0 \end{bmatrix} X'X\hat{\pi}_1 \\ &= \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\Pi}_1 & \hat{\Pi}_1'X'X \begin{bmatrix} I \\ 0 \end{bmatrix} \\ [I \quad 0]X'X\hat{\Pi}_1 & [I \quad 0]X'X \begin{bmatrix} I \\ 0 \end{bmatrix} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\pi}_1 \\ [I \quad 0]X'X\hat{\pi}_1 \end{bmatrix} \\ &= \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\Pi}_1 & \hat{\Pi}_1'X'X_1 \\ X_1'X\hat{\Pi}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\pi}_1 \\ X_1'X\hat{\pi}_1 \end{bmatrix}. \end{aligned}$$

This is Theil's TSLS estimator and its two stages are not apparent. There is an alternative to show this estimator. The structural equation (3.25) can be written as

$$\begin{aligned} y_1 &= Y_1\gamma_1 + X_1\beta_1 + u_1 \\ &= Y_1\gamma_1 + X_1\beta_1 + v_1 - V_1\gamma_1 \end{aligned}$$

where $v_1 = V_1\gamma_1 + u_1$. Then,

$$\begin{aligned} y_1 &= \gamma_1(Y_1 - V_1) + X_1\beta_1 + v_1 \\ &= [Y_1 - V_1 \quad X_1] \begin{bmatrix} \gamma_1 \\ \beta_1 \end{bmatrix} + v_1 \end{aligned}$$

$$= [X\Pi_1 \quad X_1]\begin{bmatrix}\gamma_1\\ \beta_1\end{bmatrix} + v_1$$

where $Y_1 = X\Pi_1 + V_1$. Since

$$XJ_1 = X\begin{bmatrix}\Pi_1 & I\\ & 0\end{bmatrix}$$
$$= [X\Pi_1 \quad X_1]$$

and

$$\bar{Z}_1 = E(Z_1)$$
$$= [E(Y_1) \quad E(X_1)]$$
$$= [X\Pi_1 \quad X_1], \qquad (3.41.)$$

y_1 can be finally written as

$$y_1 = XJ_1\delta_1 + v_1$$
$$= \bar{Z}_1\delta_1 + v_1. \qquad (3.42.)$$

Then, the TSLS estimator is

$$\delta_1^{LS} = (\bar{Z}_1'\bar{Z}_1)^{-1}\bar{Z}_1'y_1. \qquad (3.43.)$$

Π_1 and thereby $\bar{Z}_1$ are not known in practise. By using $\hat{\Pi}_1$ (3.38) at the first stage, $\bar{Z}_1$ is estimated as

$$\hat{\bar{Z}}_1 = [X\hat{\Pi}_1 \quad X_1]. \qquad (3.44.)$$

Therefore, (3.43) can be writtten as

$$\begin{aligned}\hat{\delta}_1 &= \left(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1\right)^{-1}\hat{\bar{Z}}_1'y_1 \\ &= \left(\hat{J}_1'X'X\hat{J}_1\right)^{-1}\hat{J}_1'X'y_1 \\ &= \left(\hat{J}_1'X'X\hat{J}_1\right)^{-1}\hat{J}_1'X'X\hat{\pi}_1.\end{aligned} \tag{3.45.}$$

For

$$\begin{aligned}\hat{\bar{Z}}_1 &= [X\hat{\Pi}_1 \quad X_1] \\ &= [X(X'X)^{-1}Y_1 \quad X_1] \\ &= X(X'X)^{-1}X'Z_1,\end{aligned}$$

$\hat{\delta}_1$ can be written as

$$\begin{aligned}\hat{\delta}_1 &= (Z_1'X(X'X)^{-1}X'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}X'y_1 \\ &= (Z_1'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}X'y_1.\end{aligned}$$

The TSLS method can be clarified with the consideration of the following model with income and money supply functions, respectively:

$$Y_{1t} = \beta_{10} + \beta_{11}Y_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \tag{3.46.}$$

$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + u_{2t}. \tag{3.47.}$$

Y_1, Y_2, X_1 and X_2 are income, stock of money, investment expenditure and government expenditure, respectively on goods and services. X_1 and X_2 are exogenous variables. The model specification must be changed for the income function since it is underidentified. The money supply equation is overidentified and can not be estimated by the ILS method since the estimate of β_{21} is not unique. Because of the possible relation between Y_1 and u_2, the OLS method gives inconsistent estimates when it is applied. The OLS method can be applied to the

money supply function by the means of instrumental variable for Y_1, that is uncorrelated with u_2. Two stages of the process is as follows:

Stage1: To get rid of the correlation with u_2, Y_1 is regressed on all the predetermined variables in the system as

$$Y_{1t} = \hat{\Pi}_0 + \hat{\Pi}_1 X_{1t} + \hat{\Pi}_2 X_{2t} + \hat{u}_t. \qquad (3.48.)$$

$\hat{u}_t$ are the usual OLS residuals and $\hat{Y}_{1t}$ is an estimate of the mean value of Y conditional upon the fixed X's:

$$\hat{Y}_{1t} = \hat{\Pi}_0 + \hat{\Pi}_1 X_{1t} + \hat{\Pi}_2 X_{2t}.$$

A reduced form regression equation (3.48) can be shown as

$$Y_{1t} = \hat{Y}_{1t} + \hat{u}_t$$

and includes two parts of which are a linear combination of the nonstochastic X variables, $\hat{Y}_{1t}$, and a random component, $\hat{u}_t$.

Stage 2: The money supply equation can be written as

$$\begin{aligned} Y_{2t} &= \beta_{20} + \beta_{21}(\hat{Y}_{1t} + \hat{u}_t) + u_{2t} \\ &= \beta_{20} + \beta_{21}\hat{Y}_{1t} + u_t^* \end{aligned}$$

where $u_t^* = u_{2t} + \beta_{21}\hat{u}_t$. In large samples $\hat{Y}_{1t}$ is uncorrelated with u_t^* asymptotically, therefore the OLS method can be applied to this equation resulting with consistent estimates of the parameters of the money supply function.

Although the TSLS estimator has several desirable properties and thus is a preferred method of equation estimation, it is neverthless extremely sensitive to multicollinearity. One of the most popular procedure for coping with the

multicollinearity problem is to use one of a family of biased, variance reduced estimation methods known as ridge regression. In the presence of multicollinearity ridge regression usually generates estimates with smaller variance than the OLS methods, thus in theory two stages ridge regression should generate estimates with smaller variance than the TSLS.

3.3.1.3. Ridge Type Estimators

A generalized ridge estimator (GRE) of δ_1 in the equation (3.42) where $E(v_1) = 0$ and $E(v_1 v_1') = w_{11}$, can be written as

$$\delta_1^{GRE} = (\bar{Z}_1'\bar{Z}_1 + HKH')^{-1}\bar{Z}_1'y_1 = H\Delta H'\delta_1^{LS} \tag{3.49.}$$

where H is the matrix of the eigenvectors of $\bar{Z}_1'\bar{Z}_1$ such that $H'H = I$ and $\Delta = \mathrm{diag}(q_j)$ a diagonal matrix of $q_j = \gamma_j(\gamma_j + k_j)^{-1}$, where the γ_j's with $j = 1, \dots, p$ are the eigenvalues of $\bar{Z}_1'\bar{Z}_1$, and the k_j's are the diagonal elements of the matrix of ridge constants K. An ordinary RE is a special case of the GRE, where $K = kI$:

$$\delta_1^{RE} = (\bar{Z}_1'\bar{Z}_1 + kI)^{-1}\bar{Z}_1'y_1. \tag{3.50.}$$

For given $\bar{Z}_1$, the ridge estimators are operational. However, Π_1 and thereby $\bar{Z}_1$ are not known in practise. The estimators (3.49) and (3.50) becomes operational with the use of $\hat{\Pi}_1$ (3.38) and $\hat{\bar{Z}}_1$ (3.44). Then the operational RE is

$$\begin{aligned}\hat{\delta}_1^{RE} &= \left(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1 + kI\right)^{-1}\hat{\bar{Z}}_1'y_1 \\ &= \left(\hat{J}_1'X'X\hat{J}_1 + kI\right)^{-1}\hat{J}_1'X'y_1 \\ &= (Z_1'X(X'X)^{-1}X'Z_1 + kI)^{-1}Z_1'X(X'X)^{-1}X'y_1 \end{aligned} \tag{3.51.}$$

and the operational GRE is

$$\hat{\delta}_1^{GRE} = (Z_1'X(X'X)^{-1}X'Z_1 + HKH')^{-1}Z_1'X(X'X)^{-1}X'y_1. \quad (3.52.)$$

Since the usual OLS estimator of Π_1 is used at the first stage, these estimators are called as second stage ridge estimators. When k or $K = 0$ as a special case, the ridge estimators turn to the TSLS estimator (3.43).

In the presence of multicollinearity, the data matrix $X'X$ is ill conditioned, then the value of $\bar{Z}_1'\bar{Z}_1 = J_1'X'XJ_1$ will decrease. Namely, the δ_1^{LS} estimator does not give reliable results when the exogenous variables are collinear. Besides this, the first stage OLS estimator is also affected because of the ill conditioning of $X'X$. Ridge type estimators for Π_1 can be used instead of the OLS estimator $\hat{\Pi}_1$. The GRE of Π_1 can be written as

$$\tilde{\Pi}_1 = (X'X + GKG')^{-1}X'y_1 \quad (3.53.)$$

where G is the matrix of eigenvectors of $X'X$ and K is the matrix of ridge constants. By using $\tilde{\Pi}_1$ (3.53), $\bar{Z}_1$ is estimated as

$$\tilde{\bar{Z}}_1 = [X\tilde{\Pi}_1 \quad X]. \quad (3.54.)$$

With the help of $\tilde{\bar{Z}}_1$ (3.54), a new estimator of δ_1 instead of the TSLS estimator is obtained

$$\tilde{\delta}_1 = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1\right)^{-1}\tilde{\bar{Z}}_1'y_1 \quad (3.55.)$$

which can be called first stage generalized ridge and second stage least squares estimator. If $\tilde{\bar{Z}}_1$ (3.54) is used in δ_1^{GRE} given in the equation (3.49),

$$\tilde{\delta}_1^{GRE} = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1 + HKH'\right)^{-1}\tilde{\bar{Z}}_1'y_1 \quad (3.56.)$$

is called two stage generalized ridge estimator. Similarly, the RE of Π_1 can be written as

$$\tilde{\Pi}_1 = (X'X + kI)^{-1}X'y_1 \qquad (3.57.)$$

where $k > 0$ is the ridge constant. $\bar{Z}_1$ is estimated using $\tilde{\Pi}_1$ in the equation (3.57) as

$$\tilde{\bar{Z}}_1 = [X\tilde{\Pi}_1 \quad X]. \qquad (3.58.)$$

First stage ridge and second stage least squares estimator is obtained by means of $\tilde{\bar{Z}}_1$ in the equation (3.58) as

$$\tilde{\delta}_1(R) = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1\right)^{-1}\tilde{\bar{Z}}_1'y_1. \qquad (3.59.)$$

δ_1^{RE} given in the equation (3.50) is found in practise by using $\tilde{\bar{Z}}_1$ in the equation (3.58):

$$\tilde{\delta}_1^{RE} = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1 + kI\right)^{-1}\tilde{\bar{Z}}_1'y_1 \qquad (3.60.)$$

and called two stage ridge estimator.

3.3.1.4. Stein Type Estimators

To consider the Bayesian interpretation, assume $y_1 \sim N(\bar{Z}_1\delta_1, \omega_{11}I)$ and the prior distribution of $\delta_1 \sim N(\bar{\delta}_1, \tau^2 A_1)$ for given Π_1 and $\bar{\delta}_1$. With given A_1, ω_{11}, τ^2 and the data, the mean of the posterior density of δ_1, namely the Bayesian estimator, is

$$\delta_1^B = \bar{\delta}_1 + \left(I - \left(A_1^{-1} + \frac{\tau^2}{\omega_{11}}\bar{Z}_1'\bar{Z}_1\right)^{-1}A_1^{-1}\right)\left(\hat{\delta}_1 - \bar{\delta}_1\right). \qquad (3.61.)$$

By considering $\delta_1 \sim N(0, \tau^2(\bar{Z}_1'\bar{Z}_1)^{-1})$ and substituting $A_1 = \bar{Z}_1'\bar{Z}_1$ and $\bar{\delta}_1 = 0$ in the equation (3.61), a special case of Bayesian estimator, a Stein-type estimator is as follows

$$\delta_1^{ST} = \left(1 - \frac{\omega_{11}}{\tau^2 + \omega_{11}}\right)\hat{\delta}_1. \tag{3.62}$$

An operational form of Stein-type estimator is as follows

$$\hat{\delta}_1^{ST} = \left(1 - \frac{g_1 \hat{v}_1' \hat{v}_1}{\hat{\pi}_1' X'X \hat{\pi}_1}\right)\hat{\delta}_1 \tag{3.63.}$$

where $\hat{\pi}_1 = (X'X)^{-1}X'y_1$, $\hat{v}_1 = y_1 - X\hat{\pi}_1$, g_1 is an arbitrary scalar and $\hat{\delta}_1$ is the TSLS estimator.

