INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9200081

Three essays on sampling techniques: Small sample performances of estimators and predictors

Chantanahom, Parisun, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Three Essays on Sampling Techniques: Small Sample Performances of Estimators and Predictors

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Parisun Chantanahom
B.A, Thammasat University, 1984
M.S., Louisiana State University, 1989
May 1991

ACKNOWLEDGEMENT

This research is accomplished through the help and advise of many persons. First and foremost, I would like to thank Professor R. Carter Hill, my advisor, for his guidance and kindness. I thank Professor W.D. McMillin for his encouragement and useful comments. Professor Hector Zapata and Professor Faik Koray have also provided many valuable remarks. I thank the dissertation committee members for their advise. I thank Melissa Waters for her clear explanation of the model used in the first topic. Ms. Samanporn Samitsombut's help with the organization of the work is affectionately appreciated. I am grateful for the assistantship granted to me by the Department of Economics for the first four years of my study.

I have never dreamt of going to graduate school and receive a doctorate degree. This accomplishment is the result of the devotion of all of my teachers in Thailand whom I am forever in debt. One regret is that none of my grandparents whom I dearly love live to share this moment ,especially grandpa Term Chantanahom who raised me during my early childhood. Lastly, my love and wholeheartedly gratification go to my parents for their affection and devotion. So here goes "This one for you mom, dad".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The dissertation addresses the issues of small sample properties of estimators and predictors. Economic analysis usually relies on the asymptotic properties of estimators and predictors which may not be the same as their asymptotic counterparts. Furthermore, some biased estimators and predictors used in economic studies have certain asymptotic properties which are not fully understood. Consequently, sampling techniques are used to explore the small sample properties and construct confidence intervals for predictors and estimators. In the dissertation, first, Monte Carlo experiments are used to find an appropriate estimation procedure for a system of simultaneous equations which involves a latent endogenous variable. Second, Monte Carlo experiments are used to explore the small sample property of the 'equity estimator' and compare it to the small sample properties of the 'traditional' estimators. Third, bootstrap sampling techniques is utilized to construct confidence intervals for the out-of-sample forecasts obtained via biased predictors which cannot be constructed in the usual way.

The findings are 1) an instrumental variables approach is an appropriate alternative estimation technique of the system of simultaneous equation involving a latent endogenous variable 2) the small sample of the equity

estimator is dependent on the vector lengths and the conditioning of the data and 3) bootstrap method produces reasonable confidence intervals for out-of-sample forecasts.

CHAPTER 1
INTRODUCTION

1.1 ORGANIZATION OF THE STUDY

In this study, we use sampling techniques to study the small sample properties of predictors and estimators when their small sample properties are not known. The asymptotic properties are sometimes used in the place of the small sample properties but they may not be the same. Consequently, sampling techniques are used to find and compare the small sample properties of estimators and predictors.

We are going to use Monte Carlo experiments and bootstrap sampling processes to help us choose an appropriate technique for estimating a particular type of simultaneous equations model. We will also use Monte Carlo experiments to evaluate the small sample performances of the alternative estimators. Lastly, we are going to engage in bootstrap sampling techniques to construct reasonable confidence intervals for out-of-sample forecasts obtained through a group of predictors.

1.2 SIMULTANEOUS EQUATIONS GENERALIZED PROBIT MODEL

The first section of the study concerns the estimation of simultaneous equations generalized probit model. Traditionally, the method of choice is that of Heckman (1978). Amemiya (1978) suggested certain estimation alternatives. Moreover, the model itself also suggests a

restriction on the parameter space. The performance of the alternatives relative to those of Heckman are unknown. In Chapter 2, we will develop Heckman's estimation method and its alternatives algebraically . In Chapter 3, we will use Monte Carlo experiments to study the small sample performance of each of the estimation procedures.

The first alternative is a generalized least squares approach to Heckman's estimation procedure. The second alternative is the instrumental variables approach. The third alternative is again the instrumental variables approach except that we make use of the estimated covariance matrix in the estimation process. The fourth alternative is a restricted least squares type estimator.

In Chapter 3, we are going to use a bargaining law determination model which is to be estimated by each of the alternatives. The model is in a simultaneous equations context. The first equation describes the determination of bargaining coverage as a proportion of employees. The second equation describes the determination of the unobservable sentiments toward enacting bargaining legislation . We will use Monte Carlo experiments to evaluate and compare each of the estimation alternatives via mean square error criteria. We will also examine the appropriateness of using the asymptotic covariance as the sample covariance for each of the estimation methods.

1.3 EQUITY ESTIMATOR

The second topic concerns the small sample performance of an estimator introduced in Krishnamurati and Rangaswamy (KR) (1987), called the 'equity estimator'. KR suggested that the equity estimator is to be used when multicollinearity is present. They claimed that the equity estimator deals with the problem of multicollinearity by treating each control variable in an equitable manner.

The properties of the equity estimator are not fully understood. KR (1987) used Monte Carlo experiments to show that under certain circumstances, the equity estimator had smaller mean square error than that of least squares and ridge regression.

In Chapter 4, we discuss the effects of multicollinearity on least squares. Then we introduce the use of traditional biased estimators, ridge regression and Stein-like estimators, when multicollinearity is present. We will describe the derivation and properties for each of the traditional biased estimators. Then we will discuss the derivation of the equity estimator and examine its characteristics.

In Chapter 5, we use actual marketing data to study the small sample performances of the equity estimator and compare them to those of least squares and the traditional biased estimator via Monte Carlo experiment.

1.4 A CONFIDENCE INTERVAL FOR OUT-OF-SAMPLE PREDICTION WHEN USING BIASED PREDICTORS.

Topic three concerns the establishing of confidence intervals for out-of-sample prediction when biased estimators are used as predictors. The confidence intervals of biased predictors forecast values cannot be obtained in the usual manner. Consequently, we use the bootstrap method to construct reasonable confidence intervals.

In Chapter 6, we discuss the effects of multicollinearity on least squares predictors. We will examine the use of biased predictors and their properties. Then we will introduce the bootstrap re-sampling method and its application on estimating confidence intervals for out-of-sample forecasts.

In Chapter 7, we apply the traditional biased estimators discussed in Chapter 4 to an actual set of data and use the resulting estimates to make out-of-sample predictions. Afterwards, we use the bootstrap method to construct reasonable confidence intervals for the forecast values.

CHAPTER 2

ON THE ESTIMATION OF A SIMULTANEOUS EQUATIONS GENERALIZED PROBIT MODEL

2.1 INTRODUCTION

Amemiya (1978) describes Heckman's approach to the problem of estimating simultaneous equations when there is a latent endogenous variable that is observed through an observable dichotomous endogenous variable.

The disturbance terms of the estimable structural equations are correlated with the dichotomous endogenous variable. Heckman suggests a two stage estimation procedure. In the first stage the dichotomous variable is replaced with a continuous proxy and least squares is applied in the second stage.

In this study, we develop alternatives to Heckman's estimation procedure and evaluate the small sample properties of each of the estimation techniques. The first alternative is to use a generalized least squares approach as suggested by Amemiya (1978). A second alternative is the instrumental variables approach. The third alternative is again the instrumental variables approach to the problem except that we make use of the estimated covariance matrix in the estimation process.

The fourth alternative that we will consider is a restricted least squares type estimator. We will see later that, based on a consistency condition, the single equation

estimation of the structural parameters yields two estimates of a key parameter which will not have the same value. Therefore, it is reasonable that we should estimate this parameter on the condition that its two estimates are the same. The restricted least squares estimator are used when we impose this constraint.

Our other concern is the performance of the covariance matrix estimators of each of the alternative structural estimators. The asymptotic covariance matrices of the various estimators that we have discussed can be obtained analytically. However, with a limited number of observations (limited in the sense that it is finite), the small sample variability of the estimation rules may not be the same as their estimated and theoretical counterparts, or the true asymptotic covariance. We will use a Monte-Carlo experiment to compare the small sample performances of the alternative estimation rules.

2.2 THE MODEL

The simultaneous equations that we explore are

$$\underset{\sim}{y}_1 = \gamma_1 \underset{\sim}{y}_2^* + \mathbf{X}_1 \underset{\sim}{\beta}_1 + \delta_1 \underset{\sim}{d} + \underset{\sim}{u}_1 \qquad (2.1a)$$

$$\underset{\sim}{y}_2^* = \gamma_2 \underset{\sim}{y}_1 + \mathbf{X}_2 \underset{\sim}{\beta}_2 + \delta_2 \underset{\sim}{d} + \underset{\sim}{u}_2 \qquad (2.1b)$$

where $\underset{\sim}{y}_1$ is a T×1 vector containing observations on an observable dependent variable, $\mathbf{X}_1$ is a $T \times K_1$ matrix of explanatory variables, $\mathbf{X}_2$ is a $T \times K_2$ matrix of explanatory variables , $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ are vectors of parameters with K_1 and

K_2 rows respectively. $\underset{-}{y}_2^*$ is a T×1 vector of values on an unobservable endogenous variable. Vectors $\underset{-}{u}_1$ and $\underset{-}{u}_2$ are disturbance terms with independent and identical bivariate normal distributions.

The dummy variable d is defined as

$$d_t = \begin{cases} 1 & y_{2t}^* > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Thus $\underset{-}{d}$ is the observable counterpart to the unobservable variable $\underset{-}{y}_2^*$. We obtain the reduced form equation for $\underset{-}{y}_1$ by substituting $\underset{-}{y}_2^*$ into Equation (2.1a),

$$\underset{-}{y}_1 = \gamma_1\left\{\gamma_2\underset{-}{y}_1 + \mathbf{X}_2\underset{-}{\beta}_2 + \delta_2\underset{-}{d} + \underset{-}{u}_2\right\} + \mathbf{X}_1\underset{-}{\beta}_1 + \delta_1\underset{-}{d} + \underset{-}{u}_1$$

$$= \frac{\mathbf{X}_1\underset{-}{\beta}_1}{1-\gamma_1\gamma_2} + \frac{\mathbf{X}_2\underset{-}{\beta}_2\gamma_1}{1-\gamma_1\gamma_2} + \frac{\underset{-}{d}(\gamma_1\delta_2+\delta_1)}{1-\gamma_1\gamma_2} + \frac{\underset{-}{u}_1+\gamma_2\underset{-}{u}_2}{1-\gamma_1\gamma_2}$$

$$= \mathbf{X}\Pi_1 + \delta_1\underset{-}{d} + \underset{-}{v}_1 \quad . \qquad (2.2)$$

Heckman (1978) has proved that for the model to be logically consistent $\delta_2 = -\gamma_2\delta_1$.

Similarly substituting y_1 into Equation (2.1b) yields the reduced form equation for y_2^*,

$$\underset{-}{y}_2^* = \gamma_2\left\{\gamma_1\underset{-}{y}_2^* + \mathbf{X}_1\underset{-}{\beta}_1 + \delta_1\underset{-}{d} + \underset{-}{u}_1\right\} + \mathbf{X}_2\underset{-}{\beta}_2 + \delta_2\underset{-}{d} + \underset{-}{u}_2$$

$$= \frac{\mathbf{X}_1\underset{-}{\beta}_1\gamma_2}{1-\gamma_1\gamma_2} + \frac{\mathbf{X}_2\underset{-}{\beta}_2}{1-\gamma_1\gamma_2} + \frac{\underset{-}{d}(\gamma_2\delta_1+\delta_2)}{1-\gamma_1\gamma_2} + \frac{\underset{-}{u}_2+\gamma_2\underset{-}{u}_1}{1-\gamma_1\gamma_2}$$

$$= X\Pi_2 + \underset{\sim}{v}_2 \qquad (2.3)$$

where $\underset{\sim}{v}_1 = \dfrac{u_1+\gamma_1 \underset{\sim}{u}_2}{1-\gamma_1\gamma_2}$ and $\underset{\sim}{v}_2 = \dfrac{\underset{\sim}{u}_2+\gamma_2 u_1}{1-\gamma_1\gamma_2}$.

Assuming that the joint density function of v_{2t} and d_t, denoted as $g(v_{2t},d_t)$ is a proper density function ,i.e.

$$\sum_{d_t=0,1} \int_{-\infty}^{\infty} g(v_{2t},d_t)\, dv_{2t} = 1 \quad .$$

By definition of d_t, the probability that $y^*_{2t}>0$ given $d_t = 1$ is one,

i.e. $$Pr\left[v_{2t} > -x'_{1t}\Pi_{21} - x'_{2t}\Pi_{22} - \Pi_{23}\right] = 1$$

$$Pr\left[v_{2t} > \ell_t\right] = 1 \quad .$$

Therefore,

$$\int_{\ell_t}^{\infty} g(v_{2t},1)\, dv_{2t} = F_t \qquad (2.4a)$$

and

$$\int_{-\infty}^{\ell_t} g(v_{2t},1)\, dv_{2t} = 0 \quad . \qquad (2.4b)$$

Similarly, the probability that $y^*_{2t} \leq 0$ given $d_t = 0$ is one.

$$Pr\left[v_{2t} \leq -x'_{1t}\Pi_{21} - x'_{2t}\Pi_{22}\right] = 1$$

$$Pr\left[v_{2t} \leq m_t\right] = 1 \quad .$$

Consequently,

$$\int_{-\infty}^{m_t} g(v_{2t},0)\,dv_{2t} = 1-F_t \qquad (2.4c)$$

$$\int_{m_t}^{\infty} g(v_{2t},0)\,dv_{2t} = 0 \quad . \qquad (2.4d)$$

For the joint density function of v_{2t},d_t to be proper, the sum of the left hand side terms of Equations (2.4a)-(2.4d) must be equal to the sum of the right hand side terms which equals to 1. This will be the case only when $\Pi_{23} = 0$ or $\gamma_2\delta_1+\delta_2 = 0$ or when the model is logically consistent. The probit model estimates the changes in the probability of the event d = 1 with respect to the variables on the right hand side of Equation (2.3). Consequently, the probability that d = 1 cannot be a determinant of the event itself.

Let σ_1^2 represent $\mathrm{Var}(v_1)$,σ_2^2 represent $\mathrm{Var}(v_2)$ and σ_{12} represent $\mathrm{Cov}(v_1,v_2)$. We can normalize by letting $\sigma_1^2=1$ since y_2^* is a dichotomous variable and thus we can identify Π_2 only up to a scalar multiple .

Equation (2.2) can be estimated by the ordinary least squares estimator. Equation (2.3) can be estimated by the probit method using d in replacement of y_2^*.

2.3 HECKMAN'S MODEL

2.3.1 FIRST STRUCTURAL EQUATION.

Substituting y_2^* in Equation (2.3) into Equation (2.1a) and solving for the structural parameters, we get

$$\underline{y}_1 = \gamma_1\left[\mathbf{X}\Pi_2 + \underline{v}_2\right] + \mathbf{X}_1\underline{\beta}_1 + \delta_1\underline{d} + \underline{u}_1$$

$$= \gamma_1\mathbf{X}\hat{\Pi}_2 + \mathbf{X}J_1\underline{\beta}_1 + \delta_1\underline{d} - \gamma_1\mathbf{X}(\hat{\Pi}_2 - \Pi_2) + \underline{u}_1 + \gamma_1\underline{v}_2$$

$$= \mathbf{X}(\hat{\Pi}_2,\ J_1)\begin{bmatrix}\gamma_1\\ \underline{\beta}_1\end{bmatrix} + \delta_1\underline{d} + \underline{w}_1$$

$$= \mathbf{X}\hat{\mathbf{H}}\begin{bmatrix}\gamma_1\\ \underline{\beta}_1\end{bmatrix} + \delta_1\hat{F} + \delta_1(\underline{d}+F-F-\hat{F}) + \underline{w}_1$$

$$= (\mathbf{X}\hat{\mathbf{H}},\ \hat{F})\begin{bmatrix}\gamma_1\\ \underline{\beta}_1\\ \delta_1\end{bmatrix} - (\hat{F}-F)\delta_1 - \delta_1(F-\underline{d}) + \underline{w}_1$$

$$= \mathbf{Z}_1\underline{\beta}_1^* + \underline{w}_1^* \qquad (2.5)$$

where

$\mathbf{X}J_1 = \mathbf{X}_1$,

$F = F(\mathbf{X}\Pi_2)$ and $\hat{F} = F(\mathbf{X}\hat{\Pi}_2)$,

F is the CDF of a standard normal distribution function,

$\underline{w}_1 = -\gamma_1\mathbf{X}(\hat{\Pi}_2 - \Pi_2) + \underline{u}_1 + \gamma_1\underline{v}_2 = \underline{v}_1 - \gamma_1\mathbf{X}(\hat{\Pi}_2 - \Pi_2)$

and $\underline{w}_1^* = \underline{w}_1 - (\hat{F}-F)\delta_1 - \delta_1(F-\underline{d})$.

Note that $$\left[\underline{v}_1\ \underline{v}_2\right] = -\left[\underline{u}_1\ \underline{u}_2\right]\Gamma^{-1}$$

where $$\Gamma^{-1} = \begin{bmatrix}-1 & \gamma_2\\ \gamma_1 & -1\end{bmatrix}^{-1} = \frac{1}{1-\gamma_1\gamma_2}\begin{bmatrix}-1 & -\gamma_2\\ -\gamma_1 & -1\end{bmatrix}.$$

Consequently,

$$\underline{v}_1 = \frac{1}{1-\gamma_1\gamma_2}\left[\underline{u}_1 + \gamma_1\underline{u}_2\right] \qquad (2.6a)$$

$$\underset{-2}{v} = \frac{1}{1-\gamma_1\gamma_2}\left[\underset{-2}{u} + \gamma_2 \underset{-1}{u}\right] \quad . \qquad (2.6b)$$

Solve (2.6a) for $\underset{-1}{u}$ and substitute into (2.6b) ,we will get $\underset{-1}{v} = \underset{-1}{u} + \gamma_1 \underset{-2}{v}$. Similarly, if we solve (2.6b) for $\underset{-2}{u}$ and substitute it into (2.6a) , we will get $\underset{-2}{v} = \gamma_2 \underset{-1}{v} + \underset{-2}{u}$.

Let $\mathrm{Cov}(u_{\cdot t}) = \phi$ then,

$$\mathrm{Cov}(v_{\cdot t}) = \Sigma = (-\Gamma^{-1})'\phi\,(-\Gamma^{-1})$$

$$= \frac{1}{(1-\gamma_1\gamma_2)^2}\begin{bmatrix} \phi_1^2+2\gamma_1\phi_{12}+\gamma_1^2\phi_2^2 & \gamma_2\phi_1^2+(1+\gamma_1\gamma_2)\phi_{12}+\gamma_1\phi_2^2 \\ \cdot & \gamma_2^2\phi_1^2+2\gamma_2\phi_{12}+\phi_2^2 \end{bmatrix}. \qquad (2.6c)$$

Since $$v_{1t} = \gamma_1 v_{2t} + u_{1t} \quad ,$$

$$v_{1t}v_{2t} = \gamma_1 v_{2t}^2 + u_{1t}v_{2t}$$

$$\sigma_{12} = E[v_{1t}v_{2t}] = \gamma_1\sigma_2^2 + E[u_{1t}v_{2t}]$$

$$E[u_{1t}v_{2t}] = E\left[u_{1t}\left[\frac{\gamma_2 u_{1t} + u_{2t}}{1-\gamma_1\gamma_t}\right]\right]$$

$$= \frac{\gamma_2\phi_1^2 + \phi_{12}}{1-\gamma_1\gamma_2} \quad .$$

Therefore,

$$\sigma_{12} = \gamma_1\sigma_2^2 + \frac{\gamma_2\phi_1^2 + \phi_{12}}{1-\gamma_1\gamma_2} \quad .$$

We know that

$$v_{1t} = \gamma_1 v_{2t} + u_{1t}$$

$$= v_{2t}\left(\frac{\sigma_{12}}{\sigma_2^2}\right) + u_{1t} - \left\{ \frac{\gamma_2\phi_1^2 + \phi_{12}}{\gamma_2^2\phi_1^2 + 2\gamma_2\phi_{12}+\phi_2^2} \right\} v_{2t}$$

$$= (\sigma_{12}/\sigma_2^2)\ v_{2t} + e_t \qquad (2.6d)$$

where e_t is normally distributed and independent of v_{2t}

$$E[e_t v_{2t}] = E\left[v_{2t}\left[u_{1t} - \frac{\gamma_2\phi_1^2 + \phi_{12}}{\gamma_2^2\phi_1^2 + 2\gamma_2\phi_{12}+\phi_2^2}\ v_{2t}\right]\right]$$

$$= 0 \quad .$$

The relationship in Equation (2.6d) is used in the calculation of the covariance matrix.

Equation (2.5) can be estimated by ordinary least squares,

$$\hat{\underline{\beta}}_1^* = (Z_1'Z_1)^{-1}\cdot Z_1'\underline{y}_1 . \qquad (2.7)$$

The asymptotic covariance of $\hat{\underline{\beta}}_1^*$ is

$$\mathrm{Cov}(\hat{\underline{\beta}}_1^*) = \left\{(Z_1'Z_1)^{-1}Z_1'\right\}\cdot \mathrm{Cov}(w_1^*)\cdot\left\{(Z_1'Z_1)^{-1}\cdot Z_1'\right\}' . \qquad (2.8)$$

2.3.2 THE $\mathrm{Cov}(w_1^*)$

$$\mathrm{Cov}(w_1^*) = E\left\{\underline{w}_1^*\ \underline{w}_1^{*\prime}\right\}$$

$$= E\left\{\ \underline{w}_1 + (\underline{d}-F)\delta_1 - (\hat{F}-F)\delta_1\right\}\cdot\left\{\ \cdot\ \right\}'$$

$$= E\left[\underline{w}_1\underline{w}_1'\right] + E\left[\ (\underline{d}-F)\delta_1\delta_1'(\underline{d}-F)'\right] + E\left[(\hat{F}-F)\delta_1\delta_1'(\hat{F}-F)'\right]$$

$$+E\left[\underline{w}_1\delta_1'(\underline{d}-F)'\right] + E\left[\delta_1(\underline{d}-F)\underline{w}_1'\right]$$

$$-E\left[\underline{w}_1\delta_1'(\hat{F}-F)'\right] - E\left[(\hat{F}-F)\delta_1 w_1'\ \right]$$

$$-E\left[(\underset{\sim}{d}-F)\delta_1\delta_1'(F-\hat{F})'\right] \; -E\left[(F-\hat{F})\delta_1\delta_1'(\underset{\sim}{d}-F)'\right] \qquad (2.9)$$

where

$$E\left[\underset{\sim}{w}_1\underset{\sim}{w}_1'\right] = E\left[\; (\underset{\sim}{u}_1 + \gamma_1\underset{\sim}{v}_2) \; - \gamma_2 X(\hat{\Pi}_2 - \Pi_2)\right]\cdot\left[\quad\cdot\quad\right]'$$

$$= E\left[\underset{\sim}{v}_1\underset{\sim}{v}_1' \; -\gamma_1\underset{\sim}{v}_1(\hat{\Pi}_2-\Pi_2)'X' \; -\gamma_1 X(\hat{\Pi}_2-\Pi_2)\underset{\sim}{v}_1'\right.$$

$$\left. +\gamma_1^2 X(\hat{\Pi}_2-\Pi_2)(\hat{\Pi}_2-\Pi_2)'X\right]$$

$$= E\left[\underset{\sim}{v}_1\underset{\sim}{v}_1'\right] \; -\gamma_1 \mathrm{Cov}(v_1,(\hat{\Pi}_2-\Pi_2)')X' - \gamma_1 X\mathrm{Cov}((\hat{\Pi}_2-\Pi_2),\underset{\sim}{v}_1')$$

$$+\gamma_1^2 X(\mathrm{Var}(\hat{\Pi}_2))X' \quad . \qquad (2.10)$$

We will now evaluate each of these terms. We assume that (v_{t1}, v_{t2}) are independently and identically distributed with zero mean and covariance

$$\Sigma \; = \; \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix} \quad .$$

Estimation of the probit model (2.3) via maximum likelihood yields the asymptotic covariance matrix of $\hat{\Pi}_2$ as the inverse of the information matrix,

$$\mathrm{Cov}(\hat{\Pi}_2) \; = \; -E\left[\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'}\right]^{-1} \quad .$$

The log likelihood function of the probit model is

$$\ln \ell = \sum_{t=1}^{T}\left[d_t \ln(F(x_t'\Pi_2)) \; + \; (1-d_t)\ln(1-F(x_t'\Pi_2))\right] \qquad (2.11)$$

and $$\frac{\partial \ln \ell}{\partial \Pi_2} = \sum_{t=1}^{T} \left[d_t \frac{f(x_t'\Pi_2)}{F(x_t'\Pi_2)} - (1-d_t) \frac{f(x_t'\Pi_2)}{1-F(x_t'\Pi_2)} \right]$$

$$= \sum_{t=1}^{T} \left[f_t (d_t - F_t)/(F_t(1-F_t)) \right] \quad .$$

Let F_t be the CDF of standard normal N(0,1) distribution evaluated at $x_t'\Pi_2$ and $f_t = F_t'$ and

$$\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'} = -\sum_{t=1}^{T} f_t \left\{ y_t \cdot \frac{f_t + (x_t'\Pi_2)F_t}{F_t} + (1-y_t) \cdot \frac{f_t - (x_t'\Pi_2)(1-F_t)}{(1-F_t)^2} \right\} x_t' x_t$$

Then,

$$\mathrm{Cov}(\hat{\Pi}_2) = \left[\sum_{t=1}^{T} \left\{ \frac{f_t^2}{(1-F_t)F_t} \right\} x_t' x_t \right]^{-1} .$$

Let $\mathbf{A} = \mathrm{diag}\left\{ \frac{f_t^2}{(1-F_t)F_t} \right\}$.

Then,

$$\mathrm{Cov}(\hat{\Pi}_2) = (\mathbf{X'AX}^{-1}) = -\mathbf{H}^{-1} \qquad (2.12)$$

Since $\hat{\Pi}_2$ is obtained by maximizing the likelihood function in (2.11), using a first order Taylor's series expansion of $\hat{\Pi}_2$ around Π_2 gives

$$\left. \frac{\partial \ln \ell}{\partial \Pi_2} \right|_{\hat{\Pi}_2} = \left. \frac{\partial \ln \ell}{\partial \Pi_2} \right|_{\Pi_2} + \left. \frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'} \right|_{\Pi_2} (\hat{\Pi}_2 - \Pi_2) \quad .$$

Because $\hat{\Pi}_2$ maximizes $\ln\ell$, we know that

$$\left.\frac{\partial \ln \ell}{\partial \Pi_2}\right|_{\hat{\Pi}_2} = 0 \quad .$$

Thus,

$$(\hat{\Pi}_2 - \Pi_2) = -\left[\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'}\right]^{-1} \cdot \frac{\partial \ln \ell}{\partial \Pi_2} \quad .$$

Under regularity conditions (see Dhrymes 1974),

$$\operatorname{plim}_{T\to\infty} -\frac{1}{T}\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'} = -E\,\frac{1}{T}\left[\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'}\right] \quad .$$

As a consequence, $(\hat{\Pi}_2 - \Pi_2)$ has the same asymptotic distribution as $-E\left[\frac{\partial^2 \ln \ell}{\partial \Pi_2 \partial \Pi_2'}\right]^{-1} \cdot \frac{\partial \ln \ell}{\partial \Pi_2}$.

Consequently,

$$\begin{aligned}
\operatorname{Cov}((\hat{\Pi}_2 - \Pi_2), \underset{-1}{v}') &= -\mathbf{H}^{-1} E\left[\frac{\partial \ln \ell}{\partial \Pi_2} \cdot \underset{-1}{v}'\right] \\
&= (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} \cdot E\left[\sum_{t=1}^{T} x_t' \frac{d_t - F_t}{F_t(1-F_t)} f_t \underset{t-1}{v}'\right] \\
&= (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} \cdot \sum_{t=1}^{T} f_t x_t E\left[\frac{d_t - F_t}{F_t(1-F_t)} \sigma_{12} \underset{-2}{v}'\right] \\
&= (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} \sum_{t=1}^{T} f_t x_t E\left[\frac{d_t \underset{-2}{v}'}{F_t(1-F_t)}\right] ,
\end{aligned}$$

note that,

$$\begin{aligned}
E\left[d_t v_{2t}\right] &= E\left[v_{2t} | v_{2t} > 0\right] \cdot \Pr\left[v_{2t} > 0\right] \\
&= \frac{f_t}{F_t} F_t = f_t \quad .
\end{aligned}$$

Thus, $$\operatorname{Cov}((\hat{\Pi}_2 - \Pi_2), \underset{-1}{v}') = \sigma_{12}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1} \sum_{t=1}^{T} \frac{f_t^2}{F_t(1-F_t)} x_t$$

$$= \sigma_{12}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A} \quad . \quad (2.13)$$

We have

$$\mathrm{Cov}(w_1) = \sigma_1^2 I_T - \gamma_1\sigma_{12}AX(X'AX)^{-1}X - \gamma_1\sigma_{12}X(X'AX)^{-1}X'A + \gamma_1^2X(X'AX)^{-1}X' \tag{2.14}$$

$E\left[(d-F)\delta_1^2(d-F)'\right]$ has its ($t\neq s$) element described as

$$\delta_1^2 E\left[(d_t-F_t)(d_s-F_s)\right] = \begin{cases} 0 \ ; \ t\neq s \\ \delta_1^2 F_t(1-F_t) \ ; \ t=s \end{cases}.$$

Therefore,

$$E\left[\delta_1^2(d-F)(d-F)'\right] = \delta_1^2 D_2 \tag{2.15}$$

where $D_2 = \mathrm{diag}(F_t(1-F_t))$

$$E\left[\underset{\sim}{w}_1\delta_1(\hat{F}-F)'\right] = E\left\{\left[\underset{\sim}{v}_1-\gamma_1X(\hat{\Pi}_2-\Pi_2)\right]\left[\delta_1(\hat{F}-F)'\right]\right\}$$

$$= E\left[\underset{\sim}{v}_1\delta_1(\hat{F}-F)'\right] - E\left[\gamma_1\delta_1X(\hat{\Pi}_2-\Pi_2)(\hat{F}-F)'\right].$$

Using the first term of Taylor's series expansion

$$\hat{F}_t \cong F_t + f_t x_t'(\hat{\Pi}_2-\Pi_2) \ .$$

Thus,

$$E\left[\underset{\sim}{v}_1\delta_1(\hat{F}-F)'\right] = \delta_1 E\left[\underset{\sim}{v}_1(\hat{\Pi}_2-\Pi_2)\right]XD_1 = \sigma_{12}\delta_1AX(X'AX)^{-1}XD_1 \ .$$

where $D_1 = \mathrm{diag}\ (f_t)$.

Consequently,

$$E\left[\underset{\sim}{w}_1\delta_1(\hat{F}-F)'\right] = \sigma_{12}\delta_1AX(X'AX)^{-1}XD_1 + \gamma_1\delta_1X(X'AX)^{-1}XD_1. \tag{2.16}$$

$$E\left[\underset{\sim}{w}_1\delta_1(\underset{\sim}{d}-F)'\right] = E\left\{\left[\underset{\sim}{v}_1-\gamma_1X_1(\hat{\Pi}_2-\Pi_2)\right]\left[\delta_1(\underset{\sim}{d}-F)'\right]\right\}$$

$$= E\left\{\left[(\sigma_{12}\underline{v}_2+\underline{e})-\gamma_1 X_1(\hat{\Pi}_2-\Pi_2)\right]\left[\delta_1(\underline{d}-F)'\right]\right\}$$

$$= E\left[\sigma_{12}\delta_1\underline{v}_2(\underline{d}-F)'\right] - E\left[\gamma_1\delta_1 X(\hat{\Pi}_2-\Pi_2)(\underline{d}-F)'\right].$$

$$E\left[\sigma_{12}\delta_1 v_{2t}(d_s-F_s)\right] = \begin{cases} 0 \ ; \ t\neq s \\ \sigma_{12}\delta_1 f_t \ ; \ t=s \end{cases}.$$

ie. $E\left[\sigma_{12}\delta_1 v_2(d-F)'\right] = \sigma_{12}\delta_1 D_1.$

$$E\left[\gamma_1\delta_1 X(\hat{\Pi}_2-\Pi_2)(\underline{d}-F)'\right] = E\left[\gamma_1\delta_1 X(X'AX)^{-1}\sum_{t=1}^{T}\left[f_t x_t \frac{d_t-F_t}{F_t(1-F_t)}\right](\underline{d}-F)'\right]$$

$$= \gamma_1\delta_1(X'AX)^{-1}XD_1'$$

$$E\left[\gamma_1\delta_1(\underline{d}-F)'\right] = \sigma_{12}\delta_1 D_1 - \gamma_1\delta_1 X(X'AX)^{-1}XD_1' . \qquad (2.17)$$

Using Taylor's series approximation

$$E\left[(\underline{d}-F)\delta_1^2(F-\hat{F})'\right] = E\left[\delta_1^2(\underline{d}-F)(\hat{\Pi}_2-\Pi_2)XD_1'\right]$$

$$= \delta_1^2 D_1 X(X'AX)^{-1}XD_1' \qquad .(2.18)$$

Thus, $Cov(w_1^*) = \sigma_1^2 I_T + (\gamma_1 I_T+\delta_1 D_1)X(X'AX)^{-1}X'(\gamma_1 I_T+\delta_1 D_1) + \delta_1^2 D_2$

$$-\sigma_{12}(\gamma_1 I_T+\delta_1 D_1)X(X'AX)^{-1}X'A$$

$$-\sigma_{12}AX(X'AX)^{-1}-X'(\gamma_1 I_T+\delta_1 D_1) + 2\delta_1\sigma_{12}D_1$$

$$-\delta_1(\gamma_1 I_T+\delta_1 D_1)X(X'AX)^{-1}XD_1'$$

$$-\delta_1 D_1 X(X'AX)^{-1}X'(\gamma_1 I_T+\delta_1 D_1) . \qquad (2.19)$$

2.3.3 THE SECOND STRUCTURAL EQUATION

Repeating the procedure in Section 2.3.1, that is substituting y_2^* in Equation (2.3) into Equation (2.1b) and solving for y_1 yields,

$$\mathbf{X}\Pi_2 + \underline{v}_2 = \gamma_2 \underline{y}_1 + \mathbf{x}_2\underline{\beta}_2 + \delta_2 \underline{d} + \underline{u}_2$$

$$\gamma_2 \underline{y}_1 = \mathbf{X}\hat{\Pi}_2 - \mathbf{x}_2\underline{\beta}_2 - \delta_2 \underline{d} - \mathbf{X}(\hat{\Pi}_2 - \Pi_2) + \underline{v}_2 - \underline{u}_2$$

$$\underline{y}_1 = \frac{\mathbf{X}\hat{\Pi}_2}{\gamma_2} - \frac{\mathbf{x}_2\underline{\beta}_2}{\gamma_2} - \frac{\delta_2 \underline{d}}{\gamma_2} - \frac{\mathbf{X}(\hat{\Pi}_2 - \Pi_2)}{\gamma_2} + \frac{(\underline{v}_2 - \underline{u}_2)}{\gamma_2}$$

$$= \mathbf{X}(\hat{\Pi}_2, -J_2) \begin{bmatrix} 1/\gamma_2 \\ \underline{\beta}_2/\gamma_2 \end{bmatrix} + \delta_1 \underline{d} + \underline{w}_2$$

$$= \mathbf{X}\hat{\mathbf{Q}} \begin{bmatrix} 1/\gamma_2 \\ \underline{\beta}_2/\gamma_2 \end{bmatrix} + \delta_1 \hat{F} + \delta_1(\underline{d} + F - F - \hat{F}) + \underline{w}_2$$

$$= (\mathbf{X}\hat{\mathbf{Q}}, \hat{F}) \begin{bmatrix} 1/\gamma_2 \\ \underline{\beta}_2/\gamma_2 \\ \delta_1 \end{bmatrix} + \underline{w}_2^*$$

$$= \mathbf{Z}_2\underline{\beta}_2^* + \underline{w}_2^* \qquad (2.20)$$

where

$$\mathbf{X}J_2 = \mathbf{X}_2 ,$$

$$\underline{w}_2 = -\frac{1}{\gamma_2}\mathbf{X}(\hat{\Pi}_2 - \Pi_2) + \frac{1}{\gamma_2}(\underline{v}_2 - \underline{u}_2) = \underline{v}_1 - \frac{1}{\gamma_2}\mathbf{X}(\hat{\Pi}_2 - \Pi_2)$$

and $\underline{w}_2^* = \underline{w}_2 - (\hat{F} - F)\delta_1 - \delta_1(F - \underline{d})$.

Note that $\delta_2 = -\gamma_2\delta_1$ has been used, and that $\underline{\beta}_2^*$ contains δ_1 and not δ_2.

From the above expression for $\underline{w}_2^*$, we can see that $\mathrm{Cov}(\underline{w}_2^*)$ is the same as $\mathrm{Cov}(\underline{w}_1^*)$ except for the term $(\gamma_1\mathbf{I}_T + \delta_1\mathbf{D}_1)$ is replaced by $(\gamma_2^{-1}\mathbf{I}_T + \delta_1\mathbf{D}_1)$.

Equation (2.20) can be estimated by ordinary least squares, that is

$$\hat{\underline{\beta}}_2^* = (Z_2'Z_2)^{-1} \cdot Z_2'\underline{y}_1 \qquad (2.21)$$

The asymptotic covariance matrix of $\hat{\beta}_2^*$ is

$$\text{Cov}(\hat{\underline{\beta}}_2^*) = \left\{(Z_2'Z_2)^{-1}Z_2'\right\} \cdot \text{Cov}(\underline{w}_2^*) \cdot \left\{(Z_2'Z_2)^{-1}Z_2'\right\}' \quad . \qquad (2.22)$$

2.3.4 THE ESTIMATION OF Σ.

The next stage is to find a consistent estimator for σ_{12} and σ_1^2 . The technique that we are going to use is that of Heckman(1978). The estimates for γ_1, γ_2 and δ_1 can be derived from Equation (2.7) and (2.21). δ_2 is derived from Equation (2.21) by using the relationship that $\delta_2 = -\gamma_2\delta_1$. Consider Equation (2.2)

$$\underline{y}_1 = X\Pi_2 + \delta_1\underline{d} + \underline{v}_1$$

$$E\left[\underline{y}_1 | X, \underline{d}\right] = X\Pi_2 + \delta_1\underline{d} + E\left[\underline{v}_1 | X, \underline{d}\right]$$

since, $v_{1t} = \sigma_{12}v_{2t} + e_t$ where $e_t \sim N(0, \sigma_e^2)$ and is independent of v_{2t}.

Thus,

$$E\left[v_{1t} | x_t, d_t\right] = \sigma_{12}E\left[v_{2t} | x_t, d_t\right] \quad .$$

If $d_t = 1$,

$$E\left[v_{2t} | d_t\right] = E\left[v_{2t} | v_{2t} > -x_t'\Pi_2\right]$$

$$= \left[\int_{-x_t'\pi_2}^{\infty} v_{2t} \cdot (2\Pi)^{-1/2} \exp(-\tfrac{1}{2}v_{2t}^2)\, dv_{2t}\right] \cdot \left(\Pr\left[v_{2t} > -x_t'\Pi_2\right]\right)^{-1}$$

$$= \frac{f(-x_t'\Pi_2)}{1-F(-x_t'\Pi_2)}$$

$$= \frac{f(x_t'\Pi_2)}{F(x_t'\Pi_2)} = \lambda_t \quad . \tag{2.23}$$

Note that,

$$E\left[v_{2t} | v_{2t} > -x_t'\Pi_2\right] = \int_{-x_t'\Pi_2}^{\infty} v_{2t} \cdot f(v_{2t} | v_{2t} > -x_t'\Pi_2)\ dv_{2t}$$

and $$f(v_{2t} | v_{2t} > -x_t'\Pi_2) = \frac{f(v_{2t})}{h(v_{2t} > -x_t'\Pi_2)} \quad ,$$

where $$h(v_{2t} > -x_t'\Pi_2) = \int_{-x_t'\Pi_2}^{\infty} f(v_{2t})\ dv_{2t}$$

$$= 1-F(-x_t'\Pi_2) \quad .$$

Similarly,

if $1-d_t = 1$, $$E\left[v_{2t} | d_t\right] = E\left[v_{2t} | v_{2t} \leq -x_t'\Pi_2\right]$$

$$= \frac{-f(-x_t'\Pi_2)}{F(-x_t'\Pi_2)}$$

$$= \frac{-f(x_t'\Pi_2)}{F(-x_t'\Pi_2)}$$

$$= \frac{-\lambda \cdot F(x_t'\Pi_2)}{F(-x_t'\Pi_2)} = \lambda_t^* \quad . \tag{2.24}$$

Consequently,

$$E\left[\underline{y}_1 | X, \underline{d}\right] = X\Pi_1 + \delta_1 \underline{d} + \sigma_{12}\left[\lambda \underline{d} + \lambda^*(1-\underline{d})\right]$$

$$= X\Pi_1 + \delta_1 \underline{d} + \sigma_{12}\left[\hat{\lambda} \underline{d} + \hat{\lambda}^*(1-\underline{d})\right] + \hat{\underline{v}}_1 .$$

(2.25)

We know that $v_{1t} = \sigma_{12} v_{2t} + e_t$, therefore

$$\mathrm{Var}(v_{1t}|x_t,d_t) = \sigma_{12}^2 \mathrm{Var}(v_{2t}|x_t,d_t) + \mathrm{Var}(e_t).$$

Thus, $\mathrm{Var}(e_t) = \sigma_1^2 - \sigma_{12}^2$.

$$\mathrm{Var}(v_{2t}|x_t,d_t) = E\left[v_{2t}^2|x_t,d_t\right] - \left\{E\left[v_{2t}|x_t,d_t\right]\right\}^2$$

If $d_t=1$, $E\left[v_{2t}^2|x_t,d_t\right] = E\left[v_{2t}^2|x_t,v_{2t}>-x_t'\Pi_2\right]$

$$= \frac{1}{1-F(-x_t'\Pi_2)} \cdot \int_{-x_t'\Pi_2}^{\infty} v_{2t} e^{(-v_{2t}^2/2)^2} dv_{2t}$$

$$= a \cdot \left\{(-x_t'\Pi_2) f(x_t'\Pi_2) + [1-F(-x_t'\Pi_2)]\right\}$$

$$= 1 + \frac{(-x_t'\Pi_2) f(-x_t'\Pi_2)}{1-F(-x_t'\Pi_2)}$$

where $a = 1/(1-F(-x_t'\Pi_2))$.