3.3.1.5. Instrumental Variable Estimators

For the model (3.32), Y_1 is correlated with u_1 because of the equations (3.37) and (3.31) and that is why the OLS estimator

$$\hat{\delta}_1 = (Z_1'Z_1)^{-1}Z_1'y_1$$

is inconsistent. Further, the distinction between Y_1 and X_1 in the Z_1 and identifiability restrictions namely all information regarding the variables which are not included are disregarded by the OLS method. There are many estimation techniques to dissipate the problems that Y_1 and u_1 are not independent and disregarding of the identifiability restrictions. One of the method is generalized instrumental variable (GIV) estimator. In this procedure, $Z_1 = [Y_1 \quad X_1]$ is estimated as $\hat{Z}_1 = [\hat{Y}_1 \quad X_1]$ where $\hat{Y}_1$ and $\hat{Z}_1$ are called instrumental variables matrix. Namely, Y_1 is substituted

with variables that have high correlation with Y_1 and no correlation with u_1 in the limit so that

$$plim_{T\to\infty}\frac{\hat{Z}_1 u_1}{T} = 0. \tag{3.64.}$$

δ_1 is chosen in the given sample data as,

$$\frac{\hat{Z}_1'\hat{u}_1}{T} = 0$$

or

$$\hat{Z}_1'\left(y_1 - Z_1\hat{\delta}_1\right) = 0. \tag{3.65.}$$

(3.65) can be written as

$$\hat{Z}_1' y_1 = \hat{Z}_1' Z_1 \hat{\delta}_1 \tag{3.66.}$$

which is called as normal equations and thus the solution for $\hat{\delta}_1$ is

$$\hat{\delta}_1^{GIV} = \left(\hat{Z}_1' Z_1\right)^{-1}\hat{Z}_1' y_1. \tag{3.67.}$$

To write the instrumental variables in a general class, $\hat{Y}_1$ can be written as SY_1 where S is the matrix chosen arbitrarily, therefore $\hat{Z}_1$ can be writtten as

$$\begin{aligned}\hat{Z}_1 &= [SY_1 \quad X_1] \\ &= SZ_1 + (I - S)Z_1 D\end{aligned} \tag{3.68.}$$

where $D = \begin{bmatrix}0 & 0\\ 0 & I\end{bmatrix}$ so that $Z_1 D = [0 \quad X_1]$. Then, diverse options for S is considered to provide the uncorrelation between SZ_1 and u_1. For $S = X(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$ the GIV

estimator reduces to the TSLS estimator as a special case. For $S = X(X'X + HKH')^{-1}X'$, the GIV estimator in the equation (3.67) can be written as

$$\hat{\delta}_1^{GIV} = [Z_1'\mathrm{S_N}Z_1 + D'Z_1'(I - \mathrm{S_N})Z_1]^{-1}[Z_1'\mathrm{S_N} + D'Z_1'(I - \mathrm{S_N})]y_1 \qquad (3.69.)$$

which is first stage ridge and second stage least squares estimator defined before.

3.3.2. Full Information Estimators

The compact model from the M structural equations

$$\begin{aligned} \mathrm{y_1} &= \mathrm{Z_1}\delta_1 + \mathrm{u_1} \\ &\vdots \\ \mathrm{y_M} &= \mathrm{Z_M}\delta_\mathrm{M} + \mathrm{u_M} \end{aligned} \qquad (3.70.)$$

can be written as

$$y = Z\delta + u \qquad (3.71.)$$

where y is an MT × 1 vector; for $Z_i = [Y_i \quad X_i]$ and $\mathrm{p} = \sum_{\mathrm{i}=1}^{\mathrm{M}}(\mathrm{m_i} + \mathrm{N_i})$.

$$Z = \begin{bmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_M \end{bmatrix}$$

is an MT × p matrix; δ is a $p \times 1$ vector and u is an $MT \times 1$ vector with $E(uu') = \Sigma \otimes I_T$ where $\otimes$ is the Kronecker product.

The matrix of the instrumental variables can be written similar to the equation (3.68) as

$$\hat{Z} = \begin{bmatrix} \hat{\mathrm{Z}}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{\mathrm{Z}}_\mathrm{M} \end{bmatrix}$$

for

$$\begin{aligned} \hat{Z}_i &= [\hat{Y}_i \quad \hat{X}_i] \\ &= SZ_i + (I_T - S)Z_iD \end{aligned}$$

where $Z_iD = [0 \quad X_i]$. $\hat{Z}$ can be written in the compact form

$$\hat{Z} = (I_M \otimes S)Z + [I_M \otimes (I_T - S)]Z \otimes D. \tag{3.72.}$$

Since the $\hat{Z}$ (3.72) neglects the joint dependence of u and the dispersion matrix Σ, it is not convenient for the model (3.70). Therefore, an $MT \times p$ weight matrix of a weighted instruments is defined as

$$\begin{aligned} W &= (\Sigma^{-1} \otimes I_T)\hat{Z} \\ &= \begin{bmatrix} \mathrm{W}_1 & & 0 \\ & \ddots & \\ 0 & & \mathrm{W}_\mathrm{M} \end{bmatrix}. \end{aligned} \tag{3.73.}$$

The normal equations given in the equation (3.66) can be generalized to

$$W'Z\delta_{\mathrm{FGIV}} = W'y \tag{3.74.}$$

to obtain the solution for full information GIV (FGIV) estimator. W in the equation (3.74) is replaced with (3.73):

$$\left(\hat{Z}'(\Sigma^{-1} \otimes I_T)Z\right)\delta_{\mathrm{FGIV}} = \hat{Z}'(\Sigma^{-1} \otimes I_T)y.$$

3SLS estimator can be obtain as a special case of the solutions of the normal equation (3.74) by choosing $S = X(X'X)^{-1}X'$ and $\Sigma = \hat{\Sigma}$ where $\hat{\sigma}_{ij} = \left(y_i - Z_i\hat{\delta}_{\mathrm{i}}\right)'\left(y_j - Z_j\hat{\delta}_{\mathrm{j}}\right)/T$ is the (i,j) th element of $\hat{\Sigma}$ and $\hat{\delta}_{\mathrm{i}}$ is the TSLS estimator of i^{th} equation. Then, W in the equation (3.73) is

$$\begin{aligned} W &= \left(\hat{\Sigma}^{-1}\otimes I_T\right)(\mathrm{I_M}\otimes \mathrm{X(X'X)^{-1}X'})\mathrm{Z} \\ &= \left(\hat{\Sigma}^{-1}\otimes \mathrm{X(X'X)^{-1}X'}\right)\mathrm{Z}. \end{aligned} \tag{3.75.}$$

Therefore, with W from the equation (3.75), the FGIV estimator turns to 3SLS estimator as the solution to the normal equation (3.74):

$$\left[\mathrm{Z'}\left(\hat{\Sigma}^{-1}\otimes \mathrm{X(X'X)^{-1}X'}\right)\mathrm{Z}\right]\delta_{\mathrm{3SLS}} = \mathrm{Z'}\left(\hat{\Sigma}^{-1}\otimes \mathrm{X(X'X)^{-1}X'}\right)\mathrm{y}.$$

Then, $\hat{\delta}_{\mathrm{3SLS}}$, where the name three comes from the estimation of $\hat{\Sigma}$ by the second stage residuals, is as follows:

$$\hat{\delta}_{\mathrm{3SLS}} = \left[Z'\left(\hat{\Sigma}^{-1}\otimes \mathrm{X(X'X)^{-1}X'}\right)Z\right]^{-1}\mathrm{Z'}\left(\hat{\Sigma}^{-1}\otimes \mathrm{X(X'X)^{-1}X'}\right)\mathrm{y}.$$

4. TWO STAGE LIU ESTIMATOR

To overcome the problem of multicollinearity, one accepted estimation procedure often suggested is biased estimation. Since the LE is preferable to the RE and the Stein type estimator, we introduce the LE at the first, second and both stages. The ordinary LE of δ_1 for the equation (3.42) can be written as

$$\begin{aligned}\delta_1^{LE} &= (\bar{Z}_1'\bar{Z}_1 + I)^{-1}\,(\bar{Z}_1'y_1 + d\delta_1^{LS}) \\ &= (\bar{Z}_1'\bar{Z}_1 + I)^{-1}(\bar{Z}_1'\bar{Z}_1 + dI)\;\delta_1^{LS},\end{aligned} \tag{4.1.}$$

where $0 < d < 1$ and δ_1^{LS} is the TSLS estimator as defined in the equation (3.43). For given Π_1 and $\bar{Z}_1$, the LE will be operational. Since Π_1 and $\bar{Z}_1$ are not known in practise, we can use $\widehat{\Pi}_1$ (3.38) and $\hat{\bar{Z}}_1$ (3.44), hence the LE in the equation (4.1) will be operational as follows:

$$\hat{\delta}_1^{LE} = \left(\hat{\bar{Z}}_1'\hat{\bar{Z}}_1 + I\right)^{-1}\left(\hat{\bar{Z}}_1'y_1 + d\hat{\delta}_1\right), \tag{4.2.}$$

where $\hat{\delta}_1$ is the TSLS estimator in the equation (3.45). This estimator (4.2) can be called second stage LE because we use $\widehat{\Pi}_1$ at the first stage. For a special case where the biasing parameter d is equal to zero, the second stage LE reduces to the TSLS estimator. We deal the LE at the second stage but with regard to the ill conditioned $X'X$ matrix, the OLS estimator at the first stage, $\widehat{\Pi}_1$, is also affected. Under this circumstances, the LE of the reduced form parameter Π_1

$$\widetilde{\Pi}_1 = (X'X + I)^{-1}\,(\,X'Y_1\, + d\widehat{\Pi}_1), \tag{4.3.}$$

where $\widehat{\Pi}_1$ is as defined in the equation (3.38), can be used instead of the OLS estimator. With the help of $\widetilde{\Pi}_1$ (4.3) and thereby

$$\tilde{\bar{Z}}_1 = [X\widetilde{\Pi}_1 \quad X_1], \tag{4.4.}$$

a first stage Liu and second stage least squares estimator of δ_1 for the equation (3.42) is defined to be

$$\delta_1(L) = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1\right)^{-1}\tilde{\bar{Z}}_1' y_1. \tag{4.5.}$$

We can also apply the LE at both stages. We obtain $\tilde{\bar{Z}}_1$ (4.4) through $\tilde{\Pi}_1$ (4.3), then substitute in the eqution (4.1) to define two stage LE as follows:

$$\tilde{\delta}_1^{LE} = \left(\tilde{\bar{Z}}_1'\tilde{\bar{Z}}_1 + I\right)^{-1}\left(\tilde{\bar{Z}}_1' y_1 + d\tilde{\delta}_1\right). \tag{4.6.}$$

4.1. Comparisons of the Two Stage Ridge Estimator and the Two Stage Liu Estimator

The performance of an estimator can be evaluated by a mean square error criterion and we can manage the problem of choosing between the TSLS estimator and biased estimators.

If an estimator $\bar{\delta}_2$ outperforms $\bar{\delta}_1$ with respect to the MSE criterion then $mse\left(\bar{\delta}_1\right) \geq mse\left(\bar{\delta}_2\right)$. Since using the mse is convenient for application, we use the estimated mse in the numerical example. The unknown parameters are replaced with their proper estimates in practise.

The model in the equation (3.42) can be turned into a canonical form by orthogonal transformation. Let $Z = \bar{Z}_1 P$, $\alpha_1 = P'\delta_1$ and P is an orthogonal matrix such that $Z'Z = P'\bar{Z}_1'\bar{Z}_1 P = \Lambda_1 = diag\left(\lambda_{11}, \dots, \lambda_{1p_1}\right)$ where λ_{1i} are the eigenvalues of $\bar{Z}_1'\bar{Z}_1$. The canonical form of the model (3.42) is

$$y_1 = Z\alpha_1 + v_1. \tag{4.7.}$$

The TSLS estimator can be written likewise in the equation (3.43):

$$\alpha_1^{LS} = \Lambda_1^{-1} Z' y_1. \tag{4.8.}$$

For the unknown Λ_1 and Z, Λ_1 is substituted by

$$\widehat{\Lambda}_1 = P'\hat{\hat{Z}}_1'\hat{\hat{Z}}_1 P$$

and Z is replaced with

$$\hat{Z} = \hat{\hat{Z}}_1 P$$

with the help of $\hat{\hat{Z}}_1$ (3.44). Then, the TSLS estimator is found as follows:

$$\hat{\alpha}_1 = \widehat{\Lambda}_1^{-1}\hat{Z}'y_1. \tag{4.9.}$$

The RE and the LE at the second stage can be written as follows:

$$\begin{aligned}\alpha_1^{RE} &= (\Lambda_1 + \mathrm{kI})^{-1}\mathrm{Z}'y_1 \\ &= \mathrm{A}_1 y_1,\end{aligned} \tag{4.10.}$$

$$\begin{aligned}\alpha_1^{LE} &= (\Lambda_1 + \mathrm{I})^{-1}(\mathrm{Z}'y_1 + d\alpha_1^{LS}) \\ &= (\Lambda_1 + \mathrm{I})^{-1}(\Lambda_1 + dI)\Lambda_1^{-1}Z'y_1 \\ &= \mathrm{A}_2 y_1.\end{aligned} \tag{4.11.}$$

In practise, α_1^{RE} and α_1^{LE} can be given as:

$$\hat{\alpha}_1^{RE} = \left(\widehat{\Lambda}_1 + \mathrm{kI}\right)^{-1}\hat{\mathrm{Z}}'y_1, \tag{4.12.}$$

$$\hat{\alpha}_1^{LE} = \left(\widehat{\Lambda}_1 + \mathrm{I}\right)^{-1}\left(\hat{\mathrm{Z}}'y_1 + d\hat{\alpha}_1\right). \tag{4.13.}$$

If $\bar{\alpha}_1$ and $\bar{\delta}_1$ are two estimators of α_1 and δ_1, respectively, there is a relation between these models such as $\bar{\alpha}_1 = P'\bar{\delta}_1$ and $MSE(\bar{\alpha}_1) = P'MSE\left(\bar{\delta}_1\right)$. The $MSEs$ and the $mses$ of the mentioned estimators are

$$MSE(\alpha_1^{LS}) = \sigma^2\Lambda_1^{-1},$$

$$MSE(\alpha_1^{LE}) = \sigma^2(\Lambda_1 + \mathrm{I})^{-1}(\Lambda_1 + dI)\Lambda_1^{-1}(\Lambda_1 + dI)(\Lambda_1 + \mathrm{I})^{-1} + (1-\mathrm{d})^2(\Lambda_1 + \mathrm{I})^{-1}\alpha_1\alpha_1'(\Lambda_1 + \mathrm{I})^{-1},$$

$$MSE(\alpha_1^{RE}) = \sigma^2(\Lambda_1 + \mathrm{kI})^{-1}\Lambda_1(\Lambda_1 + \mathrm{kI})^{-1} + \mathrm{k}^2(\Lambda_1 + \mathrm{kI})^{-1}\alpha_1\alpha_1'(\Lambda_1 + \mathrm{kI})^{-1}$$

$$mse(\alpha_1^{LS}) = Tr[\sigma^2\Lambda_1^{-1}], \tag{4.14.}$$

$$mse(\alpha_1^{LE}) = \mathrm{Tr}[\sigma^2(\Lambda_1 + \mathrm{I})^{-1}(\Lambda_1 + dI)\Lambda_1^{-1}(\Lambda_1 + dI)(\Lambda_1 + \mathrm{I})^{-1}] + (1-\mathrm{d})^2\alpha_1'(\Lambda_1 + \mathrm{I})^{-1}(\Lambda_1 + \mathrm{I})^{-1}\alpha_1, \tag{4.15.}$$

$$mse(\alpha_1^{RE}) = Tr[\sigma^2(\Lambda_1 + \mathrm{kI})^{-1}\Lambda_1(\Lambda_1 + \mathrm{kI})^{-1}] + \mathrm{k}^2\alpha_1'(\Lambda_1 + \mathrm{kI})^{-1}(\Lambda_1 + \mathrm{kI})^{-1}\alpha_1. \tag{4.16.}$$

We substitute σ^2, α_1 and Λ_1 with

$$\hat{\sigma}_1^2 = \frac{(y_1 - \hat{\mathrm{Z}}_1'\hat{\alpha}_1)'(y_1 - \hat{\mathrm{Z}}_1'\hat{\alpha}_1)}{T - p_1}, \tag{4.17.}$$

$\hat{\alpha}_1$ and $\widehat{\Lambda}_1$, respectively, in the equation (4.14), (4.15) and (4.16) to obtain the estimated $mses$ of the estimators α_1^{LS}, α_1^{LE} and α_1^{RE}. To compute the estimated mse, we should also find an appropriate biasing parameter which is given in the subsequent section.

To compare the $MSEs'$ of the α_1^{LE} and α_1^{RE}, we give the following theorems similar to the comparisons in the linear regression model by Sakallıoğlu et al. (2001). We use a result from Farebrother (1976) as a Lemma to prove the theorems.

Lemma. (Farebrother, 1976): Let A be a positive definite (p.d.) matrix, c be a nonzero vector and θ be a positive scalar. Then $\theta A - cc'$ is p.d. if and only if $\mathrm{c}'\mathrm{A}^{-1}\mathrm{c} < \theta$.

For the first theorem let d be fixed for the MSE.

Theorem.4.1.:

a. If $Bias(\alpha_1^{LE})'(A_1A_1' - A_2A_2')^{-1}Bias(\alpha_1^{LE}) < \sigma^2$ then $MSE(\alpha_1^{RE}) - MSE(\alpha_1^{LE})$ is p.d. for $0 < k < k_{1i}^*$

b. If $Bias(\alpha_1^{RE})'(A_2A_2' - A_1A_1')^{-1}Bias(\alpha_1^{RE}) < \sigma^2$ then $MSE(\alpha_1^{LE}) - MSE(\alpha_1^{RE})$ is p.d. for $0 < k_{2i}^* < k$

where $k_i = \frac{\lambda_i(1-d)}{\lambda_i+d}$, $k_{1i}^* = min\{k_i\}$ and $k_{2i}^* = max\{k_i\}$ for $i = 1, \dots, p_1$.

Proof. Let $\Delta_1 = MSE(\alpha_1^{RE}) - MSE(\alpha_1^{LE})$.

$$\begin{aligned}\Delta_1 &= \sigma^2(A_1A_1' - A_2A_2') + Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})' \\ &= \sigma^2[(\Lambda_1 + \mathrm{kI})^{-1}\Lambda_1(\Lambda_1 + \mathrm{kI})^{-1} - (\Lambda_1 + \mathrm{I})^{-1}(\Lambda_1 + dI)\Lambda_1^{-1}(\Lambda_1 + dI)(\Lambda_1 + \mathrm{I})^{-1}] \\ &+Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})' \\ &= \sigma^2\left[diag\left\{\frac{\lambda_{1i}}{(\lambda_{1i}+k)^2} - \frac{(\lambda_{1i}+d)^2}{\lambda_{1i}(\lambda_{1i}+1)}\right\}_{i=1}^{p_1}\right] \\ &+Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})'\end{aligned}$$

$A_1A_1' - A_2A_2'$ will be p.d. if and only if $\lambda_{1i}^2(\lambda_{1i}+1)^2 - (\lambda_{1i}+k)^2(\lambda_{1i}+d)^2 > 0$ or $\lambda_{1i}(\lambda_{1i}+1) - (\lambda_{1i}+k)(\lambda_{1i}+d) > 0$ for every $i = 1, \dots, p_1$. This inequality requires that $A_1A_1' - A_2A_2'$ is p.d. for $0 < k < k_{1i}$. Since $Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})'$ is p.d., Δ_1 is p.d. if $\sigma^2(A_1A_1' - A_2A_2') - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})'$ is p.d. According to the Lemma, the sufficient condition is $Bias(\alpha_1^{LE})'(A_1A_1' - A_2A_2')^{-1}Bias(\alpha_1^{LE}) < \sigma^2$.

Similarly, let $\Delta_2 = MSE(\alpha_1^{LE}) - MSE(\alpha_1^{RE})$:

$$\Delta_2 = \sigma^2(A_2A_2' - A_1A_1') + Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})'.$$

$A_2A_2' - A_1A_1'$ will be p.d. if and only if $(\lambda_{1i}+k)(\lambda_{1i}+d) - \lambda_{1i}(\lambda_{1i}+1) > 0$ for every $i = 1, \dots, p_1$. This inequality requires that $A_2A_2' - A_1A_1'$ is p.d. for

$0 < k < k_{1i}$. From the Lemma, the sufficient condition is $Bias(\alpha_1^{RE})'(A_2A_2'-A_1A_1')^{-1}Bias(\alpha_1^{RE}) < \sigma^2$.

For the second theorem let k be fixed for the MSE.

Theorem.4.2.:

a. If $Bias(\alpha_1^{RE})'(A_2A_2' - A_1A_1')^{-1}Bias(\alpha_1^{RE}) < \sigma^2$ then $MSE(\alpha_1^{LE}) - MSE(\alpha_1^{RE})$ is p.d. for $0 < d_{1i}^* < d < 1$

b. If $Bias(\alpha_1^{LE})'(A_1A_1' - A_2A_2')^{-1}Bias(\alpha_1^{LE}) < \sigma^2$ then $MSE(\alpha_1^{RE}) - MSE(\alpha_1^{LE})$ is p.d. for $0 < d < d_{2i}^* < 1$

where $d_i = \frac{\lambda_i(1-k)}{\lambda_i+k}$, $d_{1i}^* = max\{d_i\}$ and $d_{2i}^* = min\{d_i\}$ for $i = 1, \dots, p_1$.

Proof. Let $\Delta_3 = MSE(\alpha_1^{LE}) - MSE(\alpha_1^{RE})$.

$$\begin{aligned}\Delta_3 &= \sigma^2(A_2A_2'-A_1A_1') + Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})' - Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' \\ &= \sigma^2[(\Lambda_1+\mathrm{I})^{-1}(\Lambda_1+dI)\Lambda_1^{-1}(\Lambda_1+dI)(\Lambda_1+\mathrm{I})^{-1} - (\Lambda_1+\mathrm{kI})^{-1}\Lambda_1(\Lambda_1+\mathrm{kI})^{-1}] \\ &+Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})' - Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' \\ &= \sigma^2\left[diag\left\{\frac{(\lambda_{1i}+d)^2}{\lambda_{1i}(\lambda_{1i}+1)} - \frac{\lambda_{1i}}{(\lambda_{1i}+k)^2}\right\}_{i=1}^{p_1}\right] \\ &+Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})' - Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})'\end{aligned}$$

$A_2A_2' - A_1A_1'$ will be p.d. if and only if $(\lambda_{1i}+k)^2(\lambda_{1i}+d)^2 - \lambda_{1i}^2(\lambda_{1i}+1)^2 > 0$ for every $i = 1, \dots, p_1$. Two roots of $(\lambda_{1i}+k)^2(\lambda_{1i}+d)^2 - \lambda_{1i}^2(\lambda_{1i}+1)^2 = 0$ are $d_{1i} = 1 - \frac{k(\lambda_i+1)}{\lambda_i+k}$ and $d_{2i} = 1 - \frac{-2\lambda_i^2-\lambda_i k-\lambda_i)}{\lambda_i+k}$, where $0 < d_{1i} < 1$ for $0 < k < 1$.

Thus the inequality requires that $A_2A_2' - A_1A_1'$ is p.d. for $0 < d_{1i} < d < 1$. Since $Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})'$ is p.d., Δ_3 is p.d. if $\sigma^2(A_2A_2' - A_1A_1') - Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})'$ is p.d. According to the Lemma, the sufficient condition is $Bias(\alpha_1^{RE})'(A_2A_2' - A_1A_1')^{-1}Bias(\alpha_1^{RE}) < \sigma^2$.

In a similar way, let $\Delta_4 = MSE(\alpha_1^{RE}) - MSE(\alpha_1^{LE})$.

$$\Delta_4 = \sigma^2(A_1A_1' - A_2A_2') + Bias(\alpha_1^{RE})Bias(\alpha_1^{RE})' - Bias(\alpha_1^{LE})Bias(\alpha_1^{LE})'$$

$A_1A_1' - A_2A_2'$ will be p.d. if and only if $\lambda_{1i}^2(\lambda_{1i}+1)^2 - (\lambda_{1i}+k)^2(\lambda_{1i}+d)^2 > 0$ for every $i = 1, \dots, p_1$. Thus the inequality requires that $A_1A_1' - A_2A_2'$ is p.d. for $0 < d < d_{1i} < 1$. From the Lemma, the sufficient condition is $Bias(\alpha_1^{LE})'(A_1A_1' - A_2A_2')^{-1}Bias(\alpha_1^{LE}) < \sigma^2$.

4.2. Selection of the Biasing Parameter

4.2.1. Biasing Parameter for the Liu Estimator at the Second Stage

Theorem 4.3.: If $d > d_1^{con}$ then $MSE(\alpha_1^{LS}) - MSE(\alpha_1^{LE}) > 0$ where $d_1^{con} = 1 - \frac{2\sigma^2}{\alpha_1'\alpha_1}$.