Hence,

$$\mathrm{Var}(v_{2t}|x_t,v_{2t}>-x_t'\Pi_2) = 1 + (-x_t'\Pi_2)\lambda_t - \lambda_t^2 = q_t.$$

If $1-d_t=1$, $E\left[v_{2t}^2|x_t,d_t\right] = E\left[v_{2t}^2|x_t,v_{2t}\leq -x_t'\Pi_2\right]$

$$= \frac{1}{F(-x_t'\Pi_2)} \left\{(-x_t'\Pi_2)(-f(x_t'\Pi_2)) + F(-x_t'\Pi_2)\right\}$$

Hence,

$$\mathrm{Var}(v_{2t}|x_t,v_{2t}\leq -x_t'\Pi_2) = 1 + (-x_t'\Pi_2)\lambda_t^* - \lambda_t^{*2} = s_t .$$

Consequently,

$$\mathrm{Var}(v_{2t}|x_t,d_t) = d_t q_t + (1-d_t) s_t. \qquad (2.26)$$

$$\text{Var}(v_{1t}|x_t,d_t) = \sigma_1^2 - \sigma_{12}^2 + \sigma_{12}^2\left[d_t q_t + (1-d_t)s_t\right]$$

$$= \sigma_1^2\left[(1-\rho^2) + \rho^2(d_t q_t + (1-d_t)s_t\right] \qquad (2.27)$$

where $\rho = \sigma_{12}/(\sigma_1 \cdot \sigma_2) = \sigma_{12}/\sigma_1$, since we apply normalization rule on the second reduced form equation.

Using the relationship in Equation (2.25), a consistent estimator of σ_{12} can be obtained by applying OLS. If we let $\hat{\eta}_t$ denote the residual from the OLS estimation above, then a possible consistent estimator of σ_1^2 is obtained by the estimated residuals of Equation (2.25). From Equation(2.27), the estimated σ_1^2 is obtained from

$$\hat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t^2 + \hat{\sigma}_{12}^2\left[1 - \frac{1}{T}\sum_{t=1}^{T}(d_t q_t + (1-d_t)s_t\right] \qquad (2.28)$$

2.3.5 THE PARAMETERS OF INTEREST.

From Equation (2.20), the parameters that we are interested in are $\gamma_2, \underline{\beta}_2$ and δ_2 which are non-linear functions of the parameters in $\underline{\beta}_2^*$. Let

$$\hat{\underline{\beta}}_2^* = \begin{bmatrix} \hat{\gamma}_2^{-1} \\ \hat{\underline{\beta}}_2\hat{\gamma}_2^{-1} \\ \hat{\delta}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \underline{\lambda}_2 \\ \lambda_3 \end{bmatrix} .$$

Thus,

$$\hat{\underline{\alpha}}_2 = \begin{bmatrix} \hat{\gamma}_2 \\ \hat{\underline{\beta}}_2 \\ \hat{\delta}_2 \end{bmatrix} = g(\hat{\underline{\beta}}_2^*) = \begin{bmatrix} 1/\lambda_1 \\ \underline{\lambda}_2/\lambda_1 \\ -\lambda_3/\lambda_1 \end{bmatrix} . \qquad (2.29)$$

Consequently,

$$\mathrm{Cov}(\hat{\underline{\alpha}}_2) = \frac{\partial \underline{g}}{\partial \lambda'} \mathrm{Cov}(\hat{\underline{\beta}}_2^*) \frac{\partial \underline{g}'}{\partial \lambda} \qquad (2.30)$$

where

$$\frac{\partial \underline{g}}{\partial \lambda'} = \begin{bmatrix} -1/\lambda_1^2 & 0 & 0 \\ -\underline{\lambda}_2/\lambda_1^2 & \frac{1}{\lambda_1} \cdot \mathbf{I}_{K_2} & 0 \\ -\lambda_3/\lambda_1^2 & 0 & -1/\lambda_1 \end{bmatrix}$$

$$= \begin{bmatrix} -\gamma_2^2 & 0 & 0 \\ -\underline{\beta}_2\gamma_2 & \gamma_2 \mathbf{I}_{K_2} & 0 \\ -\delta_2\gamma_2 & 0 & -\gamma_2 \end{bmatrix}$$

$$= -\gamma_2 \begin{bmatrix} \gamma_2 & 0 & 0 \\ \underline{\beta}_2 & -\mathbf{I}_{K_2} & 0 \\ \delta_2 & 0 & 1 \end{bmatrix} \qquad (2.31)$$

and

$$\mathrm{Cov}(\hat{\underline{\beta}}_2^*) = (\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2'\mathrm{Cov}(\mathbf{w}_2^*)\mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1} \qquad (2.32)$$

Similarly,

$$\mathrm{Cov}(\hat{\underline{\beta}}_1^*) = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathrm{Cov}(\mathbf{w}_1^*)\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1} \qquad (2.33)$$

2.4 ALTERNATIVE ESTIMATORS.

2.4.1 GLS OF TRANSFORMED MODEL.

As mentioned in the introduction of this chapter, we will estimate the parameters of Equation (2.1) using alternatives to Heckman's estimator. The first is a generalized least squares (GLS) approach to Heckman's model

as suggested by Amemiya (1978). The first step is to obtain estimates for the covariances of w_1^* and w_2^* which are expressed in Equations (2.19) and (2.20). Our concern is in the dimension of the covariance of w_1^* and w_2^* which are T×T. For large T, performing inverse operation on $\mathrm{Cov}(w_1^*)$ and $\mathrm{Cov}(w_2^*)$ may create round-off errors and, most of all, it is computationally burdensome.

We can reduce the dimension of the problem by premultiplying both sides of Equations (2.5) and (2.20) by a set of instrumental variables. An obvious choice is the matrix of explanatory variables, X. Therefore, from Equation (2.5) we get

$$\mathbf{X}'\underset{\sim}{y}_1 = \mathbf{X}'\mathbf{Z}_1\underset{\sim}{\beta}_1^* + \mathbf{X}'\underset{\sim}{w}_1^* \quad . \tag{2.34}$$

The GLS estimator of (2.34) is expressed as

$$\hat{\underset{\sim}{\beta}}_{1G}^* = \left\{(\mathbf{X}'\mathbf{Z}_1)'(\mathbf{X}'\mathrm{Cov}(w_1^*)\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_1)\right\}^{-1}\left\{(\mathbf{X}'\mathbf{Z}_1)'(\mathbf{X}'\mathrm{Cov}(w_1^*)\mathbf{X})^{-1}(\mathbf{X}'\underset{\sim}{y}_1)\right\}. \tag{2.35}$$

Note that instead of inverting $\mathrm{Cov}(w_1^*)$ which is of dimension T, we take inverse of $\mathbf{X}'\mathrm{Cov}(w_1^*)\mathbf{X}$ which has dimension K. This way we can tremendously reduce the size of the matrix to be inverted, provide that T>>K.

The covariance of $\underset{\sim}{\beta}_{1G}^*$ is

$$\mathrm{Cov}(\hat{\underset{\sim}{\beta}}_{1G}^*) = \left\{(\mathbf{X}'\mathbf{Z}_1)'(\mathbf{X}'\mathrm{Cov}(w_1^*)\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}_1)\right\}^{-1} \tag{2.36}$$

where

$$\begin{aligned} X' \mathrm{Cov}(w_1^*) X = {} & \sigma_1^2 X'X + X'(\gamma_1 I_T + \delta_1 D_1) X (X'AX)^{-1} X (\gamma_1 I_T + \delta_1 D_1) X \\ & + \delta_1^2 X' D_2 X - 2\sigma_{12} X' (\gamma_1 I_T + \delta_1 D_1) X \\ & + 2\delta_1 \sigma_{12} X' D_1 X \\ & - \delta_1 X' (\gamma_1 I_T + \delta_1 D_1) X (X'AX)^{-1} X' D_1 X \\ & - \delta_1 X' D_1 X (X'AX)^{-1} X' (\gamma_1 I_T + \delta_1 D_1) X \quad . \end{aligned} \tag{2.37}$$

Similarly,

$$\hat{\beta}_{2G}^* = \left\{ (X'Z_2)' (X' \mathrm{Cov}(w_2^*) X)^{-1} (X'Z_2) \right\}^{-1} \left\{ (X'Z_2)' (X' \mathrm{Cov}(w_2^*) X)^{-1} (X'y_1) \right\} \tag{2.38}$$

and

$$\mathrm{Cov}(\hat{\beta}_{2G}^*) = \left\{ (X'Z_2)' (X' \mathrm{Cov}(w_2^*) X)^{-1} (X'Z_2) \right\}^{-1} \quad . \tag{2.39}$$

The expression for $X' \mathrm{Cov}(w_2^*) X$ is similar to Equation (2.35) except that we replace $(\gamma_1 I_T + \delta_1 D_1)$ with $(\gamma_2^{-1} I_T + \delta_1 D_1)$. $\hat{\alpha}_{2G}$ and its covariance are obtained by the procedures described in Equation (2.27) and (2.28) ,respectively.

2.4.2 INSTRUMENTAL VARIABLES/OLS.

Our second alternative estimator is based on the attempt to eliminate the correlation between d and w_1 and w_2 by using X as a matrix of instrumental variables, as suggested by Amemiya(1978). Heckman uses $F(X\hat{\Pi}_2)$ to replace d in order to eliminate the correlation with the disturbance terms. Substituting Equation (2.3) into Equation (2.1a) and rearrange the terms, we get

$$y_1 = \gamma_1 \left[X\Pi_2 + v_2 \right] + X_1 \beta_1 + \delta_1 d + u_1$$

$$= \left[\hat{\mathbf{XH}}, \underline{d}\right] \begin{bmatrix} \gamma_1 \\ \underline{\beta}_1 \\ \delta_1 \end{bmatrix} + \underline{w}_1$$

$$= \mathbf{Z}_1^{\circ} \underline{\beta}_1^{*} + \underline{w}_1 \quad . \tag{2.40}$$

Premultiply both sides by $\mathbf{X}'$, we get

$$\mathbf{X}'\underline{y}_1 = \mathbf{X}'\mathbf{Z}_1^{\circ}\underline{\beta}_1^{*} + \mathbf{X}'\underline{w}_1 \quad . \tag{2.41}$$

Equation (2.41) can be estimated by OLS,

$$\tilde{\underline{\beta}}_1^{*} = \left[(\mathbf{X}'\mathbf{Z}_1^{\circ})'(\mathbf{X}'\mathbf{Z}_1^{\circ})\right]^{-1}\left[(\mathbf{X}'\mathbf{Z}_1^{\circ})'(\mathbf{X}'\underline{y}_1)\right] \tag{2.42}$$

with

$$\mathrm{Cov}(\tilde{\underline{\beta}}_1^{*}) = \left[(\mathbf{X}'\mathbf{Z}_1^{\circ})'(\mathbf{X}'\mathbf{Z}_1^{\circ})\right]^{-1}(\mathbf{X}'\mathbf{Z}_1^{\circ})'\mathbf{X}'\,\mathrm{Cov}(w_1)\,\mathbf{X}(\mathbf{X}'\mathbf{Z}_1^{\circ})\left[(\mathbf{X}'\mathbf{Z}_1^{\circ})'(\mathbf{X}'\mathbf{Z}_1^{\circ})\right]^{-1} \tag{2.43}$$

where

$$\mathbf{X}'\,\mathrm{Cov}(\underline{w}_1)\mathbf{X} = \sigma_1^2\mathbf{X}'\mathbf{X} - 2\gamma_1\sigma_{12}\mathbf{X}'\mathbf{X} + \gamma_1^2(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) \quad .$$

Similarly,

$$\mathbf{X}'\underline{y}_1 = \mathbf{X}'\mathbf{Z}_2^{\circ}\underline{\beta}_2^{*} + \mathbf{X}'\underline{w}_2 \quad , \tag{2.44}$$

$$\tilde{\underline{\beta}}_2^{*} = \left[(\mathbf{X}'\mathbf{Z}_2^{\circ})'(\mathbf{X}'\mathbf{Z}_2^{\circ})\right]^{-1}\left[(\mathbf{X}'\mathbf{Z}_2^{\circ})'(\mathbf{X}'\underline{y}_1)\right] \tag{2.45}$$

and

$$\mathrm{Cov}(\tilde{\underline{\beta}}_2^{*}) = \left[(\mathbf{X}'\mathbf{Z}_2^{\circ})'(\mathbf{X}'\mathbf{Z}_2^{\circ})\right]^{-1}(\mathbf{X}'\mathbf{Z}_2^{\circ})'\mathbf{X}'\,\mathrm{Cov}(w_2)\,\mathbf{X}(\mathbf{X}'\mathbf{Z}_2^{\circ})\left[(\mathbf{X}'\mathbf{Z}_2^{\circ})(\mathbf{X}'\mathbf{Z}_2^{\circ})\right]^{-1} \tag{2.46}$$

where $\mathbf{Z}_2^{\circ} = \left[\mathbf{X}\hat{\mathbf{Q}}, \underline{d}\right]$.

$\mathbf{X}'\mathrm{Cov}(w_2)\mathbf{X}$ is the same as $\mathbf{X}'\mathrm{Cov}(w_1)\mathbf{X}$ except γ_1 is replaced by $1/\gamma_2$. $\tilde{\underline{\alpha}}_2$ and its covariance matrix are obtained by the same procedure that we use to derive $\hat{\underline{\alpha}}_2$.

2.4.3 INSTRUMENTAL VARIABLES/GLS.

Our next alternative is to apply GLS instead of OLS to Equations (2.41) and (2.44). We get

$$\tilde{\beta}^{*}_{1G} = \left[(X\dot{Z}^{\circ}_{1})'(X' \mathrm{Cov}(w_1) X)^{-1}(X\dot{Z}^{\circ}_{1})\right]^{-1}\left[(X\dot{Z}^{\circ}_{1})'(X' \mathrm{Cov}(w_1) X)^{-1}(X\dot{y}_{1})\right] \tag{2.47}$$

with

$$\mathrm{Cov}(\tilde{\beta}^{*}_{1G}) = \left[(X\dot{Z}^{\circ}_{1})'(X' \mathrm{Cov}(w_1) X)^{-1}(X\dot{Z}^{\circ}_{1})\right]^{-1}. \tag{2.48}$$

Similarly,

$$\tilde{\beta}^{*}_{2G} = \left[(X\dot{Z}^{\circ}_{2})'(X' \mathrm{Cov}(w_2) X)^{-1}(X\dot{Z}^{\circ}_{2})\right]^{-1}\left[(X\dot{Z}^{\circ}_{2})'(X' \mathrm{Cov}(w_2) X)^{-1}(X\dot{y}_{1})\right] \tag{2.49}$$

$$\mathrm{Cov}(\tilde{\beta}^{*}_{2G}) = \left[(X\dot{Z}^{\circ}_{2})'(X' \mathrm{Cov}(w_2) X)^{-1}(X\dot{Z}^{\circ}_{2})\right]^{-1}. \tag{2.50}$$

$\tilde{\alpha}_{2G}$ is derived from the same process that we use to get $\hat{\alpha}_2$.

2.4.4 RESTRICTED LEAST SQUARES.

The final alternative we consider is a restricted least squares type estimator. From Equations (2.5) and (2.20), we can see that we have two estimated values of δ_1. When β^{*}_{1} and β^{*}_{2} are estimated separately we can get two different values of δ_1. Therefore, we can use restricted least squares estimator in the estimation of β^{*}_{1} and β^{*}_{2}; the restriction imposed is the two values of δ_1 are the same.

For computational purposes we choose to impose the restriction on Equations (2.41) and (2.44). The model is expressed as,

$$\begin{bmatrix} X'\underline{y}_1 \\ X'\underline{y}_1 \end{bmatrix} = \begin{bmatrix} X'Z_1^{\circ}\underline{\beta}_1^{*} \\ X'Z_2^{\circ}\underline{\beta}_2^{*} \end{bmatrix} + \begin{bmatrix} X'\underline{w}_1 \\ X'\underline{w}_2 \end{bmatrix} \qquad (2.51)$$

or $$\underline{q} = Q\begin{bmatrix} \underline{\beta}_1^{*} \\ \underline{\beta}_2^{*} \end{bmatrix} + \underline{r}$$

where $$Q = \begin{bmatrix} X'Z_1^{\circ} & 0 \\ 0 & X'Z_2^{\circ} \end{bmatrix} .$$

The covariance of the disturbance term r is expressed as

$$Cov(r) = \begin{bmatrix} X'Cov(w_1)X & X'Cov(w_1,w_2)X \\ X'Cov(w_2,w_1)X & X'Cov(w_2)X \end{bmatrix} . \qquad (2.52)$$

$Cov(w_1)$ and $Cov(w_2)$ are expressed in Equation (2.19) and (2.20).

$$\begin{aligned} Cov(w_1,w_2) &= E\left[\underline{w}_1'\underline{w}_2\right] \\ &= E\left[(\underline{v}_1-\gamma_1 X(\hat{\Pi}_2-\Pi_2))(v_1-\frac{1}{\gamma_2}X(\hat{\Pi}_2-\Pi_2))'\right] \\ &= E\left[\underline{v}_1'\underline{v}_1\right] - E\left[\underline{v}_1(\frac{1}{\gamma_2}X(\hat{\Pi}_2-\Pi_2))'\right] \\ &\quad -E\left[\gamma_1 X(\hat{\Pi}_2-\Pi_2)\underline{v}_1'\right] + \frac{\gamma_1}{\gamma_2}\cdot E\left[X(\hat{\Pi}_2-\Pi_2)(\hat{\Pi}_2-\Pi_2)X'\right] \\ &= \sigma_1^2 I_T - \frac{1}{\gamma_2}\sigma_{12}AX(X'AX)^{-1}X' - \gamma_1\sigma_{12}X(X'AX)^{-1}X'A \\ &\quad + \frac{\gamma_1}{\gamma_2}\cdot X(X'AX)^{-1}X' \end{aligned} \qquad (2.53)$$

The only caution for this alternative is $1/\gamma_2$ should not be the same as γ_1, as the Cov(r) will become singular for all of the terms of the covariance matrix in Equation (2.52) are the same. If no restrictions are imposed , the estimation of Equation (2.51) is carried on by GLS.

$$\hat{\underline{B}}^* = \left[\underline{Q}'(\text{Cov}(r))^{-1}\underline{Q}\right]^{-1}\left[\underline{Q}'(\text{Cov}(r))^{-1}\underline{q}\right] \qquad (2.54)$$

where

$$\underline{B}^* = \begin{bmatrix} \underline{\beta}_1^* \\ \underline{\beta}_2^* \end{bmatrix}.$$

The restriction that we want to impose is that δ_1 in $\underline{\beta}_1^*$ has the same value as δ_1 in $\underline{\beta}_2^*$. Let H be a column vector with dimension (K_1+K_2+4). The elements in H have zero values except for the $(K_1+2,1)$ and $(K_1+K_2+4,1)$ positions which have values equal to 1 and -1 , respectively

The estimation of $\underline{B}^*$ with the restriction that $H\underline{B}^*=\underline{0}$ can be expressed as,

$$\hat{\underline{B}}_R^* = \hat{\underline{B}}^* - (\underline{Q}'(\text{Cov}(r))^{-1}\underline{Q})^{-1}H'(H(\underline{Q}'(\text{Cov}(r)\underline{Q})^{-1}H')^{-1}(H\hat{\underline{B}}^*). \qquad (2.55)$$

The covariance of $\hat{\underline{B}}_R^*$ is described as

$$\text{Cov}(\hat{B}_R^*) = \text{Cov}(\hat{B}^*) - \text{Cov}(\hat{B}^*)H'(H\text{Cov}(\hat{B}^*)H')^{-1}H\text{Cov}(\hat{B}^*) \qquad (2.56)$$

where

$$\text{Cov}(\hat{B}^*) = (\underline{Q}'(\text{Cov}(r))^{-1}\underline{Q})^{-1} \quad . \qquad (2.57)$$

In order to find the estimate of $\underline{\alpha}_2$ that corresponds to $\underline{\beta}_2^*$ in $\underline{B}^*$, we simply partition the matrix $\underline{B}_R^*$ in accordance to $\underline{\beta}_{2R}^*$ ($\underline{\beta}_{2R}^*$ is the vector of parameters which is the same as

$\underline{\beta}_2^*$, the subscript R denotes that it is from the restricted least squares) and then use the similar technique that we have been using for other alternatives to transform the estimates of $\underline{\beta}_{2R}^*$ and its covariance into the estimates and covariance of $\underline{\alpha}_{2R}$.

CHAPTER 3

THE MONTE CARLO EXPERIMENT OF THE SIMULTANEOUS EQUATIONS GENERALIZED PROBIT MODEL

3.1 INTRODUCTION.

In this Chapter, we perform Monte Carlo experiments on Heckman's estimation technique and some of its alternatives. With these experiments we can investigate the small sample performance of each of the estimation rules. We are concerned about the small sample properties of these estimation procedures because their small sample variability may not be reflected by their theoretical asymptotic counterparts.

The plan of this Chapter is as follows. In Section 3.2, we discuss the model describing the determinants and effects of state-wide bargaining laws simultaneously. We estimate this model the way of Heckman's estimation technique and its suggested alternatives. In Section 3.3, we explain the concept of a Monte Carlo experiment. We also portray the criteria we used to evaluate the small sample performance of each of the estimation rules. Afterwards, we apply the Monte Carlo experiment to Heckman's estimation technique and its alternatives. Then we report the small sample performance of each of the estimation procedures. Finally, we compute the true asymptotic standard errors for the parameter estimates obtained from each estimation technique and compare them to the finite sample mean square error.

3.2 THE MODEL.

The model that we study is a simultaneous equations system with one of the endogenous variables being a latent variable. The model is expressed as

$$\begin{aligned} PUBUN_t &= \gamma_1 y^*_{2t} + \beta_{11} + \beta_{12} GOVWAGE_t + \beta_{13} PRIVUN_t \\ &\quad + \beta_{14} PROPLAW_t + \beta_{15} EAG_t + \beta_{16} SOU_t \\ &\quad + \delta_1 SENT_t + u_{1t} \end{aligned}$$

$$\begin{aligned} y^*_{2t} &= \gamma_2 PUBUN_t + \beta_{21} + \beta_{22} GOVWAGE_t + \beta_{23} PRIVUN_t \\ &\quad + \beta_{27} CA1_t + \beta_{28} COPEC_t + \beta_{29} LOGMPRTY_t \\ &\quad + \beta_{2,10} NWLF_t + \delta_2 SENT_t + u_{2t} \end{aligned}$$

$$SENT_t = \begin{cases} 1 & y^*_{2t} > 0 \\ 0 & \text{elsewhere} \end{cases} \qquad (3.1)$$

where

GOVWAGE = government employee average salary

PRIVUN = percentage of all employment organized

PROPLAW = proportion of contiguous states possessing Mandatory Bargaining Law (MBL)

EAG = percentage of employment in the agricultural sector

SOU = southern states dummy variable

SENT = MBL dummy variable which is equal to 1 if the state has MBL statue

PUBUN = percentage of the public sector unionized

CA1 = number of unfair labor practices cases charged against an employee

COPEC = fraction of votes by state's delegation to the U.S. House of Representatives consistent with AFL-CIO approved position on issues of interest to organized labor

LOGMPRTY = the natural log of the ratio of the numbers of the legislature's majority party to the total number of legislators

NWLF = percentage of non-white labor force

and u_1 and u_2 are the disturbance terms.

In the model described in (3.1), extent of unionization and the legal environment regulating unionization are jointly determined. We use the data set for the year 1977 and 1982 published in the Census of Government, which includes the 48 contiguous states.

Several studies have attempted to explain the determinants of unionization and the legal environment either in a single equation context or in a simultaneous equations context. Hunt and White (1983) study the determinants of legislative support for public school teacher collective bargaining using the ordered probit method developed by McKelvey and Zavoina (1975). Saltzman (1985) examines the determinants of teacher bargaining coverage and bargaining laws via a single equation approach.

Hunt, Terza, White and Moore (1986) provide a simultaneous framework for studying the model in which

teacher's wage and unionization are the jointly dependent variables. Farber (1988) analyzes the evolution of bargaining laws for police, teachers and state workers using a Markov transition model. Freeman and Valletta (1988) examine the effects of legislative index on collective bargaining, wages and employment in municipalities. Waters (1989) studies the determinations of state-wide bargaining laws via the estimation technique developed by McKelvey and Zavoina (1975). She also studies the determinants and effects of bargaining laws in a simultaneous equations context for school teachers, police ,fire fighters and public employees.

Using the model described in Equation (3.1), we estimate the structural coefficients by applying Heckman's estimation technique and its alternatives. The estimation techniques that we are going to use are Heckman's procedure (HECKMAN), generalized least squares of the transformed model procedure (HECKGLS), an instrumental variable / GLS procedure (AMEMIYA) and the restricted least squares procedure (RLS). We omit the instrumental variable / OLS approach because the estimation procedure is contingent on an incorrect formulation of the covariance matrices.

The instrumental variable / OLS approach estimates the parameters of the first structural equation by the relationship

$$\tilde{\beta}_1^* = \left[(X'Z_1^\circ)'(X'Z_1^\circ)\right]^{-1}\left[(X'Z_1^\circ)'(X'y_1)\right] \quad .$$

This is equivalent to assuming that the term $X'Cov(\underline{w}_1)X$ is an identity matrix which is not appropriate. Nevertheless, this will be consistent.

The estimated coefficients of the model (3.1) and their standard errors are reported in Tables 3.1 and 3.2. HECKGLS, AMEMIYA and RLS estimation procedures require the knowledge of the covariance matrices $Cov(w_1^*)$, $Cov(w_2^*)$, $Cov(w_1)$ and $Cov(w_2)$; consequently, we have to estimate their components (the parameters σ_{12}, σ_1^2, γ_1, γ_2 and δ_1) in advance via Heckman's estimation technique; we call these estimates 'starting values'.

Recalling Equations (2.25) and (2.28)

$$\mathbf{E}\left[\underline{y}_1 | \mathbf{X}, \underline{d}\right] = \mathbf{X}\underline{\Pi}_1 + \delta_1 \underline{d} + \sigma_{12}\left[\hat{\lambda}\underline{d} + \hat{\lambda}^*(1-\underline{d})\right] + \hat{\nu}_1 \qquad (3.2)$$

$$\hat{\sigma}_1^2 = \frac{1}{T}\sum_t \hat{\eta}_t^2 + \hat{\sigma}_{12}^2\left[1 - \frac{1}{T}\sum_t (d_t q_t + (1-d_t)s_t\right] \qquad (3.3)$$

We estimate σ_{12} by applying least squares to Equation (3.2). The traditional practice is to estimate the parameters σ_{12} and σ_1^2 before the estimation of the structural parameters. As a consequence, both $\hat{\sigma}_{12}$ and $\hat{\sigma}_1^2$ are not dependent on either the starting values or the estimation rules employed in estimating the structural parameters.

In Table 3.1, we use the estimates from the HECKMAN procedure as the starting values. Afterwards in Table 3.2, we use the estimates in Table 3.1 as the starting values in

TABLE 3.1

The coefficient estimates of the selected estimation process.

First structural equation

(PUBUN)

	HECKMAN	HECKGLS	AMEMIYA	RLS
Y_2^*	-2.3830	6.5070	7.0723	5.5448
(γ_1)	(4.1091)	(7.5084)	(11.3249)	(5.0937)
ONE	39.2069	54.7112	57.8739	49.2666
(β_{11})	(7.8998)	(14.0054)	(21.6366)	(9.2366)
GOVWAGE	-11.4664	-11.4505	-12.4694	-11.0605
(β_{12})	(2.0541)	(2.2354)	(4.2063)	(1.7742)
PRIVUN	0.2170	0.3417	0.3914	0.5008
(β_{13})	(0.0974)	(0.1562)	(0.2535)	(0.1075)
PROPLAW	14.9908	22.4227	24.3429	24.6209
(β_{14})	(2.6561)	(6.2382)	(11.5163)	(4.8765)
EAG	-0.8328	-0.7371	-0.7371	-0.3871
(β_{15})	(0.1975)	(0.2509)	(0.3546)	(0.0949)
SOU	-7.0743	-8.2399	-8.0428	-5.9107
(β_{16})	(2.1308)	(2.4350)	(4.3218)	(1.6643)
SENT	13.3007	-29.1647	-36.7625	-32.5057
(δ_1)	(11.5765)	(32.9590)	(51.8465)	(23.6721)

TABLE 3.1(continue)

The coefficient estimates of the selected estimation process.

	Second structural equation (SENT)			
	HECKMAN	HECKGLS	AMEMIYA	RLS
PUBUN	0.2348	0.0281	0.0136	0.0614
(γ_2)	(0.1561)	(0.9855)	(0.0201)	(0.0198)
ONE	16.8075	1.8092	-0.2897	5.7772
(β_{21})	(47.6064)	(24.4659)	(8.1073)	(3.6245)
GOVWAGE	3.0002	0.4580	0.3234	0.9030
(β_{22})	(1.2822)	(1.6509)	(0.3607)	(0.3211)
PRIVUN	-0.0813	-0.0106	-0.0043	-0.0452
(β_{23})	(0.1643)	(0.0957)	(0.0238)	(0.0128)
CA1	-0.2878	-0.0587	-0.0411	-0.0960
(β_{27})	(0.2096)	(0.1835)	(0.0452)	(0.0242)
COPEC	0.0260	0.0025	-0.0014	0.0193
(β_{28})	(0.0269)	(0.0300)	(0.0081)	(0.0033)
LOGMPRTY	-6.5865	-1.1605	-0.5208	-2.3829
(β_{29})	(12.2396)	(6.8097)	(2.1201)	(0.8632)
NWLF	0.1425	0.0206	0.0051	0.0277
($\beta_{2,10}$)	(0.1986)	(0.1222)	(0.0309)	(0.0056)
SENT	-2.3558	3.2636	3.4669	1.9961
(δ_2)	(2.1972)	(3.0126)	(0.6589)	(0.7857)
$-\gamma_2\delta_1$	-3.1230	0.8195	0.5000	1.9958

Note: The values in parentheses are the asymptotic standard errors.

TABLE 3.2

The coefficient estimates of the selected estimation process (with iterations).

First structural equation

(PUBUN)

	HECKGLS	AMEMIYA	RLS
Y_2^*	5.5669	7.0737	7.8264
(γ_1)	(10.7382)	(11.3708)	(11.0188)
ONE	53.0755	57.8862	54.9272
(β_{11})	(22.1892)	(21.7167)	(22.3512)
GOVWAGE	-10.2917	-12.4828	-12.3316
(β_{13})	(3.1039)	(4.2300)	(4.4677)
PRIVUN	0.3877	0.3914	0.5074
(β_{14})	(0.2488)	(0.2545)	(0.2212)
PROPLAW	20.8582	24.3803	29.0297
(β_{15})	(8.3838)	(11.5845)	(10.8975)
EAG	-0.6673	-0.7162	-0.4313
(β_{16})	(0.3805)	(0.3559)	(0.2538)
SOU	-8.6485	-8.0402	-5.8078
(δ_1)	(3.8307)	(4.3433)	(2.7132)
SENT	-30.6036	-36.7735	-42.9409
	(54.4235)	(52.0497)	(50.4754)

TABLE 3.2 (continue)

The coefficient estimates of the selected estimation process (with iterations).

	Second structural equation (SENT)		
	HECKGLS	AMEMIYA	RLS
PUBUN	0.0159	0.0129	0.0445
(γ_2)	(0.0328)	(0.0346)	(0.0251)
ONE	-1.2999	-0.4123	10.0331
(β_{21})	(14.7979)	(15.3682)	(8.6381)
GOVWAGE	0.3374	0.3154	0.9627
(β_{22})	(0.6426)	(0.6597)	(0.6002)
PRIVUN	0.0002	-0.0038	-0.0382
(β_{23})	(0.0426)	(0.0439)	(0.0235)
CA1	-0.0418	-0.0402	-0.1119
(β_{27})	(0.0804)	(0.0838)	(0.0630)
COPEC	0.0005	0.0013	0.0195
(β_{28})	(0.0150)	(0.0152)	(0.0092)
LOGMPRTY	-0.3252	-0.4847	-3.3281
(β_{29})	(3.8427)	(4.0050)	(2.2609)
NWLF	0.0065	0.0044	0.0404
($\beta_{2,10}$)	(0.0540)	(0.0571)	(0.0220)
SENT	3.5256	3.4775	1.9093
(δ_2)	(1.2036)	(1.2042)	(1.1761)
$-\gamma_2\delta_1$	0.4866	0.4744	1.9109

correspondence with the estimation techniques used in Table 3.2. The resulting estimates are then used as the new starting values. We repeat this process until the maximum value of the absolute values of the difference between the previous starting values and the estimates of the structural parameters is less than 1.0E-04. Let $\underset{\sim}{\beta}_{1s}$ and $\underset{\sim}{\alpha}_{2s}$ denote the vectors of the previous starting values, then the stopping criterion can be expressed as

$$\max \begin{pmatrix} \underset{\sim}{\beta}_{1s} - \hat{\underset{\sim}{\beta}}_1^* \\ \underset{\sim}{\alpha}_{2s} - \hat{\underset{\sim}{\alpha}}_2 \end{pmatrix} < 1.0E\text{-}04 \qquad (3.4)$$

where $\hat{\underset{\sim}{\beta}}_1^*$ and $\hat{\underset{\sim}{\alpha}}_2$ are the vectors containing the parameters estimates of the structural parameters. The maximum number of iterations permitted is 20.

From Table 3.1, we see that different estimation methods yield vastly different estimates of the same parameter. We concentrate on four key parameters; namely, γ_1, γ_2, δ_1 and δ_2. The RLS procedure is the only procedure which gives the estimate $\hat{\delta}_2$ that conforms with the logical consistent requirement, $\delta_2 = -\gamma_2\delta_1$. Moreover, HECKMAN procedure yields the only negative estimate for the parameter γ_1 which is not obtained by other estimation procedures. RLS procedure provides statistically significant estimates for the parameters δ_2, γ_2 and δ_1. AMEMIYA and HECKMAN procedures give statistically significant estimates for the parameters δ_2 and γ_1, respectively.

From Table 3.2, we observe minimal changes in the parameter estimates via AMEMIYA procedure when the iterative

routine is introduced. Similar to Table 3.1, RLS gives the estimate of the parameter δ_2 which complies with the logical consistent requirements. Furthermore, we find that the estimated variability of the estimates of the first structural equation obtained from HECKGLS and RLS procedures increases noticeably. All three estimation techniques in Table 3.2 yield statistically significant estimates for the parameter δ_2. The RLS procedure also gives a statistically significant estimate for the parameter γ_2. None of the iterative methods take more than ten iterations before the stopping criterion, Equation (3.4), is met.

From Tables 3.1 and 3.2, we observe that different parameter estimates and measures of variability are obtained by utilizing different estimation methods. It is not possible to choose the appropriate estimation technique based on the information presented in Tables 3.1 and 3.2. As a consequence, we use Monte Carlo experiments to examine the small sample properties of each of the techniques.

3.3 MONTE CARLO EXPERIMENT.

3.3.1 MONTE CARLO SAMPLES.

A Monte Carlo experiment is a simulation exercise designed to investigate the small sample properties of estimators. In this experiment, we assume that we know the exact nature of the relationships between the endogenous variables and the explanatory variables. Consider the simultaneous equation system in Equation (2.1)

$$\underline{y}_1 = \gamma_1 \underline{y}_2^* + \mathbf{X}_1 \underline{\beta}_2 + \delta_1 \underline{d} + \underline{u}_1 \tag{3.5}$$

$$\underline{y}_2^* = \gamma_2 \underline{y}_1 + \mathbf{X}_2 \underline{\beta}_2 + \delta_2 \underline{d} + \underline{u}_2 \tag{3.6}$$

Suppose that we know the values of the structural parameters γ_1, γ_2, $\underline{\beta}_1$, $\underline{\beta}_2$, δ_1 and δ_2, we can solve for the reduced form parameters as expressed in Equations (2.2) and (2.3)

$$\underline{y}_1 = \mathbf{X}\underline{\Pi}_1 + \delta_1 \underline{d} + \underline{v}_1 \tag{3.7}$$

$$\underline{y}_2^* = \mathbf{X}\underline{\Pi}_2 + \underline{v}_2 \tag{3.8}$$

where $\underline{v}_1$ and $\underline{v}_2$ are normally distributed vectors of disturbance terms with mean vector $\underline{0}$ and covariance matrix Σ,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}$$

The reduced form parameters $\underline{\Pi}_1$ and $\underline{\Pi}_2$ are defined by the relationship

$$\Pi = -\mathbf{B}\Gamma^{-1} \tag{3.9}$$

where

$$\Pi = [\underline{\Pi}_1, \underline{\Pi}_2] \quad ,$$

$$\Gamma = \begin{bmatrix} -1 & \gamma_2 \\ \gamma_1 & -1 \end{bmatrix}$$

and

$$\mathbf{B}' = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} & 0 & 0 & 0 & 0 \\ \beta_{21} & \beta_{22} & \beta_{23} & 0 & 0 & 0 & \beta_{27} & \beta_{28} & \beta_{29} & \beta_{2,10} \end{bmatrix}$$

If we know the values of σ_1^2 and σ_{12}, we can derive the endogenous variables y_1 and y_2^* where the variable d is obtained by

$$d_t = \begin{cases} 1 & y_{2t}^* > 0 \\ 0 & \text{otherwise} \end{cases} .$$

Using a normal random number generator, we construct N samples of (T×2) matrix **V**, $\mathbf{V} = [\underline{v}_1, \underline{v}_2]$, which are normally distributed with mean vectors $\underline{0}$ and covariance matrices Σ. Let **W** be a (T×2) matrix whose elements are generated from a N(0,1) random number generator and let $\Sigma^{1/2}$ be a square matrix such that

$$\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}.$$

We construct a matrix of disturbace terms **V** with mean vectors $\underline{0}$ and covariance matrix Σ by the relationship

$$\mathbf{V} = \mathbf{W} \cdot \Sigma^{1/2}$$

We utilize these N matrices of disturbance terms to produce N samples of $\underline{d}$ and $\underline{y}_1$. We name each of the samples of $\underline{d}$ and $\underline{y}_1$ a 'Monte Carlo' sample.

Let $\hat{\beta}_{1,1}^{*j}$ be an estimator of the i-th element of the parameter vector $\underline{\beta}_1^*$ where the super-script j denotes that the estimator is applied to the j-th Monte Carlo sample, j = 1,2,...,N. We evaluate the small sample performance of $\hat{\beta}_{1,1}^*$ by its biasedness, variability and risk (MSE).

1. Biasedness.

We observe the biasedness of an estimate from the difference between the actual parameter value and the average estimated value obtained from our Monte Carlo experiments. The bias of a parameter is defined as

$$BIAS_{1,i} = \frac{\sum_j \hat{\beta}^{*j}_{1,i}}{N} - \beta^*_{1,i} \quad . \tag{3.10}$$

2. Variability.

We measure the variability of the estimator $\hat{\underline{\beta}}^{*j}_{1,i}$ by its standard deviation which is defined as

$$SD(\hat{\beta}^*_{1,i}) = \left\{ \frac{\sum_j (\hat{\beta}^{*j}_{1,i} - \bar{\beta}^*_{1,i})^2}{N} \right\}^{1/2} \tag{3.11}$$

where $\bar{\beta}^*_{1,i} = \sum_j \beta^{*j}_{1,i} / N$.