Proof. The difference between the $MSEs$ of α_1 and α_1^{LE} is as follows:

$$\begin{aligned}\Delta_5 &= MSE(\alpha_1^{LS}) - MSE(\alpha_1^{LE}) \\ &= \sigma^2[\Lambda_1^{-1} - (\Lambda_1+I)^{-1}(\Lambda_1+dI)\Lambda_1^{-1}(\Lambda_1+dI)(\Lambda_1+I)^{-1}] \\ &-(1-d)^2(\Lambda_1+I)^{-1}\alpha_1\alpha_1'(\Lambda_1+I)^{-1} \\ &= (1-d)(\Lambda_1+I)^{-1}A_3(\Lambda_1+I)^{-1}\end{aligned}$$

where $A_3 = 2\sigma^2 I + (1+d)\sigma^2\Lambda_1^{-1} - (1-d)\alpha_1\alpha_1'$. Sakallıoğlu et al. (1996) shows that $\Delta_5 > 0$ if and only if $\alpha_1'\left[I + \frac{1+d}{2}\Lambda_1^{-1}\right]^{-1}\alpha_1 < \frac{2\sigma^2}{1-d}$. To obtain $\Delta_5 > 0$, we check whether $A_3 > 0$. Since $(1+d)\sigma^2\Lambda_1^{-1} > 0$, we look for the condition to provide $2\sigma^2 I - (1-d)\alpha_1\alpha_1' > 0$. Then, the sufficient condition for $2\sigma^2 I - (1-d)\alpha_1\alpha_1' > 0$ is $d > 1 - \frac{2\sigma^2}{\alpha_1'\alpha_1}$ as it is stated in the theorem.

d_1^{con} is estimated by substituting α_1 and σ^2 with $\hat{\alpha}_1$ and $\hat{\sigma}_1^2$ (4.17), respectively.

Liu (1993) proposed the estimate of d and its generalized form, at which the mse of the LE is minimized. By this way, we can write that $mse(\alpha_1^{LE})$ is minimized at

$$\hat{d}_1^{min} = 1 - \hat{\sigma}_1^2 \frac{\sum_{i=1}^{p_1} \frac{1}{\lambda_{1i}(\lambda_{1i}+1)}}{\sum_{i=1}^{p_1} \frac{\hat{\alpha}_{1i}^2}{(\lambda_{1i}+1)^2}} \tag{4.18.}$$

for the canonical model (4.7). The equation (4.18) can be generalized to

$$\hat{d}_1^{min}(h) = 1 - h\hat{\sigma}_1^2 \frac{\sum_{i=1}^{p_1} \frac{1}{\lambda_{1i}(\lambda_{1i}+1)}}{\sum_{i=1}^{p_1} \frac{\hat{\alpha}_{1i}^2}{(\lambda_{1i}+1)^2}} \tag{4.19.}$$

where $h > 0$ and and $\hat{\sigma}_1^2$ is as in the equation (4.17).

4.2.2. Biasing Parameter for the Liu Estimator at the First Stage

For the canonical form of the model in the equation (3.37), let $D = XQ$, $\eta_1 = Q'\Pi_1$ and Q is an orthogonal matrix such that $D'D = \mathrm{Q'X'XQ} = \Lambda_2 = \mathrm{diag}(\lambda_{21}, \dots, \lambda_{2p_2})$ where λ_{2i} are the eigenvalues of $X'X$. The canonical form of the model is

$$Y_1 = D\eta_1 + V_1. \tag{4.20.}$$

The OLS estimator of η_1 is

$$\hat{\eta}_1 = \Lambda_2^{-1} D' Y_1. \tag{4.21.}$$

Also, the LE of η_1 is

$$\hat{\eta}_1^{LE} = (\Lambda_2 + \mathrm{I})^{-1}(D'Y_1 + d\hat{\eta}_1),$$

where $\hat{\eta}_1$ is as in the equation (4.21).

Since there is a relation between these models such as $\bar{\eta}_1 = Q'\overline{\Pi}_1$ and $MSE(\bar{\eta}_1) = Q'MSE(\overline{\Pi}_1)Q$ for any two estimators $\bar{\eta}_1$ and $\overline{\Pi}_1$, the $MSEs$ and the

$mses$ are as follows:

$$MSE(\hat{\eta}_1) = \sigma^2\Lambda_2^{-1},$$
$$MSE(\hat{\eta}_1^{\mathrm{LE}}) = \sigma^2(\Lambda_2 + \mathrm{I})^{-1}(\Lambda_2 + \mathrm{dI})\Lambda_2^{-1}(\Lambda_2 + \mathrm{dI})(\Lambda_2 + \mathrm{I})^{-1} + (1-\mathrm{d})^2(\Lambda_2 + \mathrm{I})^{-1}\eta_1\eta_1'(\Lambda_2 + \mathrm{I})^{-1},$$

$$mse(\hat{\eta}_1) = Tr[\sigma^2\Lambda_2^{-1}],$$

$$mse(\hat{\eta}_1^{\mathrm{LE}}) = Tr[\sigma^2(\Lambda_2 + \mathrm{I})^{-1}(\Lambda_2 + \mathrm{dI})\Lambda_2^{-1}(\Lambda_2 + \mathrm{dI})(\Lambda_2 + \mathrm{I})^{-1}] + (1-\mathrm{d})^2\eta_1'(\Lambda_2 + \mathrm{I})^{-1}(\Lambda_2 + \mathrm{I})^{-1}\eta_1.$$

In a similar way to the previous subsection, we give a sufficient condition with a theorem.

Theorem 4.4.: If $d > d_2^{con}$ then $MSE(\hat{\eta}_1) - MSE(\hat{\eta}_1^{\mathrm{LE}}) > 0$ where $d_2^{con} = 1 - \frac{2\sigma^2}{\eta_1'\eta_1}$.

Proof.

$$\begin{aligned}\Delta_6 &= MSE(\hat{\eta}_1) - MSE(\hat{\eta}_1^{\mathrm{LE}}) \\ &= \sigma^2[\Lambda_2 - (\Lambda_2 + \mathrm{I})^{-1}(\Lambda_2 + \mathrm{dI})\Lambda_2^{-1}(\Lambda_2 + \mathrm{dI})(\Lambda_2 + \mathrm{I})^{-1}] \\ &-(1-\mathrm{d})^2(\Lambda_2 + \mathrm{I})^{-1}\eta_1\eta_1'(\Lambda_2 + \mathrm{I})^{-1} \\ &= (1-\mathrm{d})(\Lambda_2 + \mathrm{I})^{-1}\mathrm{A}_4(\Lambda_2 + \mathrm{I})^{-1}\end{aligned}$$

where $\mathrm{A}_4 = 2\sigma^2\mathrm{I} + (1+\mathrm{d})\sigma^2\Lambda_2^{-1} - (1-\mathrm{d})\eta_1\eta_1'$. If $\mathrm{A}_4 > 0$ then $\Delta_6 > 0$. We check whether $2\sigma^2\mathrm{I} - (1-\mathrm{d})\eta_1\eta_1' > 0$ since $(1+\mathrm{d})\sigma^2\Lambda_2^{-1} > 0$. Then, the sufficient condition to provide $2\sigma^2\mathrm{I} - (1-\mathrm{d})\eta_1\eta_1' > 0$ is $d > 1 - \frac{2\sigma^2}{\eta_1'\eta_1}$.

η_1 and σ^2 is replaced with $\hat{\eta}_1$ and

$$\hat{\sigma}_2^2 = \frac{(Y_1 - D\hat{\eta}_1)'(Y_1 - D\hat{\eta}_1)}{T - p_2}, \qquad (4.22.)$$

respectively to obtain $\hat{d}_2^{con}$.

We can state that $mse(\hat{\eta}_1^{\mathrm{LE}})$ is minimized at

$$\hat{d}_2^{min} = 1 - \hat{\sigma}_2^2 \frac{\sum_{i=1}^{p_2} \frac{1}{\lambda_{2i}(\lambda_{2i}+1)}}{\sum_{i=1}^{p_2} \frac{\hat{\eta}_{1i}^2}{(\lambda_{2i}+1)^2}}$$

and the generalized type of the estimate is

$$\hat{d}_2^{min}(h) = 1 - h\hat{\sigma}_2^2 \frac{\sum_{i=1}^{p_2} \frac{1}{\lambda_{2i}(\lambda_{2i}+1)}}{\sum_{i=1}^{p_2} \frac{\hat{\eta}_{1i}^2}{(\lambda_{2i}+1)^2}},$$

where $h > 0$ and $\hat{\sigma}_2^2$ (4.22) for the canonical model (4.20).

4.2.3. Biasing Parameter of the Two Stage Liu Estimator

We confront with the choice of d at both stages for the two stage LE. For the first stage, the selection of the parameter is the same with the LE at the first stage. Namely, $\hat{d}_3^{con1} = \hat{d}_2^{con}$ and $\hat{d}_3^{min1} = \hat{d}_2^{min}$. Then, we can determine the selection of the parameter for the second stage. We consider the canonical form of the model in the equation (3.42):

$$y_1 = C\theta_1 + v_1. \tag{4.23.}$$

where $C = \bar{Z}_1 R$, $\theta_1 = R'\delta_1$ and R is an orthogonal matrix such that $C'C = R'\bar{\mathrm{Z}}_1'\bar{\mathrm{Z}}_1 R = \Lambda_3 = diag(\lambda_{31}, \dots, \lambda_{3p_1})$, λ_{3i} are the eigenvalues of $\bar{\mathrm{Z}}_1'\bar{\mathrm{Z}}_1$. The TSLS estimator is

$$\theta_1^{LS} = \Lambda_3^{-1} C' y_1.$$

The two stage LE is

$$\theta_1^{LE} = (\Lambda_3 + \mathrm{I})^{-1}(\mathrm{C}'y_1 + d\theta_1^{LS}).$$

Besides this, when Λ_3 and C are not known these are substituted with

$$\widetilde{\Lambda}_3 = R'\tilde{\tilde{Z}}_1'\tilde{\tilde{Z}}_1 R,$$

where $\tilde{\tilde{Z}}_1$ is as in the equation (4.4) and

$$\tilde{C} = \tilde{\tilde{Z}}_1 R.$$

Then, the TSLS estimator and two stage LE can be written as follows:

$$\tilde{\theta}_1 = \widetilde{\Lambda}_3^{-1}\tilde{C}'y_1.$$

Similarly, θ_1^{LE} can be written in practise,

$$\tilde{\theta}_1^{\mathrm{LE}} = \left(\widetilde{\Lambda}_3 + \mathrm{I}\right)^{-1}\left(\tilde{\mathrm{Z}}'\mathrm{y}_1 + \mathrm{d}\tilde{\theta}_1\right).$$

There is a relation between these models such as $\bar{\theta}_1 = R'\bar{\delta}_1$ and $MSE(\bar{\theta}_1) = R'MSE\left(\bar{\delta}_1\right)R$ for $\bar{\theta}_1$ and $\bar{\delta}_1$, any two estimators of θ_1 and δ_1, respectively. The $MSEs$ of the mentioned estimators are

$$MSE(\theta_1^{LS}) = \sigma^2\Lambda_3^{-1},$$

$$MSE(\theta_1^{LE}) = \sigma^2(\Lambda_3 + \mathrm{I})^{-1}(\Lambda_3 + \mathrm{dI})\Lambda_3^{-1}(\Lambda_3 + \mathrm{dI})(\Lambda_3 + \mathrm{I})^{-1} + (1-\mathrm{d})^2(\Lambda_3 + \mathrm{I})^{-1}\theta_1\theta_1'(\Lambda_3 + \mathrm{I})^{-1},$$

$$mse(\theta_1^{LS}) = Tr[\sigma^2\Lambda_3^{-1}], \qquad (4.24.)$$

$$mse(\theta_1^{LE}) = Tr[\sigma^2(\Lambda_3 + \mathrm{I})^{-1}(\Lambda_3 + \mathrm{dI})\Lambda_3^{-1}(\Lambda_3 + \mathrm{dI})(\Lambda_3 + \mathrm{I})^{-1}]$$

$$+(1-\mathrm{d})^2\theta_1'(\Lambda_3+\mathrm{I})^{-1}(\Lambda_3+\mathrm{I})^{-1}\theta_1. \quad (4.25.)$$

The estimated $mses$ of the θ_1^{LS} and θ_1^{LE} are obtained by replacing σ^2, θ_1 and Λ_3 with

$$\tilde{\sigma}^2=\frac{(y_1-\tilde{C}\tilde{\theta}_1)'(y_1-\tilde{C}\tilde{\theta}_1)}{T-p_1}, \quad (4.26.)$$

$\tilde{\theta}_1$ and $\tilde{\Lambda}_3$, respectively in the equations (4.24) and (4.25). A sufficient condition is given like the previous subsection:

Theorem 4.5.: If $d>d_3^{con2}$, $MSE(\theta_1^{LS})-MSE(\theta_1^{LE})>0$ where $d_3^{con2}=1-\frac{2\sigma^2}{\theta_1'\theta_1}$.

Proof.

$$\begin{aligned}\Delta_7 &= MSE(\theta_1^{LS})-MSE(\theta_1^{LE})\\ &=\sigma^2[\Lambda_3-(\Lambda_3+\mathrm{I})^{-1}(\Lambda_3+d\mathrm{I})\Lambda_3^{-1}(\Lambda_3+d\mathrm{I})(\Lambda_3+\mathrm{I})^{-1}]\\ &-(1-\mathrm{d})^2(\Lambda_3+\mathrm{I})^{-1}\theta_1\theta_1'(\Lambda_3+\mathrm{I})^{-1}\\ &=(1-\mathrm{d})(\Lambda_3+\mathrm{I})^{-1}A_5(\Lambda_3+\mathrm{I})^{-1}\end{aligned}$$

where $A_5=2\sigma^2\mathrm{I}+(1+\mathrm{d})\sigma^2\Lambda_3^{-1}-(1-\mathrm{d})\theta_1\theta_1'$. To provide $\Delta_7>0$, we should check whether $A_5>0$. Since $(1+\mathrm{d})\sigma^2\Lambda_3^{-1}>0$, the sufficient condition to obtain $2\sigma^2\mathrm{I}-(1-\mathrm{d})\theta_1\theta_1'>0$ is $d>1-\frac{2\sigma^2}{\theta_1'\theta_1}$.

To estimate d_3^{con2}, $\tilde{\theta}_1$ and $\tilde{\sigma}^2$ (4.26) are put into place of θ_1 and σ^2, respectively.

As the estimate of d, which minimizes the mse of the LE, is introduced in Liu (1993), for the canonical model (4.23), $mse\left(\tilde{\theta}_1^{LE}\right)$ is minimized at

$$\hat{d}_3^{min2}=1-\tilde{\sigma}^2\frac{\sum_{i=1}^{p_1}\frac{1}{\lambda_{3i}(\lambda_{3i}+1)}}{\sum_{i=1}^{p_1}\frac{\tilde{\theta}_{1i}^2}{(\lambda_{3i}+1)^2}}$$

and the generalized type is

$$\hat{d}_3^{min2}(h) = 1 - h\tilde{\sigma}^2 \frac{\sum_{i=1}^{p_1} \frac{1}{\lambda_{3i}(\lambda_{3i}+1)}}{\sum_{i=1}^{p_1} \frac{\tilde{\theta}_{1i}^2}{(\lambda_{3i}+1)^2}},$$

where $h > 0$ and $\tilde{\sigma}^2$ is as in the equation (4.26).

4.3. Numerical Example

We take the model of U.S. economy constructed by Klein (1950). Klein Model I, which is based on annual data in the period between 1920-1941. It comprises three behavioral equations which are consumption, investment and labor, respectively as follows:

$$C_t = \delta_0 + \delta_1 P_t + \delta_2 P_{t-1} + \delta_3 (W_t + W_t') + u_t,$$
$$I_t = \delta_0' + \delta_1' P_t + \delta_2' P_{t-1} + \delta_3' K_{t-1} + u_t',$$
$$W_t = \delta_0'' + \delta_1'' X_t + \delta_2'' X_{t-1} + \delta_3'' (t - 1931) + u_t''.$$

Also, three identities such as

$$X_t = C_t + I_t + G_t,$$
$$P_t = X_t - W_t - T_t,$$
$$K_t = K_{t-1} + I_t$$

are included.

C is aggregate consumption, P is total profits, W is total wages paid by private industry, W' is the government wage bill, I is net investment, K is the stock of capital goods, X is the total production of private industry, G_t is government nonwage expenditure and T is business taxes in a year.

The six endogenous variables are C, P, W, I, K and X. The eight predetermined variables are G, T, t, W', K_{-1}, P_{-1}, X_{-1} and intercept.

The consumption function can be written in the form of the equation (4.7) as follows:

$$\begin{bmatrix} C_1 \\ \vdots \\ C_{21} \end{bmatrix} = \begin{bmatrix} P_1 & W_1 + W_1' & \vdots & 1 & P_0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{21} & W_{21} + W_{21}' & \vdots & 1 & P_{20} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_3 \\ \dots \\ \delta_0 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_{21} \end{bmatrix}$$

where the subscript 21 refers to the last of the 21 annual observations of the period 1921 to 1941. Z_1 consists of two submatrices; Y_1 contains the values of the explanatory jointly dependent variables and X_1 contains the values of the explanatory predetermined variables

The investment function can be shown similarly:

$$\begin{bmatrix} I_1 \\ \vdots \\ I_{21} \end{bmatrix} = \begin{bmatrix} P_1 & \vdots & 1 & P_0 & K_0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{21} & \vdots & 1 & P_{20} & K_{20} \end{bmatrix} \begin{bmatrix} \delta_1' \\ \cdots \\ \delta_0' \\ \delta_2' \\ \delta_3' \end{bmatrix} + \begin{bmatrix} u_1' \\ \vdots \\ u_{21}' \end{bmatrix}$$

and finally the labor function is

$$\begin{bmatrix} W_1 \\ \vdots \\ W_{21} \end{bmatrix} = \begin{bmatrix} X_1 & \vdots & 1 & X_0 & -10 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{21} & \vdots & 1 & X_{20} & 10 \end{bmatrix} \begin{bmatrix} \delta_1'' \\ \cdots \\ \delta_0'' \\ \delta_2'' \\ \delta_3'' \end{bmatrix} + \begin{bmatrix} u_1'' \\ \vdots \\ u_{21}'' \end{bmatrix}.$$

The numerical results for the estimators are given by using the canonical forms and both of the three functions are examined separately. First of all, we investigate the mse performances of the LE to be compared with the TSLS. The estimates of the biasing parameters are given as it is handled in the subsection 4.2.

For the first function, in the case of the TSLS estimation, $\hat{\sigma}_1^2 = 3.9562$ and $\hat{mse}(\alpha_1^{LS}) = 6.7130$. We investigate which estimator gives better outcomes in the sense of the mse. The condition number of X is 1.95×10^8 so there is a severe multicollinearity. For this reason, using the OLS at the first stage gives unreliable results and the LE is preferred to use at the first stage. For this function we encounter with a multivariate regression model (4.20) while computing the results for the LE at the first stage. So, $\hat{\sigma}_2^2$ is shown as $\hat{\sigma}_2^2(1)$ and $\hat{\sigma}_2^2(2)$ and also similar expressions are used for d_2, $\hat{d}_2^{con}$ and $\hat{d}_2^{min}$. For the LE at the first stage: $\hat{\sigma}_2^2(1) = 4.7656$ and $\hat{\sigma}_2^2(2) = 3.0774$. Thus, the estimate of the biasing parameter for the first stage is chosen as $d_2(1) = 0.9976$ and $d_2(2) = 0.9991$ so as to be greater than $\hat{d}_2^{con}(1) = 0.9956$ and $\hat{d}_2^{con}(2) = 0.9971$, respectively (Theorem 4.4). Then, the condition number of $\tilde{\tilde{Z}}_1$ becomes 8.22×10^4. Consequently, it is concluded that the LE should be applied at the second stage, too. For the LE at the second stage $\tilde{\sigma}^2 = 3.9564$. Since the estimate of the second biasing parameter should be greater than $\hat{d}_3^{con2} = 0.9712$, it can be chosen as $d = 0.9722$ (Theorem 4.5) and eventually $\hat{mse}(\theta_1^{LE}) = 6.5676$ shows that it is less than $\hat{mse}(\alpha_1^{LS})$.

We can also use the estimate of the biasing parameter $\hat{d}_2^{min}(1) = 0.5992$ and $\hat{d}_2^{min}(2) = 0.6511$, at which the mse is minimized, for the first stage. The condition number of the $\tilde{\tilde{Z}}_1$ becomes 8.2702×10^4. Then, for the second stage the estimate of biasing parameter is $\hat{d}_3^{min}(2) = 0.9602$ and $\hat{mse}(\theta_1^{LE}) = 6.6734$.

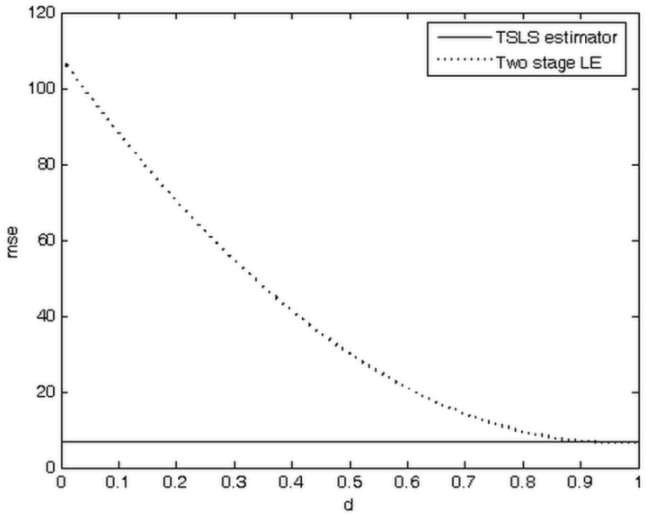


Figure 4.1(a). Estimated mse values of the TSLS estimator and the two stage LE for the first equation.

By means of Figure 4.1(a), the biasing parameter d is plotted against the estimated mse values of the TSLS estimator and the two stage LE for the first function. Firstly, we use the estimate of $d_2(1) = 0.9976$ and $d_2(2) = 0.9991$ for the first stage. Then, the horizontal axis shows the possible biasing parameter estimates of the second stage.

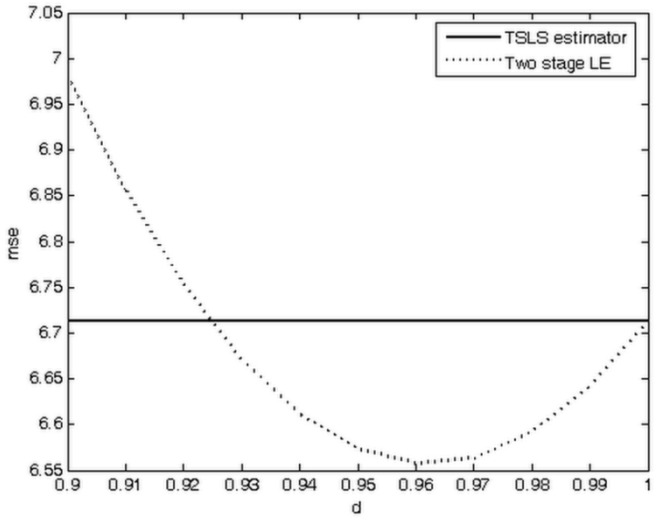


Figure 4.1(b). Estimated mse values of the TSLS estimator and the two stage LE for the first equation.

Figure 4.1(b) only shows some part of the Figure 4.1(a), where the estimated mse values of the two stage LE is smaller than the estimated mse values of the TSLS estimator. Namely, the only difference for the Figure 4.1(b) is the interval of d, which is taken as $0.9 < d < 1$.

The case of the TSLS estimation for the second function results in $\hat{\sigma}_1^2 = 2.4198$ and $\hat{mse}(\alpha_1^{LS}) = 99.6343$. Due to excessive condition number of X, the LE is used at the first stage and the estimate of σ_2^2 is $\hat{\sigma}_2^2 = 4.7653$. $\hat{d}_2^{con} = 0.9962$ and by choosing $d = 0.9982$, namely $d > \hat{d}_2^{con}$ (Theorem 4.4), the condition number of $\tilde{\tilde{Z}}_1$ becomes 3.5290×10^7. Thus, the LE should be also used at the second stage and the estimate of σ^2 is $\tilde{\sigma}^2 = 2.4199$. For chosen $d = 0.9892$, which is greater than $\hat{d}_3^{con2} = 0.9882$ (Theorem 4.5), $\hat{mse}(\theta_1^{LE}) = 97.5836$. It is apparent that the estimated mse diminishes.

In the case of biasing parameter which minimizes mse, i.e. $\hat{d}_2^{min} = 0.5992$, the condition number of $\tilde{\tilde{Z}}_1$ is 33233×10^7. For $\hat{d}_3^{min2} = 0.7534$, $\hat{mse}(\theta_1^{LE}) = 80.1448$.