3. Risk(MSE).

We compute two types of risk. First, we estimate the risk for the individual estimator $\hat{\beta}^*_{1,i}$. Second, we calculate the overall risk of applying the estimation technique to the model (model error).

The average risk (MSE) of the estimator $\hat{\beta}^*_{1,i}$ is defined as

$$MSE_{1,i} = \frac{\sum_j (\hat{\beta}^{*j}_{1,i} - \beta^*_{1,i})^2}{N} \quad . \tag{3.12}$$

Let $\hat{\underline{\beta}}^*_1$ and $\hat{\underline{\alpha}}_2$ be estimators of the parameter vectors $\underline{\beta}^*_1$ and $\underline{\alpha}_2$, respectively. $\underline{\beta}^*_1$ is the vector of the structural parameters of the first equation and $\underline{\alpha}_2$ is the vector of the

parameters of interest of the second structural equation. Furthermore, let $\hat{\underline{\beta}}_1^{*j}$ and $\hat{\underline{\alpha}}_2^{j}$ denote that we apply the estimation rule to the j-th Monte Carlo sample. We then define the model error risk for the first and second structural equation as

$$RISK_1 = \sum_{j=1}^{N} (\hat{\underline{\beta}}_1^{*j} - \underline{\beta}_1^{*})'(\hat{\underline{\beta}}_1^{*j} - \underline{\beta}_1)/N \qquad (3.13)$$

and

$$RISK_2 = \sum_{j=1}^{N} (\hat{\underline{\alpha}}_2^{j} - \underline{\alpha}_2)'(\hat{\underline{\alpha}}_2^{j} - \underline{\alpha}_2)/N \qquad (3.14)$$

respectively.

3.3.2 MONTE CARLO EXPERIMENTS RESULTS.

In this section, we report the results from the Monte Carlo experiments. As discussed in the preceding section, we assign the parameter values and generate the data in the experiments. The parameter values selected to be the actual parameter values are those of the Heckman's estimation technique in Table 3.1. We choose these values for the purpose of defined asymptotic standard errors which we discuss later on. Nevertheless, we have to calculate a new estimate for the parameter δ_2 to ensure that the consistency requirement $\delta_2 = -\delta_1\gamma_2$ is met.

Assigning the actual parameter values is not as simple as it appears to be. We discovered that some values of the parameters are not usable in the experiments. By being not usable, we mean that some parameter values lead to negative estimated values for the asymptotic variances for some of

the parameters, which is an extreme undesirable property. We find that the parameter values of the HECKMAN procedure must be scaled down so that they are usable for all of the estimation techniques. We divide the parameter estimates of the second structural equation by 8, which is the smallest value that eradicates the problem of negative estimated asymptotic variances. However, we also have used the square root of the variable PROPLAW in place of its original value to reduce its variation. This solves the problem of its approximated asymptotic variance being negative. The true parameter values used in the Monte Carlo experiments are presented in the first column of Table 3.3.

We have also experimented using the estimates obtained from other estimation techniques as the actual parameter values. The estimates for the techniques which incorporate generalized least squares have one feature in common. Refering to Section 3.3.1, we use the structural parmaters to derived the reduced form parameters. The reduced form parameter vectors Π_2 that are obtained from the techniques which incorporate generalized least squares result in the product $X\Pi_2$ being less than zero for all observations. Keeping this feature in mind and bringing to mind how the variable d is generated in the Monte Carlo experiments, we realize that by using the parameter estimates from other techniques besides Heckman's, the generated d variable will be very likely to take on the values of zero which create very little scattering of the data generated. Moreover, as

we shall see in the next section, when $X\Pi_2$ is less than zero for every observation, the asymptotic standard errors of the techniques using the instrumental variables approach are not defined.

For the covariance matrix used in the Monte Carlo experiments, the values of σ_1^2 and σ_{12} are the estimates obtained via applying Equations (3.2) and (3.3) to the original data set. In order to be consistent with the scaling of the parameters of the second structural equation, the estimate of σ_{12} is divided by the same constant, 8. Consequently, the covariance matrix used is

$$\begin{bmatrix} 51.5406 & -0.6336 \\ -0.6336 & 1 \end{bmatrix} .$$

The constant 1 is not changed for $\sigma_2^2 = 1$ due to identification condition (see Section 2.2).

While we were performing Monte Carlo experiments, we came across the problem of unusable estimated parameter values often. In other words, many Monte Carlo samples give negative estimated variances for at least one of the parameter estimates. Thus, we eliminated such samples and generated replacements. We find that far more than a thousand Monte Carlo samples must be generated to obtain a thousand sets of parameter estimates of which all asymptotic variances are positive. The total number of Monte Carlo samples generated for the HECKMAN, HECKGLS, AMEMIYA and RLS procedures are 1646, 2004, 2476 and 4235, respectively. The

difference in the number of total Monte Carlo samples generated suggests that each procedures can accept different combinations of the parameter estimates and the estimates of the covariance matrix components.

As described in Section 2.4.4, the RLS procedure is not defined whenever $\gamma_1 = 1/\gamma_2$. By utilizing the term $Cov(w_1,w_2)$ as one of the elements in the RLS procedure, when $\gamma_1 = 1/\gamma_2$ the terms $Cov(w_1)$, $Cov(w_2)$ and $Cov(w_1,w_2)$ are all identical and hence causes the covariance of the vector of disturbance terms $\underline{r}$, Equation (2.52) to become singular. Some Monte Carlo samples yield the estimates of γ_1 and γ_2 which are nearly identical and makes the RLS procedure undefined. Once we encounter such a sample, we drop that particular sample and generate its replacement. The problem of undefined RLS procedure is not a serious one; of the 4235 Monte Carlo samples generated for the RLS procedure, only 8 samples cause this problem.

As outlined earlier, we use a Monte Carlo experiment for each of the estimation techniques to study their small sample properties. We obtain a thousand sets of estimates for the structural parameters for each of the estimation techniques and present their average values together with the true parameter values in Table 3.3. The last row of Table 3.3 is the average values of the negative of the product between the estimates of δ_1 and γ_2 in order to test how strongly the logical consistency requirement is implemented.

TABLE 3.3

Average values of the parameter estimates obtained through Monte Carlo experiment.

First structural equation

(PUBUN)

	ACTUAL	HECKMAN	HECKGLS	AMEMIYA	RLS
Y_2^*	-2.3830	6.2237	16.0281	-6.6288	-31.4698
ONE	39.2069	48.8574	65.4825	30.5729	-16.5942
GOVWAGE	-11.4664	-13.0767	-10.5118	-11.0677	-6.3286
PRIVUN	0.2170	0.2305	0.1365	0.2249	0.3759
PROPLAW	14.9908	14.3617	12.9156	13.8236	7.0950
EAG	-0.8328	-0.7823	-0.8784	-0.7885	-0.3796
SOU	-7.0743	-6.7742	-7.9864	-6.6523	-3.7791
SENT	13.3007	-7.4779	-34.2523	29.6112	106.5000

Second structural equation

(SENT)

PUBUN	0.0294	0.1253	-0.0030	-0.0053	-0.0194
ONE	2.1009	-16.8970	-2.4216	-1.9342	-1.9567
GOVWAGE	0.3750	0.2226	-0.0858	-0.0398	-0.3577
PRIVUN	-0.0102	-0.0387	0.0021	0.0027	0.0069
CA1	-0.0360	-0.0323	0.0069	-0.00001	0.0382
COPEC	0.0033	-0.0327	0.0005	-0.00007	-0.0033
LOGMPRTY	-0.8233	2.5657	0.2310	0.1242	0.2706
NWLF	0.0178	-0.0019	-0.0039	-0.0005	-0.0139
SENT	-0.3904	5.9432	3.1670	3.2365	4.0383
$-\gamma_2\delta_1$		10.8299	-4.0703	0.1139	4.0372

For the first structural equation, the AMEMIYA and RLS procedures yield the correct signs for both of the key parameters γ_1 and δ_1, on the average. However, almost all estimation techniques, with the exception of the RLS procedure, give the average estimated values of other parameters in the first structural equation besides γ_1 and δ_1 which closely resemble the true parameter values.

For the second structural equation, none of the estimation techniques being considered yield satisfying estimates of the structural parameters. None of the techniques give the correct signs for the average values of the estimates of γ_2 and δ_2. Furthermore, the average estimated values for the structural parameters do not closely approximate the true parameter values. Nevertheless, the RLS procedure still guarantees the logical consistency requirements as indicated by the term $-\delta_1\gamma_2$.

In Table 3.4, we report the bias of the estimates along with the calculated standard errors and mean square errors obtained from the Monte Carlo experiments. The traditional HECKMAN procedure gives the smallest standard errors and mean square errors for the estimates of all parameters in the first structural equation. However, the HECKMAN procedure does not produce estimates with the lowest bias for all estimates.

On the contrary, the estimates of the HECKGLS procedure have the smallest total mean square error in the second structural equation. But not all parameter estimates of the

TABLE 3.4

Bias, standard errors and mean square errors of the parameter estimates obtained through Monte Carlo experiments.

First structural equation

(PUBUN)

	HECKMAN	HECKGLS	AMEMIYA	RLS
Y_2^*	8.6067	18.4111	-4.2458	-29.0868
	(44.2635)	(820.7)	(172.200)	(240.600)
	2033.30	**6.7E05**	**3.0E04**	**5.9E04**
ONE	9.6505	26.2756	-8.6350	-55.8016
	(58.3161)	(1147.3)	(233.500)	(335.200)
	3493.90	**1.3E06**	**5.5E04**	**1.2E05**
GOVWAGE	0.3897	0.9546	0.3987	5.1378
	(2.8662)	(21.7914)	(5.8266)	(7.2360)
	8.3669	**475.800**	**34.1078**	**78.7570**
PRIVUN	0.0135	-0.0805	0.0079	0.1589
	(0.1333)	(2.1326)	(0.2742)	(0.3803)
	0.0179	**4.5544**	**0.0753**	**0.1698**
PROPLAW	-0.6291	-2.0752	-1.1672	-7.8958
	(4.0087)	(51.1505)	(9.4199)	(15.3988)
	16.4654	**2620.70**	**90.0973**	**299.50**
EAG	0.0505	-0.0456	0.0443	0.4532
	(0.2666)	(1.8826)	(0.5960)	(0.8348)
	0.0736	**3.5462**	**0.3572**	**0.9023**
SOU	0.2996	-0.9121	0.4220	3.3952
	(2.7454)	(16.9835)	(8.1891)	(6.7119)
	7.6268	**289.300**	**67.2400**	**55.9075**
SENT	-20.7786	-47.5530	16.3105	93.1687
	(116.700)	(2394.7)	(469.900)	(667.600)
	1.4E04	**5.7E06**	**2.2E05**	**4.5E05**
Total MSE	2.67E5	7.73E6	3.06E5	6.29E5

The values in parentheses are the standard errors and the values in bold are the mean square errors.

TABLE 3.4(continue)

Bias, standard errors and mean square errors of the parameter estimates obtained through Monte Carlo experiments.

Second structural equation (SENT)

	HECKMAN	HECKGLS	AMEMIYA	RLS
PUBUN	0.0959	-0.0323	0.0347	-0.0487
	(2.9997)	(0.0975)	(0.0964)	(0.7017)
	9.0071	**0.0106**	**0.0105**	**0.4947**
ONE	-18.9980	-4.5225	-4.0352	-4.0572
	(490.600)	(14.6193)	(17.6756)	(36.5004)
	2.4E05	**234.200**	**328.700**	**1348.700**
GOVWAGE	-0.1524	-0.4608	-0.4148	-0.7327
	(2.5906)	(2.9405)	(0.8832)	(11.2709)
	6.7345	**8.8592**	**0.9520**	**127.6**
PRIVUN	-0.0285	0.0123	0.0129	0.0171
	(1.0295)	(0.0576)	(0.0655)	(0.3287)
	1.0607	**0.0035**	**0.0080**	**1.9498**
CA1	0.0683	0.0429	0.0360	0.0742
	(1.8425)	(0.2286)	(0.0817)	(1.3944)
	3.3996	**0.0541**	**0.0045**	**0.1083**
COPEC	-0.0360	-0.0028	-0.0033	0.0065
	(0.9631)	(0.0183)	(0.0114)	(0.1540)
	0.9288	**3.4E-4**	**1.4E-4**	**0.0238**
LOGMPRTY	3.3890	1.0543	0.9475	1.0939
	(86.2970)	(3.3222)	(3.9191)	(12.2268)
	7458.7	**12.1483**	**16.2567**	**150.70**
NWLF	-0.0197	-0.0218	-0.0183	-0.0317
	(0.5178)	(0.0779)	(0.0501)	(0.5316)
	0.2685	**0.0065**	**0.0028**	**0.2836**
SENT	6.3336	3.5574	3.6268	4.4287
	(136.700)	(10.0609)	(6.5986)	(44.1104)
	1.9E04	**113.900**	**56.6960**	**1965.3**
Total MSE	1192.4	369.1825	402.6306	3596.1602

HECKGLS procedure have the lowest mean square error; some of the parameters estimates obtained from the AMEMIYA procedure have lower mean square errors than those of the HECKGLS procedure . The AMEMIYA and HECKGLS procedures are both outstanding techniques to be used in estimating the second structural equation compared to the traditional HECKMAN procedure.

To sum up, one must be careful in choosing the appropriate technique for the problem at hand since there is no clear-cut rule. The rule of thumb is that the traditional HECKMAN procedure ought to be used when the attention is on the first structural equation but the AMEMIYA or HECKGLS procedure ought to be used when the attention is on the second structural equation. Therefore the researcher has to weigh the importance of the first structural equation against the second structural equation. Nevertheless, the AMEMIYA procedure is an excellent alternative to the HECKMAN procedure since its total mean square error in the second structural equation is roughly 0.003 times of the mean square error of the HECKMAN procedure. The total mean square error of the AMEMIYA procedure in the first structural equation is 15 times that of the HECKMAN procedure.

If we add up the total mean square errors of the first and second structural equation, the HECKMAN procedure has the lowest overall mean square error. Nevertheless, using the overall mean square error to evaluate the performances of the estimation technique is misleading for the mean

square errors of the first structural equation are overwhelmingly larger than those of the second structural equation simply due the difference in the absolute values of the estimates. Consequently, the technique that best perform in estimating the first structural equation is likely to be chosen regardless of its performance in the second structural equation. Next we study the distributions of the estimates from each of the estimation techniques to give us a more thorough understanding in the characteristics of the estimators. We find the descriptive statistics for the four key parameters, γ_1, γ_2, δ_1 and δ_2 and present them in Tables 3.5 through 3.8 together with the distribution plots.

For the parameters γ_1 and δ_1, only the estimates obtained from the HECKMAN procedure show well formed distributions. The frequency distribution plots of the estimates of γ_1 and δ_1 obtained from the HECKGLS, AMEMIYA and RLS procedures are sketchy at the very least and are scattered over tremendous ranges.

The frequency distribution plots of the estimates of γ_2 and δ_2 obtained from the HECKGLS and AMEMIYA procedures more closely resemble one another than suggested by the descriptive statistics. However, the peak of their frequency distribution plots are not concentrated around the actual parameter values unlike those of the HECKMAN procedure. The disadvantage of the estimates obtained from the HECKMAN procedure are their very large variation. Therefore, there is evidence of trade-off between biasness and variability.

TABLE 3.5

The description of the distribution of the estimates of the parameter γ_1.

γ_1 =-2.3830

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	6.2237	16.0281	-6.6288	-31.4689
std	44.2635	820.7	172.2	240.6
skewness	-1.1725	-0.3302	-4.6104	-7.6885
kurtosis	59.0450	131.7	62.1709	125.0
max	558.9	12673	873.6	1288.5
min	-553.0	-12304	-2348.7	-4359.0
median	4.5604	2.9103	-0.7752	-12.6582

FIGURE 3.1

The frequency distribution of γ_1

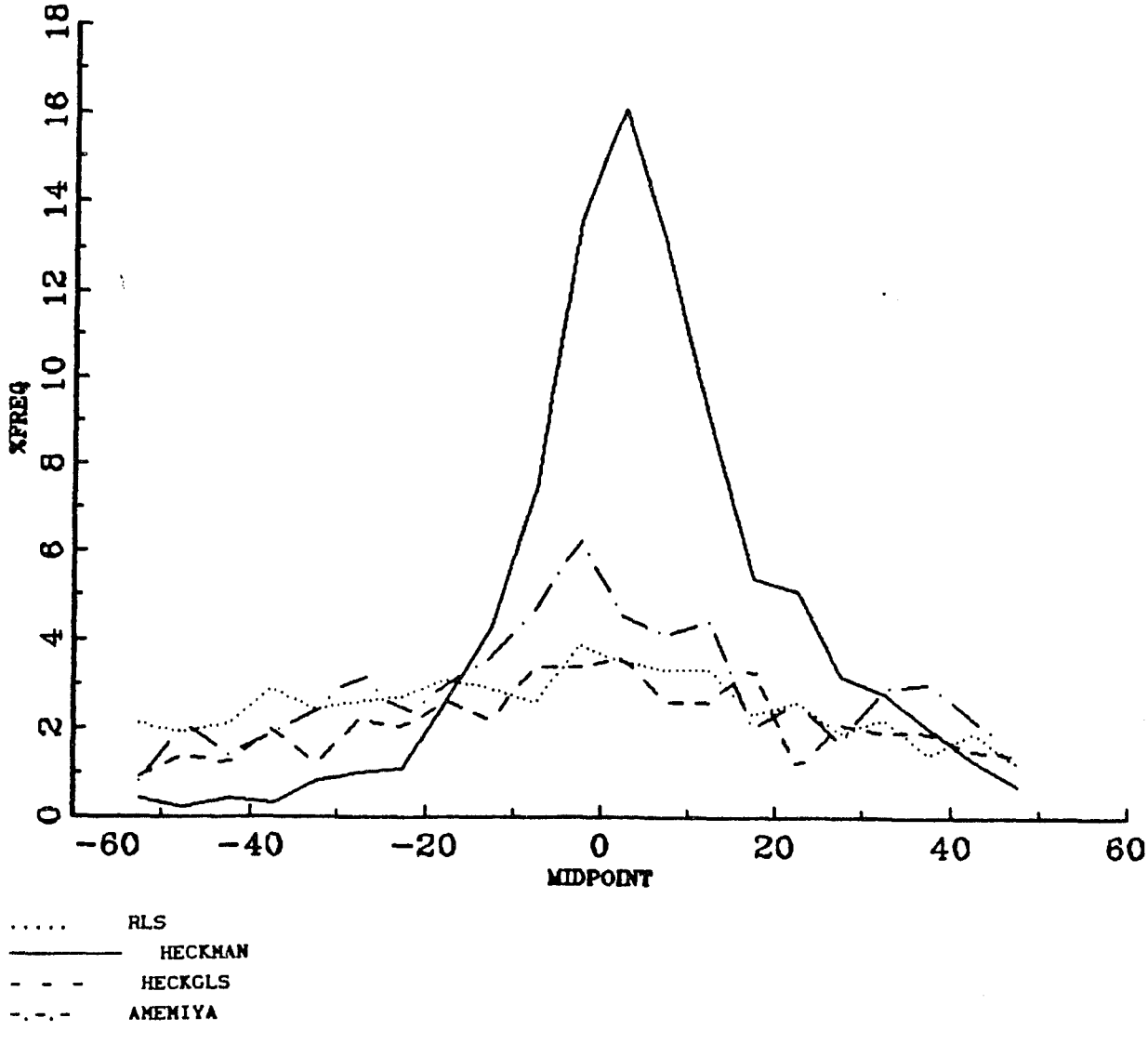


TABLE 3.6

The description of the distribution of the estimates of the parameter δ_1.

$\delta_1 = 13.3007$

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	-7.4779	-34.2523	29.6112	106.5
std	116.7	2394.7	469.9	667.6
skewness	1.0686	1.8480	4.3409	8.0464
kurtosis	56.0987	144.9	55.5316	136.7
max	1414	39380	6469.5	12526
min	-1427	-35394	-2377	-3249.5
median	-2.6165	-2.8487	13.2910	57.7578

FIGURE 3.2

The frequency distribution of δ_1

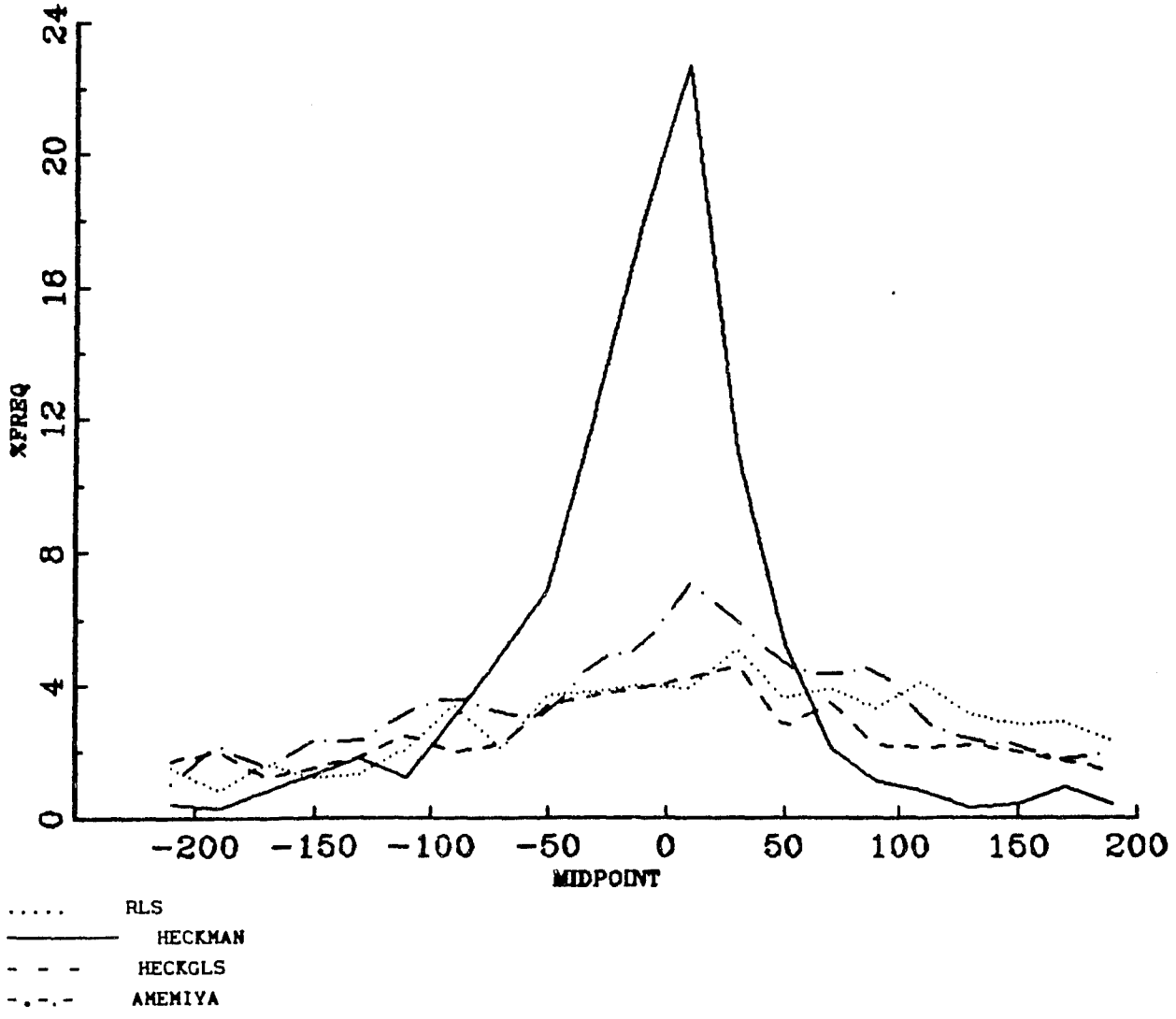


TABLE 3.7

The description of the distribution of the estimates of the parameter γ_2.

$\gamma_2 = 0.0294$

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	0.1253	-0.0030	-0.0053	-0.0194
std	2.9997	0.0975	0.0964	0.7017
skewness	30.6855	-17.3648	-12.5335	-25.0754
kurtosis	963.0	420.7	278.2	744.3
max	93.9723	0.4298	0.9238	4.1869
min	-5.1406	-2.4652	-2.1557	-20.5875
median	0.0319	0.0003	0.0001	-0.0022

FIGURE 3.3

The frequency distribution of γ_2

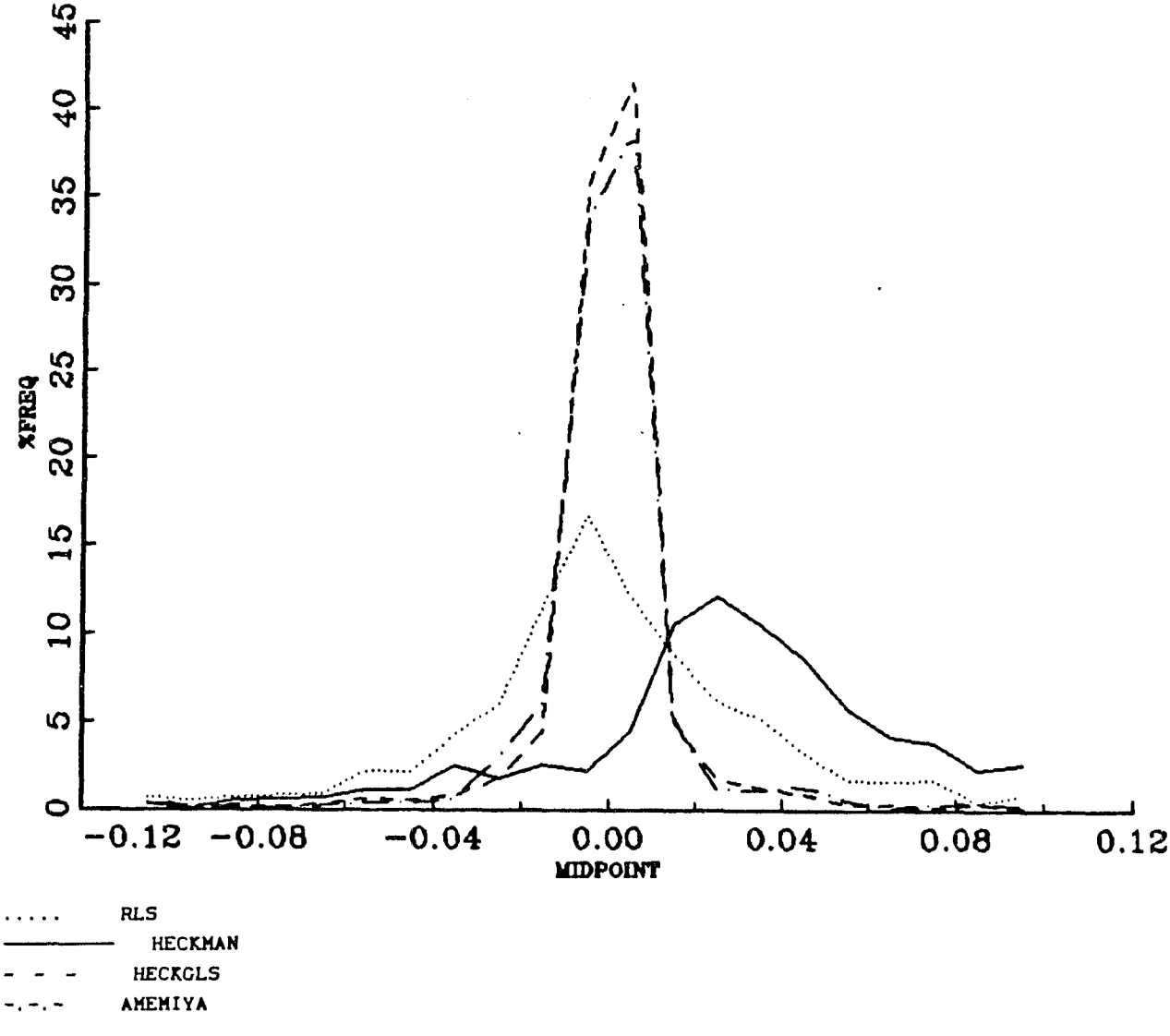


TABLE 3.8

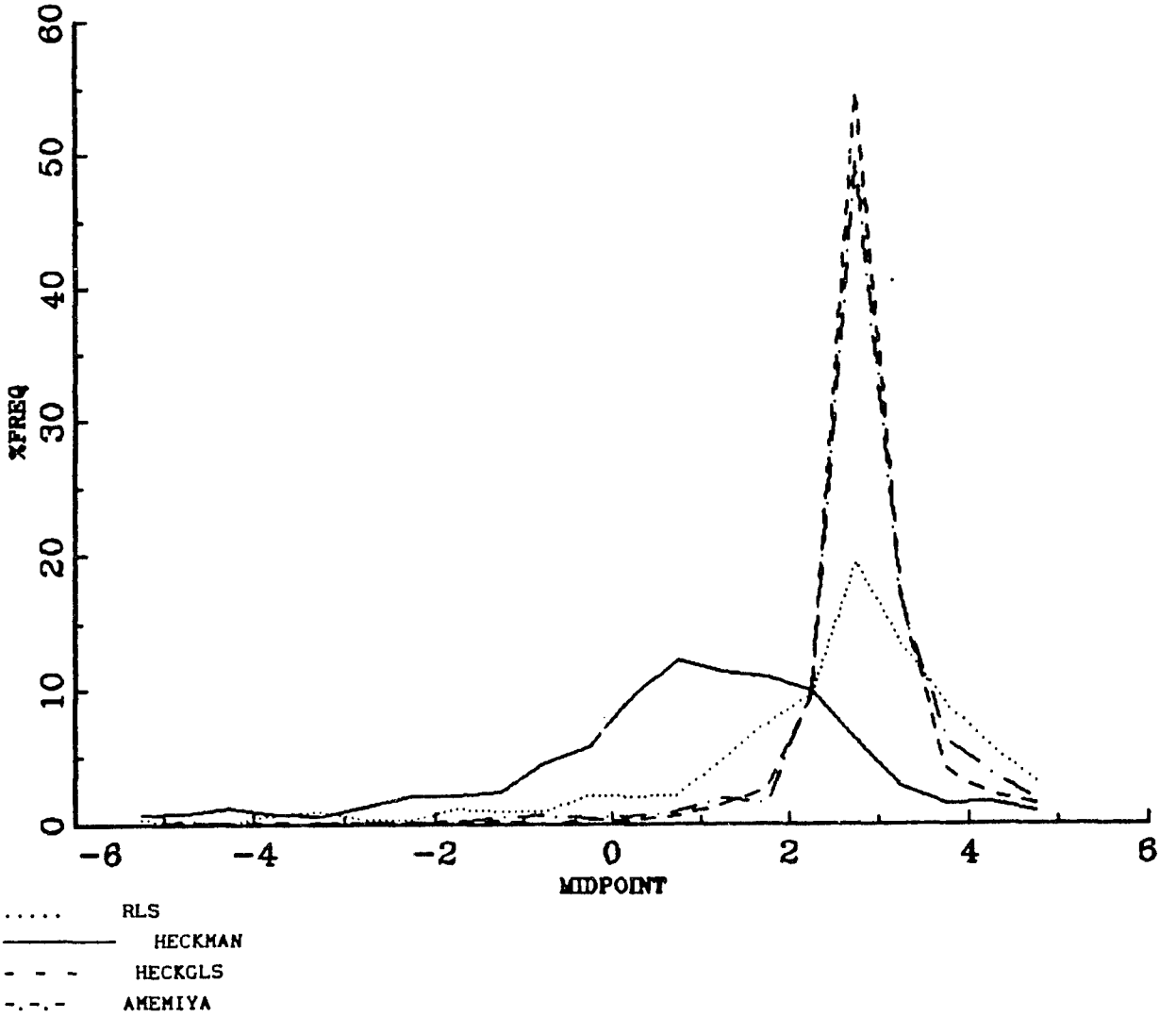
The description of the distribution of the estimates of the parameter δ_2.

δ_2 =-0.3904

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	5.9432	3.1670	3.2365	4.0383
std	136.7	10.0609	6.5986	44.1104
skewness	31.0881	25.0969	19.0264	27.0495
kurtosis	980.0	723.5	475.7	819.1
max	4302	295.9	174.7	1329
min	-109.4	-23.1698	-36.3709	-194.9
median	1.0253	2.7668	2.8138	2.8342

FIGURE 3.4

The frequency distribution of δ_2



3.3.3 THE CHARACTERISTICS OF THE MONTE CARLO SAMPLES USED.

Recall that not all Monte Carlo samples generated can be used in the experiments for some of the samples lead to negative estimated asymptotic variances of the parameter estimates. Each estimation technique needs different groups of Monte Carlo samples to come up with a thousand sets of parameter estimates for which all have positive estimated asymptotic variances. Therefore, the Monte Carlo samples used for each estimation technique contains useful information regarding their characteristics.

Recall that the parameters σ_{12} and σ_1^2 are estimated prior to the estimation of the structural parameters, as a consequence, the estimates of σ_{12} and σ_1^2 are not dependent on the estimation techniques used in estimating the structural parameters. We get the same estimates of σ_{12} and σ_1^2 if the same Monte Carlo samples are used regardless of the estimation techniques used in estimating the structural parameters. Accordingly, the information concerning the characteristics of the generated Monte Carlo samples that yield positive estimates of the asymptotic variances for all structural parameters for each estimation techniques are captured by the estimates of σ_{12} and σ_1^2.

In Tables 3.9 and 3.10, the descriptive statistics of the estimates of σ_{12} and σ_1^2 obtained through performing a Monte Carlo experiment for each of the estimation techniques are presented along with the frequency distribution plots.

TABLE 3.9

The description of the distribution of the estimates of the parameter σ_{12} .

$\sigma_{12} = -0.6636$

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	0.3230	-3.9887	-3.1900	-2.5698
std	23.3971	23.0758	21.4482	18.7341
skewness	0.2486	0.7524	2.0927	-0.3606
kurtosis	9.4950	12.1878	27.1614	10.6162
max	177.4	177.4	230.3	115.1
min	-130.9	-133.2	-99.0670	-113.3
median	0.9879	-3.4266	-2.2313	-1.1599
mse	6.9436	5.8408	5.4693	7.4507

FIGURE 3.5

The frequency distribution of σ_{12}

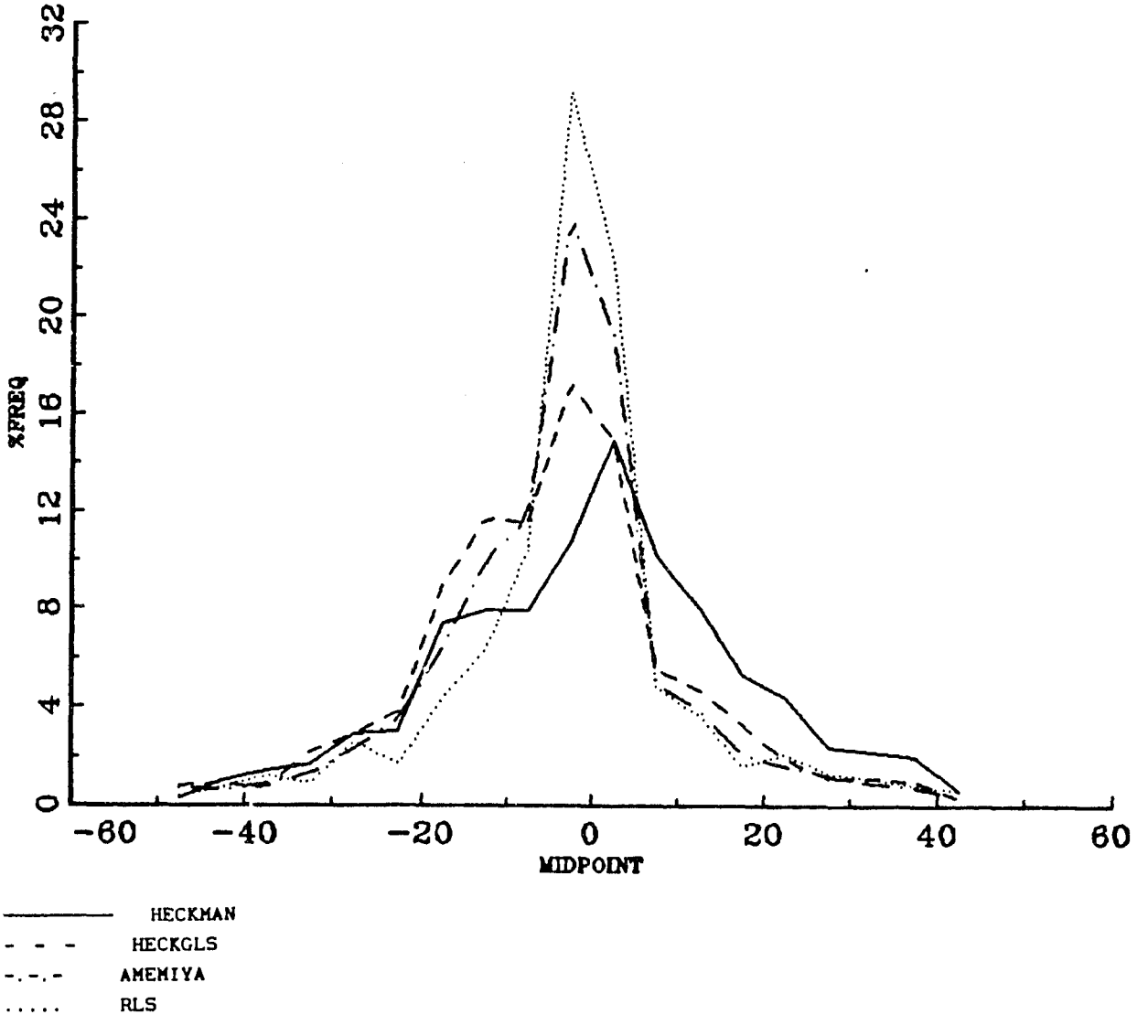


TABLE 3.10

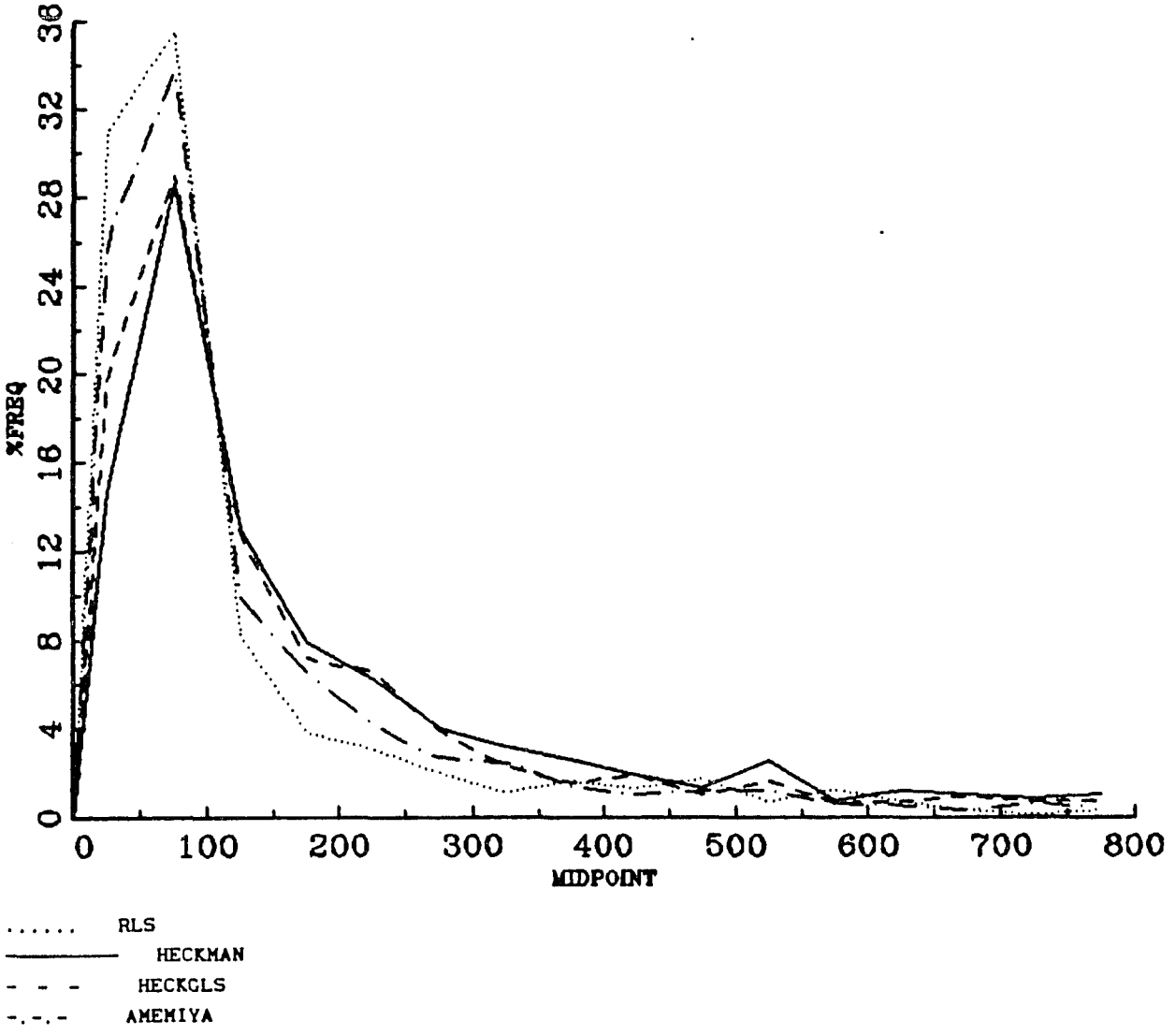
The description of the distribution of the estimates of the parameter σ_1^2.