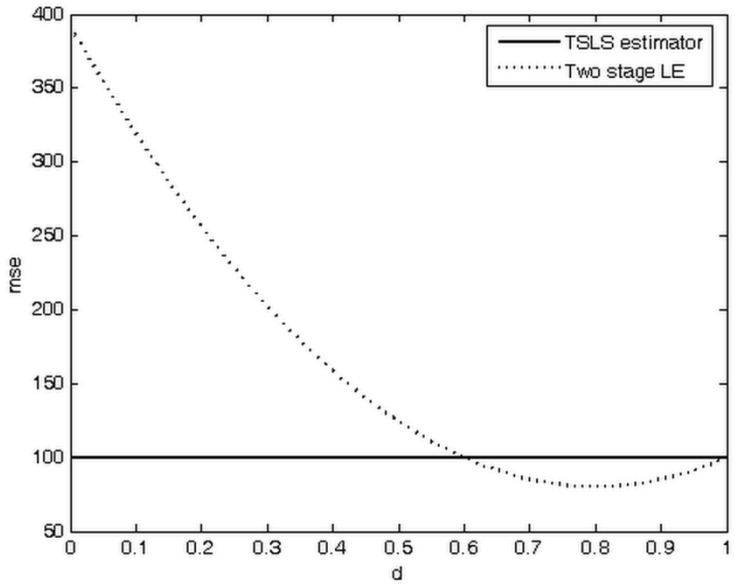


Figure 4.2. Estimated mse values of the TSLS estimator and the two stage LE for the second equation.

The estimate of biasing parameter $d = 0.9982$ ($d > \hat{d}_2^{con}$) is used for the first stage. Afterwards, the estimated mse values are plotted depending on possible biasing parameter estimates of the second stage.

The TSLS estimation outcomes for the third function are: $\hat{\sigma}_1^2 = 2.7035$ and $m\hat{s}e(\alpha_1^{LS}) = 7.4963$. Since there is a severe multicollinearity for X, the LE is used at the first stage. For the LE at the first stage: $\hat{\sigma}_2^2 = 14.2391$. The estimate of biasing parameter for the LE at the first stage is chosen as $d = 0.9987$ so as to be greater than $\hat{d}_2^{con} > 0.9967$ (Theorem 4.4). The condition number of $\tilde{\tilde{Z}}_1$ becomes 4.1445×10^5 that shows the LE should be used at the second stage, too. For the LE at the second stage: $\tilde{\sigma}^2 = 2.7036$. Since the second biasing parameter should be greater than $\hat{d}_3^{con2} = -1.1767$, it can be chosen as $d = 0.90$ (arbitrarily chosen so as to be d > 0) and $m\hat{s}e(\theta_1^{LE}) = 6.4543$. The results exhibit the decrease in the estimated mse.

When the biasing parameter that minimizes mse, i.e. $\hat{d}_2^{min} = 0.6549$, is used, the condition number of $\tilde{\tilde{Z}}_1$ is 4.1246×10^5 so we deduce to use the LE at the second stage, too. For $\hat{d}_3^{min2}(h) = 0.8830$ where $h = 0.01$, $m\hat{s}e(\theta_1^{LE}) = 6.4083$.

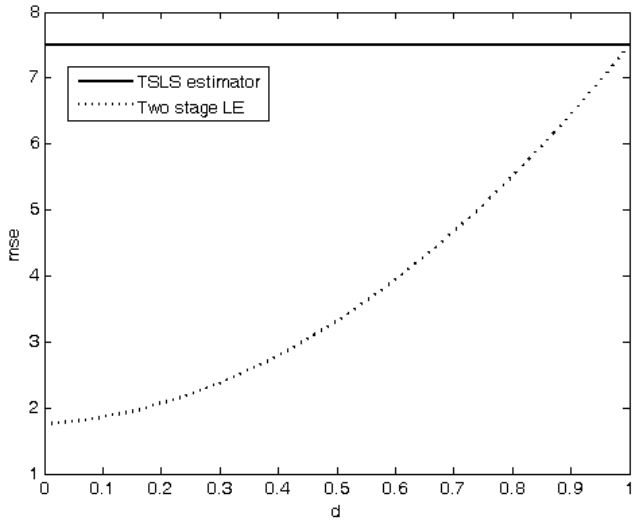


Figure 4.3. Estimated mse values of the TSLS estimator and the two stage LE for the third equation.

The estimated mse values of the TSLS estimator and the two stage LE are plotted via Figure 4.3, for the third function $d = 0.9987$, which is greater than $\hat{d}_2^{con}$, is used for the first stage than the possible biasing parameter estimates of the second stage is given in the horizontal axis.

Since the values chosen for the biasing parameter are very high (namely it is close to 1), the estimated coefficients of the two stage LE are very close to the estimated coefficients of the TSLS estimator. Therefore, we didn't asses the new estimator in the context of estimated coefficients. We only show that the two stage LE outperforms the TSLS estimator in the sense of the mse.

Afterwards, the mse performances of the LE and the RE are examined. The following practical results for the Theorem 4.1 and Theorem 4.2 are obtained. σ_1^2 and α_1 are substituted by $\hat{\sigma}_1^2$ and $\hat{\alpha}_1$ to compute the estimated mse values.

Firstly, the computations are carried out for the first function and $\hat{\sigma}_1^2 = 3.9562$. Let $d = 0.9$ be fixed. The values of k_i are yielded by using Theorem 4.1:

k_i: 0.0404, 0.0992, 0.0998, 0.0999.

$m\hat{s}e(\alpha_1^{LE}) = 6.98$ and for k =0.03 ($0 < k < k_{1i}^* = 0.0404$), $m\hat{s}e(\alpha_1^{RE}) = 10.64$ therefore α_1^{LE} has a smaller estimated mse value than α_1^{RE}. Since the condition in Theorem 4.1(b) is not satisfied, the estimated mse value of α_1^{RE} for $0 < k_{2i}^* = 0.0999 < k$ is not smaller than $m\hat{s}e(\alpha_1^{LE})$.

Let $k = 0.02$ be fixed. From the Theorem 4.2 the values of d_i are:

d_i: 0.9483, 0.9795, 0.97992, 0.97999.

For $d = 0.93$ ($0 < d < d_{2i}^* = 0.9483 < 1$), $m\hat{s}e(\alpha_1^{LE}) = 6.67$. Since $m\hat{s}e(\alpha_1^{RE}) = 10.63$, the estimated mse value of α_1^{LE} is smaller. For the Theorem 4.2(a), the condition is not satisfied so α_1^{RE} does not have smaller estimated mse value than α_1^{LE} for $0 < d_{1i}^* = 0.97999 < d < 1$.

Secondly, numerical outcomes are given for the second function and

$\hat{\sigma}_1^2 = 2.4198$. Let $d = 0.9$ be fixed and $m\hat{s}e(\alpha_1^{LE}) = 85.059$. Using Theorem 4.1 the values of k_i are generated as:

k_i: 0.0026, 0.09995, 0.099996, 0.099999.

For $k = 0.001$ $(0 < k < k_{1i}^* = 0.0026)$, $m\hat{s}e(\alpha_1^{RE}) = 3780.94$. Thus, α_1^{LE} has a smaller estimated mse value than α_1^{RE}. However, the condition for the part(b) of the Theorem 4.1 is not satisfied and the estimated mse value of α_1^{RE} is not smaller than the estimated mse value of α_1^{LE} for $0 < k_{2i}^* = 0.099999 < k$.

Let $k = 0.02$ be fixed. $m\hat{s}e(\alpha_1^{RE}) = 1316.85$. The values of d_i are written from the Theorem 4.2 as follows:

d_i: 0.5376, 0.9797, 0.97998, 0.97999.

For $d = 0.52$ $(0 < d < d_{2i}^* = 0.5376 < 1)$, $m\hat{s}e(\alpha_1^{LE}) = 118.48$. $m\hat{s}e(\alpha_1^{LE})$ is smaller than $m\hat{s}e(\alpha_1^{RE})$ as it is claimed in the Theorem 3.2(b). α_1^{RE} does not have smaller estimated mse value than α_1^{LE} for $0 < d_{1i}^* = 0.97999 < d < 1$ because the condition is not satisfied.

Finally, numerical computations for the third function are obtained and $\hat{\sigma}_1^2 = 2.7035$. Let $d = 0.9$ be fixed. Theorem 4.1 results in the following values of k_i:

k_i: 0.0287, 0.09985, 0.09996, 0.09999.

For $k = 0.01$ $(0 < k < k_{1i}^* = 0.0287$, $m\hat{s}e(\alpha_1^{RE}) = 19.58$. $m\hat{s}e(\alpha_1^{LE}) = 6.450$ is smaller than $m\hat{s}e(\alpha_1^{RE})$ that accords with the assertation of the part (a) of the Theorem 4.1. Nevertheless, the condition for the part(b) of the Theorem 4.1 is not satisfied so as to obtain smaller estimated mse value of α_1^{RE} than α_1^{LE} for $0 < k_{2i}^* = 0.09999 < k$.

Let $k = 0.02$ be fixed and $m\hat{s}e(\alpha_1^{RE}) = 18.57$. The values of d_i are yielded by using Theorem 4.2:

d_i: 0.9286, 0.97989, 0.97997, 0.97999.

For $d = 0.91$ $(0 < d < d_{2i}^* = 0.9286 < 1)$, $m\hat{s}e(\alpha_1^{LE}) = 6.55$ which is smaller than $m\hat{s}e(\alpha_1^{RE})$. But $m\hat{s}e(\alpha_1^{RE})$ can not be obtained as smaller than $m\hat{s}e(\alpha_1^{LE})$ for $0 < d_{1i}^* = 0.97999 < d < 1$ since the condition of the Theorem 4.2(a).

5. INEQUALITY RESTRICTIONS IN SIMULTANEOUS EQUATIONS MODEL

In empirical studies, there is a need to the prior and sample information for parameter estimation to sustain structural consistency in a linear regression model. Liew (1976b) offered ICLS estimation in this sense. In the presence of multicollinearity, inequality restrictions are applied to the ridge regression estimation by Toker et al. (2013). Inequality constrained ridge regression (ICRR) estimator is proposed to deal with this problem.

Inequality restrictions are also required to maintain structural consistency in a system of simultaneous equations. Liew (1976a) presented two stage and three stage least squares estimation with inequality restrictions on parameters. We intend to apply inequality restrictions to ridge regression estimation and propose inequality constrained two stage and three stage ridge regression estimators to handle multicollinearity.

5.1. Inequality Constrained Least Squares Estimator

An inequality constrained model is

$$y = X\beta + \varepsilon, \tag{5.1.}$$

$$A\beta \geq c, \tag{5.2.}$$

where y is an $n \times 1$ vector of observations on the dependent variable; X has a full column rank of $n \times p$ nonstochastic explanatory variables matrix; β is a $p \times 1$ vector of unknown parameters; A is an $m \times p$ fixed matrix; c is a vector of m components. ε is an $n \times 1$ vector of disturbances with expectation $E(\varepsilon) = 0$ and dispersion matrix $Var(\varepsilon) = \sigma^2 I$.

The primal-dual relation of Dorn (1960) and Mangasarian (1962) is as follows

primal:

$$min\ z = f(\beta) \tag{5.3.}$$

subject to

$$A\beta \geq c,$$

β is unrestricted in sign and
dual:

$$max\ Q = f(\beta_0) - \beta_0{}'\nabla f(\beta_0) + c'\lambda \tag{5.4.}$$

subject to

$$A'\lambda = \nabla f(\beta_0),\ \lambda \geq 0,$$

where $f(\beta)$ is a differentiable and convex function of β and $\nabla f(\beta_0)$ is the gradient vector evaluated at β_0 which is a solution to the primal problem and λ is a dual vector.