σ_1^2 =51.5406

	HECKMAN	HECKGLS	AMEMIYA	RLS
mean	370.8	373.9	328.9	259.1
std	984.0	1080.1	1422.8	669.6
skewness	10.7419	9.4935	15.8972	6.5635
kurtosis	175.1	133.6	325.1	58.47
max	19810	19810	33119	8251.3
min	28.97	28.97	31.01	31.80
median	123.8	106.1	73.2940	60.1892
mse	186.5	76.7011	169.7	1466.9

FIGURE 3.6

The frequency distribution of σ_1^2



The estimates of σ_{12} and σ_1^2 obtained via the Monte Carlo samples used in the experiment of the HECKMAN procedure are very dispersed which suggests that the HECKMAN procedure is able to handle a wide range of fluctuation in the data. The HECKGLS procedure, however, appears to perform in the same range data fluctuation as that of the HECKMAN procedure.

The procedures that use the generalized least squares approach show that they are sensitive to the fluctuation in the data set. All the generated Monte Carlo samples that produce positive estimates of the asymptotic variances of the procedures employing the generalized least squares approach give estimates of σ_{12} and σ_1^2 that are concentrated around the actual values. Furthermore, the RLS procedure which imposes an additional restriction in the estimation process yield estimates of σ_{12} and σ_1^2 with great precision. The evidence indicates that the Monte Carlo samples which produce the estimates of σ_{12} and σ_1^2 which do not agree with the actual values are very likely to be rejected by the estimation techniques employing the generalized least squares approach.

3.3.4 THE ASYMPTOTIC STANDARD ERRORS.

In this Section, we address the question concerning the ability of the standard errors obtained from the Monte Carlo experiments in approximating the actual asymptotic standard errors. We mentioned in the introduction that the small

sample variability may not be the same as the theoretical asymptotic variability. We approximate the small sample variability through the standard deviations of the parameter estimates obtained through the Monte Carlo experiments. Then we calculate true asymptotic standard errors by substituting the actual parameter values into the asymptotic covariance matrix equations. The means of calculating the actual asymptotic standard errors for the HECKMAN, HECKGLS, AMEMIYA and RLS procedures are described below.

1.HECKMAN procedure.

Recall that we estimate the parameters of the first structural equation by the relationship

$$\hat{\underline{\beta}}_1^* = (Z_1'Z_1)^{-1}Z_1'\underline{y}_1 \quad . \tag{3.15}$$

Using the relationship in equation (2.5)

$$\underline{y}_1^* = Z_1\underline{\beta}_1^* + \underline{w}_1^* \quad .$$

Consequently, we get

$$\begin{aligned}\hat{\underline{\beta}}_1^* &= (Z_1'Z_1)^{-1}Z_1'(Z_1\underline{\beta}_1^* + \underline{w}_1^*) \\ &= (Z_1'Z_1)^{-1}(Z_1'Z_1)\underline{\beta}_1^* + (Z_1'Z_1)^{-1}Z_1'\underline{w}_1^* \\ &= \underline{\beta}_1^* + (Z_1'Z_1)^{-1}Z_1'\underline{w}_1^* \quad . \end{aligned} \tag{3.16}$$

Next, we want to determine the limiting distribution of the sequence $\sqrt{T}\cdot(\hat{\underline{\beta}}_1^* - \underline{\beta}_1^*)$ as $T\to\infty$. We have

$$\begin{aligned}\sqrt{T}\cdot(\hat{\underline{\beta}}_1^* - \underline{\beta}_1^*) &= \sqrt{T}\cdot(Z_1'Z_1)^{-1}Z_1'\underline{w}_1^* \\ &= (Z_1'Z_1/T)^{-1}Z_1'\underline{w}_1^*/\sqrt{T} \quad . \end{aligned} \tag{3.17}$$

The asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\underline{\beta}}_1^* - \underline{\beta})$ is

$$\text{plim}\left[\frac{Z_1'Z_1}{T}\right]^{-1}\frac{Z_1'}{\sqrt{T}}\ \text{Cov}(w_1^*)\ \frac{Z_1}{\sqrt{T}}\ \left[\frac{Z_1'Z_1}{T}\right]^{-1}$$

Now

$$\begin{aligned}
&\text{plim}\ \tfrac{1}{T}\ (Z_1'Z_1) \\
&= \text{plim}\ \tfrac{1}{T}\ \left[\ X\hat{\Pi}_2\ ,\ X_1\ ,\ \hat{F}\right]'\left[\ \cdot\ \right] \\
&= \text{plim}\ \tfrac{1}{T}\begin{bmatrix} \hat{\Pi}_2'X'X\hat{\Pi}_2 & \hat{\Pi}_2X'X_1 & \hat{\Pi}_2X' \\ X_1'X\hat{\Pi}_2 & X_1'X_1 & X_1'\hat{F} \\ \hat{F}X\hat{\Pi}_2 & \hat{F}'X_1 & \hat{F}'\hat{F} \end{bmatrix} \qquad . \qquad (3.18)
\end{aligned}$$

We substitute the following expression for the expression in equation as an approximation (3.18)

$$\tfrac{1}{T}\begin{bmatrix} \Pi_2'X'X\Pi_2 & \Pi_2X'X_1 & \Pi_2X' \\ X_1'X\Pi_2 & X_1'X_1 & X_1'F \\ FX\Pi_2 & F'X_1 & F'F \end{bmatrix}$$

Similarly, we use $(1/\sqrt{T})\left[X\Pi_2,\ X_1\ ,\ F\right]$ as the proxy for $\text{plim}(Z_1'/\sqrt{T})$. With the expression for $\text{Cov}(w_1^*)$ described in Equation(2.19), we are now able to calculate the asymptotic standard errors for the parameters of the first structural equation by substituting in the actual parameter values. The calculated true asymptotic covariance matrix of the first structural equation is expressed as

$$\text{Cov}(\hat{\underline{\beta}}_1^*) = \left\{(Z_1'Z_1)^{-1}Z_1'\right\}\cdot\text{Cov}(w_1^*)\cdot\left\{(Z_1'Z_1)^{-1}Z_1'\right\}'$$

with the term $\left[X\Pi_2,\ X_1\ ,F\ \right]$ replacing the matrix Z where $F_t = F(x_t'\Pi_2)$ and F is the cumulative distribution function of the normal distribution. The standard errors for the second structural equation are also acquired in the same manner.

2.HECKGLS procedure.

Recalling Equation(2.35), the estimates for the parameters of the first structural equation are obtained by the relationship

$$\hat{\beta}^{*}_{1G} = \left[(X'Z_1)'(X'\mathrm{Cov}(w_1^{*})X)^{-1}(X'Z_1)\right]^{-1}\left[(X'Z_1)'(X'\mathrm{Cov}(w_1^{*})X)^{-1}X'y_1\right] .$$

$$= \beta^{*}_1 + S^{-1}\left[(X'Z_1)'(X'\mathrm{Cov}(w_1^{*})X)(X'w_1^{*})\right] \qquad (3.19)$$

where $S = (X'Z_1)'(X'\mathrm{Cov}(w_1^{*})X)^{-1}(X'Z_1)$.

Similar to the case of HECKMAN procedure, we want to find the limiting distribution for the sequence $\sqrt{T}(\hat{\beta}^{*}_{1G} - \beta^{*}_1)$ as $T\to\infty$. We have

$$\sqrt{T}(\hat{\beta}^{*}_{1G} - \beta^{*}_1) = V^{-1}\left[(X'Z_1/\sqrt{T})'(X'\mathrm{Cov}(w_1^{*})X)^{-1}X'w_1^{*}\right]$$

where

$$V = \left[(X'Z_1/\sqrt{T})'(X'\mathrm{Cov}(w_1^{*})X)^{-1}(X'Z_1/\sqrt{T})\right]$$

The asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\beta}^{*}_{1G} - \beta^{*}_1)$ is described as

$$\mathrm{plim}\ V^{-1}\left\{X'Z_1/\sqrt{T})'(X'\mathrm{Cov}(w_1^{*})X)^{-1}X'w_1^{*}\right\}\cdot\left\{\quad\right\}'V^{-1}$$

$$= \mathrm{plim}\ V^{-1} \qquad . \qquad (3.20)$$

Again, the term plim $(X'Z_1/\sqrt{T})$ is substituted by

$$\frac{X'}{\sqrt{T}}\left[X\Pi_2,\ X_1,\ F\right]$$

The expression for the asymptotic covariance of HECKGLS procedure is described in Equation (2.36) as

$$\mathrm{Cov}(\hat{\beta}^{*}_{1G}) = \left\{(X'Z_1)'(X'\mathrm{Cov}(w_1^{*})X)^{-1}(X'Z_1)\right\}^{-1} \quad .$$

As we did for the HECKMAN procedure, we use the true parameter values in the expression for $Cov(w_1^*)$ in Equation (3.20) and the term $\left[X\Pi_2,\ X_1\ ,\ F\right]$ replacing the matrix Z_1 to obtain the true asymptotic standard errors for the parameters of the first structural equation. We find the asymptotic standard errors for the second structural equation in the same manner.

3. AMEMIYA procedure.

The instrumental variable/GLS or AMEMIYA procedure estimates the parameters of the first structural equation by the relationship

$$\tilde{\beta}_{1G}^{*} = \left[(X'Z_1^{\circ})'(X'Cov(w_1)X)^{-1}(X'Z_1^{\circ})\right]^{-1}\left[(X'Z_1^{\circ})'(X'Cov(w_1)X)^{-1}(X'\underline{y}_1)\right]$$

$$= \beta_1^{*} + W^{-1}\left[(X'Z_1^{\circ})'(X'Cov(w_1)X)^{-1}(X'\underline{w}_1)\right] \qquad (3.21)$$

where $W = \left[(X'Z_1^{\circ})'(X'Cov(w_1)X)^{-1}(X'Z_1^{\circ})\right]$.

Now we have

$$\sqrt{T}(\tilde{\beta}_{1G}^{*} - \beta_1^{*}) = V^{-1}\left[(X'Z_1^{\circ}/\sqrt{T})'(X'Cov(w_1)X)^{-1}(X'Z_1^{\circ}/\sqrt{T})\right]$$

where $V = \left[(X'Z_1^{\circ}/\sqrt{T})'(X'Cov(w_1)X)^{-1}(X'Z_1^{\circ}/\sqrt{T})\right]$.

We write the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\beta}_{1G}^{*} - \beta_1^{*})$ as

$$plim V^{-1}(X'Z_1^{\circ}/\sqrt{T})'(X'Cov(w_1)X)^{-1}X'\underline{w}_1\underline{w}_1'X(X'Cov(w_1)X)^{-1}(X'Z_1^{\circ}/\sqrt{T})V^{-1}$$

$$= plim V^{-1}.$$

The matrix Z_1° is described as

$$Z_1^{\circ} = \left[X\hat{H}\ ,\ d\right]$$

$$= \left[X\hat{\Pi}_2,\ X_1\ ,\ d\right]$$

Therefore,

$$X'Z_1^{\circ}/\sqrt{T} = \frac{1}{T}^{1/2}\left[X'X\hat{\Pi}_2,\ X'X_1\ ,\ X'd\right]$$

We use the following expression instead of plim $X'Z_1^{\circ}/\sqrt{T}$ as an approximate

$$\frac{1}{T}^{1/2}\left[X'X\Pi_2,\ X'X_1\ ,X'd\right]$$

where $d_t = \begin{cases} 0 & \text{if } x_t'\Pi_2<0 \\ 1 & \text{if } x_t'\Pi_2\geq 0 \end{cases}$.

By substituting in the actual parameter values in the expression for $\mathrm{Cov}(w_1)$ in Equation (2.48) together with the approximate of plim$X'Z_i^{\circ}/\sqrt{T}$, we obtain the asymptotic variance-covariance matrix of the estimates for the parameters of the first structural equation which is expressed as

$$\mathrm{Cov}(\tilde{\beta}_{1G}^{*}) = \left[(X'Z_1^{\circ})'(X'\mathrm{Cov}(w_1)X)^{-1}(X'Z_1^{\circ})\right]^{-1}$$

The asymptotic covariance matrix for the estimates of the parameters of the second structural equation are acquired in similar manner.

4. RLS procedure.

The restricted least squares procedure estimates the parameters of both structural equations and imposes the restriction that the estimates of δ_1 in both structural equations are the same, simultaneously (Section 2.4.4).

The covariance matrix of the restricted least squares estimator is expressed as

$$\mathrm{Cov}(\hat{\mathbf{B}}_R^*) = \mathrm{Cov}(\hat{\mathbf{B}}^*) - \mathrm{Cov}(\hat{\mathbf{B}}^*)\mathbf{H}'(\mathbf{H}\mathrm{Cov}(\hat{\mathbf{B}}^*)\mathbf{H}')^{-1}\mathbf{H}\mathrm{Cov}(\hat{\mathbf{B}}^*) \qquad (3.22)$$

where

$$\mathrm{Cov}(\hat{\mathbf{B}}^*) = \left[\mathbf{Q}'(\mathrm{Cov}(\mathbf{r}))^{-1}\mathbf{Q}\right]^{-1} \qquad (3.23)$$

and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{X}'\mathbf{Z}_1^\circ & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'\mathbf{Z}_2^\circ \end{bmatrix},$$

$$\mathrm{Cov}(\mathbf{r}) = \begin{bmatrix} \mathbf{X}'\mathrm{Cov}(\mathbf{w}_1)\mathbf{X} & \mathbf{X}'\mathrm{Cov}(\mathbf{w}_1,\mathbf{w}_2)\mathbf{X} \\ \cdot & \mathbf{X}'\mathrm{Cov}(\mathbf{w}_2)\mathbf{X} \end{bmatrix}.$$

We have

$$\hat{\mathbf{B}}^* = \left[\mathbf{Q}'(\mathrm{Cov}(\mathbf{r}))^{-1}\mathbf{Q}\right]^{-1}\left[\mathbf{Q}'(\mathrm{Cov}(\mathbf{r}))^{-1}\underline{q}\right]$$

where

$$\underline{q} = \mathbf{Q}\mathbf{B}^* + \underline{r}$$

$$\begin{bmatrix} \mathbf{X}'\underline{y}_1 \\ \mathbf{X}'\underline{y}_1 \end{bmatrix} = \mathbf{Q}'\begin{bmatrix} \underline{\beta}_1^* \\ \underline{\beta}_2^* \end{bmatrix} + \begin{bmatrix} \mathbf{X}'\underline{w}_1 \\ \mathbf{X}'\underline{w}_2 \end{bmatrix}.$$

Following the same procedure as other estimation procedures, we can show that

$$\sqrt{T}(\hat{\mathbf{B}}^* - \mathbf{B}^*) = \left[\mathbf{Q}'/\sqrt{T}(\mathrm{Cov}(\mathbf{r}))^{-1}\mathbf{Q}/\sqrt{T}\right]^{-1}\left[\mathbf{Q}'/\sqrt{T}(\mathrm{Cov}(\mathbf{r}))^{-1}\underline{r}\right]$$

and the asymptotic variance-covariance matrix for $\sqrt{T}(\hat{\mathbf{B}}^* - \mathbf{B})$ is described as

$$\mathrm{plim}\left[\mathbf{Q}'/\sqrt{T}(\mathrm{Cov}(\mathbf{r}))^{-1}\mathbf{Q}/\sqrt{T}\right]^{-1}.$$

As for the previous procedures, we approximate for the term plim $Q/\sqrt{T}$ in which we use $\left[X\Pi_2,\ X_1,\ d\right]$ and $\left[X\Pi_2,\ -X_2,\ d\right]$ to replace Z_1° and Z_2° in the calculation of the true asymptotic covariance matrix, respectively. Note that

$$d_t = \begin{cases} 0 \text{ if } x_t'\Pi_2 < 0 \\ 1 \text{ otherwise} \end{cases} .$$

It is simple to show that the asymptotic covarince matrix of $\hat{B}_R^*$ is the same as the expression in Equation (3.22). We calculate the true covariance matrices of the first and second structural equations all together by substituting the true parameter values into Equation (3.22).

In Table 3.11, we present the actual asymptotic standard errors, the standard errors obtained through the Monte Carlo experiments and their percentage differences. Let A denote actual asymptotic standard error of a parameter estimate and s denote standard error obtained through Monte Carlo experiments, the percentage difference between A and S is defined as $[(A-S)/S]\times 100$.

The standard errors of the HECKMAN procedure obtained via the Monte Carlo experiments underestimate the asymptotic standard errors for all parameters in the first structural equation except for the parameter associated with the variable PROPLAW; in contrast, the asymptotic standard errors of the second structural equation are overestimated except for the parameter associated with the variable GOVWAGE. The standard errors obtained through the Monte Carlo experiments of the HECKGLS procedure underestimate the asymptotic standard errors of the estimates of the

TABLE 3.11

The calculated actual asymptotic standard errors.

First structural equation
(PUBUN)

	HECKMAN	HECKGLS	AMEMIYA	RLS
Y_2^*	160.9000	1613.30	70.5308	28.3506
	(44.2635)	(820.7)	(172.200)	(240.600)
	-72.9	**-49.1**	**144**	**749**
ONE	211.3000	2118.70	28.6507	12.0963
	(58.3161)	(1147.3)	(233.500)	(335.200)
	-72.4	**-45.9**	**714**	**2671**
GOVWAGE	2.8781	3.4568	5.4986	2.4331
	(2.8662)	(21.7914)	(5.8266)	(7.2360)
	-0.4	**530.4**	**6.0**	**197**
PRIVUN	0.1361	0.1436	0.2572	0.1134
	(0.1333)	(2.1326)	(0.2742)	(0.3803)
	-2.1	**1385**	**6.6**	**235**
PROPLAW	1.7385	13.7361	11.1073	4.9587
	(4.0087)	(51.1505)	(9.4199)	(15.3988)
	130.6	**272.0**	**15.2**	**210**
EAG	0.2700	0.4513	0.6941	0.3129
	(0.2666)	(1.8826)	(0.5960)	(0.8348)
	-1.3	**317.0**	**-14.1**	**167**
SOU	2.8105	9.5070	9.0700	3.9415
	(2.7454)	(16.9835)	(8.1891)	(6.7119)
	-2.3	**78.6**	**-9.7**	**70.3**
SENT	422.2000	4236.30	60.1810	24.6451
	(116.700)	(2394.7)	(469.900)	(667.600)
	-72.3	**-43.5**	**680.0**	**2609**

The values in parentheses are the standard errors obtained from the Monte Carlo experiments and the values typed in bold are the percentage difference between the estimated standard errors and the true asymptotic values.

TABLE 3.11(continue)
The calculated actual asymptotic standard errors.

Second structural equation
(SENT)

	HECKMAN	HECKGLS	AMEMIYA	RLS
PUBUN	0.2340	2.5017	0.6506	0.0325
	(2.9997)	(0.0975)	(0.0964)	(0.7017)
	1811	**-96.1**	**-85.2**	**2059**
ONE	41.8935	295.0	61.2016	4.5566
	(490.600)	(14.6193)	(17.6756)	(36.5004)
	1071	**-95.0**	**-71.1**	**701**
GOVWAGE	2.8328	31.8027	8.4539	0.4373
	(2.5906)	(2.9405)	(0.8832)	(11.2709)
	-8.6	**-90.8**	**-89.7**	**2477**
PRIVUN	0.1226	0.8912	0.2302	0.0166
	(1.0295)	(0.0576)	(0.0655)	(0.3287)
	739.7	**-93.5**	**-71.6**	**1880**
CA1	0.3019	3.0348	0.6995	0.0307
	(1.8425)	(0.2286)	(0.0817)	(1.3944)
	510.3	**-92.5**	**-88.3**	**4442**
COPEC	0.0312	0.2797	0.0582	0.0034
	(0.9631)	(0.0183)	(0.0114)	(0.1540)
	2986	**-93.5**	**-80.4**	**4429**
LOGMPRTY	9.9664	71.4107	20.0309	1.2070
	(86.2970)	(3.3222)	(3.9191)	(12.2268)
	765.9	**-95.4**	**-80.4**	**912.9**
NWLF	0.1835	1.5171	0.3364	0.0193
	(0.5178)	(0.0779)	(0.0501)	(0.5316)
	182.2	**-94.9**	**-85.1**	**175.4**
SENT	23.8048	253.80	23.4299	1.1321
	(136.700)	(10.0609)	(6.5986)	(44.1104)
	474.3	**-96.0**	**-71.8**	**3796**

parameters γ_1, δ_1 and the intercept term in the first structural equation while they underestimate all of the asymptotic standard errors in the second structural equation. The standard errors obtained through the Monte Carlo experiments for the AMEMIYA procedure overestimate the asymptotic standard errors of parameters associated with the variables EAG and SOU while underestimate asymptotic standard errors of the others parameters in the first structural equation. Similar to the case for the HECKGLS procedure, the standard errors obtained via the Monte Carlo experiments for the second structural equation underestimate the asymptotic standard errors of the second structural equation. Finally, the standard errors obtained through the Monte Carlo experiments for the RLS procedure overestimate the asymptotic standard errors for all parameters in both structural equations.

Several remarks can be made based on Table 3.11. First, the standard errors obtained from the Monte Carlo experiments of the HECKMAN procedure give the best estimates of the asymptotic standard errors in the first structural equation while those of the AMEMIYA procedure give the best estimates of the asymptotic standard errors in the second structural equation. Second, by using the generalized least squares approach with the HECKMAN procedure, we increase the asymptotic standard errors in both of the structural equations. Third, the asymptotic standard errors from the AMEMIYA procedure are less than those of the HECKGLS

procedure for most of the parameter estimates in both structural equations. Lastly, the RLS procedure yields the smallest asymptotic standard errors for the estimates in the second structural equation even though this is not reflected via the Monte Carlo experiments.

From Table 3.11, the Monte Carlo experiments show that the small sample properties of the estimation techniques, namely the HECKMAN, HECKGLS, AMEMIYA and RLS procedure, are quite dissimilar to their asymptotic theoretical counterparts. A plausible explanation is the fact that not all Monte Carlo samples can be used in the experiments which may lead to the failure of capturing all aspects of the characteristics of each estimation technique.

3.3.5 AN ALTERNATIVE ESTIMATION TECHNIQUE OF THE COVARIANCE MATRIX.

Recall that we estimate the parameter σ_{12} by applying least squares to Equation (3.2) which is expressed as

$$E\left[\underset{\sim}{y}_1, | X, \underset{\sim}{d}\right] = X\underset{\sim}{\Pi}_1 + \delta_1 \underset{\sim}{d} + \sigma_{12}\left[\hat{\lambda}\underset{\sim}{d} + \hat{\lambda}^*(1 - \underset{\sim}{d})\right] + \hat{\underset{\sim}{v}}_1$$

where

$$\lambda_t = f(x'_{t-2}\Pi_2)/F(x'_{t-2}\Pi_2) \text{ and}$$

$$\lambda_t^* = -\lambda_t F(x'_{t-2}\Pi_2)/F(-x'_{t-2}\Pi_2).$$

We calculate the values of $\hat{\lambda}_t$ and $\hat{\lambda}_t^*$ by substituting in the estimate of $\underset{\sim}{\Pi}_2$ obtained from the probit estimation process, Equation (2.3). An argument can be made that the estimate of σ_{12} obtained this way may not be effcient since we do not

utilize the estimated structural parameters. In this section, we introduce an alternative approach to the traditional practice of estimating the components of the covariance matrix and use the Monte Carlo experiments to evaluate its performance.

Referring to Section 3.3.1, we show how to calculate the reduced form parameters given the structural parameters. Consequently, by using the estimated structural parameters we can derive the reduced form parameters which can be used in the estimation of the components of the covariance matrix. The benefit of this approach is that additional information concerning the structural parameters are permitted into the estimation process. Consider the equation

$$E\left[\underline{y}_1 | \mathbf{X}, \underline{d}\right] = \mathbf{X}\underline{\Pi}_1 + \delta_1 \underline{d} + \sigma_{12}\left[\hat{\lambda}\underline{d} + \hat{\lambda}^*(1 - \underline{d})\right] + \hat{\underline{v}}_1$$

Let denote the derived reduced form parameters as

$$\hat{\Pi} = [\hat{\underline{\Pi}}_{1D}, \hat{\underline{\Pi}}_{2D}] \quad .$$

Furthermore, let $\hat{\delta}_1$ denote the estimate of δ_1. Substituting $\hat{\delta}_1$ and $\hat{\underline{\Pi}}_2$ into Equation (3.2), we get

$$\begin{aligned} E\left[\underline{y}_1 | \mathbf{X}, \underline{d}\right] &= \mathbf{X}\hat{\underline{\Pi}}_{1D} + \mathbf{X}(\underline{\Pi}_1 - \hat{\underline{\Pi}}_{1D}) + \hat{\delta}_1 \underline{d} + (\delta_1 - \hat{\delta}_1)\underline{d} \\ &\quad + \sigma_{12}\left[\hat{\lambda}\underline{d} + \hat{\lambda}^*(1 - \underline{d})\right] + \hat{\underline{v}}_1 \\ &= \mathbf{X}\hat{\underline{\Pi}}_{1D} + \hat{\delta}_1 \underline{d} + \sigma_{12}\left[\hat{\lambda}\underline{d} + \hat{\lambda}^*(1 - \underline{d})\right] \\ &\quad + \left[\mathbf{X}(\underline{\Pi}_1 - \hat{\underline{\Pi}}_{1D}) + (\delta_1 - \hat{\delta}_1)\underline{d} + \hat{\underline{v}}_1\right]. \end{aligned}$$

By rearranging the known values to the left hand side we get

$$E\left[\underline{y}_1 | \mathbf{X}, \underline{d}\right] - \mathbf{X}\hat{\underline{\Pi}}_{1D} - \hat{\delta}_1 \underline{d} = \sigma_{12}\left[\hat{\lambda}\underline{d} + \hat{\lambda}^*(1 - \underline{d})\right] + \underline{\eta} \qquad (3.24)$$

where

$$\underline{\eta} = \mathbf{X}(\underline{\Pi}_1 - \hat{\underline{\Pi}}_{1D}) + (\delta_1 - \hat{\delta}_1)\underline{d} + \hat{\underline{v}}_1 ,$$

$$\hat{\lambda}_t = f(x'_{t-2D}\hat{\Pi}_{2D})/F(x'_{t-2D}\hat{\Pi}_{2D}) \text{ and}$$

$$\hat{\lambda}^*_t = -\hat{\lambda}_t F(x'_{t-2D}\hat{\Pi}_{2D})/F(-x'_{t-2D}\hat{\Pi}_{2D}) .$$

The alternative estimation technique of the parameter σ_{12} is the method that applies least sqaures to Equation (3.24). Let $\hat{\underline{\eta}}$ be the vector of residuals obtained from the application of least squares to Equation (3.24). We estimate the parameter σ_1^2 from the relationship

$$\hat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t^2 + \hat{\sigma}_{12}^2\left[1 - \frac{1}{T}\sum_{t=1}^{T}(d_t q_t + (1 - d_t)s_t\right] \quad .$$

We estimate q_t and s_t by

$$\hat{q}_t = 1 + (-x'_{t-2D}\hat{\Pi}_{2D})\hat{\lambda}_t - \hat{\lambda}_t^2,$$

and

$$\hat{s}_t = 1 + (-x'_{t-2D}\hat{\Pi}_{2D})\hat{\lambda}^*_t - \hat{\lambda}^{*2}_t,$$

respectively.

In Table 3.12, we compare the estimates of standard errors, and the estimates of the parameters σ_{12} and σ_1^2 obtained from the traditional approach to those obtained from the derived reduced form parameters approach using the original data. Both techniques give nearly identical estimates for the parameter σ_1^2 and slightly different estimates for the parameter σ_{12}. The standard errors obtained from the two estimation techniques are quite similar except for the variable PROPLAW, as a result, we

TABLE 3.12

The standard errors of the parameter estimates and the estimates of the components of the covaraince matrix.

	Parameter estimates Covariance matrix traditional	derived reduced form parameters
σ_{12}	-5.0689	-3.9177
σ_1^2	51.0689	51.7824
	Standard errors	
	First structural equation	
	traditional	derived reduced form parameters
Y_2^*	4.1091	4.0618
ONE	7.8998	8.0240
GOVWAGE	2.0541	2.1934
PRIVUN	0.0974	0.1027
PROPLAW	2.6561	0.8909
EAG	0.1975	0.2133
SOU	2.1308	2.2343
SENT	11.5765	8.9890
	Second structural equation	
PUBUN	0.1561	0.1640
ONE	47.6064	46.2756
GOVWAGE	1.2822	1.3283
PRIVUN	0.1643	0.1644
CA1	0.2096	0.2255
COPEC	0.0269	0.0284
LOGMPRTY	12.2396	12.1893
NWLF	0.1986	0.2037
SENT	2.1972	1.8327

find no major changes in level of statistical significance of the parameter estimates.

In Table 3.13, we present the descriptive statistics of the estimated values of σ_{12} and σ_1^2 obtained from the Monte Carlo experiments via the traditional approach and the derived reduced form parameters approach, both approaches are applied to the traditional HECKMAN procedure. Therefore, the estimates of σ_{12} and σ_1^2 obtained from the Monte Carlo experiment via the traditional approach are exactly identical to those obtained in the study of small sample performances of the HECKMAN procedure.

Once again, in the process of performing Monte Carlo experiment on the derived reduced form parameters approach, when we find the Monte Carlo sample that has negative estimates of the asymptotic variances, we disregard that particular sample and generate its replacement. We repeat this process until we get a thousand estimates of σ_{12} and σ_1^2. However, we discovered that some Monte Carlo samples cause the right hand side of Equation (3.24) to become zero, hence rendering least squares inapplicable. We also delete such samples from the Monte Carlo experiment. All together, we have to generate 1498 Monte Carlo samples in the experiment. Note that the number of Monte Carlo samples generated is not the same as that of the traditional approach which is equal to 1646. Thus, the two approaches of estimating the covariance matrix are able to use different sets of samples.

TABLE 3.13

The descriptive statistics of the estimates of σ_{12} and σ_1^2 obtained from the Monte Carlo experiments of the traditional and the derived reduced form parameters approach.

$\sigma_{12} = -0.6336$

	traditional	derived reduced form parameters
mean	0.3230	-74.7188
std	23.3971	2487
skewness	0.2486	-31.5951
kurtosis	9.4950	1001
max	177.4	305.2
min	-130.9	-78600
median	0.9876	4.5295
mse	547.9	6.18E6

$\sigma_1^2 = 51.5406$

	traditional	derived reduced form parameters
mean	370.8	2.25E4
std	981.0	2.27E5
skewness	10.7419	18.5008
kurtosis	175.1	380.3
max	19810	5.17E6
min	28.97	31.7216
median	186.5	282.7
mse	1.06E6	5.20E10

From Table 3.14, we observe that the estimates obtained from using the derived reduced form parameters approach are by no means more accurate than those obtained from those obtained from the traditional approach. However, by using the covariance estimates obtained from the derived reduced form parameters less Monte Carlo samples are rejected which may lead to better estimates of the structural parameters.

In Table 3.15, we present the structural parameter estimates and their standard errors obtained from applying the HECKMAN procedure and its alternative on the original data using the derived reduced form parameters approach in estimating σ_{12} and σ_1^2. The structural parameters used in deriving the reduced form parameter are from the HECKMAN procedure. The results in Table 3.15 are quite similar to those in Table 3.1 where the components of the covariance matrix are obtained via the traditional approach.

The Monte Carlo experiments using the estimates of σ_{12} and σ_1^2 obtained via the derived reduced form parameter approach in the process are also performed on the HECKGLS, AMEMIYA and RLS procedure; once more, the derived reduced form parameters are estimated by using the structural parameter estimates obtained through the HECKMAN procedure. We discover that a great deal of generated Monte Carlo samples have to be omitted since they yield negative estimates of the parameter variances. Moreover, the mean square errors of all parameter estimates are increased by

TABLE 3.14

Bias, standard errors and mean square errors of the parameter estimates obtained through Monte Carlo experiments.

	First structural equation (PUBUN)	
	traditional	derived reduced form parameters
Y_2^*	8.6067	4.3714
	(44.2635)	(27.8230)
	2033.30	**793.20**
ONE	9.6505	4.1659
	(58.3161)	(37.7928)
	3493.30	**1445.60**
GOVWAGE	0.3897	0.1980
	(2.8662)	(3.0157)
	8.3669	**9.1335**
PRIVUN	0.0135	0.0067
	(0.1333)	(0.1393)
	0.0179	**0.0195**
PROPLAW	-0.6291	-0.4164
	(4.0087)	(4.2892)
	16.4654	**18.5704**
EAG	0.0505	0.0372
	(0.2666)	(0.3071)
	0.0736	**0.0957**
SOU	0.2996	0.2496
	(2.7454)	(3.0022)
	7.6268	**9.0756**
SENT	-20.7786	-9.1241
	(116.700)	(75.5782)
	1.4E04	**5795.4**
Total	1.96E4	8071.1

The values in parentheses are the standard errors and the bold values are the mean square errors.

TABLE 3.14(continue)

Bias, standard errors and mean square errors of the parameter estimates obtained through Monte Carlo experiments.

	Second structural equation (SENT)	
	traditional	derived reduced form parameters
PUBUN	0.0959	0.0032
	(2.9997)	(0.9124)
	9.0071	**0.8325**
ONE	-18.9980	-3.3335
	(490.600)	(55.8953)
	2.4E05	**3135.40**
GOVWAGE	-0.1524	-0.3825
	(2.5906)	(4.7641)
	6.7345	**22.8430**
PRIVUN	-0.0285	-0.0069
	(1.0295)	(0.4198)
	1.0607	**0.1763**
CA1	0.0683	0.0126
	(1.8425)	(0.5047)
	3.3996	**0.2549**
COPEC	-0.0360	-0.0048
	(0.9631)	(0.1332)
	0.9288	**0.0178**
LOGMPRTY	3.3890	0.6892
	(86.2970)	(12.6449)
	7458.7	**160.4**
NWLF	-0.0197	-0.0047
	(0.5178)	(0.3327)
	0.2685	**0.1107**
SENT	6.3336	2.4245
	(136.700)	(43.0134)
	1.9E04	**1856**
Total	2.64E4	5176

TABLE 3.15

The coefficient estimates of the selected estimation process.

First structural equation

(PUBUN)

	HECKMAN	HECKGLS	AMEMIYA	RLS
Y_2^*	-2.3830	6.2804	6.9705	6.1254
(γ_1)	(4.0618)	(7.7579)	(5.6104)	(5.5335)
ONE	39.2069	54.3126	57.2841	49.7700
(β_{11})	(8.0240)	(14.5804)	(10.8814)	(10.1303)
GOVWAGE	-11.4664	-11.4714	-11.9181	-10.8677
(β_{12})	(2.1934)	(2.3614)	(1.9606)	(1.9182)
PRIVUN	0.2170	0.3402	0.3928	0.5126
(β_{13})	(0.1027)	(0.1642)	(0.1278)	(0.1194)
PROPLAW	14.9908	22.3594	22.8074	24.5160
(β_{14})	(0.8909)	(6.6273)	(5.2942)	(5.2528)
EAG	-0.8328	-0.7388	-0.7129	-0.3478
(β_{15})	(0.2133)	(0.2643)	(0.1800)	(0.1186)
SOU	-7.0743	-8.2288	-8.1551	-5.8438
(β_{16})	(2.2343)	(2.5701)	(2.0843)	(1.8550)
SENT	13.3007	-28.1923	-36.1480	-34.8884
(δ_1)	(8.9890)	(34.5827)	(25.9471)	(25.8097)

TABLE 3.15(continue)

The coefficient estimates of the selected estimation process.

Second structural equation

(SENT)

	HECKMAN	HECKGLS	AMEMIYA	RLS
PUBUN	0.2348	0.2751	0.0259	0.0609
(γ_2)	(0.1640)	(0.0132)	(0.0129)	(0.0205)
ONE	16.8075	1.7121	1.7064	6.1455
(β_{21})	(46.2756)	(2.7913)	(2.8438)	(3.6117)
GOVWAGE	3.0002	0.4508	0.4523	0.8799
(β_{22})	(1.3283)	(0.1853)	(0.1775)	(0.3138)
PRIVUN	-0.0813	-0.0101	-0.0123	-0.0434
(β_{23})	(0.1644)	(0.0109)	(0.0102)	(0.0132)
CA1	-0.2878	-0.0578	-0.0568	-0.0958
(β_{27})	(0.2255)	(0.0207)	(0.0199)	(0.0243)
COPEC	0.0260	0.0024	0.0028	0.0172
(β_{28})	(0.0284)	(0.0034)	(0.0031)	(0.0038)
LOGMPRTY	-6.5865	-1.1308	-1.1060	-2.4701
(β_{29})	(12.1893)	(0.7761)	(0.7727)	(0.8865)
NWLF	0.1425	0.0200	0.0165	0.0326
($\beta_{2,10}$)	(0.2037)	(0.0139)	(0.0135)	(0.0072)
SENT	-2.3558	3.2666	3.2955	2.1233
(δ_2)	(1.8327)	(0.3367)	(0.3269)	(0.7645)
$-\gamma_2\delta_1$	-3.1230	0.7752	0.9362	2.1247

Note: The values in parentheses are the asymptotic standard errors.

many fold compared to those presented in Table 3.4. As a consequence, the estimates of σ_{12} and σ_1^2 obtained via the derived reduced form parameters approach should not be used in the HECKGLS, AMEMIYA and RLS procedure since they increase the variability. The reason is that the estimates of σ_{12} and σ_1^2 obtained via the derived reduced form parameters approach are not very accurate in estimating the actual values of σ_{12} and σ_1^2.

It is possible to use the reduced form parameters derived from the structural parameter estimates of the HECKMAN procedure as the initial starting values to calculate $\hat{\sigma}_{12}$ and $\hat{\sigma}_1^2$ in the derived reduced form parameters approach. This can be implemented in the HECKGLS, AMEMIYA and RLS procedure. The reduced form parameters are updated by using the recently obtained structural parameter estimates and then the estimates of σ_{12} and $\hat{\sigma}_1^2$ are re-calculated. We proceed with this iterative method until certain convergency criteria are met. Possible problems of estimating the structural parameters by this iterative method are that there are no guaranties that convergence exists and, even when there is convergence, maybe not all estimates of the variances are positive. Furthermore, using Monte Carlo experiment to find the small sample properties of such estimation technique is time consuming.

3.4 CONCLUSIONS.

1. We have introduced three alternatives estimation techniques (HECKGLS, AMEMIYA and RLS) to the traditional HECKMAN procedure. The alternative estimation techniques all utilize generalized least squares methods These alternatives yield estimation results that are quite alike. With the implementation of the iterative routine, we have shown that these alternatives estimation techniques have a tendency to produce estimates that converge to the same values. However, the results obtained are very distinctive from those of the HECKMAN procedure.

2. We have used Monte Carlo experiments to study the small sample properties of the HECKMAN procedure and its alternatives. In the course of the experiments, we have discovered that not all Monte Carlo samples generated can be used for they do not yield positive estimates of variances. From the Monte Carlo experiments, we have suggested the AMEMIYA procedure as an alternative to the traditional HECKMAN procedure.

3. We have calculated the theoretical standard errors of the estimates obtained from the traditional HECKMAN procedure and its alternatives, namely the HECKGLS, AMEMIYA and RLS procedure. We have found that the small samples properties reflected by the Monte Carlo experiments are vastly different from their theoretical counterparts.

4. We have introduced an alternative approach in estimating the covariance matrix called the derived reduced

form parameters approach. The estimates of the components of the covariance matrix obtained from the derived reduced form parameters are by no means more accurate than those obtained from the traditional approach. However, it provides the opportunity of including iterative routines in the estimation of the structural parameters as well as the covariance matrix of the HECKGLS, AMEMIYA and RLS procedure.