To introduce the least squares estimator with inequality restrictions on parameters, Liew (1976b) followed Dorn (1960) and Mangasarian (1962) to introduce the primal-dual relation similar to the equations (5.3) and (5.4)
primal:

$$\min_{\beta}\ z = \left(\frac{1}{2}\right)(y - X\beta)'(y - X\beta)$$

subject to

$$A\beta \geq c\ \ (or\ A\beta - v = c),$$

where β is unrestricted in sign and v is a nonnegative m component surplus vector, dual:

$$\max_{\lambda} Q = \left(\frac{1}{2}\right)(\mathrm{y'y} - \beta'\mathrm{X'X}\beta) + \mathrm{c'}\lambda$$

subject to

$$A'\lambda + X'y = (X'X)\beta,\ \lambda \geq 0, \tag{5.5.}$$

where λ is an m component dual vector and β is a solution to the primal problem. Solution can be obtained by reducing the primal dual relation to the fundamental problem of Dantzig-Cottle (1967, 1974) and calculating with Lemke (1962) or Dantzig-Cottle (1967, 1974) algorithms as:

$$v = W\lambda + q \tag{5.6.}$$

or

$$[\,I \quad -W\,]\begin{bmatrix} v \\ \lambda \end{bmatrix} = q$$

subject to

$$v'\lambda = 0,\ v \geq 0 \text{ and } \lambda \geq 0, \tag{5.7.}$$

where

$$W = A(X'X)^{-1}A',$$
$$q = A\hat{\beta} - c,$$
$$\hat{\beta} = (X'X)^{-1}X'y. \tag{5.8.}$$

A sufficient condition of the solution requires the matrix $X'X$ to be p.d. and $\lambda^*,\ v^*$ are nonnegative complementary solutions of the problem. By replacing λ^* in equation (5.5), Liew (1976b) obtained the ICLS estimator, β_{IC}^*, as:

$$\begin{aligned}\beta_{IC}^* &= (X'X)^{-1}X'y + (X'X)^{-1}A'\lambda^* \\ &= \hat{\beta} + (X'X)^{-1}A'\lambda^*. \end{aligned} \tag{5.9.}$$

If the parameters of the model are not restricted, λ^* becomes zero and β_{IC}^* becomes equal to $\hat{\beta}$ (5.8) which is the OLS estimator. If the parameters are constrained by equalities all the elements of λ^* in the complementary solution becomes positive ($\lambda^* \gg 0$) which implies $v^* = 0$ and

$$\lambda^* = -(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - c).$$

Thus, β_{IC}^* reduces to the equality constrained least squares estimator as:

$$\beta_C^* = \hat{\beta} + (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(c - A\hat{\beta}).$$

5.2. Inequality Constrained Ridge Regression Estimator

To overcome deficiency of multicollinearity, RE is developed for the model (5.1) by Hoerl and Kennard (1970a) as the solution to the following problem:

$$(y - X\beta)'(y - X\beta) + k(\beta'\beta - d) \tag{5.10.}$$

where $(y - X\beta)'(y - X\beta)$ is minimized subject to $\beta'\beta = d$, where d is a constant, and k is a Lagrangian multiplier. Toker et al. (2013) apply the idea of inequality restrictions to the ridge regression to cope with multicollinearity. Therefore, this objective function (5.10) is adopted to the primal-dual relation of Dorn (1960) and Mangasarian (1962) as follows:

primal:

$$\min_{\beta} z = \left(\frac{1}{2}\right)[(y - X\beta)'(y - X\beta) + k(\beta'\beta - d)]$$

subject to

$$A\beta \geq c \ \ (or\ A\beta - v = c),$$

where β is unrestricted in sign and v is a nonnegative m component surplus vector and d is a constant,
dual:

$$\max_{\lambda} Q = \left(\frac{1}{2}\right)(y'y - \beta'X'X\beta - k\beta'\beta - kd) + c'\lambda$$

subject to

$$A'\lambda + X'y = (X'X + kI)\beta,\ \lambda \geq 0 \qquad (5.11.)$$

where λ is an m component dual vector and β is a solution to the primal problem.

The primal-dual relation reduce to the fundamental problem as it is in the equations (5.6)-(5.7):

$$v = W\lambda + q$$

subject to

$$v'\lambda = 0,\ v \geq 0 \text{ and } \lambda \geq 0$$

where

$$W = A(X'X + kI)^{-1}A',$$
$$q = A\hat{\beta}(k) - c,$$
$$\hat{\beta}(k) = (X'X + kI)^{-1}X'y. \quad (5.12.)$$

λ^* and v^* are nonnegative complementary solutions of the problem and by replacing λ^* in the equation (5.11), the ICRR estimator is obtained as:

$$\beta_{IC}^*(k) = (X'X + kI)^{-1}X'y + (X'X + kI)^{-1}A'\lambda^*$$
$$= \hat{\beta}(k) + (X'X + kI)^{-1}A'\lambda^*. \quad (5.13.)$$

λ^* becomes zero when the parameters of the model are not restricted and thus $\beta_{IC}^*(k)$ is reduced to $\hat{\beta}(k)$ (5.12) which is the RE of Hoerl and Kennard (1970a).

All the elements of λ^* in the complementary solution becomes positive ($\lambda^* \gg 0$) when the parameters are constrained by equalities and so it implies $v^* = 0$ and

$$\lambda^* = -(A(X'X + kI)^{-1}A')^{-1}(A\hat{\beta}(k) - c). \quad (5.14.)$$

Thus, by the equations (5.13) and (5.14), $\beta_{IC}^*(k)$ in the equation (5.13) becomes equal to equality constrained RE of Groβ (2003) as

$$\beta_C^*(k) = \hat{\beta}(k) + (X'X + kI)^{-1}A'(A(X'X + kI)^{-1}A')^{-1}(c - A\,\hat{\beta}(k)).$$

5.3. Inequality Constrained Two Stage Least Squares Estimator

Liew (1976a) considered the model which has been simplified by deleting all observation subscripts

$$y_j = Y_j\gamma_j + X_j\beta_j + \varepsilon_j,\ j = 1, \ldots, r, \quad (5.15.)$$

where γ_j and β_j are parameter vectors of the j^{th} structural equation; y_j is j^{th} dependent variable; Y_j is jointly dependent variables included in the j^{th} equation; X_j is predetermined variables included in the j^{th} equation. The condensed form of the equation (5.15) is

$$y_j = Z_j\delta_j + \varepsilon_j,\ j = 1,\dots,r, \tag{5.16.}$$

where $Z_j = [Y_j \quad X_j]$ and $\delta_j = \begin{bmatrix}\gamma_j \\ \beta_j\end{bmatrix}$.

$$y = Z\delta + \varepsilon \tag{5.17.}$$

is the whole model which is formed by all r equations where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_r \end{bmatrix},$$

$$Z = \begin{bmatrix} Z_1 & & & 0 \\ & Z_2 & & \\ & & \ddots & \\ 0 & & & Z_r \end{bmatrix},$$

$$\delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_r \end{bmatrix},$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_r \end{bmatrix}.$$

$\bar{X}$ is defined as $\bar{X} = (I \otimes X)$ where $\otimes$ is the Kronecker product and X is a matrix of predetermined variables of the model. By premultiplying $\bar{X}'$ to the equation (5.17), the model becomes

$$\bar{X}'y = \bar{X}'Z\delta + \bar{X}'\varepsilon, \tag{5.18.}$$

where $E(\bar{X}'\varepsilon) = 0$ and $V(\bar{X}'\varepsilon) = \Sigma \otimes n\Sigma x'x$. For a two stage parameter estimation, Σ is cancelled out in equations since Σ is a diagonal matrix and a cancelled form of Σ, or identity matrix I, is used. $\Sigma x'x$ can be estimated consistently by $n^{-1}(X'X)$. Consequently, $V(\bar{X}'\varepsilon) = \mathrm{I} \otimes (X'X)$. To introduce the two stage least squares estimator with inequality restrictions on parameters, the primal-dual relation is given by Liew (1976a) similar to the way of Dorn (1962) and Mangasarian (1962)
primal:

$$\min_{\delta} z = \left(\frac{1}{2}\right)(\bar{X}'y - \bar{X}'Z\delta)'(I \otimes (X'X)^{-1})(\bar{X}'y - \bar{X}'Z\delta)$$

subject to

$$A\delta \geq c \ \ (or\ A\delta - v = c),$$

where A is an l_1 by l_2 matrix, c is an l_1 component vector, v is a nonnegative l_1 component surplus vector and δ is an l_2 component vector which is unrestricted in sign,
dual:

$$\max_{\lambda} Q = \left(\frac{1}{2}\right)((\bar{X}'y)'(I \otimes (X'X)^{-1})\bar{X}'y - \delta'Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z\delta) + c'\lambda$$

subject to

$$A'\lambda + Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'y = Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z\delta,\ \lambda \geq 0\,, \qquad (5.19.)$$

where λ is an l_1 component dual vector and δ is a solution to the primal problem. Solution can be obtained by reducing the primal dual relation to the fundamental problem of Dantzig-Cottle (1967, 1974) and calculating with Lemke (1962) or Dantzig-Cottle (1967, 1974) algorithms as:

$$v = W\lambda + q$$

or

$$[\,I \quad -W]\begin{bmatrix} v \\ \lambda \end{bmatrix} = q$$

subject to

$$v'\lambda = 0,\ v \geq 0 \text{ and } \lambda \geq 0$$

where

$$W = A(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A',$$
$$q = A\hat{\delta} - c,$$
$$\hat{\delta} = (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'y). \qquad (5.20.)$$

λ^* and v^* are the complementary solutions of the fundamental problem. By replacing λ^* in the equation (5.19), Liew (1976a) obtained the two stage least squares estimator when δ is constrained by inequalities, δ_{IC}^*, as follows

$$\delta_{IC}^* = \hat{\delta} + (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A'\lambda^*.$$

If the parameters are unconstrained, λ^* becomes zero and δ_{IC}^* reduces to $\hat{\delta}$ (5.20) which is Jorgenson's two stage least squares estimator vector (Jorgenson, 1962). Besides this, when the parameters are constrained by equalities, all the elements of the λ^* in the complementary solution become positive. When $\lambda^* \gg 0$, it implies $v^* = 0$, thus:

$$\lambda^* = -(A(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A')^{-1}(A\hat{\delta} - c)$$

and the δ_{IC}^{*} turns out to be equality constrained two stage least squares estimator as

$$\delta_{IC}^{*} = \hat{\delta} + (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A' \times (A(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A')^{-1}(c - A\hat{\delta}).$$

A three stage least squares estimation differs from the two stage least squares estimation for the assumption of $V(\bar{X}'\varepsilon)$, namely for $V(\bar{X}'\varepsilon) = (\Sigma \otimes n\Sigma x'x)$, S is put into place of Σ, where $S_{ij} = \frac{1}{n}e_i'e_j$ and e_i is the difference between the actual y_i vector and a two stage predicted vector $\hat{y}_i$. Liew (1976a) introduced the inequality constrained three stage least squares estimator as follows:

$$\delta_{IC}^{**} = \tilde{\delta} + (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A'\lambda^{*},$$

where

$$\tilde{\delta} = (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z)^{-1}(Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'y) \qquad (5.21.)$$

is Jorgenson's three stage least squares estimator vector. If $\lambda^{*} = 0$, namely the model is unconstrained, δ_{IC}^{**} reduces to $\tilde{\delta}$ (5.21). Moreover, if all the elements of λ^{*} are positive, that is the model is equality constrained, δ_{IC}^{**} reduces to equality constrained three stage least squares estimators:

$$\delta_{C}^{**} = \tilde{\delta} + (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z)^{-1} \times A'(A(Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z)^{-1}A')^{-1}(c - A\tilde{\delta}).$$

5.4. Inequality Constrained Two Stage Ridge Regression Estimator

On the purpose of eliminating the deficiencies of multicollinearity, we apply inequality restrictions to the ridge regression and offer an alternative estimator to the ICLS estimator. Consider the model in the equations (5.15-5.18) to introduce two

stage ridge regression estimation with inequality restrictions on parameters. The primal-dual relation is as follows:

primal

$$\min_{\delta} z = \left(\frac{1}{2}\right)\left((\bar{X}'y - \bar{X}'Z\delta)'(I \otimes (X'X)^{-1})(\bar{X}'y - \bar{X}'Z\delta) + k(\delta'\delta - d)\right)$$

subject to

$$A\delta \geq c \ \ (or\ A\delta - v = c)$$

where δ is an l_2 component vector which is unrestricted in sign, d is a constant and A, c and v are an l_1 by l_2 matrix, an l_1 component vector and a nonnegative l_1 component surplus vector, respectively and
dual:

$$\max_{\lambda} Q = \left(\frac{1}{2}\right)((\bar{X}'y)'(I \otimes (X'X)^{-1})\bar{X}'y - \delta'Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z\delta - k\delta'\delta - kd) + c'\lambda$$

subject to

$$A'\lambda + Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'y = (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z + kI)\delta,\ \lambda \geq 0$$

where λ is an l_1 component dual vector and δ is a solution to the primal problem. In a similar manner of finding the two stage least squares estimator with inequality restrictions previously, the two stage ridge regression estimator when δ is constrained by inequalities is

$$\delta_{IC}^{*}(k) = \hat{\delta}(k) + (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}A'\lambda^{*},$$

where

$$\hat{\delta}(k) = (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'y)$$

is the two stage ridge regression estimator vector (see Vinod and Ullah, 1981, p. 276). $\delta_{IC}^{*}(k)$ reduces to $\hat{\delta}(k)$, if the parameters are not restricted. Besides this, when the parameters are constrained by equalities, we propose an equality constrained two stage ridge regression estimator, $\delta_{C}^{*}(k)$, as

$$\begin{aligned}\delta_{C}^{*}(k) \;&= \hat{\delta}(k) + (Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1} \\ &\quad \times A'(A(Z'\bar{X}(I \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}A')^{-1}(c - A\hat{\delta}(k)).\end{aligned}$$

The assumption of $V(X'\varepsilon)$ for the three stage ridge regression estimation is different from the two stage ridge regression estimation, namely Σ is replaced by S for $V(X'\varepsilon) = (\Sigma \otimes n\Sigma x'x)$, where $S_{ij} = \frac{1}{n}e_i'e_j$ and e_i is the difference between the actual y_i vector and a two stage predicted vector $\hat{y}_i$. The inequality constrained three stage ridge regression estimator is as follows:

$$\delta_{IC}^{**}(k) = \tilde{\delta}(k) + (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}A'\lambda^{*},$$

where

$$\tilde{\delta}(k) = (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}(Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'y)$$

is the three stage ridge regression estimator (see Vinod and Ullah 1981, p. 287). If $\lambda^{*} = 0$, namely model is unconstrained, $\delta_{IC}^{**}(k)$ reduces to $\tilde{\delta}(k)$. If all the elements of λ^{*} are positive, namely model is equality unconstrained, we define $\delta_{C}^{**}(k)$ as equality constrained three stage ridge regression estimator:

$$\delta_C^{**}(k) = \tilde{\delta}(k) + (Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}$$
$$\times A'(A(Z'\bar{X}(S^{-1} \otimes (X'X)^{-1})\bar{X}'Z + kI)^{-1}A')^{-1}(c - A\tilde{\delta}(k)).$$

6. CONCLUSIONS

Biased estimation is an approach to estimate the structural coefficients in the presence of multicollinearity. Vinod and Ullah (1981) examined ridge type and Stein type estimators in the estimation of coefficients in a simultaneous equations model. We take into consideration that some improved estimation for the coefficients in a structural equation can be applied at the first and/or second stage of the TSLS. One of the improved estimator is the LE which is named corresponding to the stages where it is implemented. Two different choices of the biasing parameter of the LE are derived. One of the biasing prameter is defined so as to minimize the mse of the LE. The other one is determined to provide the sufficient condition that the LE outperforms the TSLS estimator in the sense of the MSE. Furthermore, the mean square errors of the mentioned estimators are expressed in order to compare their performances. The comparison of the RE and the LE are carried out and the intervals of k and d are given where the LE or the RE outperforms the other. Numerical example based on widely analysed data set Klein (1950) is illustrated. The estimated biasing parameters and the $mses$ are given according to the outcomes of the data analysis. It is demonstrated that the theoretical results are supported by the numerical example. Researchers who are intended to use the LE can infer that the two stage LE performs better than the TSLS estimator.