In this Chapter, we have examined several aspects of the estimations of the simultaneous generalized probit model as well as introduced several alternative estimation procedures. The matter is not yet settled and additional investigations are needed. Furthermore, new estimation techniques could be developed. One plausible technique is the hybrid between the HECKMAN procedure and the AMEMIYA procedure. Recall that Heckman uses $F(\mathbf{X}\hat{\Pi}_2)$ to replace the variable d in order to eliminate its correlation with the disturbance terms while Amemiya suggests the use of instrumental variables approach or pre-multiplying the vector $\underline{d}$ by the matrix of explanatory variables X. The combination of both methods is to use a proxy for $\underline{d}$ called $\hat{\underline{d}}$ where

$$\hat{d}_t = \begin{cases} 1 \text{ if } (\mathbf{X}\hat{\Pi}_2) > 0 \\ 0 \text{ elsewhere} \end{cases} .$$

The variable $\hat{d}$ can also be pre-multiplied by the matrix of explanatory variables X to further eliminate the correlation with the disturbances.

CHAPTER 4
EQUITY ESTIMATOR

4.1 INTRODUCTION

In this chapter, we will discuss the biased estimator developed by Krishnamurati and Rangaswamy (1987) hereinafter denoted KR . The estimator is called the "equity estimator." KR claim that the equity estimator is superior to the simple ridge estimator on the basis of mean square error comparisons in Monte Carlo experiments. Consequently, the equity estimator may be useful in the presence of multicollinearity. We are going to study the various properties of the equity estimator and compare them to the ordinary least squares within the context of multicollinear data. First, we define multicollinearity and discuss its effects on ordinary least squares in Section 4.2. We study the use of biased estimators as alternatives to least squares in the presence of multicollinearity in Section 4.3. The traditional biased estimators that we are going to consider are the ridge regression estimator and a Stein-like principal components estimator introduced in Section 4.4.

Finally, in Section 4.5 we introduce the equity estimator and examine its characteristics. The small sample properties of ridge regression estimator, principal components estimator and equity estimator can be compared via a Monte Carlo experiment which will be presented in the following chapter.

4.2 MULTICOLLINEARITY AND ITS EFFECTS ON ORDINARY LEAST SQUARES

Consider the model

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e} \tag{4.1}$$

where $\underline{y}$ is a (T×1) vector of observations on a dependent variable, $\mathbf{X}$ is a fixed (T×K) full rank matrix of observations on exogenous variables, $\underline{\beta}$ is a (K×1) vector of unknown parameters and $\underline{e}$ is a (T×1) vector of disturbance terms which are identically and independently distributed as $N(0,\sigma^2)$.

Exact multicollinearity is present when at least one of the explanatory variables is a linear combination of the remaining explanatory variables. The matrix $\mathbf{X}$ is not of full column rank and $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist. Furthermore, for each exact linear dependence among the columns of $\mathbf{X}$ one of the eigenvalues of $\mathbf{X}'\mathbf{X}$ is zero. In practice, exact multicollinearity is rare except for the case where too many dummy variables are included or the sample size T<K, and we exclude the possibility of its ocurrence.

Let the columns of $\mathbf{X}$ matrix be denoted by $\underline{x}_i$, i = 1,2,...,K. Then near exact multicollinearity exists if

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_K\underline{x}_K \cong \underline{0} \tag{4.2}$$

Alternatively, near exact multicollinearity exists if at least one of the eigenvalues of $\mathbf{X}'\mathbf{X}$ has value approximately equal to zero.

The effect of multicollinearity on the least squares estimator is revealed through a transformation of the LS estimator. Let **P** be the matrix whose columns are the orthonormal characteristic vectors of (**X'X**) corresponding to the ordered characteristic roots of (**X'X**) which are contained in the diagonal matrix $\Lambda = \text{diag}(\lambda_1,\ldots,\lambda_K)$, such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K$. Let **Q** be the matrix consisted of the orthonormal characteristic roots of (**XX'**) associated with the K nonzero eigenvalues of (**XX'**). Muticollinearity exists when one or more of the eigenvalues of (**X'X**) are near zero.

The least squares (LS) estimator is

$$\underline{b} = (\mathbf{X'X})^{-1}\mathbf{X'}\underline{y} \tag{4.3}$$

The covariance matrix of $\underline{b}$ is

$$\text{Cov}(\underline{b}) = \sigma^2(\mathbf{X'X})^{-1} = \sigma^2\mathbf{P}\Lambda^{-1}\mathbf{P'}$$

$$= \sigma^2\sum_{j=1}^{K}\lambda_j^{-1}\underline{p}_j\underline{p}_j' \tag{4.4}$$

The variance of a particular b_j can be written as

$$\text{Var}(b_j) = \left\{\frac{p_{j1}^2}{\lambda_1} + \frac{p_{j2}^2}{\lambda_2} + \ldots + \frac{p_{jK}^2}{\lambda_K}\right\}\sigma^2 \tag{4.5}$$

The effects of multicollinearity are clearly observed from Equation (4.5). Small eigenvalues tend to increase the variation of b_j. Nevertheless, it is not necessary that small eigenvalues will result in great variations in all of

the estimated coefficients since the p_{ij}^2 , i and j = 1,2,...,K, are weights assigned to each component of the variance as indicated in equation (4.5). Even though some values of the eigenvalues could be extremely small, the variation of a particular coefficient may not be large if the corresponding weight is small or σ^2 is small.

Multicollinearity results in the large sampling variances of the estimated coefficients and the values of the coefficients are often too large in absolute values with some having wrong expected signs. Consequently, we observe the following: " First, the direct result is that the separate effects of explanatory variables involved may not be estimated precisely. Second, given the above, coefficients may not appear significantly different from zero and may be excluded from the analysis, not because the associated variable has no effect but because the sample is inadequate to isolate it. This situation may occur despite possibly high R^2 or F values, indicating a model that fits the data well. Third, estimated coefficients may be sensitive to the addition or deletion of a few observations or the deletion of an apparently insignificant variable."(Judge ,et al., 1982, p.610)

4.3 BIASED ESTIMATION

Consider the model in Equation (4.1)

$$\underline{y} = X\underline{\beta} + \underline{e}$$

where $\mathbf{X}$ is a $(T \times K)$ fixed matrix of rank $K \leq T$ and $e \sim N(0, \sigma^2 I_K)$. Let's suppose that the matrix $\mathbf{X}$ and the vector $\underline{y}$ are standardized by subtracting variable means and dividing by variable standard deviations.

The least squares estimator $b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underline{y}$ has covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$; furthermore, it is unbiased since $E[\underline{b}] = \underline{\beta}$. Out of the class of unbiased estimators, the least squares estimator is best, where best implies minimum variance.

We have shown that in the presence of multicollinearity some elements of $\underline{b}$ may be variable and imprecisely estimated. Consequently, we consider biased estimators of $\underline{\beta}$, $\underline{\delta}$, that may have smaller variation than $\underline{b}$, and thus may provide estimates closer to the true parameter values than the LS estimator.

As a basis for evaluating estimator performance, we consider the weighted squared error loss measure

$$L(\underline{\beta}, \underline{\delta}, \mathbf{D}) = (\underline{\delta} - \underline{\beta})'\mathbf{D}(\underline{\delta} - \underline{\beta}) \tag{4.6}$$

where $\mathbf{D}$ is a positive definite and symmetric matrix. The sampling performance of $\underline{\delta}$ is evaluated by its risk function

$$R(\underline{\beta}, \underline{\delta}, \mathbf{D}) = E\left[(\underline{\delta} - \underline{\beta})'\mathbf{D}(\underline{\delta} - \underline{\beta})\right] \tag{4.7}$$

The most common choices for the weight matrix $\mathbf{D}$ in (4.6) are $\mathbf{D} = \mathbf{I}$, which defines the risk of estimation to be the mean square error, and $\mathbf{D} = \mathbf{X}'\mathbf{X}$ which corresponds to mean square error of in-sample prediction.

For $\mathbf{D} = \mathbf{I}$, Equation(4.7) is written as

$$R(\underline{\beta},\underline{\delta},\mathbf{I}) = E\left[(\underline{\delta} - \underline{\beta})'(\underline{\delta} - \underline{\beta})\right]$$

Let us rewrite $(\underline{\delta} - \underline{\beta})$ as

$$(\underline{\delta} - \underline{\beta}) = (\underline{\delta} - E(\underline{\delta})) + (E(\underline{\delta}) - \underline{\beta})$$

Note that the expression $(E(\underline{\delta}) - \underline{\beta})$ is the bias vector. Consequently, the mean square error loss can be expressed as

$$\begin{aligned} R(\underline{\beta},\underline{\delta},\mathbf{I}) &= E\left[(\underline{\delta} - E(\underline{\delta})) + (E(\underline{\delta}) - \underline{\beta})\right]'\left[\cdot\right] \\ &= E\left[(\underline{\delta} - E(\underline{\delta}))'(\underline{\delta} - E(\underline{\delta}))\right] \\ &\quad + 2\,E\left[(\underline{\delta} - E(\underline{\delta}))'(\underline{\delta} - E(\underline{\delta}))\right] \\ &\quad + E\left[(E(\underline{\delta}) - \underline{\beta})'(E(\underline{\delta}) - \underline{\beta})\right] \end{aligned} \tag{4.8}$$

The expression $(\underline{\delta} - E(\underline{\delta}))$ has zero mean and thus its expectation is a null vector.

Therefore, (4.8) can be rewritten as

$$\begin{aligned} R(\underline{\beta},\underline{\delta},\mathbf{I}) &= E\left[(\underline{\delta} - E(\underline{\delta}))'(\underline{\delta} - E(\underline{\delta}))\right] \\ &\quad + E\left[(E(\underline{\delta}) - \underline{\beta})'(E(\underline{\delta}) - \underline{\beta})\right] \\ &= \mathrm{tr}(\mathrm{Cov}(\underline{\delta})) + \mathrm{tr}(\mathrm{bias}(\underline{\delta}))(\mathrm{bias}(\underline{\delta}))' \end{aligned} \tag{4.9}$$

For $\underline{\delta} = \underline{b}$, $R(\underline{\beta},\underline{b},\mathbf{I}) = \mathrm{tr}(\mathrm{Cov}(\underline{b}))$ since $\underline{b}$ is unbiased.

The biased estimator $\underline{\delta}$ is superior to the least squares estimator when $[\mathrm{tr}(\mathrm{Cov}(\underline{b})) - \mathrm{tr}(\mathrm{Cov}(\underline{\delta}))] > \mathrm{tr}(\mathrm{bias}(\underline{\delta}))(\mathrm{bias}(\underline{\delta}))'$. That is the increase of risk due to

biasness is less than the decrease of risk due to reduced variability. In the following sections we consider the alternative biased estimation rules that may yield lower estimator risk than the LS estimator in the presence of multicollinearity.

4.4 RIDGE REGRESSION ESTIMATOR

Consider the model in Equation (4.1)

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e}$$

where $\mathbf{X}$ is a fixed $(T \times K)$ matrix of rank $K \leq T$ and $\underline{e} \sim N(0, \sigma^2 \mathbf{I})$. The matrix $\mathbf{X}$ and the vector $\underline{y}$ are standardized by subtracting variable means and dividing by variable standard deviation.

The generalized ridge estimator introduced by Hoerl and Kennard (1970 a,b) is

$$\hat{\underline{\beta}}^{*}(\mathbf{K}) = \left[\mathbf{X}'\mathbf{X} + \mathbf{PKP}'\right]^{-1}\mathbf{X}'\underline{y} \qquad (4.10)$$

where $\mathbf{P}$ is previously defined and $\mathbf{K} = \text{diag}\ (k_1, \ldots, k_K)$, $k_i \geq 0$; $i = 1, 2, \ldots, K$.

If $\mathbf{PKP}' = k\mathbf{I}_K$, $k \geq 0$, then $\hat{\beta}^{*}(\mathbf{K})$ is reduced to the simple ridge estimator.

$$\hat{\underline{\beta}}^{*}(k) = \left[\mathbf{X}'\mathbf{X} + k\mathbf{I}_K\right]^{-1}\mathbf{X}'\underline{y} \qquad (4.11)$$

Given the risk function described in Equation (4.7), with $\mathbf{D} = \mathbf{I}$, the resulting risk is called MSE (mean squared error) in much statistical literature. Hoerl and Kennard (1970a) found that there always exists a $k>0$ such that the

ridge estimator has smaller risk than the least squares estimator.

Theobald (1974), using the generalized risk function in Equation (4.7), showed that a sufficient condition for $\hat{\underline{\beta}}^{*}(k)$ to have a smaller risk than the least squares estimator is that $k < 2\sigma^2/(\underline{\beta}'\underline{\beta})$.

The sufficient conditions for $\hat{\underline{\beta}}^{*}(k)$ to have smaller risk than least square provided in Hoerl and Kennard (1970a) and Theobald (1974) are dependent on the unknown parameters $\underline{\beta}$ and σ^2. Consequently, they are not empirically applicable. The constant k has to be estimated using the data. Therefore, if the estimated k depends on $\underline{y}$ it is stochastic, and the properties of ridge regression are no longer valid.

Hoerl, Kennard and Baldwin (1975) suggested a k value of

$$k(y) = \frac{K\hat{\sigma}^2}{\underline{b}'\underline{b}} \qquad (4.12)$$

where k(y), a function of the data, is the sample analogue of $K\sigma^2/(\underline{\beta}'\underline{\beta})$ and $\hat{\sigma}^2=(\underline{y}-X\underline{b})'(\underline{y}-X\underline{b})/(T-K)$. Lawless and Wang (1976) and Dempster, Schatzoff and Wermuth (1977) proposed alternatives to k(y) in Equation (4.12). Monte Carlo experiments are performed for each of the estimation rules to show that the ridge estimator may have smaller mean square error than the least squares estimator.

Several studies examined the risk properties of these simple ridge estimators and determined the conditions under which they have lower risk than the least squares estimator.

Sidik (1975) showed that it was possible to reduce the MSE of the simple ridge estimator by slightly increasing k. Deegan (1975), using a bias minimization technique, showed a way to find a probable upper bound for k. Farebrother (1976), Lee and Trivedi (1982) provided additional insights on the conditions for k such that the simple ridge estimator will have smaller MSE risk than the least squares rule. Thisted (1977, 1978a) and Casella (1977) exhibited conditions under which $\hat{\underline{\beta}}^{*}(k(y))$, for several choices of estimation rules for k(y), is minimax. Furthermore, they also showed that the minimaxity of the simple ridge estimator is dependent on the eigenvalues of $\mathbf{X'X}$. For $K \geq 4$, ridge regression is minimax if the eigenvalues are all equal. Thisted (1978b) studied the minimax condition for generalized ridge of the form in equation (4.10) and concluded that simple ridge estimators cannot maintain the necessary condition for minimaxity. Ullah, Vinod and Kadiyala (1978) presented alternatives choices for k that improve the performance of the simple ridge estimator. For collection of articles on ridge regression and its application during 1962-79, see Hoerl and Kennard (1981).

Hoerl and Kennard (1970a) also considered a generalized ridge estimator of the form $[\mathbf{X'X} + k\mathbf{B}]^{-1}\mathbf{X'}\underline{y}$ where $\mathbf{B}$ is a symmetric, positive definite matrix. This form can be rewritten as $[\mathbf{I} + k\mathbf{C}]^{-1}\underline{b}$, where $\mathbf{C} = (\mathbf{X'X})^{-1}\mathbf{B}$. Strawderman (1978) presented a ridge estimator of this form described as

$$\underline{\delta}(\underline{b},s) = [I + k(y)C]^{-1}\underline{b} \qquad (4.13)$$

where $s = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})$. The purpose of constructing this alternative was to find estimator of this form that are minimax and have the same properties as usual ridge regression.

Given the risk function in Equation (4.7), K≥3, and the usual assumptions of the classical normal linear regression, Strawderman showed that the ridge rule

$$\underline{\delta}(\underline{b},s) = \left[I + \frac{asD^{-1}X'X}{\underline{b}'X'X\underline{b} + gs + h}\right]^{-1}\underline{b} \qquad (4.14)$$

is minimax, where $0 \le a \le \frac{2(K-2)}{(T-K+2)} \cdot \frac{1}{\lambda_{max}[D^{-1}X'X]}$ and $h \ge 0$, $g \ge 2K/(T-K+2)$. $\lambda_{max}[D^{-1}X'X]$ is the largest eigenvalue of $D^{-1}X'X$.

4.5 STEIN-LIKE PRINCIPAL COMPONENTS ESTIMATOR

Marquardt (1970) studied the properties of principal components estimator in the form of generalized inverse estimation technique with some eigenvalues being zero. He showed that the resulting estimator is a linear transformation of the least squares estimator and the biasedness of the estimator depends on how close to zero are some of the eigenvalues. Marquardt also developed a condition under which the generalized inverse estimator has less mean square error risk than that of least squares rule. Farebrother (1972) explored the properties of the principal

components estimator under the minimum weighted estimated mean square error criteria. He suggested a possible solution to the problem by calculating the mean square error for all of the possible combinations of components to be deleted. In the case that K is very large, this method is impractical.

Johnson, Reimer and Rothrock (1973) formulated the principal components estimator in the form of restricted least squares estimator. Consequently, they showed that the restrictions implied by the principal components estimator can be tested before being imposed in the estimation process. Greenberg (1975) discussed the trade off between reduced variance and increased biasedness when the components with small eigenvalues are dropped in principal components estimation. He suggested that the components to be dropped or the implied restrictions should be chosen by examining the eigenvalues and their associated eigenvectors. The test for the statistical significance of the restrictions can also be used in choosing the set of restrictions to be imposed. Mittelhammer and Baritelle (1977) studied two criteria for selecting the components to be deleted and their small sample properties. The first criteria considered is to delete the components associated with small eigenvalues. The second criteria is to test the statistical significance of the components before deleting. The small sample properties of the criteria are obtained through Monte Carlo experiment. They found that the performance of the principal components estimator using the

two criteria of components selection decreases as the data become more highly collinear. Nevertheless, the principal components estimator using the two criteria has lower mean square error than least squares estimator under certain conditions.

Fomby and Hill (1987) suggested some alternative criteria of choosing the components to be deleted. They rejected Pidot's criterion of deleting a component when its associated eigenvalue is less than the average root or 1 when the matrix of the explanatory variables is in correlation form. Instead the components to be deleted should be selected on the basis of the variance reduction potential which can be obtained by decomposing the covariance matrix of the least squares estimator. Another alternative is to test the restrictions implied by the principal components estimator. They also suggested the use of a Stein-like estimator to combine the non-sample information or restrictions with the sample information. Stein-like estimator is known to dominate least squares under squared error loss if certain conditions are met.

Consider the model in Equation (4.1)

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e}$$

where $\mathbf{X}$ is a fixed $(T \times K)$ matrix of rank $K \leq T$ and $\underline{e} \sim N(0, \sigma^2\mathbf{I})$. The matrix $\mathbf{X}$ and the vector $\underline{y}$ are standardized by subtracting variable means and dividing by variable standard deviations. Suppose we have exact nonsample information

relative to a particular parameter or linear combination of parameters that may be stated as

$$\mathbf{R}\underline{\beta} = \underline{r} \tag{4.15}$$

where $\underline{r}$ is a (J×1) vector of known elements and $\mathbf{R}$ is a known (J×K) prior information design matrix of rank J≤K. The restricted least squares estimator is

$$\underline{b}^* = \underline{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\underline{b} - \underline{r}) \tag{4.16}$$

In 1961, James and Stein exhibited a *non*-linear estimator that combined nonsample information with sample information and dominated the least squares-maximum likelihood estimator, and thus demonstrated its inadmissibility. The James-Stein estimator may be written in general form as

$$\underline{\beta}^* = \left[1 - \frac{as}{(\mathbf{R}\underline{b} - \underline{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\underline{b} - \underline{r})}\right](\underline{b} - \underline{b}^*) + \underline{b}^* \tag{4.17}$$

The estimator $\underline{\beta}^*$ has risk less than or equal to that of least squares if J≥3 and

$$0 < a < \frac{2}{T-K+2}\left[\frac{tr\{[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}}{\eta} - 2\right]$$

where η is the largest eigenvalue of the expression in the brackets { }, with s given in (4.13).

The expression in Equation (4.17) can also be written as

$$\underline{\beta}^* = \left[1 - \frac{c}{u}\right](\underline{b} - \underline{b}^*) + \underline{b}^*$$

$$= \left[1 - \frac{c}{u} \right] \underline{b} + \underline{b}^* \frac{c}{u} \qquad (4.18)$$

where $u = \dfrac{(R\underline{b} - \underline{r})'[R(X'X)^{-1}R']^{-1}(R\underline{b} - \underline{r})}{J\hat{\sigma}^2}$ is the likelihood ratio test statistic on the restrictions $R\underline{\beta} = \underline{r}$ and $c = a(T-K)/J$. The random variable u has a central F distribution with J and (T-K) degree of freedom if the restrictions are correct.

If the restrictions are strongly supported by the data, u will be small and the weight assigned to $\underline{b}^*$ will be large. If the restrictions are not supported by the data, u will be large and more weight will be assigned to $\underline{b}$.

From Equation (4.18), if $c > u$, then $\underline{\beta}^*$ is no longer a convex combination of $\underline{b}$ and $\underline{b}^*$. For a specific case where $R = I$ and $\underline{r} = \underline{0}$, the estimation rule in Equation (4.18) changes the sign of the least squares estimator when $c > u$. As a consequence, the rule in Equation (4.18) ought not to be used. As it turns out

$$\underline{\beta}^+ = \left[1 - (c/u) \right]_+ (\underline{b} - \underline{b}^*) + \underline{b}^* \qquad (4.19)$$

where $\left[1 - (c/u) \right]_+ = \max \left[1 - (c/u), 0 \right]$, uniformly improves on $\underline{\beta}^*$ which is shown in Adkins and Hill (1989). Therefore, $\underline{\beta}^*$ is inadmissible. $\underline{\beta}^+$ is called positive Stein rule.

The Stein-like rule (4.19) can be applied to the principal components estimator, since the principal components estimator is a form of restricted least squares.

Again, consider the Equation (4.1)

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e}$$

We can transform **X** into the matrix of principal components by post multiplying **X** by **P** where **P** is previously defined. That is

$$\underline{y} = \mathbf{XPP}'\underline{\beta} + \underline{e}$$

$$= \mathbf{Z}\underline{\theta} + \underline{e} \qquad (4.20)$$

The i-th column of **Z**, $\underline{z}_i$, is equal to $\mathbf{X}\underline{p}_i$ where $\underline{p}_i$ is the i-th column of **P**. $\underline{z}_i$ has the property that

$$\underline{z}_i'\underline{z}_i = \underline{p}_i'\mathbf{X}'\mathbf{X}\underline{p}_i = \lambda_i$$

Suppose the columns of **X** obey J linear independent restrictions so that **X** has rank K-J and consequently $\lambda_{K-J+1} = \ldots = \lambda_K = 0$. Therefore, we can partition the matrix **Z** into two parts, $[\mathbf{Z}_1|\mathbf{Z}_2]$, according to whether the associated eigenvalues are zero or not. Since the last J eigenvalues are zero, the corresponding principal components are null vectors. Thus, we can rewrite Equation(4.15) as

$$\underline{y} = \mathbf{Z}\underline{\theta} + \underline{e}$$

$$= [\mathbf{Z}_1 \quad \mathbf{Z}_2]\begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix} + \underline{e}$$

$$= \mathbf{Z}_1\underline{\theta}_1 + \underline{e} \qquad (4.21)$$

We can see that $\mathbf{Z}_2$ is deleted from the model which is equivalent to specifying $\underline{\theta}_2 = \underline{0}$. From Equation(4.15), we have

$$\underline{\theta} = \mathbf{P}'\underline{\beta}$$

$$\begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{bmatrix} \underline{\beta} \qquad (4.22)$$

Thus, specifying $\underline{\theta}_2 = \underline{0}$ implies restricting $\mathbf{P}_2'\underline{\beta} = \underline{0}$. Applying Stein-like rule to the principal components estimator, denoted by $\hat{\underline{\beta}}_{PC}$, is equivalent to applying Stein-like rule to Equation (4.1) with the restrictions that $\mathbf{P}_2'\underline{\beta} = \underline{0}$.

4.6 THE EQUITY ESTIMATOR

Krishanmurati and Rangaswamy (1987) suggest an alternative to traditional biased estimators called the equity estimator. The name equity comes from the suggestion that "... when no specific information is available about the relative effects, each control variable must be treated in an equitable manner in determining its relative impact on the response variable" (Rangaswamy et al., 1985, p.18). KR (1987) perform Monte Carlo experiments to evaluate the small sample properties of the equity estimator. They suggest that the equity estimator should be used when the R-square of the OLS is less than 0.7. Rangaswamy and Krishamurati (1991) use Jackknife and bootstrap resampling techniques to study the small sample property of the equity estimator. They claim that the equity estimator ought to be used for solving the problem of resource allocation . Moreover, they say that the

equity estimator gives estimates with more 'valid' signs. However, they admit that the equity estimator performs poorly when evaluated by the prediction mean square error criterion.

The development of the equity estimator is divided into several steps which we describe below. Following Rangaswamy et al.(1985) and Krishnamurati and Rangaswamy (1987), the matrix **X** and the vector $\underline{y}$ are standardized by subtracting variable means and dividing by variable standard deviations.

Step 1. Find an orthonormal matrix $\tilde{Z}$ such that it is maximally correlated with **X**, which is equivalent to maximizing the correlation between **X** and $\tilde{Z}$. This may be stated as

$$\max \phi = \sum_{j=1}^{K} \phi_j$$

$$\text{subject to } \tilde{Z}'\tilde{Z} = I_K$$

where $\phi_j = \sum_{j=1}^{K} z_{tj}x_{tj}$; t = 1,2,...,T.

The solution is given by $\tilde{Z} = QP'$ where **P** is previously defined and **Q** is the matrix of orthonormal eigenvectors of (**XX'**).

Step 2. Find the relationship between $\tilde{Z}$ and $\underline{y}$ which is achieved by regressing $\underline{y}$ against $\tilde{Z}$. We get

$$\hat{\underline{\eta}} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\underline{y} \qquad (4.23)$$

Step 3. Scale the $\hat{\underline{\eta}}_j$'s in order to obtain coefficients in terms of the **X** variables by constructing

$$\underline{w} = \hat{\eta}_1 \underline{x}_1 + \ldots + \hat{\eta}_K \underline{x}_K = X\hat{\underline{\eta}} \tag{4.24}$$

Then regress $\underline{y}$ against $\underline{w}$ to get a scaling constant $\hat{\nu}$.

$$\hat{\nu} = (\underline{w}'\underline{w})^{-1}\underline{w}'\underline{y} \tag{4.25}$$

The equity estimator is expressed as

$$\hat{\underline{\beta}}_E = \hat{\nu} \cdot \hat{\underline{\eta}} \tag{4.26}$$

KR do not show any statistical properties of the equity estimator. They rely on Monte Carlo experiments to examine the performances of the equity estimator. They even neglect to provide adequate intuition about the equity estimator. Therefore, we are going to investigate each step of the development of the equity estimator.

Step 1. Consider the matrix X which can be expressed as

$$X = Q\Lambda^{1/2}P' \tag{4.27}$$

by singular matrix decomposition, where the matrices P and Q are as previously defined. Near exact multicollinearity exists if one or more of the eigenvalues of X'X has value approximately equal to zero. The matrix $\tilde{Z}$ eliminates the problem of near exact multicollinearity by suppresing the $\Lambda^{1/2}$ component from Equation (4.27). That is $\tilde{Z} = QI_K P'$, and $\Lambda^{1/2}$ is replaced by an identity matrix. In effect, the variability of the regressors is ignored, and made equal, by transforming the data to be orthonormal. That is $\tilde{Z} = QP'$

Since $Q = XP\Lambda^{-1/2}$ from (4.27)

$$\tilde{Z} = XP\Lambda^{-1/2}P'$$

$$= X(X'X)^{-1/2} \qquad (4.28)$$

where $P\Lambda^{-1/2}P' = (X'X)^{-1/2}$ is a symmetric positive definite matrix such that

$$(X'X)^{-1/2}(X'X)^{-1/2} = (X'X)^{-1},$$

$$(X'X)^{1/2} = P\Lambda^{1/2}P',$$

$$(X'X)^{1/2}\cdot(X'X)^{-1/2} = I,$$

and $(X'X)^{1/2}(X'X)^{1/2} = (X'X)$.

Consequently,

$$\tilde{Z}'\tilde{Z} = (X'X)^{-1/2}X'X(X'X)^{-1/2} = I_K.$$

Another way to view the matrix $\tilde{Z}$ is from a graphical perspective. Consider equation (4.20)

$$\underline{y} = Z\underline{\theta} + \underline{e}$$

We know that $\underline{z}_i$ has the property that $\underline{z}_i'\underline{z}_i = \lambda_i$. Suppose that the matrix X has the vectors $\underline{a}_1,\ldots,\underline{a}_K$ as the unit basis vectors in the original coordinate system. We then transform the matrix X into a new coordinate system with basis vectors $\underline{p}_1,\ldots,\underline{p}_K$, where $P = [\underline{p}_1\ldots\underline{p}_K]$, which are in the direction of the axes of the data ellipsoid. Fomby et al. (1984) show that " λ_j can be represented geometrically as the sum of squares of the projections of the T points x_1, ... ,x_T onto the $\underline{p}_j$ axes. Thus the characteristic roots of X'X measure the variability of the data in the direction of the axes of the ellipsoid" (p.290).

The i-th column of the matrix $\tilde{Z}$, $\tilde{\underline{z}}_i$, when expressed in the term of the new basis vectors $\underline{p}_i$'s is equal to the i-th column of the matrix Z, $\underline{z}_i$, weighted by $\lambda_i^{-1/2}$. This

rescaling effort results in the variability of the data with respect to the basis vectors $\underline{p}_i$ being equal in every direction, with the direction associated with the most variability being the one that is scaled towards the origin the most. That is $\tilde{\underline{z}}_i = \underline{z}_i\lambda_i^{-1/2}$, therefore $\tilde{\underline{z}}_i'\tilde{\underline{z}}_i = \underline{z}_i'\underline{z}_i\cdot\lambda_i^{-1} = \lambda_i\lambda_i^{-1} = 1$.

Thus, the first step of equity estimation amounts to a re-scaling of the original data to equal variability in all directions of a transformed observation space.

Step 2.

Consider the equation

$$\begin{aligned} \underline{y} &= \mathbf{X}\underline{\beta} + \underline{e} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{1/2}\underline{\beta} + \underline{e} \\ &= \tilde{\mathbf{Z}}\underline{\eta} + \underline{e} \end{aligned} \tag{4.29}$$

where $\underline{\eta} = (\mathbf{X}'\mathbf{X})^{1/2}\underline{\beta}$

From equation (4.29), we see that $\hat{\underline{\eta}}$ is the maximum likelihood estimator of $\underline{\eta}$, a nonsingular transformation of the parameter vector of interest $\underline{\beta}$. Since $\hat{\underline{\eta}} = (\mathbf{X}'\mathbf{X})^{1/2}\underline{b}$, $\hat{\underline{\eta}} \sim N(\underline{\eta}, \sigma^2\mathbf{I}_K)$. Thus $E(\hat{\underline{\eta}}) = (\mathbf{X}'\mathbf{X})^{1/2}\underline{\beta}$ and $\mathrm{Cov}(\hat{\underline{\eta}}) = \sigma^2\mathbf{I}_K$. As an estimator of $\underline{\beta}$, $\hat{\underline{\eta}}$ is biased and inconsistent unless $(\mathbf{X}'\mathbf{X}) = \mathbf{I}_K$. The estimator $\hat{\underline{\eta}}$ shifts the effects of multicollinearity from the estimator covariance matrix to its bias. Specifically,

$$R[\hat{\underline{\eta}},\underline{\beta},\mathbf{D}=\mathbf{I}] = \mathrm{tr}(\mathrm{Cov}(\hat{\underline{\eta}})) + \sum_{i=1}^{K} \mathrm{bias}^2(\hat{\eta}_i)$$

$$= \sigma^2 K + \underline{\beta}' \left[(X'X)^{1/2} - I_K \right]' \cdot \left[(X'X)^{1/2} - I_K \right] \underline{\beta} \qquad (4.30)$$

where $\sigma^2 K$ is the variance component of risk and the second term is the squared bias component. The bias component can be re-expressed as

$$\sum_{i=1}^{K} \text{bias}^2(\hat{\eta}_i) = \underline{\beta}' \left[(X'X)^{1/2} - I_K \right]' \cdot \left[(X'X)^{1/2} - I_K \right] \underline{\beta}$$

$$= \underline{\beta}' (X'X)^{1/2} \left[I_K - (X'X)^{-1/2} \right] \left[I_K - (X'X)^{-1/2} \right] (X'X)^{1/2} \underline{\beta}$$

$$= \underline{\eta}' \left[I_K - (X'X)^{-1/2} \right]^2 \underline{\eta}$$

but $\left[I_K - (X'X)^{-1/2} \right]^2 = P \left(I_K - \Lambda^{-1/2} \right)^2 P'$, so

$$\sum_{i=1}^{K} \text{bias}^2(\hat{\eta}_i) = \underline{\theta}_E' (\text{diag}[(1-\lambda_1^{-1/2})^2, \ldots, (1-\lambda_K^{-1/2})^2]) \underline{\theta}_E$$

$$= \theta_{1E}^2 (1-\lambda_1^{-1/2})^2 + \ldots + \theta_{KE}^2 (1-\lambda_K^{-1/2})^2 \qquad (4.31a)$$

where $\underline{\theta}_E = P' \underline{\eta} = P' (X'X)^{1/2} \underline{\beta} = P' (P \Lambda^{1/2} P') \underline{\beta} = \Lambda^{1/2} P' \underline{\beta} = \Lambda^{1/2} \underline{\theta}$,

(4.31b)

and θ is the vector of parameters from the principal components model. Consequently, the total squared bias of $\hat{\underline{\eta}}$ as an estimator of $\underline{\beta}$ is dependent not only on $\underline{\beta}$ but also the characteristic roots and vectors of $(X'X)$ that characterize the nature of the multicollinearity in the data.

By rescaling the basis vectors $\underline{p}_i$'s, we can write the linear statistical model in terms of $\underline{\theta}_E$, as

$$\underline{y} = X\underline{\beta} + \underline{e}$$

$$= XPP'\underline{\beta} + \underline{e}$$

$$= Z\underline{\theta} + \underline{e}$$

$$= Z\Lambda^{-1/2}\Lambda^{1/2}\underline{\theta} + \underline{e}$$

$$= \hat{Z}\underline{\theta}_E + \underline{e}$$

$$= XP\Lambda^{-1/2}\Lambda^{1/2}P'\underline{\beta} + \underline{e} \qquad (4.32a)$$

and

$$\hat{\underline{\theta}}_E = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'\underline{y}$$

$$= \hat{Z}'\underline{y} = \Lambda^{-1/2}Z'\underline{y} \qquad (4.32b)$$

Since $\hat{Z} = Z\Lambda^{-1/2}$ and $\hat{Z}'\hat{Z} = I$

From Equation (4.31b) and (4.32), we can see that the estimate θ_{iE} is a decreasing function of λ_i. We have shown that $\text{Var}(\hat{\eta}_i) = \sigma^2$; it follows that $\text{Var}(\hat{\theta}_{iE}) = \sigma^2$ also. The least squares estimator of $\underline{\theta}$, $\hat{\underline{\theta}} = (Z'Z)^{-1}Z'\underline{y} = \Lambda^{-1}Z'\underline{y}$, has covariance matrix $\text{Cov}(\hat{\underline{\theta}}) = \sigma^2\Lambda^{-1}$ or $\text{Var}(\hat{\theta}_i) = \sigma^2/\lambda_i$.

Step 3. KR do not give any explanation of the meaning of the vector $\underline{w}$ in (4.24). They use it as a vehicle to obtain the scaling constant $\hat{\nu}$, which is simply the ratio of the correlations between $(X,\underline{y})$ and $(\tilde{Z},\underline{y})$. That is

$$\hat{\nu} = \frac{\underline{y}'X(X'X)^{-1/2}X'\underline{y}}{\underline{y}'XX'\underline{y}}$$

$$= \frac{\underline{y}'X\ \tilde{Z}'\underline{y}}{\underline{y}'X\ X'\underline{y}}$$

$$= \frac{\underline{y}'\mathbf{X}\cdot\hat{\underline{\eta}}}{\underline{y}'\mathbf{X}\mathbf{X}'\underline{y}} \tag{4.33}$$

Since,

$$\underline{w} = \mathbf{X}\hat{\underline{\eta}}$$

$$= \mathbf{X}(\mathbf{X}'\mathbf{X})^{1/2}\underline{b}$$

$$= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\underline{y}$$

$$= \mathbf{X}\,\tilde{\mathbf{Z}}'\underline{y}$$

Therefore,

$$(\underline{w}'\underline{w}) = \underline{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\underline{y}$$

$$= \underline{y}'\mathbf{X}\mathbf{X}'\underline{y}$$

and

$$\underline{w}'\underline{y} = \underline{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\underline{y}$$

$$= \underline{y}'\tilde{\mathbf{Z}}\mathbf{X}'\underline{y}$$

From equation (4.33) we can see that $\hat{\nu}$ is dependent on $\hat{\underline{\eta}}$. Consequently, the scaling variable $\hat{\nu}$ is random and depends on the eigenvalues of $\mathbf{X}'\mathbf{X}$ which characterize the structure of multicollinearity.

The scalar $\hat{\nu}$ can also be written as

$$\hat{\nu} = \hat{\underline{\eta}}'\mathbf{X}'\underline{y}/\hat{\underline{\eta}}'\mathbf{X}'\mathbf{X}\hat{\underline{\eta}}$$

$$= \frac{\underline{b}'(\mathbf{X}'\mathbf{X})^{1/2}\mathbf{X}'\underline{y}}{\underline{b}'(\mathbf{X}'\mathbf{X})^{1/2}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{1/2}\underline{b}}$$

$$= \frac{\underline{b}'(X'X)^{3/2}\underline{b}}{\underline{b}'(X'X)^{2}\underline{b}} \tag{4.34}$$

The vector $\underline{w}$ can be viewed as the predicted values of $\underline{y}$ using the estimator $\hat{\underline{\eta}}$. That is

$$\hat{\underline{y}}_{(\hat{\underline{\eta}})} = X \cdot \hat{\underline{\eta}}$$

As a consequence,

$$\hat{\nu} = \frac{\hat{\underline{y}}'_{(\hat{\eta})}\underline{y}}{\underline{y}'\underline{y}}$$

If $\underline{y} = \hat{\underline{y}} = X \cdot \hat{\eta}$, then $\hat{\nu} = 1$ which happens when $X'X = I_K$. Following Rao(1973,p. 74)

$$\sup_{\underline{b}} \hat{\nu} = d_1, \quad \inf_{\underline{b}} \hat{\nu} = d_K$$

where $d_1 \geq d_2 \geq \ldots \geq d_K$ are the roots of

$$|(X'X)^{3/2} - d(X'X)^2| = |(X'X)^2| \cdot |(X'X)^{-1/2} - \lambda I| = 0$$

Hence $d_1 = \lambda_K^{-1/2}$ and $d_K = \lambda_1^{-1/2}$. Consequently, $(1/\lambda_1)^{1/2} \leq \hat{\nu} \leq (1/\lambda_K)^{1/2}$. The scale factor $\hat{\nu}$ may substantially increase or decrease the value of $\hat{\underline{\eta}}$.

The equity estimator is described as

$$\hat{\underline{\beta}}_E = \hat{\nu}\hat{\underline{\eta}}$$

The finite sample properties of $\hat{\underline{\beta}}_E$ are very difficult to determine for $\hat{\nu}$ is dependent on $\hat{\underline{\eta}}$. We have shown that $\hat{\underline{\eta}}$ is a biased and inconsistent estimator of $\underline{\beta}$. The scaling constant $\hat{\nu}$ does not change the direction of $\hat{\underline{\eta}}$. When a biased estimator is used, we hope that the increased biasedness is

less than the reduction of the estimator variability. The estimator $\hat{\underline{\eta}}$ will have less variability than $\underline{b}$ if $\sigma^2(X'X)^{-1} - \sigma^2 I_T$ is positive definite. When $\hat{\nu}$ is used to scale $\hat{\eta}$, the values of $\hat{\nu} < 1$ will reduce the variability and $\hat{\nu} > 1$ will increase its variability. Therefore, we cannot be certain that $\hat{\underline{\beta}}_E$ will have lower variability than $\underline{b}$ that more than offsets its biasedness and results in less overall risk than the LS estimator $\underline{b}$.

If we assume that $\lim_{T\to\infty}(X'X/T) = \tilde{Q}$ is finite and nonsingular, then

$$\begin{aligned}
\operatorname{plim}_{T\to\infty} \hat{\underline{\beta}}_E &= \operatorname{plim}_{T\to\infty} \hat{\nu} \cdot \operatorname{plim}_{T\to\infty} \hat{\underline{\eta}} \\
&= \operatorname{plim}_{T\to\infty} \frac{T^{3/2}\underline{b}'(X'X/T)^{3/2}\underline{b}}{T^2\ \underline{b}'(X'X/T)^2\ \underline{b}} \quad \times \\
&\qquad \operatorname{plim}_{T\to\infty} T^{1/2}(X'X/T)^{1/2}\underline{b} \\
&= \frac{\underline{\beta}'\tilde{Q}^{3/2}\underline{\beta}}{\underline{\beta}'\tilde{Q}^2\ \underline{\beta}} \cdot \tilde{Q}^{1/2}\underline{\beta} \neq \underline{\beta} \qquad (4.35)
\end{aligned}$$

Therefore, $\hat{\underline{\beta}}_E$ is inconsistent for all $\underline{\beta} \neq 0$ unless $\tilde{Q} = I$.

Several observations can be made about the equity estimator.

1.The sampling performance of the equity estimator is extremely sample specific. Its bias and mean square error depend of the values of the unknown parameter β and the design matrix X.

2. The Monte Carlo experiment in KR indicates that as the severity of multicollinearity increases, measured by the

increase of the ratio of the largest to the smallest eigenvalues, the MSE risk ($\mathbf{D} = \mathbf{I}$) and in-sample prediction risk ($\mathbf{D} = \mathbf{X'X}$) of the equity estimator decrease.

3. The Monte Carlo experiment in KR also shows that for all values of σ^2, the model error, the equity estimator has in-sample prediction risk greater than or equal to that of the simple ridge estimator. But the MSE risk of the equity estimator is always less than that of the ridge estimator for all value of σ^2. Consequently, if the risk of the equity estimator is plotted against the model error, the performance of the equity estimator relative to least squares and other alternative estimators depends on the choice of the weight matrix **D**. The equity estimator attempts to use additional information involving the eigenvalues of $\mathbf{X'X}$ to help in the estimation of the unknown parameters β when multicollinearity is present. However, the distribution and exact finite sample properties of the equity estimator are still unknown. Thus, the study about its small sample properties must rely on simulations. In chapter 5, we are going to investigate the small sample properties of the equity estimator and compare them to those of traditional biased estimators. Monte Carlo experiments will be employed which may help us to gain more knowledge about the properties of the equity estimator.

CHAPTER 5

THE SMALL SAMPLE PERFORMANCE OF THE EQUITY ESTIMATOR AND ITS ALTERNATIVES

5.1 INTRODUCTION.

In this Chapter, we use Monte Carlo experiments to assess the small sample performance of the equity estimator and some of its alternatives. We compare the performances of the alternative estimation procedures using various squared error loss criteria. The plan of the Chapter is as follows: In Section 5.2, we discuss the price promotion model used in the study as a basis for the Monte Carlo experiments and define the data set. Afterwards, we perform collinearity diagnostics on the data in order to check the severity of the multicollinearity problem. In Section 5.3, we re-introduce the equity estimator and some of its alternatives, which include the least squares estimator, two Stein-like estimators and ridge regression. In Section 5.4, we describe the nature of the Monte Carlo experiments as well as the method used in constructing the values of the dependent variable. In addition, we report the findings of the Monte Carlo experiments.

5.2 PRICE PROMOTION MODEL.

5.2.1 PRICE PROMOTION MODEL.

A price promotion model is used as a basis for studying the small sample properties of the equity estimator and its

alternatives. This particular model relates the unit sales of a target brand of a product to its own price, the prices of its competitors, price discounts and dummy variables for the types of advertising being used. The model is

$$\begin{aligned} \ln s_{1t} = {} & \alpha_0 + \alpha_1 R_{1t} + \alpha_2 d_{2t} + \alpha_3 I_{MAD1t} \\ & + \alpha_4 I_{DIS1t} + \alpha_5 I_{DISMAD1t} \\ & + \sum_{i=2,4} (\beta_{12} d_{it} + \beta_{13} I_{MADit} \\ & + \beta_{14} I_{DISit} + \beta_{15} I_{DISMADit}) \\ & + \sum_{n=2}^{13} \mu_n MN_{nt} + e_{1t} \qquad . \qquad (5.1) \end{aligned}$$

The variables are

S_{it} = unit sales of brand i

R_{it} = regular price of brand i

P_{it} = actual price of brand i

$d_{it} = (R_{it} - P_{it})/R_{it}$ = price discount of brand i

I_{MADit} = major ad only indicator

I_{DISit} = display only indicator

$I_{DISMADit}$ = display and major ad indicator

MN_{nt} = month effect

e_{it} = disturbance terms

where i = 1 for target brand and i = 2,3 and 4 for competitive brands. The analysis of the price promotion model is based on weekly data from a Nielson SCANTRACK^R Major Market. The product class studied is canned tuna and the target brand (brand #1) and brand #2 are major brands. The data set contains 52 store-weeks of data for a chain store.

5.2.2 COLLINEARITY DIAGNOSTICS.

We report the condition numbers for the actual data, mean centered data and centered and normalized to unit length data in Table 5.1. The condition number associated with the eigenvalue λ_j is defined as the square root of the ratio λ_1/λ_j , $j = 1,2,\ldots,K$. Centering the data improves the conditioning of the regressors by changing the origin of the data ellipsoid. The uncentered data exhibits severe collinearity due to the fact that the means of the explanatory variables are far from the origin. For further discussion about centering data see Belsley(1984) and Hill(1987).

We reduce the problem of multicollinearity by centering and normalizing the data to unit length the data as shown by the condition numbers in the third column of Table 5.1. Multicollinearity appears not to be a problem when the data is centered and normalized to unit length. By scaling the data to unit length, we eliminate the variability induced by the choices of units of measurement which in turn affect the measured degree of multicollinearity.

5.3 THE EQUITY ESTIMATOR AND ITS ALTERNATIVES.

5.3.1 EQUITY ESTIMATOR.

In order to conform with KR(1987), we apply the equity estimator to the centered and normalized to unit length

TABLE 5.1

The condition numbers for the untransformed and transformed data.

actual	centered	centered and normalized
1.000	1.000	1.000
2.961	1.037	1.249
3.056	1.105	1.277
3.201	1.388	1.341
4.088	1.589	1.382
4.579	1.783	1.545
5.134	1.873	1.600
5.394	1.965	1.689
5.659	2.000	1.714
5.761	2.025	1.791
5.849	2.032	1.832
5.854	2.080	1.865
5.991	2.166	1.982
6.239	2.406	2.148
6.932	2.690	2.390
7.571	2.787	2.505
8.118	3.342	2.665
9.642	3.582	3.387
10.331	4.727	3.488
13.623	5.034	4.287
14.677	6.921	4.678
20.680	9.085	5.201
27.692	12.340	7.095
35.576	14.270	7.357
41.130	33.682	11.501
97.459	67.876	26.973
258.501		

data. Let $\bar{x}_k$ denote the mean of the explanatory variable x_k and $\bar{y}$ denote the mean of the endogenous variable y. The variables x_k and y are centered and normalized to unit length by

$$x_{tck} = \frac{x_{tk} - \bar{x}_k}{\sqrt{\sum_t (x_{tk} - \bar{x}_k)^2}} \qquad (5.2a)$$

and

$$y_{tc} = \frac{y_t - \bar{y}}{\sqrt{\sum_t (y_t - \bar{y})^2}} \quad , \qquad (5.2b)$$

where t = 1,2,...,T and k = 2,3,...,K.

Let $\underset{\sim}{y}_c = [y_{1c}, y_{2c}, \ldots, y_{Tc}]'$ and $X_c = [x_{.c2}, x_{.c3}, \ldots, x_{.cK}]$. We call the matrix $(X_c'X_c)$ the correlation matrix of X abbreviated as r_{xx}. The elements of the matrix r_{xx} are the coefficients of correlation among the exogenous variables excluding the intercept term. That is

$$r_{xx} = \begin{bmatrix} 1 & r_{23} & \cdots & r_{2K} \\ & 1 & \cdots & r_{3K} \\ & & & \vdots \\ & & & 1 \end{bmatrix} \quad , \qquad (5.3)$$

where r_{ij} is the coefficient of correlation between the variables x_i and x_j; i and j = 2,3,...,K. Note that r_{xx} is a symmetric square matrix with diagonal elements equal to 1.

Similarly, the vector $(X_c'\underset{\sim}{y}_c)$, denoted as r_{xy}, is a (K-1) dimension vector with its elements being the coefficients of correlation between the endogenous variable y and the exogenous variable omitting the intercept term.

Consequently, the equity estimator can be described as

$$\hat{\underline{\beta}}_{Ec} = \hat{\nu}_c \hat{\underline{\eta}}_c = (\tilde{Z}_c' \tilde{Z}_c)^{-1} \tilde{Z}_c' \underline{y}_c$$

where

$$\tilde{Z}_c = X_c (X_c' X_c)^{-1/2} \quad .$$

$$\hat{\nu}_c = (\underline{w}_c' \underline{w}_c)^{-1} \underline{w}_c' \underline{y}_c$$

and

$$\underline{w}_c = X_c \hat{\underline{\eta}}_c \quad .$$

The estimate $\hat{\underline{\beta}}_{Ec}$ is associated with the centered and normalized variables. Therefore, we have to convert $\hat{\underline{\beta}}_{Ec}$ back into the original parameter space. Let $\hat{\beta}_{Eck}$ be the k-th element of the vector $\hat{\underline{\beta}}_{Ec}$, then $\hat{\underline{\beta}}_{Eck}$ can be transformed to the original parameter space by the relationship

$$\hat{\beta}_{Ek} = \hat{\beta}_{Eck} \frac{S_y}{S_k} \qquad (5.4)$$

where S_y and S_k are the standard deviations of the dependent variable y and the regressor x_k, respectively.

Let $\hat{\underline{\beta}}_{EO} = [\hat{\beta}_{E2}, \hat{\beta}_{E3}, \ldots, \hat{\beta}_{EK}]'$. We observe that the vector $\hat{\underline{\beta}}_{EO}$ does not contain the estimate for the intercept term. We compute the intercept term denoted as $\hat{\beta}_{E1}$ by

$$\hat{\beta}_{E1} = \bar{y} - \hat{\beta}_{E2} \bar{x}_2 - \ldots - \hat{\beta}_{EK} \bar{x}_K \quad . \qquad (5.5)$$

Finally, the equity estimator expressed in the original parameter space is

$$\hat{\underline{\beta}}_E = \begin{bmatrix} \hat{\beta}_{E1} \\ \hat{\underline{\beta}}_{EO} \end{bmatrix} \quad . \qquad (5.6)$$

5.3.2 THE ALTERNATIVES.

In this section, we present the biased estimator alternatives to the equity estimator. The estimators considered are the least squares estimator, two Stein-like estimators and a ridge regression estimator.

1. Least squares.

The least squares estimator is

$$\underline{b} = (X'X)^{-1}X'\underline{y} \quad . \tag{5.7}$$

2. Stein-like principal component estimator or PC-Stein.

We elect to apply PC-Stein estimator to the mean centered data set due to the better conditioning of the data. The mean centered data is written as

$$x_{tDk} = x_{tk} - \bar{x}_k \tag{5.8a}$$

and

$$y_{tD} = y_t - \bar{y} \tag{5.8b}$$

where $t = 1,2,\ldots,T$ and $k = 2,3,\ldots,K$. Let $X_D = [x_{.D2}, x_{.D3}, \ldots, x_{.DK}]$ and $\underline{y}_D = [y_{1D}, y_{2D}, \ldots, y_{TD}]'$. In addition, let $\underline{b}_D$ denote the least squares estimate of the centered data, that is

$$\underline{b}_D = (X_D'X_D)^{-1}X_D'\underline{y}_D \quad . \tag{5.9}$$

Let P_D be the matrix whose columns are the orthogonal characteristic vectors of $(X_D'X_D)$ corresponding to the ordered characteristic roots of $(X_D'X_D)$ which are contained in the diagonal matrix $\Lambda_D = \text{diag}(\lambda_{1D}, \lambda_{2D}, \ldots, \lambda_{(K-1),D})$ such

that $\lambda_{1D} \geq \lambda_{2D} \geq \ldots \geq \lambda_{(K-1),D}$. Following the procedure outlined in Section 4.5, we are able to construct a restriction matrix R such that

$$R\underline{\beta}_D = P_{2D}\underline{\beta}_D \tag{5.10}$$

where $\underline{\beta}_D$ is the parameter vector $\underline{\beta}_D$ without the intercept term. For our study, we disregard the small eigenvalues which contribute in total roughly 5% of the total variation. In other words, we treat the last n eigenvalues of the vector Λ_D such that $\sum_{i=n}^{K-1} \lambda_{i,D} / \sum_{i=1}^{K-1} \lambda_{i,D} \cong 0.05$ as neglectable. This selection method permits the eigenvalues that are relatively small to be eliminated while maintaining most of the variability in the data. The variability is reflected in the eigenvalues, Equation (4.5). As a result, the number of restrictions used is 10,

$$R\underline{\beta}_D = \underline{r} = \underline{0}$$

where $\underline{r}$ is a null vector of dimension J, J = 10.

With the restriction above, we express the restricted least squares estimator as

$$\underline{b}_{DR} = \underline{b}_D - (X_D'X_D)^{-1}R'[R(X_D'X_D)^{-1}R']^{-1}R\underline{b}_D \tag{5.11}$$

Referring to Section 4.5, the positive rule PC-Stein estimator is

$$\underline{\beta}_D^+ = \left[1 - \frac{c}{u} \right]_+ (\underline{b}_D - \underline{b}_{DR}) + \underline{b}_{DR} \tag{5.12}$$

where

$$\left[1 - \frac{c}{u} \right]_+ = \max\left[1 - \frac{c}{u}, 0 \right]$$

and

$$u = \frac{(R\underline{b} - \underline{r})'[R(X_D'X_D)^{-1}R']^{-1}(R\underline{b} - \underline{r})}{J\hat{\sigma}^2}$$

which is the likelihood ratio test statistic of the restrictions $R\underline{\beta}_D = \underline{0}$. The constant c is specified as

$$c = \frac{a(T-K1)}{J}$$

where K1 = K-1 and $0 < a < a_{max}$.

$$a_{max} = \frac{2}{(T-K1+2)}\left[\frac{tr\{[R(X_D'X_D)^{-1}R']^{-1}R(X_D'X_D)^{-1}D(X_D'X_D)R'\}}{\eta} - 2\right].$$

η is the largest eigenvalue of the term in the brackets { }. D is a weight matrix which is positive definite. There are several choices of D. We select D=I and $D = (X_D'X_D)$. Note that the matrix D mentioned above need not be the same as the matrix D in the weight squared error loss function in Equation (4.6). With the weight being $X_D'X_D$, we are evaluating the estimator based on its predictive ability. With the weight being I, we are evaluating the estimator based on its mean square error or its ability of estimating the parameter.

When we use $D = (X_D'X_D)$, the constant a_{max} is reduced to

$$a_{max} = \frac{2(J-2)}{(T-K1+2)} \quad .$$

We find the shrinkage constant a by averaging 0 and a_{max}; hence,

$$a = \frac{a_{max}}{2} \quad .$$

We have shown in Section 4.5 that the value of a such that $0 < a < a_{max}$ yields Stein-like estimator that has risk less than or equal to those of least squares when the number of restriction is greater than three.

Similar to the case of the estimate $\hat{\underline{\beta}}_{EO}$ of the equity estimator, the vector $\underline{\beta}_D^+$ does not contain the estimate for the intercept term. The intercept can be estimated by the relationship stated in Equation (5.6).

3. Stein-like estimator (Stein).

The Stein estimator is almost identical to PC-Stein estimator except for the restrictions. The restrictions are

$$\underline{\beta}_D = \underline{0}$$

which implies that we shrink all parameters in $\underline{\beta}_D$ towards the origin. In other words, the restriction matrix **R** is an identity matrix; therefore, J = K1.

Once again, we employ two weight matrices, **D=I** and $\mathbf{D}=(\mathbf{X}_D'\mathbf{X}_D)$. When we use **D=I**, the constant a_{max} is simplified to

$$a_{max} = \frac{2\ tr(\mathbf{X}_D'\mathbf{X}_D)^{-1}\lambda_{(K-1),D}}{(T-K1+2)} \quad .$$

When we use $\mathbf{D}=\mathbf{X}_D'\mathbf{X}_D$, the constant a_{max} is reduced to

$$a_{max} = \frac{2(K1-2)}{(T-K1+2)} \quad .$$

4. Ridge regression.

Ridge regression is commonly applied to data presented in the correlation matrix form. Let $\underline{b}_c$ be the least squares estimate of the standardized data, ie.

$$\underset{\sim}{b}_c = r_{xx}^{-1} \cdot r_{xy} \quad .$$

Let $\hat{\sigma}_c^2 = (\underset{\sim}{y}_c - X_c \underset{\sim}{b}_c)'(\underset{\sim}{y}_c - X_c \underset{\sim}{b}_c)/(T-K1)$. The ridge regression is expressed as

$$\hat{\underset{\sim}{\beta}}^*_{c0}(k) = \{ r_{xx} + kI_{K1} \}^{-1} r_{xy} \quad ,$$

where

$$k = \frac{K1\hat{\sigma}_c^2}{\underset{\sim}{b}_c' \underset{\sim}{b}_c} \quad .$$

The estimator $\hat{\underset{\sim}{\beta}}^*_{c0}(k)$ is associated with the centered and normalized variables and, as a consequence, have to be converted back to the original parameter space. Subsequently, we have to estimate the intercept term as well. The procedures for converting the estimate into the original parameter space and calculating the intercept term are described in Section 5.3.1.

5.4 MONTE CARLO EXPERIMENTS

5.4.1 PARAMETERS ESTIMATION.

Before performing the Monte Carlo experiments, we apply the equity estimator and its alternatives to the original price-promotion data used in our study. We expect the estimates of α_1 and β_{ij}, i = 2,3,4 and j = 2,4,5 to be negative; brands #2, #3, and #4 do not use major ad only promotions. The parameter α_1 measures the own price effects on units sold of the target brand and the parameters β_{ij} measure the price discount effects and promotion effects of

competitive brands on units sold of the target brand. The estimates of the parameters α_2, α_3, α_4 and α_5 are expected to have positive signs since these variables reflect the own price discount effects and own promotion effects on units sold of the target brand. In addition, we expect that the parameters associated with the variable indicating the use of display and major promotion campaign of a particular brand to be larger in absolute value that the parameters associated with the variables indicating display only and major ad only campaign. In other words, we expect the use of combination of promotion campaigns to be more effective than the use of just one campaign. This is called the inequality restriction.

We present the parameter estimates obtained from the equity estimator and its alternatives in Table 5.2. Brands #2, #3 and #4 do not use major ad only promotion. The findings are

1. The Stein (D=I) and PC-Stein (D=I) estimator yield identical results to those of least squares. The reason is that the constant a_{max} becomes 0 and the two Stein-like estimators offer no improvement over the least squares estimator.

2. The least squares, Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) estimator yield incorrect signs for the estimates of the parameter α_1, β_{22}, β_{34}, β_{35} and β_{42}. Among these estimates, the ones that we are most concerned with are the estimates of α_1 which is the parameter measuring own price

TABLE 5.2

The estimated parameter values of the equity estimator and its alternatives.

Parameter	Equity	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS	HCA.
α_0	10.105	9.513	2.071	3.019	0.945	3.55
α_1	-2.012	-1.347	6.619	5.619	7.900	-1.25
α_2	1.451	1.645	1.294	1.217	1.544	2.01
α_3	0.489	0.326	0.242	0.240	0.289	1.53
α_4	0.162	0.259	0.201	0.203	0.240	1.73
α_5	0.968	1.341	1.214	1.402	1.448	2.53
β_{22}	-0.575	0.377	0.400	0.278	0.477	-0.67
β_{24}	-0.105	-0.287	-0.229	-0.309	-0.273	-0.15
β_{25}	-0.316	-0.495	-0.471	-0.562	-0.562	-0.28
β_{32}	-1.572	-3.040	-3.533	-3.081	-4.217	-1.01
β_{34}	0.198	0.159	0.119	0.103	0.142	-0.40
β_{35}	-0.190	0.313	0.600	0.308	0.716	-0.58
β_{42}	-0.880	-0.404	1.303	1.073	1.555	-0.60
β_{44}	-0.231	-0.306	-0.347	-0.306	-0.414	-0.08
β_{45}	-0.357	-0.509	-0.714	-0.819	-0.853	-0.15
μ_2	0.080	-0.096	-0.220	-0.066	-0.262	0.10
μ_3	0.071	-0.069	-0.420	-0.321	-0.502	0.04
μ_4	-0.168	-0.372	-1.056	-0.905	-1.261	0.14
μ_5	-0.088	-0.304	-1.015	-0.785	-1.212	0.23
μ_6	0.211	0.002	-0.973	-0.769	-1.162	0.28
μ_7	-0.363	-0.517	-1.561	-1.466	-1.863	0.20
μ_8	-0.720	-0.858	-1.937	-1.892	-2.312	0.17
μ_9	-0.039	-0.266	-1.427	-1.208	-1.703	0.17
μ_{10}	-0.174	-0.547	-1.863	-1.614	-2.224	0.18
μ_{11}	0.064	-0.072	-1.445	-1.185	-1.724	0.15
μ_{12}	-0.252	-0.254	-1.358	-1.217	-1.621	0.17
μ_{13}	-0.264	-0.131	-1.095	-0.975	-1.307	0.14

note: HCA is the true parameter values used in the Monte Carlo experiments.

effects on units sold of our target brand, which must be negative. It appears that the presence of multicollinearity in the actual data and mean centered data cause these estimates to have incorrect signs.

3. The equity estimator and ridge regression estimator yield positive signs for the estimates of the parameter β_{34} which expected to have a negative sign. Only the equity estimator gives the correct sign for the parameter β_{22}.

4. For all of the estimators considered, the inequality restriction is achieved only if the parameters associated with the variables indicating promotion campaigns of a brand have the correct signs.

From Table 5.2, we observe that only the equity estimator yields estimates of key parameters with acceptable signs, except for the paramter β_{34}. Nevertheless, we do not have any information about the performance of the equity estimator in small samples compared to its alternatives. In order to resolve this issue, we introduce Monte Carlo experiments into our study.

5.4.2 THE NATURE OF MONTE CARLO EXPERIMENTS.

A Monte Carlo experiment is a simulation exercise designed to investigate the small sample properties of estimators. In this experiment, we assume that the exact nature of the relationship between the endogenous variable and its explanatory variables are known. Consider the model

$$\underline{y} = X\underline{\beta} + \underline{e} \quad , \tag{5.13}$$

where $\underline{y}$ is a (T×1) vector of observations on a dependent variable, $\mathbf{X}$ is a given (T×K) matrix of exogenous variables and $\underline{e}$ is a (T×1) random vector which is normally distributed with mean vector $\underline{0}$ and covariance matrix $\sigma^2 I_T$.

We assume that the real values of $\underline{\beta}$ and σ^2 are known. Normal random number generator is used to construct N samples of the vector $\underline{e}$ which are all normally distributed with mean vector $\underline{0}$ and covariance matrix $\sigma^2 I_T$. Subsequently, we add the generated vectors to the vector $X\underline{\beta}$ and obtain N repeated samples of the vector $\underline{y}$.

The next stage is to select the vector of actual parameter values to be used in generating N repeated samples of $\underline{y}$. We use the parameter estimates presented in Hill, Cartwright and Arbaugh (1991), which utilizes seemingly unrelated regressions technique, as the vector of actual parameter values since they have the correct signs and follow the inequality restrictions. The use of estimates obtained from the estimator considered in Table 5.2 may give advantage to the estimator whose estimated parameters are selected. We use the estimate of σ^2 obtained via the least squares estimator as the true value, in this case, σ^2 = 0.169431.

The number of repeated samples of $\underline{y}$ generated is 500. We use the Monte Carlo experiment to study three aspects of the small sample performance of the equity estimator and its alternatives.

1. We evaluate the small sample performances of the

equity estimator and its alternatives via the weighted square error loss measure, Equation (4.6). The weight matrices used are I and X'X. With the identity matrix as the weight matrix, we get mean square error loss and the estimator is evaluated on its ability to estimate the parameters vector. With the X'X as the weight matrix, we get the mean square error of prediction loss and the estimator is evaluated on its ability to predict the in-sample observations.

We perform Monte Carlo experiments utilizing various parameter lengths. Let $\underline{\beta}_0$ denote the estimates obtained from the column labeled HCA in Table 5.2 and c be any constant, then the parameter $\underline{\beta}$ in Equation (5.13) is defined as

$$\underline{\beta} = c\underline{\beta}_0 \tag{5.14}$$

c is the parameter length. We alter the parameter length of $\underline{\beta}$ by altering the constant c. Finally, we calculate the signal to noise ratio (R) associated with the values of c and σ^2 by the relationship

$$R = \frac{c^2\underline{\beta}'\underline{\beta}}{\sigma^2} \tag{5.15}$$

2. As a second part of the study, we perform the Monte Carlo experiments on the equity estimator and its alternatives using various vector lengths in the data generating process, just like in the previous part, but we are going to focus on the estimation of a single parameter instead of the estimation of the whole parameter vector. We

select to study the estimation of the parameter α_2, which is the own discount effects. The estimates of α_2 obtained from the equity estimator and its alternatives are evaluated based on their mean square error and bias.

3. Finally, we discuss the performances of the equity estimator when the data is not well conditioned. We have shown that when the data is centered and normalized to unit length the degree of multicollinearity decreases dramatically. Therefore, by applying the equity estimator to the centered and normalized to unit length data, we fail to test the performances of the equity estimator when there is severe multicollinearity. Consequently, we apply the equity estimator to the actual data and mean centered data in the content of Monte Carlo experiments. In our experiments, we use various parameter lengths in the data generating process. The performance of the equity estimator is then evaluated based on the mean square errors, prediction mean square errors, mean square errors of the estimates of α_2 and the bias of the estimates of α_2.

5.4.3 MONTE CARLO EXPERIMENTS RESULTS.

1.Mean square and prediction mean square error criteria.

We examine the small sample properties of the equity estimator and its alternatives by varying the vector length in order to alternate the signal to noise ratio.

In Tables 5.3 and 5.4, we present the mean square

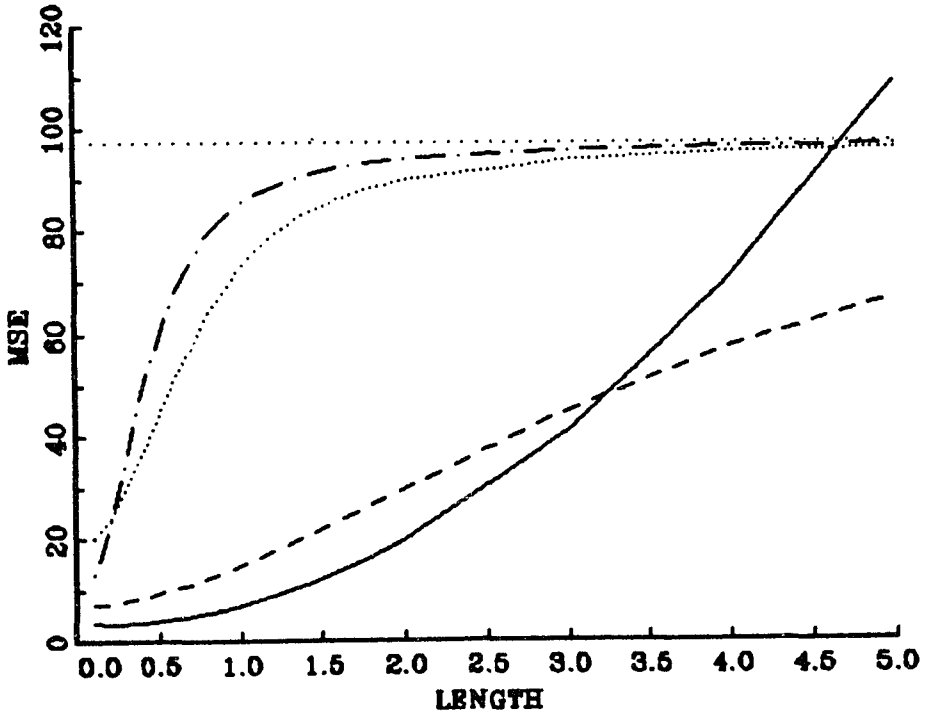
TABLE 5.3

The average mean square errors of the equity estimator and its alternatives associated with various vector lengths.

length(R)	Equity	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS
0.1(1.934)	3.434	7.240	13.131	20.163	97.160
0.2(7.734)	3.385	7.556	22.911	23.977	97.160
0.4(30.94)	3.636	8.679	50.333	36.889	97.160
0.6(69.60)	4.378	10.303	69.361	51.878	97.160
0.8(123.7)	5.517	12.336	79.599	64.133	97.160
1.0(193.4)	7.017	14.721	85.271	72.893	97.160
1.2(278.4)	8.868	17.399	88.644	78.936	97.160
1.4(379.0)	11.064	20.300	90.785	83.115	97.160
1.6(495.0)	13.603	23.357	92.219	86.067	97.160
1.8(626.4)	16.485	26.509	93.224	88.206	97.160
2.0(773.4)	19.709	29.702	93.953	89.796	97.160
3.0(1740)	40.947	44.974	95.715	93.769	97.160
4.0(3094)	70.713	57.467	96.344	95.229	97.160
5.0(4834)	109.004	66.755	96.637	95.918	97.160

FIGURE 5.1

The average mean square errors of the equity estimator and its alternatives associated with various vector lengths.



---RIDGE ———EQUITY -.-.- STEIN .....PC-STEIN . . . OLS

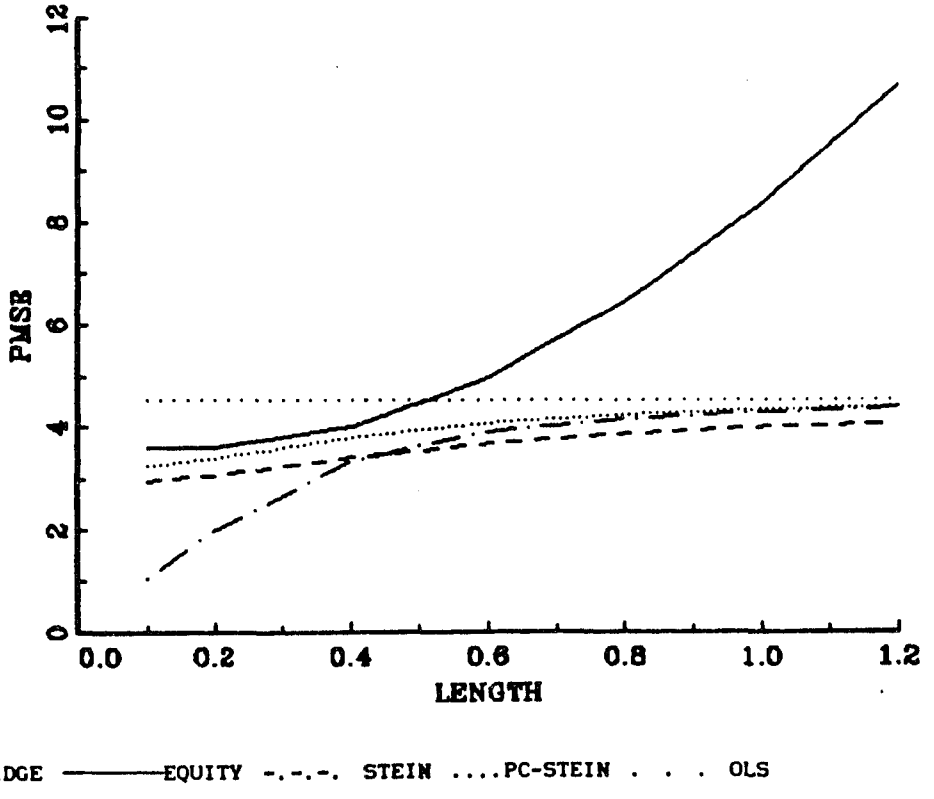
TABLE 5.4

The average prediction mean square errors of the equity estimator and its alternatives associated with various vector lengths.

length(R)	Equity	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS
0.1(1.934)	3.588	2.948	1.040	3.229	4.536
0.2(7.734)	3.589	3.069	1.983	3.382	4.536
0.4(30.94)	4.002	3.388	3.338	3.784	4.536
0.6(69.60)	4.978	3.661	3.901	4.081	4.536
0.8(123.7)	6.431	3.851	4.156	4.246	4.536
1.0(193.4)	8.330	3.979	4.287	4.337	4.536
1.2(278.4)	10.665	4.067	4.362	4.392	4.536
1.4(379.0)	13.430	4.130	4.408	4.427	4.536
1.6(495.0)	16.625	4.179	4.438	4.451	4.536
1.8(626.4)	20.249	4.217	4.459	4.468	4.536
2.0(773.4)	24.301	4.248	4.474	4.480	4.536
3.0(1740)	50.973	4.349	4.510	4.510	4.536
4.0(3093)	88.329	4.405	4.522	4.521	4.536
5.0(4833)	136.366	4.440	4.528	4.526	4.536

FIGURE 5.2

The average prediction mean square errors of the equity estimator and its alternatives associated with various vector lengths.



errors (MSE) and prediction mean square error (PMSE) of the equity estimator and its alternatives when the vector lengths are increasing. From Table 5.3 and Figure 5.1, we observe that the equity estimator has the lowest MSE when the vector length is less than 3. As the vector length surpasses 3, the MSE of the equity estimator increases rapidly. Unlike the other biased estimators considered, the MSE of the equity estimator exceeds that of the least squares estimator for high values of vector length. The PC-Stein ($D=X_D'X_D$) estimator produces MSE's that are slightly lower than those of the Stein ($D=X_D'X_D$) estimator. The Stein (D=I) and PC-Stein (D=I) estimator yield the same values of MSE as those of the least squares estimator for all values of vector length.

From Table 5.4 and Figure 5.2, we observe that the equity estimator has the highest PMSE among the biased estimators considered. Furthermore, the PMSE of the equity estimator exceed that of the least squares estimator for vector lengths in excess of 0.6. We note that at large vector length, the PMSE's of the equity estimator are much higher than those of its alternatives. The Stein ($D=X_D'X_D$) estimator gives the lowest PMSE for vector lengths less than 0.4, while the ridge regression estimator gives the lowest PMSE for vector lengths greater than 0.4. The Stein (D=I) and PC-Stein (D=I) yield identical PMSE's to those of the least squares estimator for all vector lengths.

2. <u>The estimation of α_2</u>.

We are now going to study the performance of the equity estimator and its alternatives in estimating a parameter of interest which is α_2 or own discount effects. Again, we conduct the study when the vector length used in the data generating process is increasing. The variable d_2, the own discount variable, is one of the variables that have small eigenvalues which indicates strong linear relationship with the other variables. From Table 5.5 and Figure 5.3, we observe that the equity estimator produces the smallest MSE for all values of vector length greater than 0.1. In addition, for vector lengths exceeding 0.1, the MSE's obtained from the equity estimator are much smaller than those of its alternatives. Nevertheless, the bias of the equity estimator increases steadily as the vector length surpasses 0.6 as demonstrated in Table 5.6 and Figure 5.4.

3. Applying the equity estimator to data with various degree of multicollinearity.

We apply the equity estimator to generated data that are not centered and normalized to unit length. The purpose is to study the performance of the equity estimator when the data is not well conditioned. As discussed earlier, the equity estimator is designed to handle the problem of estimation in the presence of multicollinearity and, as a consequence, should perform well even if there is severe multicollinearity. By centering and normalizing to unit length, we have drastically lessened the degree of multicollinearity. Thus, by applying the equity estimator to

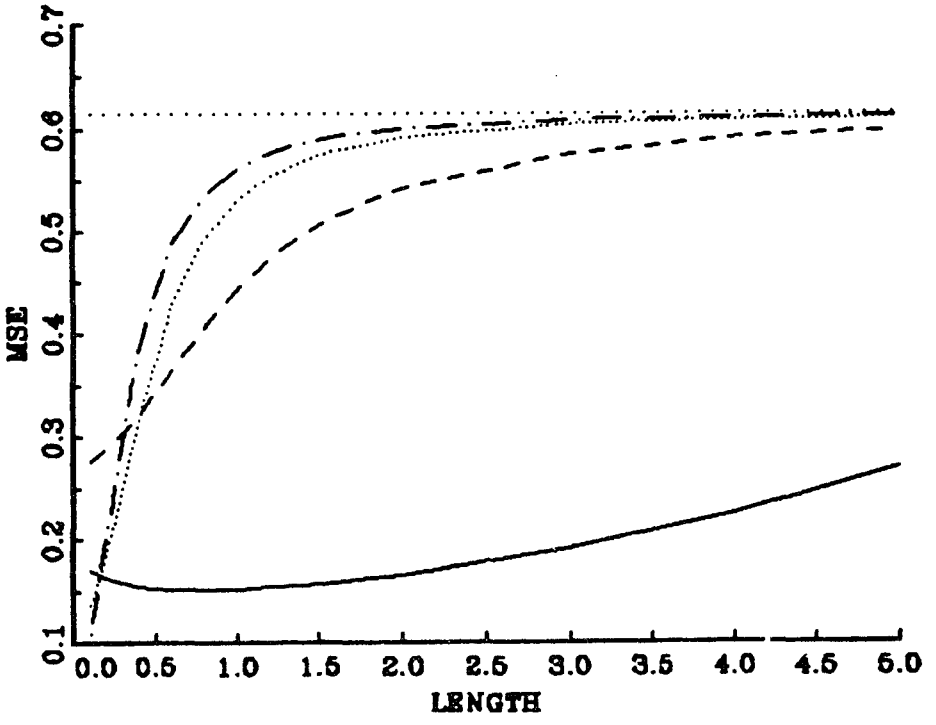
TABLE 5.5

The average mean square errors of the estimates of α_2 of the equity estimator and its alternatives associated with various vector lengths.

length(R)	Equity	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS
0.1(1.934)	0.171	0.276	0.108	0.137	0.614
0.2(7.734)	0.163	0.287	0.209	0.186	0.614
0.4(30.94)	0.154	0.322	0.391	0.318	0.614
0.6(69.60)	0.151	0.365	0.487	0.427	0.614
0.8(123.7)	0.151	0.406	0.535	0.492	0.614
1.0(193.4)	0.152	0.442	0.561	0.529	0.614
1.2(278.4)	0.154	0.472	0.576	0.553	0.614
1.4(379.0)	0.156	0.496	0.586	0.568	0.614
1.6(495.0)	0.158	0.514	0.592	0.578	0.614
1.8(626.4)	0.162	0.529	0.597	0.585	0.614
2.0(773.4)	0.165	0.541	0.600	0.591	0.614
3.0(1740)	0.190	0.575	0.608	0.604	0.614
4.0(3093)	0.224	0.590	0.610	0.608	0.614
5.0(4833)	0.269	0.598	0.612	0.610	0.614

FIGURE 5.3

The average mean square errors of the estimates of α_2 of the equity estimator and its alternatives associated with various vector lengths.



----RIDGE ————EQUITY -.-.-. STEIN ....PC-STEIN . . . OLS

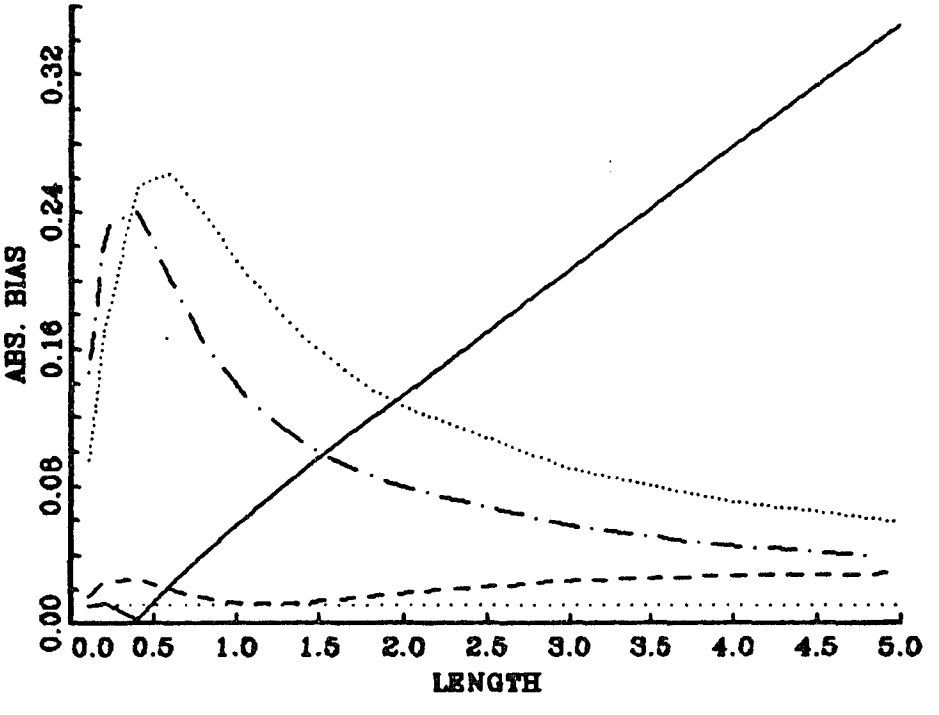
TABLE 5.6

The average bias of the estimates of α_2 of the equity estimator and its alternatives associated with various vector lengths.

length(R)	Equity	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS
0.1(1.934)	0.010	-0.016	-0.146	-0.095	-0.011
0.2(7.734)	0.012	-0.024	-0.231	-0.173	-0.011
0.4(30.94)	-0.002	-0.026	-0.241	-0.138	-0.011
0.6(69.60)	-0.022	-0.020	-0.199	-0.262	-0.011
0.8(123.7)	-0.040	-0.015	-0.164	-0.239	-0.011
1.0(193.4)	-0.057	-0.012	-0.138	-0.211	-0.011
1.2(278.4)	-0.073	-0.012	-0.120	-0.187	-0.011
1.4(379.0)	-0.089	-0.012	-0.105	-0.167	-0.011
1.6(495.0)	-0.104	-0.014	-0.094	-0.151	-0.011
1.8(626.4)	-0.119	-0.016	-0.086	-0.137	-0.011
2.0(773.4)	-0.133	-0.018	-0.079	-0.126	-0.011
3.0(1740)	-0.206	-0.025	-0.057	-0.090	-0.011
4.0(3093)	-0.278	-0.028	-0.045	-0.071	-0.011
5.0(4833)	-0.349	-0.029	-0.039	-0.059	-0.011

FIGURE 5.4

The average bias of the estimates of α_2 of the equity estimator and its alternatives associated with various vector lengths.