To maintain the structural consistency in a linear regression model and in a system of simultaneous equations, the use of prior and sample estimation is required. That's the reason why the inequality constrained estimation is proposed. ICLS estimator and two stage and three stage least squares estimators under inequality restrictions are suggested by Liew (1976a) and Liew (1976b.). We extend the results of Liew (1976a, 1976b) and propose ICRR estimator and inequality constrained two stage and three stage ridge regression estimators. The use of these estimators are motivated from the fact that biased estimation is preferable to the OLS estimation under multicollinearity. Many other biased estimators can be drawn on in a similar manner to utilize the inequality restrictions.

REFERENCES

AGUNBIADE, D. A. and IYANIWURA, J. O., 2010. Estimation Under Multicollinearity: A Comparative Approach Using Monte Carlo Methods. Journal of Mathematics and Statistics, 6(2): 183-192.

AGUNBIADE, D. A., 2011. Effect of Multicollinearity and the Sensitivity of the Estimation Methods in Simultaneous Equation Model. Journal of Modern Mathematics and Statistics, 5(1): 9-12.

AKDENİZ, F. and KAÇIRANLAR, S., 1995. On the Almost Unbiased Generalized Liu Estimator and Unbiased Estimation of the Bias and MSE. Communications in Statistics-Theory and Methods, 24(7): 1789–1797.

AMEMIYA, T., 1974. The Nonlinear Two-Stage Least-Stage Estimator. Journal of Econometrics, 2: 105–10.

ANDERSON, T., 2005. Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators. Journal of Econometrics, 127: 1–16.

ANDERSON, T., KUNITOMO, N. and MORIMUNE, K., 1986. Comparing Single Equation Estimators in a Simultaneous Equation System. Econometric Theory, 2: 1–32.

BASMANN, R. L., 1957. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. Econometrica, 25: 77–83.

CAPPS O. J. and GRUBBS, W., 1991. A Monte Carlo Study of Collinearity in Linear Simultaneous Equation Models. Journal of Statistical Computation and Simulation, 39(3): 139-162.

COTTLE, R. W. and DANTZIG, G. B., 1974. Complementary Pivot of Mathematical Programming. In: Dantzig GB, Eaves BC (Ed) Studies in Optimization, Vol. 10. Mathematical Association of America, Washington.

DANTZIG, G. B. and COTTLE, R. W., 1967. Positive (Semi-) Definite Matrices and Mathematical Programming. In: Abadie, J (Ed) Nonlinear Programming, North Holland Publishing Co, Amsterdam, pp 55-73.

DORN, W.S., 1960. A Duality Theorem for Convex Programs. IBM Journal of Research, 4: 407-413.

FAREBROTHER, R. W., 1976. Further Results on the Mean Square Error of Ridge Regression. Journal of the Royal Statistical Society, Series B (Methodological), 38(3): 248–250.

_____, 1999. Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900. New York, Springer.

FARRAR, D. E. and GLAUBER, R. R., 1967. Multicollinearity in Regression Analysis: The Problem Revisited. The Review of Economics and Statistics, 49(1): 92-107.

GROß, J., 2003. Restricted Ridge Estimation. Stat. Prob. Lett, 65: 57-64.

HOERL, A. E. and KENNARD, R.W., 1970a. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12: 55-67.

_____, 1970b. Ridge Regression: Application to Nonorthogonal Problems. Tecnometrics, 12: 69-82.

_____, 1976. Ridge Regression: Iterative Estimation of the Biasing Parameter. Communications in Statistics - Theory and Methods, A5(1): 77-78.

HOERL, A. E., KENNARD, R. W., and BALDWIN, K. F., 1975. Ridge Regression: Some Simulation. Communication in Statistics, 4(2): 105-123.

INTRILIGATOR, M. D., 1978. Econometric Methods. New York, McGraw Hill.

JAMES, W. and STEIN, C., 1961. Estimation with Quadratic Loss.Proc. Fourth Berkeley Symp. Math. Statist. Probab. 1: 361-380. Univ. California Press, 767p.

JORGENSON, D. W., 1962. Lecture Notes: Economics 241 Part IV: The Multivariate Structural Model. Committee on Econometrics and Mathematical Economics, Institute of Business and Economic Research, University of California, Berkeley.

KADIYALA, K. R. and NUNNS J. R., 1976. Estimation of a Simultaneous System of Equations When the Sample is Undersized. Unpublished Report, Indiana: Purdue University.

KEJIAN, L., 1993. A New Class of Biased Estimate in Linear Regression. Communications Statistics – Theory Methods, 22(2): 393-402.

KLEIN, L. R., 1950. Economic Fluctuation in the United States, 1921-1941. New York, John Wiley and Sons, Inc.

KOUTSOYIANNIS, A., 1977. Ekonometri Kuramı, (Çev. Ü. Şenesen, G. G. Şenesen). Verso Yayıncılık, Ankara, 688s.

LAWLESS, J. F. and WANG, P., 1976. A Simulation Study of Ridge and Other Regression Estimators. Communication in Statistics, 7: 139-164.

LEAMER, E. E., 1975. A Result on Sign of Restricted Least Squares Estimates. Journal of Econometrics, 3: 387-390.

LEMKE, C. E., 1962. A Method of Solution for Quadratic Programs. Management Science, 8(4): 442-453.

LIEW, C. K., 1976a. A Two-Stage Least-Squares Estimation with Inequality Restrictions on Parameters. The Review of Economics and Statistics, LVIII(2): 234-238.

_____, 1976b. Inequality Constrained Least-Squares Estimation. Journal of the American Statistical Associations Theory and Method Section, 71(355): 746-750.

LIU, K., 1993. A New Class of Biased Estimate in Linear Regression. Communication in Statistics - Theory and Methods, 22 (2): 393-402.

MAASOUMI, E., 1978. A Modified Stein Like Estimator for the Reduced Form Coefficients of Simultaneous Equations. Econometrica, 46, 695-703.

_____, 1980. A Ridge - Like Method for Simultaneous Estimation of Simultaneous Equations. Journal of Econometrics, 12(2): 161-176.

MANGASARIAN, O. L., 1962. Duality in Nonlinear Programming. Quarterly of Applied Mathematics, 20: 300-302.

MARQUARDT, D. W., 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation. Technometrics, 12: 591-612.

MAYER, L. W. and WILKIE, T. A., 1973. On Biased Estimation in Linear Models. Technometrics, 15: 497-508.

MCDONALD, G. C. and GALARNEAU, D. I., 1975. A Monte Carlo Evaluation of Some Ridge - Type Estimators. Journal of the American Statistical Association, 70: 407-416.

MEHTA, J. S. and SWAMY, P. A. V. B., 1978. The Existence of Moments of Spme Simple Bayes Estimators of Coeffficients in a Simultaneous Equations Model. Journal of Econometric, 7: 1-13.

MULLET, G. M., 1976. Why Regression Coefficients Have the Wrong Sign. Journal of Quality Technology, 8: 121-126.

NAGAR, A. L., 1960. A Monte Carlo Study of Alternative Simultaneous Equation Estimators. Econometrica, Vol. 28, No. 3, pp. 573-590.

NAMBOODIRI, N. K., CARTER, L. F. and BLALOCK, H. M., 1975. Applied Multivariate Analysis and Experimental Design. New York, McGraw Hill.

RAJ, B., 1980. A Monte Carlo Study of Small-Sample Properties of Simultaneous Equation Estimators with Normal and Nonnormal Disturbances Source. Journal of the American Statistical Association, Vol. 75, No. 369 pp. 221-229.

RAO, C. R., 1973. Linear Statistical Inference and Its Applications. 2nd ed. London:Wiley, 625p

SAKALLIOĞLU, S., KAÇIRANLAR, S. and AKDENİZ, F., 1996. Anoteon Combining Ridge and Least Squares Estimator. Journal of Institute of Mathematics and Computer Science (Math. Series), 9(2): 193-198.

_____, 2001. Mean Squared Error Comparisons of Some Biased Regression Estimators. Communications in Statistics, Theory and Methods, 30(2): 347-361.

SARGAN, J. D., 1958. Estimation of Economic Relationships Using Instrumental Variables. Econometrica, 67: 557–586.

SRIVASTAVA, V. K., TEWARIA, R. and PANDEY, K. N., 1978. Properties of Approximate Maximum Risk k-class Estimators When Disturbances are Small. Sankhya, Ser.C, 39: 110-127.

STEIN, C., 1956. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. Proceeding of the Third Berkeley

Symposium on Mathematical Statistics and Probability. Univeristy of California Press, Berkeley, 197-206.

THEIL, H., 1953a. Repeated Least-Squares Applied to Complete Equation Systems. Centraal Planbureau Memorandum.

_____, 1953b. Estimation and Simultaneous Correlation in Complete Equation Systems. Centraal Planbureau Memorandum. (Reprinted In: Raj, B., Koerts, J. (Eds.), 1992 Henri Theil's Contributions to Economics and Econometrics, Vol. 1, Kluwer, Dordrecht.)

_____, 1954. Estimation of Parameters in Econometric Models. Bulletin of the International Statistical Institute, 34: 122–129.

_____, 1961. Economic Forecasts and Policy (2nd rev. ed.), (North-Holland, Amsterdam). Contributions to Economic Analysis No. XV, (First Published 1958).

_____, 1963. On the Use of Incomplete Prior Information in Regression Analysis, J. Am. Stat. Assoc. 58: 401–414.

_____, 1971. Principal Econometrics, New York, John Wiley and Sons.

THEOBALD, C. M., 1974. Generalizations of Mean Square Error Applied to Ridge Regression. Journal of the Royal Statistical Society, Series B, 36: 103-105.

TOKER, S., ÜSTÜNDAĞ ŞİRAY, G. and KAÇIRANLAR, S., 2013. Inequality Constrained Ridge Regression Estimator. Statistics and Probability Letters, 83: 2391-2398.

ULLAH, A. and SRIVASTAVA, V. K., 1988. On the Improved Estimation of Structural Coefficients, Sankhya. The Indian Journal of Statistics, 50: 111-118.

VINOD H. D. and ULLAH A., 1981. Recent Advances in Regression Methods. Marcel Dekker, New York, Inc, 361p.

VISCO, I., 1978. On Obtaining the Ridge Sing of a Coefficient Estimate by Omitting a Variable From the Regression. Journal of Econometrics, 6: 115-118.

WAGNER, H. M., 1958. A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations. Econometrica, Vol. 26, No. 1, pp. 117-133.

ZELLNER, A. and VANDAELE, W., 1975. Bayes - Stein Estimators for k-Means, Regression and Simultaneous Equation Models. In Studies in Bayesian Econometrics and Statistics (Eds. S. E. Fienberg and A. Zellner), North-Holland Publishing Company, Amsterdam, 627-53.

CURRICULUM VITAE

Selma TOKER was born in Adana in 1985. She was graduated from Adana Anatolian High School in 2003. She started undergraduate education in Faculty of Science and Letters at Çukurova University in 2004. She graduated from the Department of Mathematics and the Department of Statistics with a double major programme in 2008. She got a Master's Degree in Statistics in 2011 and she joined the Doctor of Philosophy programme at the same department in the same year. She has been working as a research assistant in the Department of Statistics at Çukurova University since 2008.