----RIDGE ————EQUITY -.-.-. STEIN ....PC-STEIN . . . OLS

the centered and normalized to unit length data, we are estimating the data set that is fairly well conditioned which may not indicate the usefulness of the equity estimator in coping with multicollinearity problem.

We are going to study the small sample properties of the equity estimator when applied to the actual data, mean centered data and centered and normalized to unit length data. First, we compare the parameter estimates of the original data set. Second, within the context of the Monte Carlo experiments, we study the small sample properties of the equity estimator when applied to various type of data as we increase the vector length in the data generating process.

3.1 Parameter estimates.

In Table 5.7, we present the parameter estimates obtained through applying the equity estimator to the centered and normalized to unit length data, the actual data and mean centered data. We find that the equity estimator yields vastly different estimates of the same parameter when different data conditioning tools are used. By applying the equity estimator to the actual data, we find that many of the estimates have incorrect signs.

3.2 Increasing vector length.

From Table 5.8 and Figures 5.5a and 5.5b, we find that the MSE's obtained by applying the equity estimator to the actual data have the lowest value when R is extremely small. Moreover, we notice that as the data used in the estimation

TABLE 5.7

The estimated parameter values of the equity estimator applied to various types of conditioned data.

Parameter	centered and normalized	actual	centered
α_0	10.105	3.163	7.938
α_1	-2.012	3.223	-0.018
α_2	1.451	0.381	0.570
α_3	0.489	0.119	0.243
α_4	0.162	1.008	0.291
α_5	0.968	0.457	1.085
β_{22}	-0.575	0.284	-0.086
β_{24}	-0.104	0.700	-0.183
β_{25}	-0.316	0.406	-0.486
β_{32}	-1.571	0.053	-0.638
β_{34}	0.198	0.886	0.372
β_{35}	-0.190	0.171	-0.358
β_{42}	-0.880	0.073	-0.129
β_{44}	-0.231	0.148	-0.239
β_{45}	-0.357	0.042	-0.406
μ_2	0.080	0.265	0.106
μ_3	0.071	0.317	0.108
μ_4	-0.168	0.154	-0.186
μ_5	-0.089	0.144	-0.098
μ_6	0.211	0.300	0.210
μ_7	-0.363	0.195	-0.356
μ_8	-0.720	0.109	-0.729
μ_9	-0.040	0.183	-0.081
μ_{10}	-0.174	0.084	-0.214
μ_{11}	0.064	0.144	-0.016
μ_{12}	-0.251	0.130	-0.334
μ_{13}	-0.264	0.179	-0.306

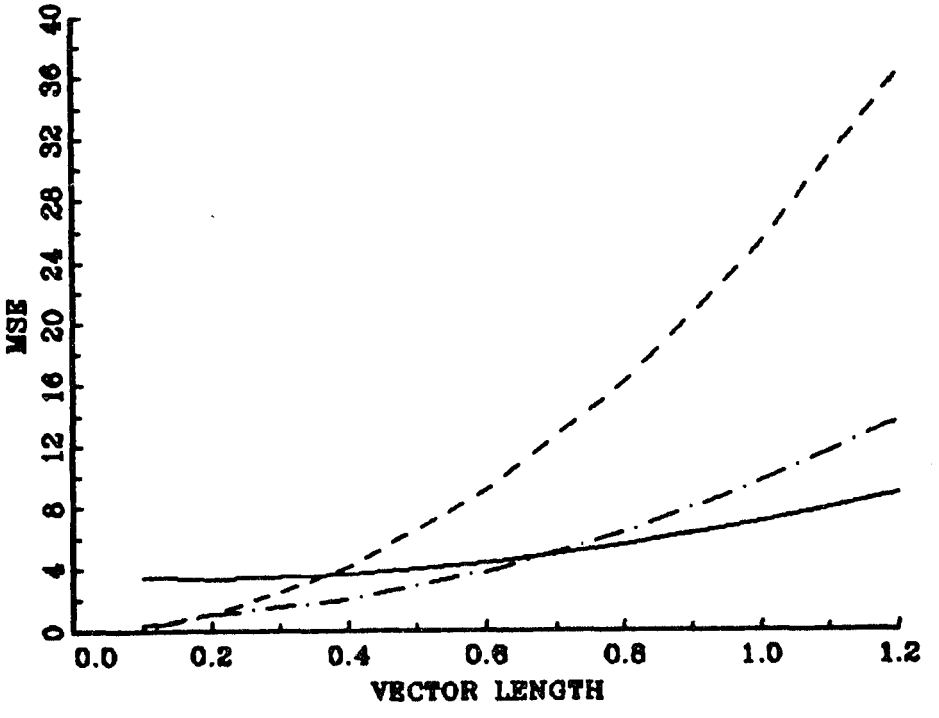
TABLE 5.8

The average mean square errors of the equity estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	3.434	0.289	0.092
7.734	0.20	3.385	1.035	1.058
30.94	0.40	3.636	4.044	2.000
69.60	0.60	4.378	9.061	3.773
123.7	0.80	5.517	16.085	6.313
193.4	1.00	7.017	25.116	9.599
278.4	1.20	8.868	36.154	13.623
379.0	1.40	11.064	49.199	18.383
495.0	1.60	13.603	64.250	23.879
626.4	1.80	16.485	81.309	30.109
773.4	2.00	19.708	100.374	37.084
1740	3.00	40.947	225.803	82.907
3094	4.00	70.713	401.403	147.087
4834	5.00	109.004	627.174	229.613

FIGURE 5.5a

The average mean square errors of the equity estimator applied to various types of conditioned data.

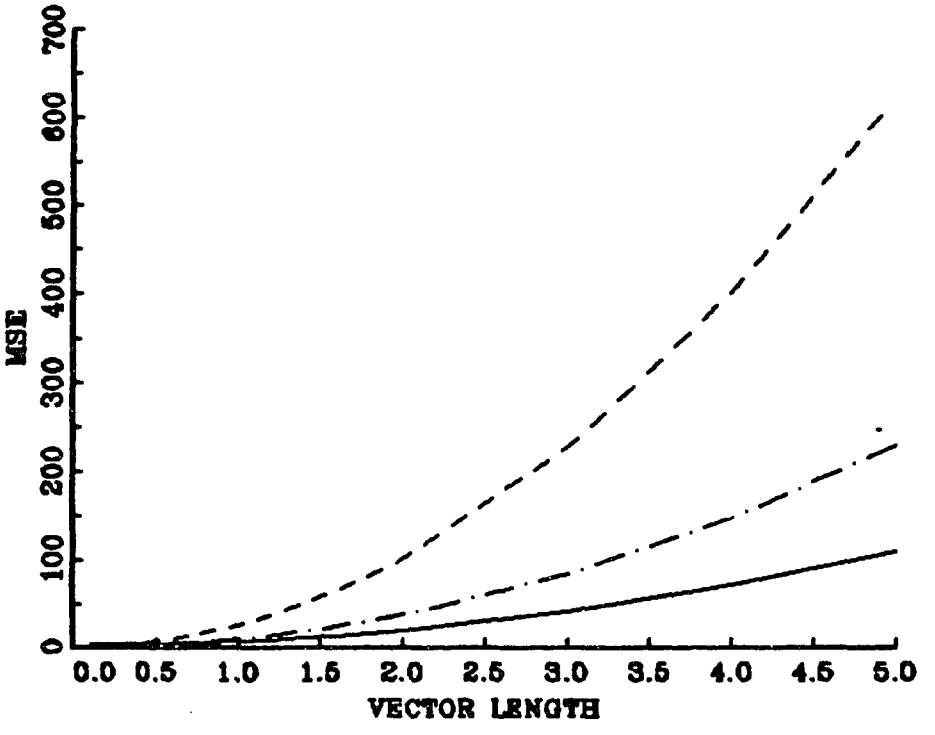


———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

FIGURE 5.5b

The average mean square errors of the equity estimator applied to various types of conditioned data.



——————— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

process become more well conditioned, the MSE's obtained via the equity estimator increases less drastically as R increases. We arrive at similar conclusions in the case of the PMSE's obtained from applying the equity estimator to the various type of conditioned data as presented in Table 5.9 and Figures 5.6a and 5.6b.

From Table 5.10 and Figure 5.7, we discover that the MSE's of the estimates of α_2 obtained through the equity estimator are very much dependent on the conditioning of the data. When the data are not centered and normalized to unit length, the MSE's obtained are remarkably low at small values of vector length. However, the MSE's increase at an accelerating pace as the values of vector length become higher. We find resemblance between the MSE's obtained from applying the equity estimator to the actual data and the mean centered data. Recall that when we examine the MSE's and PMSE's obtained via applying the equity estimator of the model as a whole, we find similarity between the MSE's and PMSE's obtained from the mean centered data and centered and normalized to unit length data instead.

From Table 5.11 and Figure 5.8a, we find that the bias of the estimates of α_2 obtained from applying the equity estimator to the actual data are much higher than those obtained from the normalized data especially around the vector length 0.5 where the bias acquired from the normalized data reaches its minimum. As the vector length surpasses 1, the bias obtained from the actual data and the

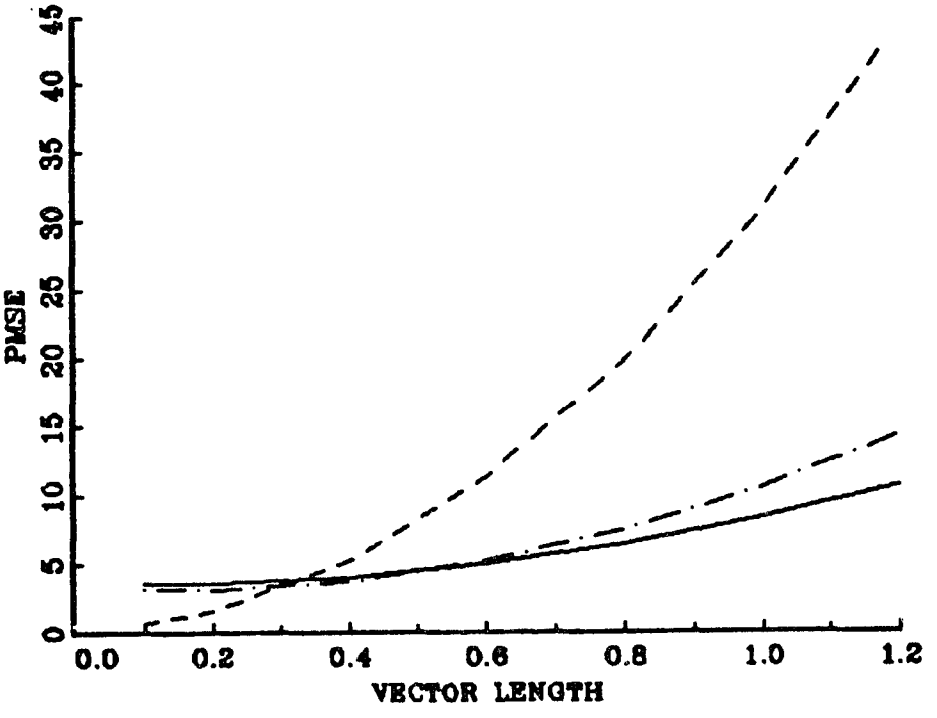
TABLE 5.9

The average prediction mean square errors of the equity estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	3.588	0.651	3.220
7.734	0.20	3.589	1.522	3.119
30.94	0.40	4.002	5.167	3.658
69.60	0.60	4.978	11.254	5.177
123.7	0.80	6.431	19.775	7.484
193.4	1.00	8.330	30.731	10.514
278.4	1.20	10.665	44.122	14.247
379.0	1.40	13.430	59.946	18.673
495.0	1.60	16.625	78.204	23.791
626.4	1.80	20.249	98.897	29.598
773.4	2.00	24.301	122.023	36.093
1740	3.00	50.973	274.168	78.883
3094	4.00	88.329	487.166	138.849
4834	5.00	136.366	761.016	215.987

FIGURE 5.6a

The average prediction mean square errors of the equity estimator applied to various types of conditioned data.

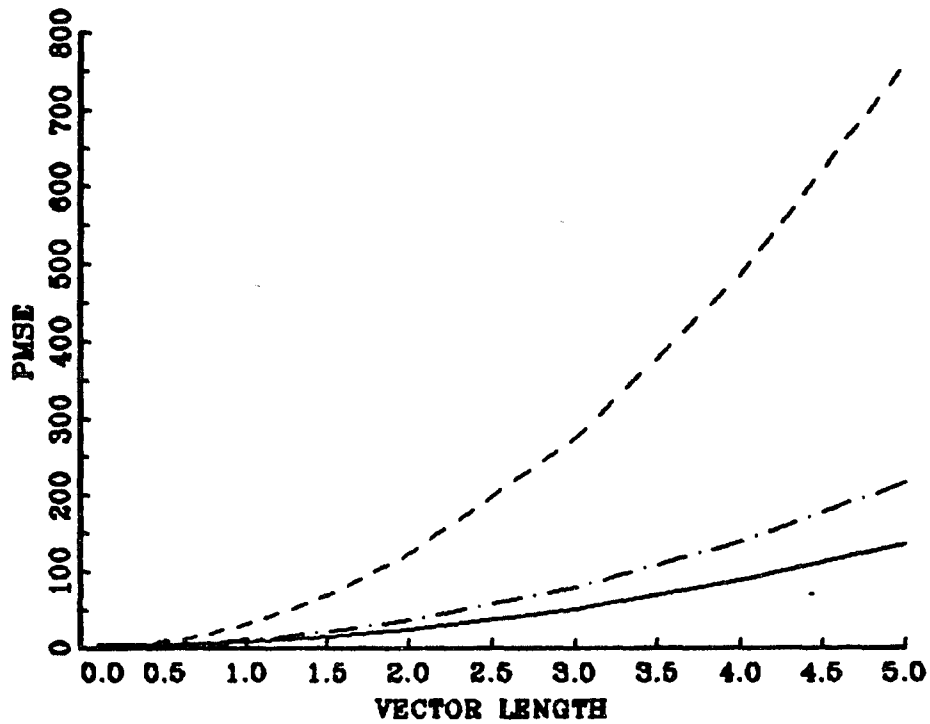


———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

FIGURE 5.6b

The average prediction mean square errors of the equity estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

TABLE 5.10

The average mean square errors of the estimates of α_2 of the equity estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.171	0.031	0.047
7.734	0.20	0.164	0.122	0.092
30.94	0.40	0.154	0.488	0.302
69.60	0.60	0.151	1.097	0.670
123.7	0.80	0.151	1.949	1.192
193.4	1.00	0.152	3.045	1.863
278.4	1.20	0.154	4.384	2.685
379.0	1.40	0.156	5.967	3.656
495.0	1.60	0.159	7.793	4.777
626.4	1.80	0.162	9.862	6.047
773.4	2.00	0.165	12.175	7.467
1740	3.00	0.190	27.390	16.809
3094	4.00	0.224	48.690	29.889
4834	5.00	0.269	76.076	46.706

FIGURE 5.7

The average mean square errors of the estimates of α_2 of the equity estimator applied to the actual and mean centered data relative to those of the centered and normalized data.

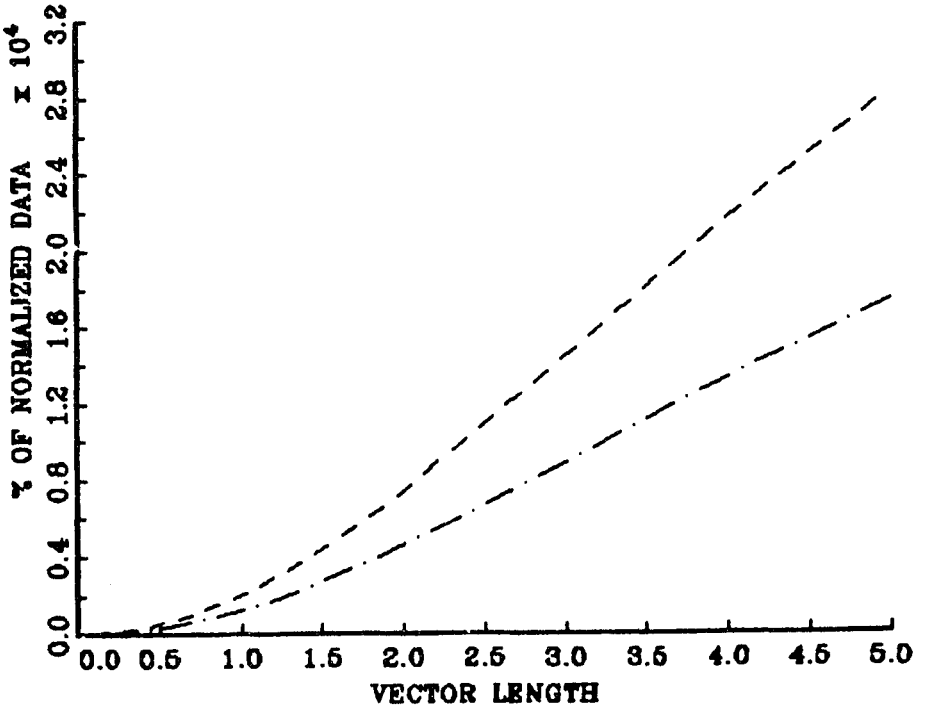


TABLE 5.11

The bias of the estimates of α_2 of the equity estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.010	-0.172	-0.121
7.734	0.20	0.013	-0.348	-0.250
30.94	0.40	-0.002	-0.698	-0.526
69.60	0.60	-0.022	-1.047	-0.804
123.7	0.80	-0.040	-1.396	-1.081
193.4	1.00	-0.057	-1.745	-1.357
278.4	1.20	-0.073	-2.094	-1.631
379.0	1.40	-0.089	-2.442	-1.906
495.0	1.60	-0.104	-2.791	-2.180
626.4	1.80	-0.119	-3.140	-2.455
773.4	2.00	-0.134	-3.489	-2.729
1740	3.00	-0.206	-5.233	-4.097
3094	4.00	-0.278	-6.978	-5.465
4834	5.00	-0.349	-8.722	-6.833

FIGURE 5.8a

The bias of the estimates of α_2 of the equity estimator applied to the actual and mean centered data relative to those of the centered and normalized data.

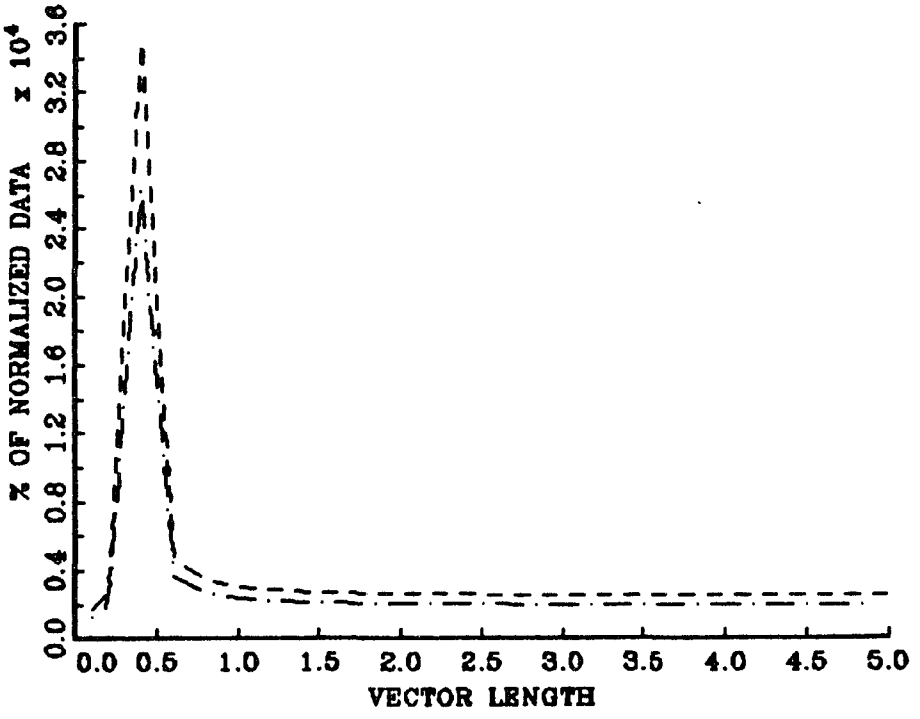
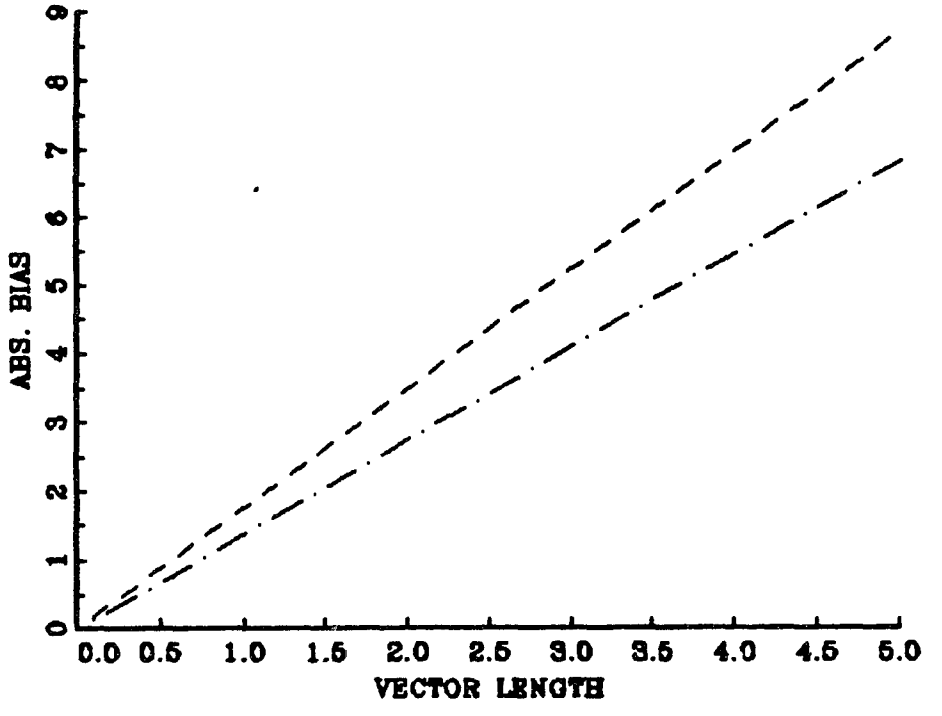


FIGURE 5.8b

The bias of the estimates of α_2 of the equity estimator applied to the actual and mean centered data relative to those of the centered and normalized data.



----- ACTUAL -.-.-. CENTERED

mean centered data appear to be a constant function of those acquired from the normalized data. Figures 5.4 and 5.8b demonstrate that the bias of the estimates of α_2 obtained by applying the equity estimator to various data show linear relationship with the vector length.

We now apply the ridge regression estimator, the Stein estimators and the PC-Stein estimators to various conditioned data set. The results are presented in Tables 5.12 - 5.16. Table 5.12 presents the parameter estimates obtained from applying the ridge regression estimator to various conditioned data. We discover that the estimates in Table 5.12 have more similarity across the data set than those of the equity estimator shown in Table 5.7. In addition, the estimates in Table 5.12 follow the inequality restriction. Nevertheless, only the estimate of α_1 obtained via the normalized data has the correct sign.

From Table 5.13 and Figure 5.9, we find that the application of the ridge regression estimator to the mean centered data produces the smallest MSE's. However, from Table 5.14 and Figure 5.10, the PMSE's vary slightly across data. The PMSE's of the normalized data are the smallest at small vector length. The differences among the PMSE's of various conditioned data set dwindle as the vector length increases. Unlike the case of the overall MSE, the MSE's of the estimates of α_2 obtained through the normalized data shows the best performance as shown in Table 5.15 and Figure 5.11. There is minimal difference between the MSE's of the

TABLE 5.12

The estimated parameter values of the ridge regression estimator applied to various types of conditioned data.

Parameter	centered and normalized	actual	centered
α_0	9.513	4.690	8.159
α_1	-1.347	3.668	0.136
α_2	1.645	1.545	1.446
α_3	0.326	0.335	0.364
α_4	0.259	0.318	0.271
α_5	1.341	1.441	1.433
β_{22}	0.377	0.430	0.511
β_{24}	-0.287	-0.288	-0.303
β_{25}	-0.495	-0.510	-0.531
β_{32}	-3.040	-2.740	-2.786
β_{34}	0.160	0.189	0.169
β_{35}	0.313	0.231	0.263
β_{42}	-0.404	-0.275	-0.182
β_{44}	-0.306	-0.268	-0.315
β_{45}	-0.509	-0.609	-0.604
μ_2	-0.096	-0.085	-0.173
μ_3	-0.069	-0.152	-0.169
μ_4	-0.372	-0.762	-0.556
μ_5	-0.304	-0.717	-0.486
μ_6	0.002	-0.480	-0.210
μ_7	-0.517	-1.070	-0.749
μ_8	-0.858	-1.464	-1.126
μ_9	-0.266	-0.901	-0.524
μ_{10}	-0.547	-1.414	-0.883
μ_{11}	-0.072	-0.871	-0.403
μ_{12}	-0.254	-0.804	-0.538
μ_{13}	-0.131	-0.620	-0.376

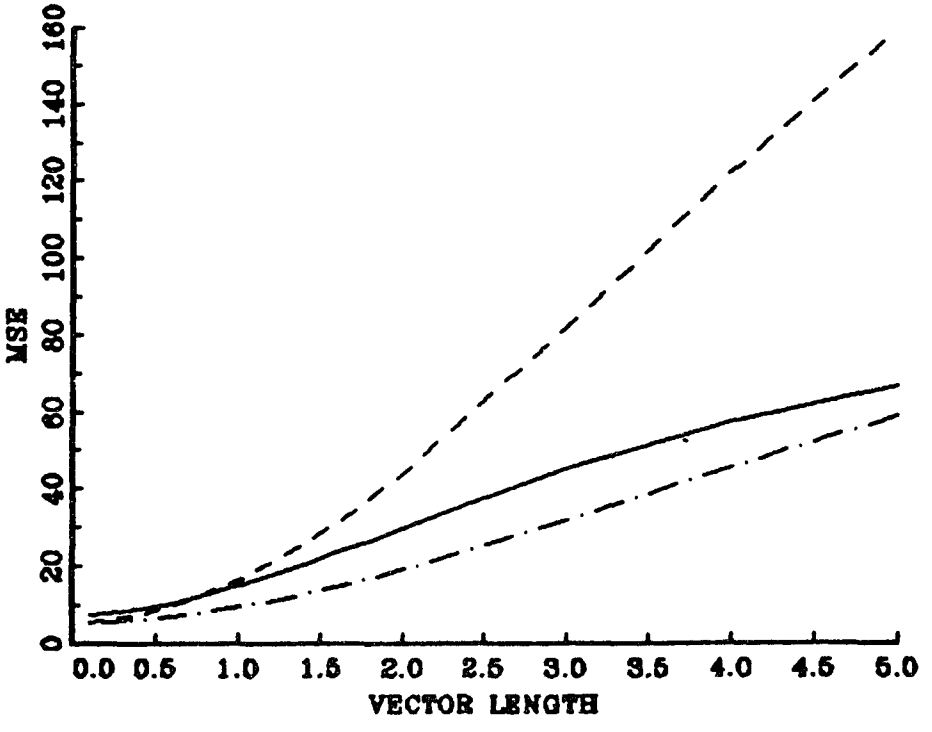
TABLE 5.13

The average mean square errors of the ridge regression estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	7.240	5.546	5.332
7.734	0.20	7.556	5.894	5.471
30.94	0.40	8.679	7.247	5.998
69.60	0.60	10.303	9.432	6.825
123.7	0.80	12.336	12.398	7.918
193.4	1.00	14.721	16.097	9.250
278.4	1.20	17.399	20.480	10.802
379.0	1.40	20.300	25.498	12.555
495.0	1.60	23.357	31.096	14.490
626.4	1.80	26.509	37.219	16.590
773.4	2.00	29.702	43.810	18.838
1740	3.00	44.974	81.611	31.664
3093	4.00	57.437	122.029	45.531
4833	5.00	66.755	159.033	58.661

FIGURE 5.9

The average mean square errors of the ridge regression estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

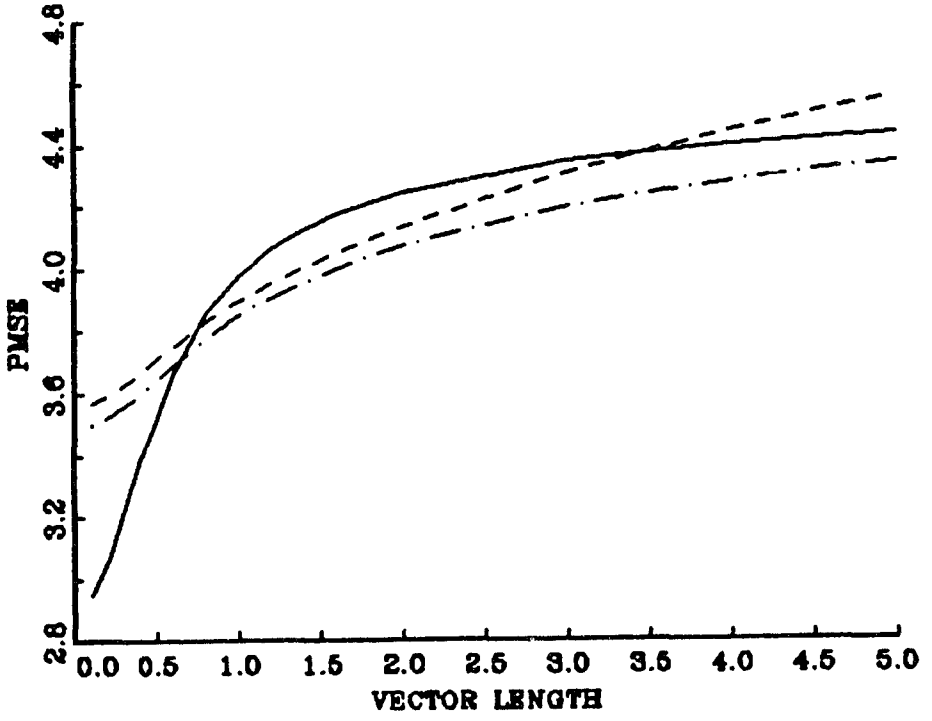
TABLE 5.14

The average prediction mean square errors of the ridge regression estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	2.948	3.570	3.499
7.734	0.20	3.068	3.593	3.523
30.94	0.40	3.388	3.666	3.601
69.60	0.60	3.661	3.750	3.691
123.7	0.80	3.851	3.827	3.774
193.4	1.00	3.979	3.894	3.846
278.4	1.20	4.067	3.953	3.906
379.0	1.40	4.130	4.005	3.958
495.0	1.60	4.179	4.053	4.002
626.4	1.80	4.217	4.096	4.041
773.4	2.00	4.248	4.137	4.075
1740	3.00	4.349	4.310	4.200
3094	4.00	4.405	4.449	4.284
4834	5.00	4.440	4.561	4.347

FIGURE 5.10

The average prediction mean square errors of the ridge regression estimator applied to various tyes of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

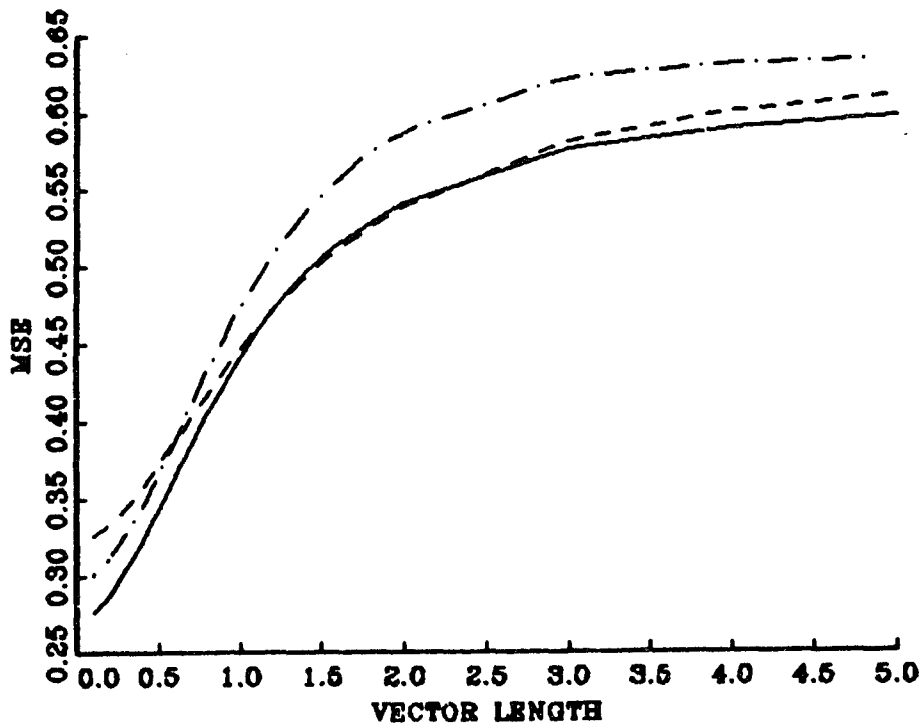
TABLE 5.15

The average mean square error of the estimates of α_2 of the ridge regression estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.276	0.326	0.301
7.734	0.20	0.287	0.333	0.311
30.94	0.40	0.322	0.357	0.345
69.60	0.60	0.365	0.388	0.389
123.7	0.80	0.406	0.419	0.434
193.4	1.00	0.442	0.447	0.474
278.4	1.20	0.472	0.472	0.508
379.0	1.40	0.496	0.493	0.535
495.0	1.60	0.514	0.511	0.557
626.4	1.80	0.529	0.526	0.575
773.4	2.00	0.541	0.539	0.588
1740	3.00	0.576	0.581	0.623
3094	4.00	0.590	0.601	0.632
4834	5.00	0.597	0.611	0.635

FIGURE 5.11

The average mean square error of the estimates of α_2 of the ridge regression estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

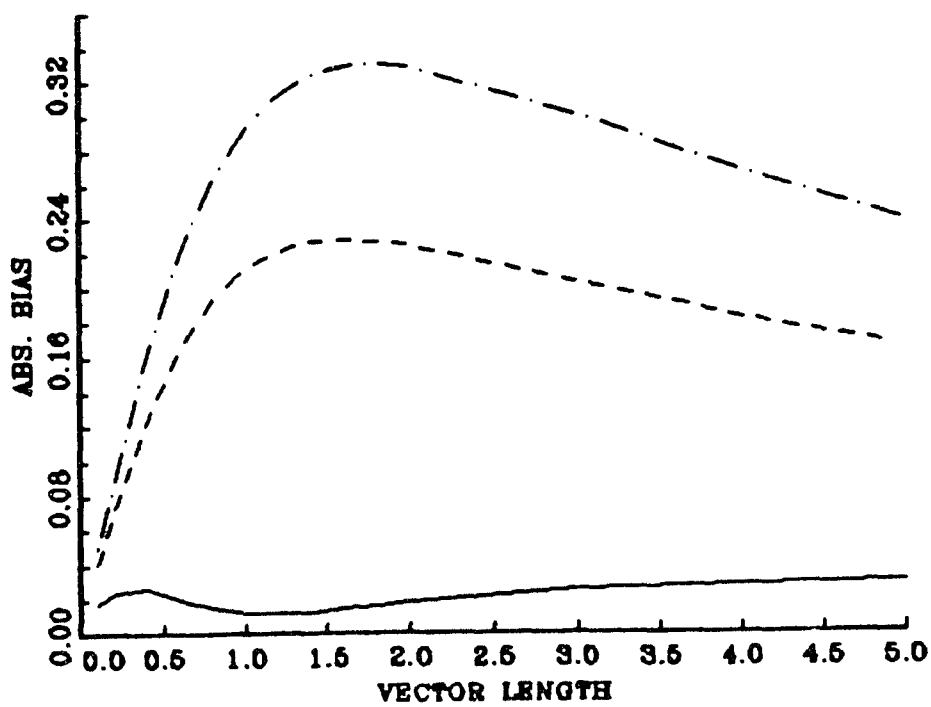
TABLE 5.16

The bias of the estimates of α_2 of the ridge regression estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	-0.017	-0.041	-0.050
7.734	0.20	-0.024	-0.072	-0.090
30.94	0.40	-0.026	-0.125	-0.163
69.60	0.60	-0.020	-0.165	-0.221
123.7	0.80	-0.015	-0.193	-0.263
193.4	1.00	-0.012	-0.211	-0.293
278.4	1.20	-0.012	-0.221	-0.313
379.0	1.40	-0.012	-0.227	-0.324
495.0	1.60	-0.014	-0.228	-0.330
626.4	1.80	-0.016	-0.227	-0.331
773.4	2.00	-0.018	-0.225	-0.329
1740	3.00	-0.025	-0.203	-0.301
3094	4.00	-0.028	-0.182	-0.268
4834	5.00	-0.029	-0.166	-0.239

FIGURE 5.12

The bias of the estimates of α_2 of the ridge regression estimator applied to various types of conditioned data.



—— —— CENTERED AND NORMALIZED

----- - ACTUAL -.-.-. CENTERED

actual data and the normalized data. The estimates of α_2 acquired from the normalized data produces the smallest bias as shown in Table 5.16 and Figure 5.12.

The parameter estimates obtained via applying the Stein (D=X'X) estimator to various conditioned data are presented in Table 5.17. The estimates from the mean centered data and normalized data are identical. The estimates from the actual data are very alike those of the least squares shown in Table 5.2. From Table 5.18 and Figure 5.13, we discover that applying the Stein (D=X'X) estimator to the normalized data or mean centered data produces the smallest MSE's for all vector lengths. Similar deductions can be made for the cases of the overall PMSE and the MSE of the estimates of α_2 which are outlined in Tables 5.19-5.20 and Figures 5.14-5.15. From Table 5.21 and Figure 5.16, we observe that the estimates of α_2 obtained from the actual data give lower bias than those of the mean centered data or the normalized data.

In Table 5.22, we report the parameter estimates obtained by applying the PC-Stein (D=X'X) estimator to various data. The parameter estimates change moderately as we change the data compared to the estimates of the equity estimator. From Tables 5.23-5.25 and Figures 5.17-5.19, we discover that there are little differences when different conditioned data are used in the estimation process. The bias of the estimates of the parameter α_2 of the normalized data shown in Table 5.26 and Figure 5.20 are the smallest.

TABLE 5.17

The estimated parameter values of the Stein ($D=X'X$) estimator applied to various types of conditioned data.

Parameter	centered and normalized	actual	centered
α_0	2.071	0.944	2.071
α_1	6.619	7.891	6.619
α_2	1.294	1.542	1.294
α_3	0.242	0.289	0.242
α_4	0.201	0.240	0.201
α_5	1.214	1.447	1.214
β_{22}	0.400	0.477	0.400
β_{24}	-0.229	-0.273	-0.229
β_{25}	-0.471	-0.561	-0.471
β_{32}	-3.533	-4.212	-3.533
β_{34}	0.119	0.142	0.119
β_{35}	0.600	0.715	0.600
β_{42}	1.303	1.553	1.303
β_{44}	-0.347	-0.413	-0.347
β_{45}	-0.714	-0.852	-0.714
μ_2	-0.220	-0.262	-0.220
μ_3	-0.420	-0.501	-0.420
μ_4	-1.056	-1.259	-4.056
μ_5	-1.015	-1.210	-1.015
μ_6	-0.973	-1.160	-0.973
μ_7	-1.561	-1.259	-1.561
μ_8	-1.937	-1.160	-1.937
μ_9	-1.427	-1.701	-1.427
μ_{10}	-1.863	-2.221	-1.863
μ_{11}	-1.445	-1.722	-1.445
μ_{12}	-1.358	-1.619	-1.358
μ_{13}	-1.095	-1.305	-1.095

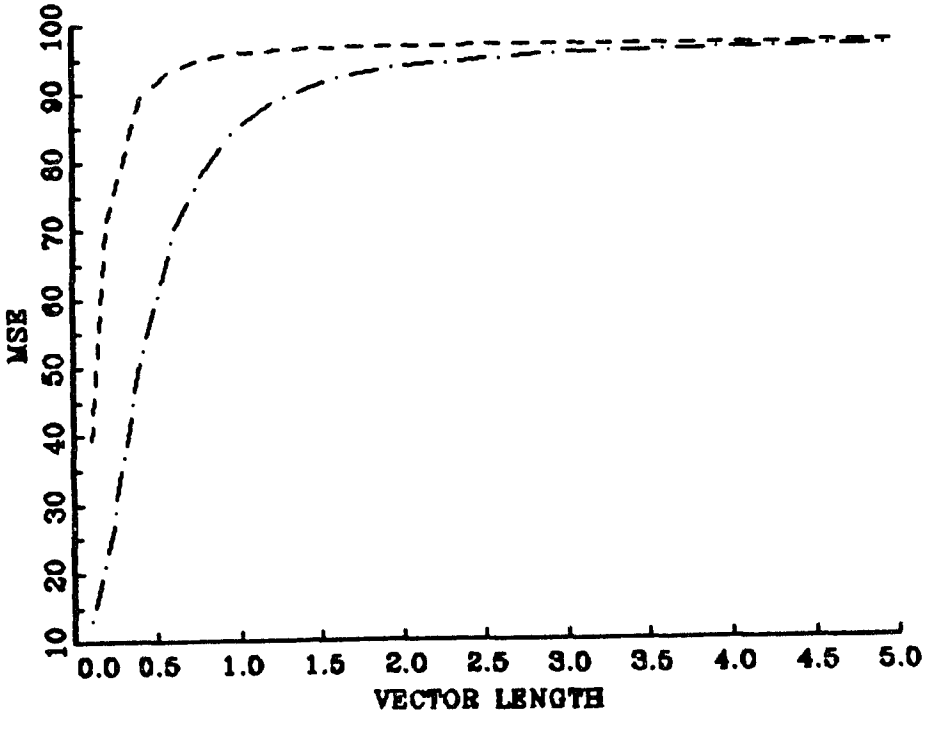
TABLE 5.18

The average mean square errors of the Stein (D=X'X)estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	13.131	39.495	13.131
7.734	0.20	22.911	71.489	22.911
30.94	0.40	50.334	89.397	50.334
69.60	0.60	69.362	93.583	69.362
123.7	0.80	79.601	95.124	79.601
193.4	1.00	85.273	95.851	85.273
278.4	1.20	88.645	96.249	88.645
379.0	1.40	90.786	96.490	90.786
495.0	1.60	92.221	96.647	92.221
626.4	1.80	93.225	96.755	93.225
773.4	2.00	93.954	96.832	93.954
1740	3.00	95.717	97.015	95.717
3094	4.00	96.345	97.079	96.345
4834	5.00	96.638	97.109	96.638

FIGURE 5.13

The average mean square errors of the Stein (D=X'X)estimator applied to various types of conditioned data.



------- ACTUAL -.-.-. CENTERED

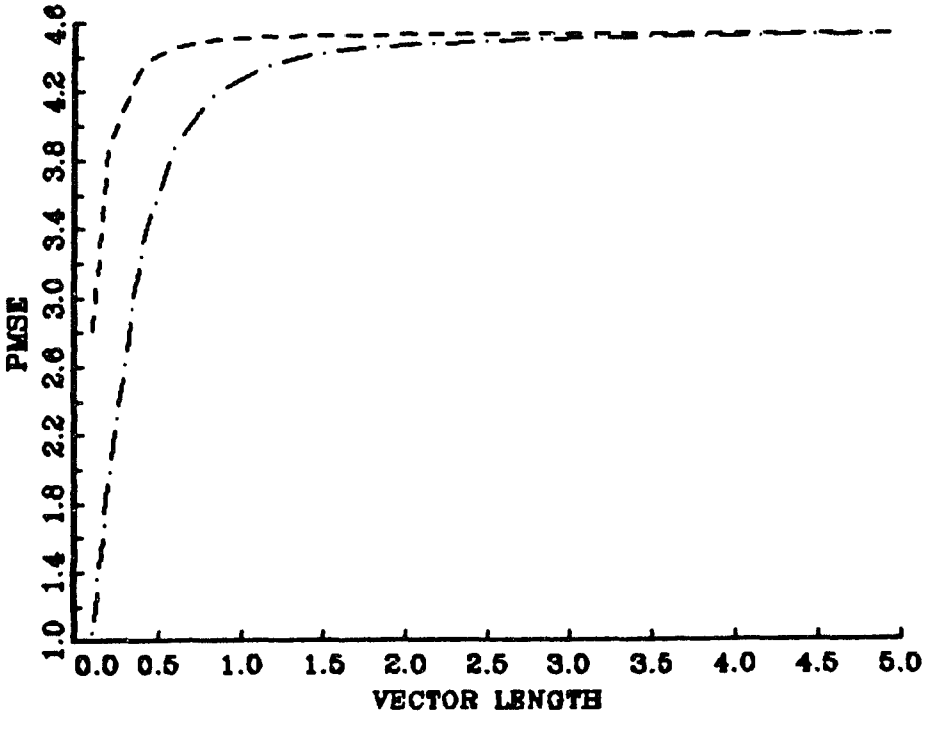
TABLE 5.19

The average prediction mean square errors of the Stein (D=X'X) estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	1.040	2.811	1.040
7.734	0.20	1.983	3.914	1.983
30.94	0.40	3.338	4.368	3.338
69.60	0.60	3.901	4.463	3.901
123.7	0.80	4.156	4.497	4.156
193.4	1.00	4.287	4.513	4.287
278.4	1.20	4.362	4.521	4.362
379.0	1.40	4.408	4.526	4.408
495.0	1.60	4.438	4.529	4.438
626.4	1.80	4.459	4.531	4.459
773.4	2.00	4.474	4.532	4.474
1740	3.00	4.510	4.535	4.510
3094	4.00	4.523	4.536	4.523
4834	5.00	4.528	4.537	4.528

FIGURE 5.14

The average prediction mean square errors of the Stein (D=X'X)estimator applied to various types of conditioned data.



------- ACTUAL -.-.-. CENTERED

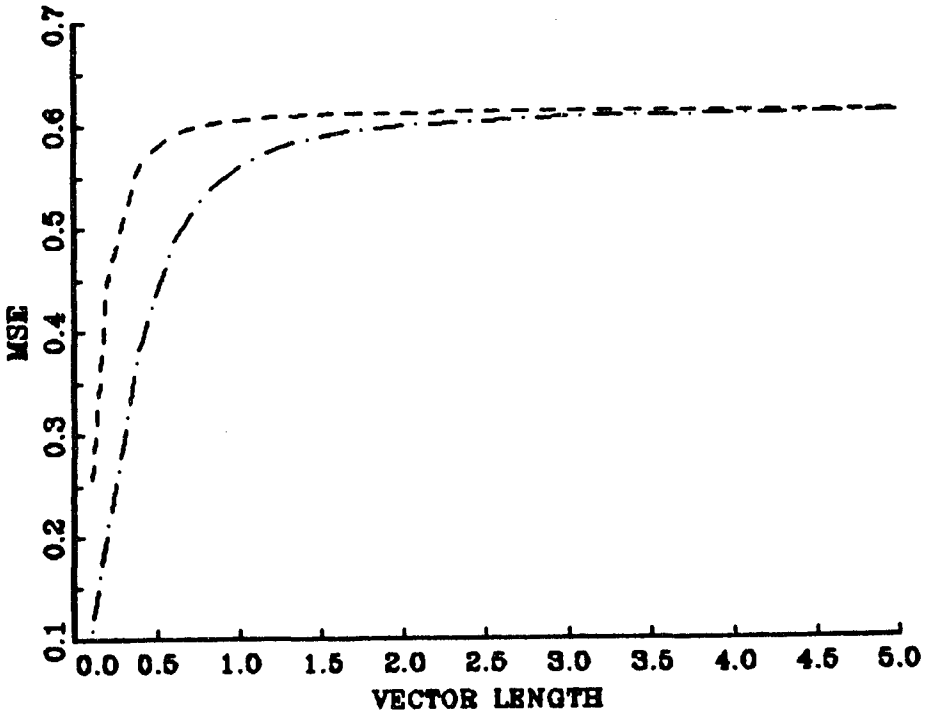
TABLE 5.20

The average mean square error of the estimates of α_2 of the Stein ($D=X'X$) estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.108	0.257	0.108
7.734	0.20	0.209	0.456	0.209
30.94	0.40	0.391	0.566	0.391
69.6	0.60	0.487	0.592	0.487
123.7	0.80	0.535	0.601	0.535
193.4	1.00	0.561	0.606	0.561
278.4	1.20	0.576	0.608	0.576
379.0	1.40	0.586	0.610	0.586
495.0	1.60	0.592	0.611	0.592
626.4	1.80	0.597	0.611	0.597
773.4	2.00	0.600	0.612	0.600
1740	3.00	0.608	0.613	0.608
3094	4.00	0.610	0.613	0.610
4834	5.00	0.612	0.613	0.612

FIGURE 5.15

The average mean square error of the estimates of α_2 of the Stein ($D=X'X$) estimator applied to various types of conditioned data.



------- ACTUAL -.-.-. CENTERED

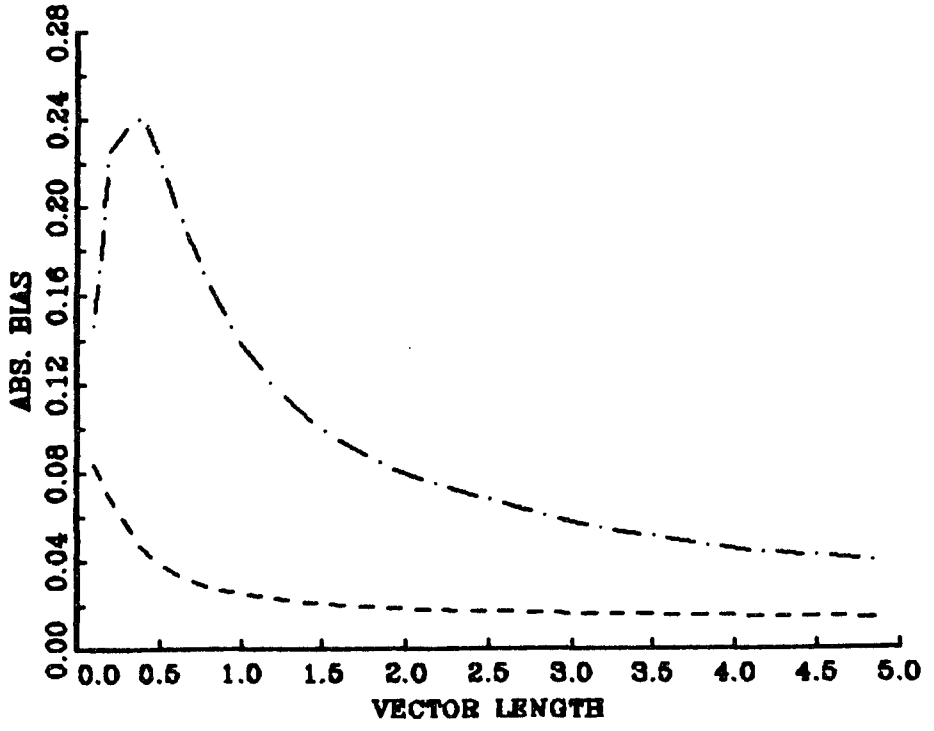
TABLE 5.21

The bias of the estimates of α_2 of the Stein ($D=X'X$) estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	-0.146	-0.084	-0.146
7.734	0.20	-0.231	-0.068	-0.231
30.94	0.40	-0.241	-0.044	-0.241
69.60	0.60	-0.199	-0.034	-0.199
123.7	0.80	-0.164	-0.028	-0.164
193.4	1.00	-0.138	-0.025	-0.138
278.4	1.20	-0.120	-0.023	-0.120
379.0	1.40	-0.105	-0.021	-0.105
495.0	1.60	-0.094	-0.020	-0.094
626.4	1.80	-0.086	-0.019	-0.086
773.4	2.00	-0.079	-0.018	-0.079
1740	3.00	-0.057	-0.016	-0.057
3094	4.00	-0.045	-0.015	-0.045
4834	5.00	-0.039	-0.014	-0.039

FIGURE 5.16

The bias of the estimates of α_2 of the Stein ($D=X'X$) estimator applied to various types of conditioned data.



------- ACTUAL -.-.-. CENTERED

TABLE 5.22

The estimated parameter values of the PC-Stein (D=X'X) estimator applied to various types of conditioned data.

Parameter	centered and normalized	actual	centered
α_0	5.042	1.167	3.019
α_1	3.539	7.516	5.619
α_2	1.750	1.430	1.217
α_3	0.190	0.297	0.240
α_4	0.165	0.266	0.203
α_5	1.235	1.364	1.402
β_{22}	-0.101	0.438	0.278
β_{24}	-0.241	-0.232	-0.310
β_{25}	-0.441	-0.513	-0.562
β_{32}	-2.950	-3.847	-3.081
β_{34}	0.134	0.172	0.103
β_{35}	0.242	0.648	0.308
β_{42}	0.733	1.416	1.073
β_{44}	-0.254	-0.399	-0.306
β_{45}	-0.669	-0.779	-0.819
μ_2	-0.087	-0.176	-0.066
μ_3	-0.662	-0.362	-0.321
μ_4	-0.824	-1.163	-0.905
μ_5	-0.735	-1.135	-0.785
μ_6	-0.615	-0.951	-0.769
μ_7	-1.299	-1.643	-1.466
μ_8	-1.662	-2.126	-1.892
μ_9	-1.026	-1.519	-1.208
μ_{10}	-1.430	-2.025	-1.614
μ_{11}	-0.986	-1.573	-1.185
μ_{12}	-1.046	-1.490	-1.217
μ_{13}	-0.877	-1.201	-0.975

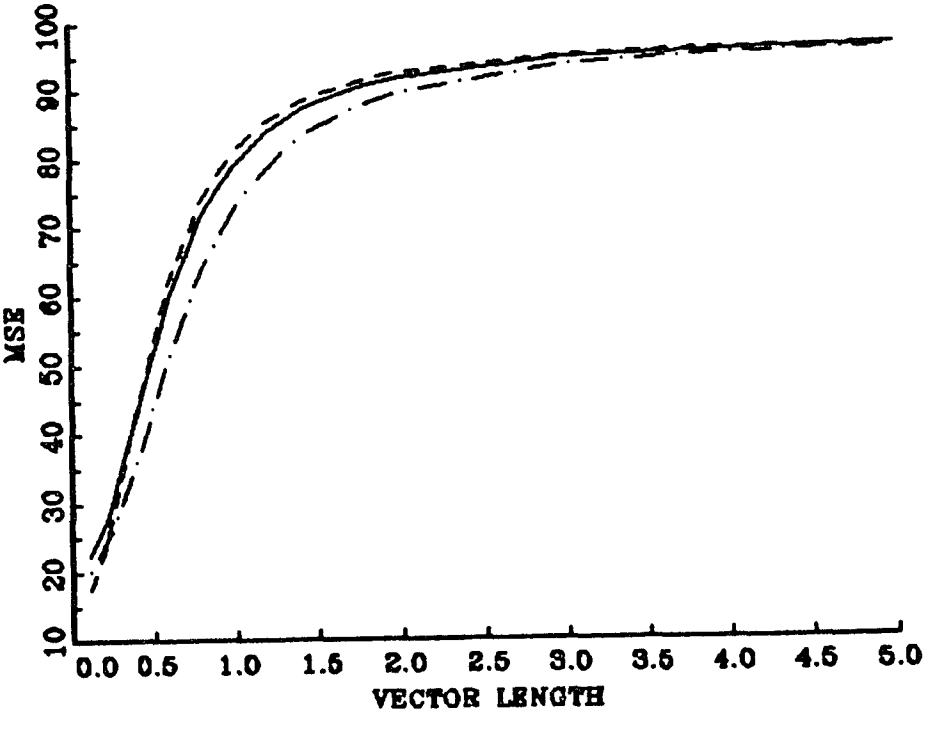
TABLE 5.23

The average mean square errors of the PC-Stein (D=X'X)estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	22.383	17.589	20.163
7.734	0.20	27.463	24.241	23.977
30.94	0.40	43.831	44.530	36.889
69.60	0.60	60.189	62.613	51.878
123.7	0.80	71.780	74.188	64.133
193.4	1.00	79.215	81.185	72.893
278.4	1.20	83.985	85.542	78.936
379.0	1.40	87.145	88.381	83.115
495.0	1.60	89.319	90.316	86.067
626.4	1.80	90.867	91.685	88.206
773.3	2.00	92.005	92.687	89.796
1740	3.00	94.802	95.134	93.769
3094	4.00	95.817	96.014	95.229
4834	5.00	96.294	96.426	95.918

FIGURE 5.17

The average mean square errors of the PC-Stein (D=X'X)estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

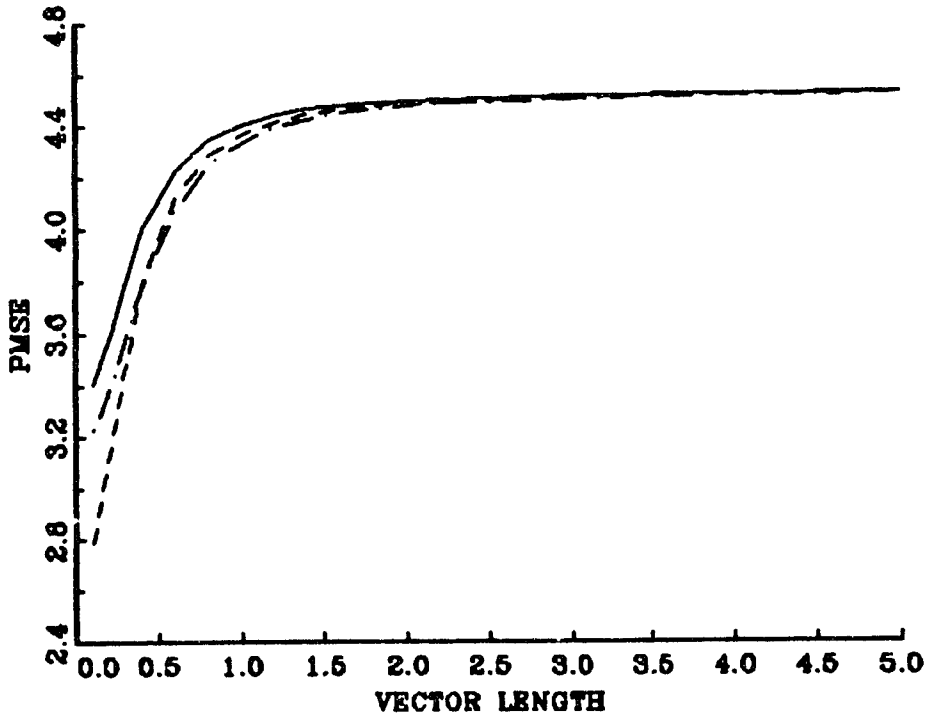
TABLE 5.24

The average prediction mean square errors of the PC-Stein (D=X'X)estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	3.402	2.790	3.229
7.734	0.20	3.585	3.135	3.382
30.94	0.40	3.997	3.798	3.784
69.60	0.60	4.226	4.125	4.081
123.7	0.80	4.342	4.282	4.246
193.4	1.00	4.405	4.366	4.337
278.4	1.20	4.442	4.414	4.392
379.0	1.40	4.466	4.445	4.427
495.0	1.60	4.481	4.466	4.451
626.4	1.80	4.492	4.480	4.468
773.4	2.00	4.500	4.490	4.480
1740	3.00	4.519	4.515	4.510
3094	4.00	4.526	4.524	4.521
4834	5.00	4.530	4.528	4.526

FIGURE 5.18

The average prediction mean square errors of the PC-Stein (D=X'X)estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

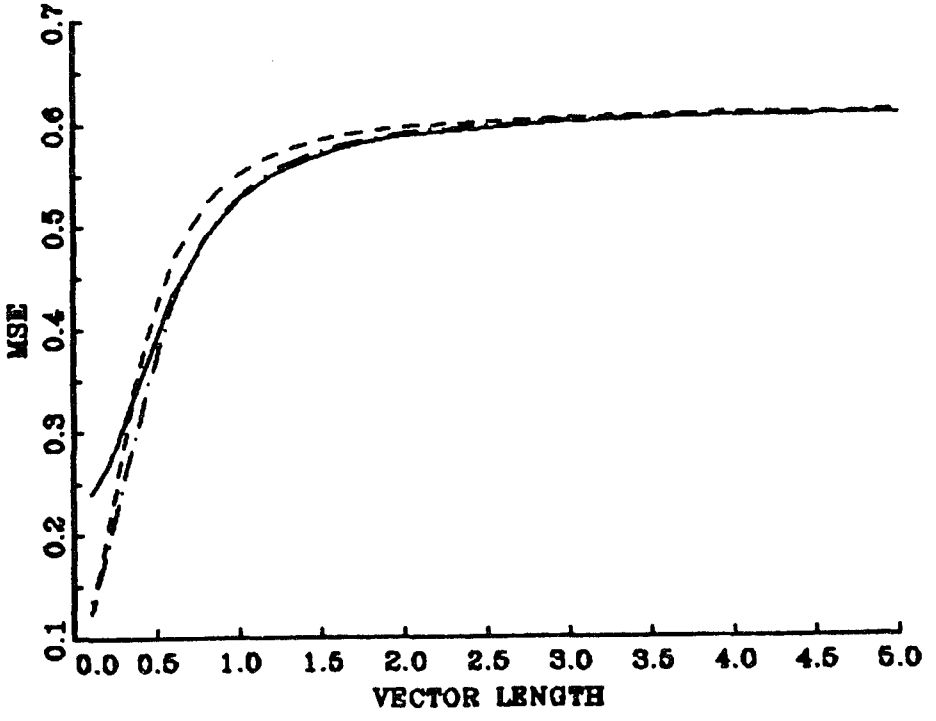
TABLE 5.25

The average mean square error of the estimates of α_2 of the PC-Stein ($D=X'X$) estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.240	0.125	0.137
7.734	0.20	0.266	0.204	0.186
30.94	0.40	0.350	0.370	0.318
69.60	0.60	0.432	0.470	0.427
123.7	0.80	0.489	0.522	0.492
193.4	1.00	0.526	0.551	0.529
278.4	1.20	0.549	0.569	0.553
379.0	1.40	0.564	0.580	0.568
495.0	1.60	0.575	0.588	0.578
626.4	1.80	0.583	0.593	0.585
773.4	2.00	0.588	0.597	0.591
1740	3.00	0.602	0.606	0.604
3094	4.00	0.607	0.610	0.608
4834	5.00	0.609	0.611	0.610

FIGURE 5.19

The average mean square error of the estimates of α_2 of the PC-Stein ($D=X'X$) estimator applied to various types of conditioned data.



———— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

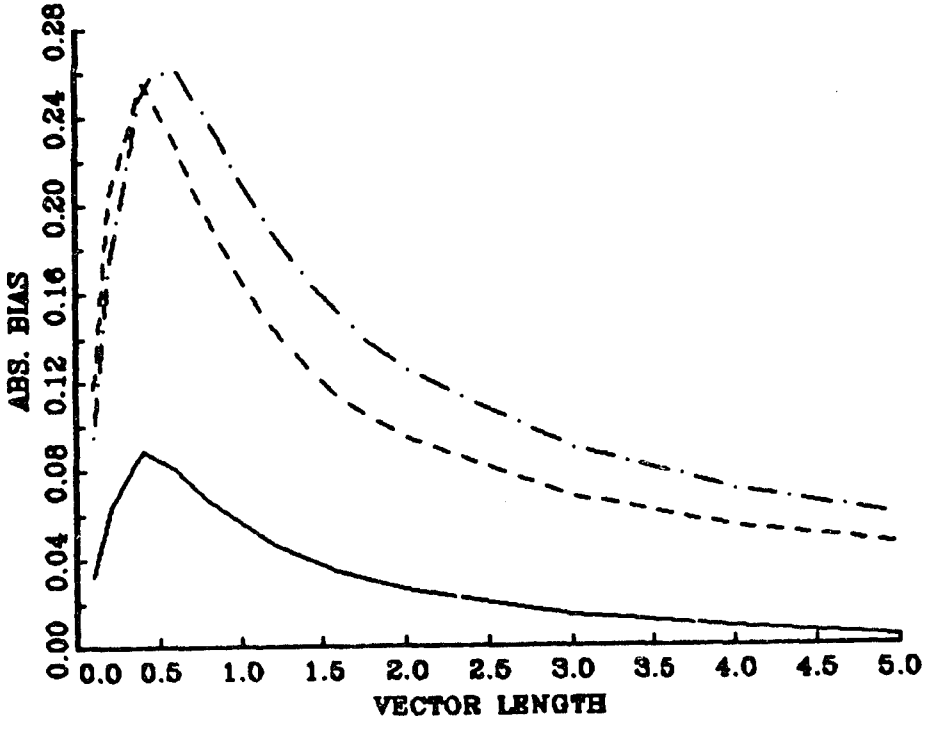
TABLE 5.26

The bias of the estimates of α_2 of the PC-Stein (D=X'X) estimator applied to various types of conditioned data.

R	length	centered and normalized	actual	centered
1.934	0.10	0.032	-0.118	-0.095
7.734	0.20	0.063	-0.205	-0.173
30.94	0.40	0.089	-0.256	-0.256
69.60	0.60	0.081	-0.228	-0.262
123.7	0.80	0.067	-0.194	-0.239
193.4	1.00	0.056	-0.166	-0.211
278.4	1.20	0.047	-0.144	-0.187
379.0	1.40	0.040	-0.128	-0.167
495.0	1.60	0.034	-0.114	-0.151
626.4	1.80	0.030	-0.104	-0.137
773.4	2.00	0.026	-0.095	-0.126
1740	3.00	0.014	-0.068	-0.090
3094	4.00	0.008	-0.054	-0.071
4834	5.00	0.004	-0.046	-0.059

FIGURE 5.20

The bias of the estimates of α_2 of the PC-Stein (D=X'X) estimator applied to various types of conditioned data.



——— CENTERED AND NORMALIZED

------- ACTUAL -.-.-. CENTERED

5.5 CONCLUSIONS.

1. When the equity estimator is applied to the original data, non-generated data, it yields similar estimates to those obtained via the ridge regression estimator. Both estimators give theoretically acceptable signs for most of the estimates.

2. From the Monte Carlo experiments, the performance of the equity estimator judged by the MSE criterion is extremely impressive but the equity estimator performs poorly under the PMSE criteria. Intuitively, this finding is not surprising. The nature of the equity estimator is that it trades biasedness for reduction in the variability of the parameter estimates. As a result, the predictability of the equity estimator is not very impressive.

3. The equity estimator estimates the parameter of interest α_2, the own discount effects, extremely well for all vector lengths. Furthermore, the equity estimator shows similar characteristics to those of the ridge regression estimator. However, the equity estimator produces high bias when the vector length is very large which demonstrates the trade-off between biasedness and variability.

4. When the data are not conditioned, actual data, the equity estimator yields incorrect signs for most of the parameter estimates. Moreover, the equity estimator gives vastly different estimates for different type of conditioned data.

5. The equity estimator does not perform well when applied to the actual data except when the vector length is small under both MSE and PMSE criteria.

6. The equity estimator estimates the parameter α_2 well based on the MSE criterion; however, the data have to be centered and normalized to unit length. Without centering and normalizing the data to unit length, the equity estimator performs well only if the length of the vector is small.

From the study, we learn that the equity estimator can produce satisfactory results. However, its performance is dependent heavily on the vector length. As the vector length increases, the bias appears to be increased proportionally which overshadows the gain by reducing the variability of the estimates at high vector length. Unlike its competitors, the MSE and PMSE of the equity estimator seem to have no boundaries. The trade off between biasedness and variability is evident when the equity estimator is evaluated under the prediction mean square error criteria. The equity estimator is by no means a good predictor. The choice of the weight matrix in the squared error loss function is very important to the performance of the equity estimator.

Moreover, we observe that the performance of the equity estimator depends on the conditioning of the data. Without proper data conditioning, the equity estimator yields high MSE and PMSE at high values of vector length compared to those of its traditional biased estimator alternatives. The

MSE of the estimates of the parameter α_2 of the equity estimator obtained from utilizing the mean centered data or the actual data can be as high as 300 times the MSE obtained via the centered and normalized data. Consequently, the evidence does not support the claim that the equity estimator deals with the problem of multicollinearity efficiently.

The Monte Carlo experiments indicate that the equity estimator performs well only when the degree of multicollinearity is substantially reduced by means of centering and normalizing the data. Contradictory to the equity estimator, the performances of the traditional biased estimator considered do not change as drastically as the performance of the equity estimator when different conditioned data are used even though the performances of the biased estimators examined can be improved through the selection of the data conditioning. Therefore, the use of the equity estimator adds another risk factor into the estimation process. The risk is that we do not know whether the data have been adequately conditioned so that the application of the equity estimator will produce satisfactoty results.

The dependency of the equity estimator on the conditioning of the data can be linked to the transformation of the matrix of exogenous variables. The equity estimator transforms the matrix of exogenous variables by post multiplying with the matrix $(X'X)^{-1/2}$, Equation (4.28). We

have shown that the equity estimator is biased and inconsistent unless $\mathbf{X'X} = \mathbf{I}$, Section 4.6. When the data are conditioned, the degree of multicollinearity is reduced. In other words, the matrix $\mathbf{X'X}$ more closely resembles the identity matrix. As the data become better conditioned, the transformation becomes less drastic which reduces the biasedness and ,hence, improves the performance of the equity estimator. KR (1987) have dealt the problem of biasedness by scaling the parameter estimates by the constant ν, Equation (4.25). The Monte Carlo experiments show that this scaling technique does not rectify the bias problem especially when the vector length is high or when the data are not well conditioned. Overall, the equity estimator must be used with caution. Its performance depends heavily on the size of the underlying parameters as well as the conditioning of the data. One significant conclusion is that the equity estimator does not solve the problem of multicollinearity as claimed. For the equity estimator to perform well, the data must be moderately well conditioned which raises the question of how can we tell whether the data are well conditioned enough to be estimated by the equity estimator.

CHAPTER 6

A CONFIDENCE INTERVAL FOR OUT-OF-SAMPLE PREDICTION WHEN USING BIASED PREDICTORS

6.1 INTRODUCTION

In this chapter we will construct confidence intervals for out-of-sample prediction when multicollinearity is present and biased predictors are used. Least squares may not be a good predictor when the explanatory variables are linearly related. Biased estimators can be used as predictors in the presence of multicollinearity. However, obtaining the confidence interval of the forecast values cannot be done in the usual way. Bootstrap re-sampling methods will be used to construct a reasonable confidence interval for the each of the prediction techniques under consideration.

In Section 6.2, we discuss the effects of multicollinearity on out-of-sample prediction. The effects are not fully understood. The conventional rule is that as long as the out-of-sample data have the same pattern of multicollinearity as the in-sample data ,then multicollinearity is not a problem for prediction . However, it is very difficult to find a set of data that has the mentioned characteristic.

In Section 6.3, we discuss the use of biased estimators as predictors in the presence of multicollinearity. Several studies, have shown that some type of biased estimators can

be used as predictors and improve on the out-of-sample performance of least squares. The biased estimators that we are going to study are described in Chapters 4 and 5. They are ridge regression and two Stein-like estimators.

In Section 6.4, we will discuss the bootstrap method. Using the bootstrap method will help us to construct the confidence interval for our forecasts from the prediction technique which we are considering. We use bootstrap to trace the distributions of the forecasted values which enable us to construct a reasonable confidence interval. Actual data will be used in order to empirically construct confidence intervals which will be carried out in Chapter 7.

6.2 THE EFFECTS OF MULTICOLLINEARITY ON OUT-OF-SAMPLE PREDICTION

Consider the model

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e} \tag{6.1}$$

where $\underline{y}$ is a (T×1) vector of observations on a dependent variable, $\mathbf{X}$ is a fixed (T×K) full rank matrix of observations of the exogenous variables, $\underline{\beta}$ is a (K×1) vector of unknown parameters and $\underline{e}$ is a (T×1) vector of disturbance terms which are identically and independently distributed as $N(0,\sigma^2)$.

We are interested in predicting an (m×1) vector of future values of y, $\underline{y}_0$, which are related to $\mathbf{X}_0$, a known (m×K) matrix of future values of $\mathbf{X}$.

Consequently, $\underline{y}_0$ can be expressed as

$$\underline{y}_0 = \mathbf{X}_0\underline{\beta} + \underline{e}_0 \tag{6.2}$$

where $\underline{e}_0 \sim N(\underline{0}, \sigma^2\mathbf{I}_m)$ and $E[\underline{e}\cdot\underline{e}_0'] = 0$.

Let $\underline{\delta}$ be an estimator of $\underline{\beta}$. Similar to Chapter 4, we consider the weighted squared error loss measure

$$L(\underline{\beta},\underline{\delta},\mathbf{D}) = (\underline{\delta} - \underline{\beta})'\mathbf{D}(\underline{\delta} - \underline{\beta}) \tag{6.3}$$

as a basis for evaluating estimator performance where $\mathbf{D}$ is a positive definite and symmetric matrix. The sampling performance of $\underline{\delta}$ is evaluated by its risk function

$$R(\underline{\beta},\underline{\delta},\mathbf{D}) = E\left[(\underline{\delta} - \underline{\beta})'\mathbf{D}(\underline{\delta} - \underline{\beta})\right] \tag{6.4}$$

As we are interested in the out-of-sample mean square error of prediction, the weight matrix that we wish to consider is $\mathbf{D} = \mathbf{X}_0'\mathbf{X}_0$ so that

$$\begin{aligned} R(\underline{\beta},\underline{\delta},\mathbf{X}_0'\mathbf{X}_0) &= E\left[(\underline{\delta} - \underline{\beta})'\mathbf{X}_0'\mathbf{X}(\underline{\delta} - \underline{\beta})\right] \\ &= E\left[(\mathbf{X}_0\underline{\delta} - \mathbf{X}_0\underline{\beta})'(\mathbf{X}_0\underline{\delta} - \mathbf{X}_0\underline{\beta})\right] \\ &= E\left[(\hat{\underline{y}}_0 - E(\underline{y}_0))'(\hat{\underline{y}}_0 - E(\underline{y}_0))\right] \end{aligned} \tag{6.5}$$

The out-of-sample mean square error of prediction for the ordinary least squares estimator can be written as

$$\begin{aligned} R(\hat{\underline{y}}_0 = \mathbf{X}_0\underline{b},\ E(\underline{y}_0)) &= E\left[\mathbf{X}_0\underline{b} - E(\underline{y}_0)\right]'\left[\mathbf{X}_0\underline{b} - E(\underline{y}_0)\right] \\ &= E\left[\mathbf{X}_0(\underline{b} - \underline{\beta})\right]'\left[\mathbf{X}_0(\underline{b} - \underline{\beta})\right] \\ &= \text{tr } E\left[\mathbf{X}_0(\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'\mathbf{X}_0'\right] \\ &= \text{tr } \mathbf{X}_0E(\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'\mathbf{X}_0' \end{aligned}$$

$$= \sigma^2 \text{tr } X_0 (X'X)^{-1} X_0'$$

$$= \sigma^2 \text{tr} (X_0' X_0) (X'X)^{-1} \qquad (6.6)$$

The effects of multicollinearity on the out-of-sample prediction of least squares is not fully understood, but it is known that the predictive performance of a regression equation can be adversely affected by multicollinearity. For example, see Askin (1982), Marquandt and Snee (1975) and Montgomery and Peck (1982).

From Equation (6.6), the prediction risk for least squares estimator depends on the cross-product matrices for X_0 and X. We are going to decompose the out-of-sample prediction risk of least squares into several components.

Let us define $X = [\underset{-T}{i}, X_1]$ and $X_0 = [\underset{-m}{i}, X_2]$ where $\underset{-T}{i}$ and $\underset{-m}{i}$ are unit vectors of dimension T and m respectively. X_1 and X_2 are the matrices of slope regressor which can be transformed into deviation from mean form as

$$X_1^* = M_1 X_1 = (I_T - \underset{-T}{i}\underset{-T}{i}'/T) X_1 \qquad (6.4a)$$

and

$$X_2^* = M_2 X_2 = (I_m - \underset{-m}{i}\underset{-m}{i}'/m) X_2 \qquad (6.4b)$$

We write the difference in the means of the X_1 and X_2 regressors as

$$\underset{-}{d} = (\underset{-2}{\bar{x}} - \underset{-1}{\bar{x}}) = X_2' \underset{-m}{i}/m - X_1' \underset{-T}{i}/T \qquad (6.7c)$$

The prediction risk in equation (6.6) can be written as

$$R(\underset{-}{\hat{y}}_0 = x\underset{-}{b}, E(\underset{-}{y}_0)) = \sigma^2 \left[\frac{m}{T} + \text{tr} \left\{ (X_2^* + \underset{-m}{i}\underset{-}{d}')' (X_2^* + \underset{-m}{i}\underset{-}{d}) (X_1^{*\prime} X_1^*)^{-1} \right\} \right]$$

$$= \sigma^2\left[\frac{m}{T} + \mathrm{tr}(X_2^{*\prime}X_2^*)(X_1^{*\prime}X_1^*)^{-1} + m\underline{d}'(X_1^{*\prime}X_1^*)^{-1}\underline{d}\right]$$

$$= \sigma^2\left[\frac{m}{T} + \mathrm{tr}P_1'P_2\Lambda_2P_2'P_1\Lambda_1^{-1} + m\underline{d}'P_1\Lambda_1^{-1}P_1'\underline{d}\right] \qquad (6.8)$$

where $X_1^{*\prime}X_1^* = P_1\Lambda_1P_1'$ and $X_2^{*\prime}X_2^* = P_2\Lambda_2P_2'$. P_1 and P_2 are matrices whose columns are the characteristic vectors and Λ_1 and Λ_2 are the diagonal matrices of characteristic roots of $X_1^{*\prime}X_1^*$ and $X_2^{*\prime}X_2^*$ respectively. Λ_1 and Λ_2 are arranged such that $\lambda_{i,1} \geq \lambda_{i,2} \ldots \geq \lambda_{i,K}$, $i = 1,2$.

The expression (6.8) decomposes the prediction risk into several components. The least squares prediction risk is directly related to σ^2, the precision of the population regressor function. The least squares predictor is also directly related to m, the number of out-of-sample observations, and inverse related to T, the number of in-sample observations. The expression (6.8) is also affected by the matrices P_1 and P_2 which indicate the directions of the major and minor axes of the centered data X_1 and X_2. Other factors which affect the expression (6.8) are the matrices Λ_1 and Λ_2 which reflect both variation and multicollinearity among X_1 and X_2, and the distance $\underline{d}$ which indicates the difference between the centers of the in-sample and out-of-sample regressors.

The data ellipsoids are rotationally equivalent if $P_1 = P_2$ so that $P_1'P_2 = I$ and they are variationally equivalent if $\Lambda_1 = \Lambda_2$. If the in-sample and out-of-sample data are rotationally equivalent, equation (6.8) can be rewritten as

$$R(\hat{\underline{y}}_0 = \underline{X}\underline{b}, E(\underline{y}_0)) = \sigma^2\left[\frac{m}{T} + tr\Lambda_2\Lambda_1^{-1} + m\underline{d}'\underline{P}_1\Lambda_1^{-1}\underline{P}_1\underline{d}\right] \qquad (6.9)$$

If the in-sample and out-of-sample data are both rotationally and variationally equivalent, the risk function reduces to

$$R(\hat{\underline{y}}_0 = \underline{X}\underline{b}, E(\underline{y}_0)) = \sigma^2\left[\frac{m}{T} + tr(K-1) + m\underline{d}'\underline{P}_1\Lambda_1^{-1}\underline{P}_1\underline{d}\right] \qquad (6.10)$$

As we have shown, if the in-sample and out-of-sample data are similar enough or have the same pattern, which means that they are rotationally and variationally equivalent, prediction appears to have no serious risk.

Form Equation (6.10) we can see that the risk function also depends on the term $\underline{d}'\underline{P}_1$. If $\underline{d} = \underline{0}$, the risk function becomes

$$R(\hat{\underline{y}}_0 = \underline{X}\underline{b}, E(\underline{y}_0)) = \sigma^2\left[\frac{m}{T} + (K-1)\right] \qquad (6.11)$$

If $\underline{d} \neq \underline{0}$, the prediction risk is dependent on the orientation and length of $\underline{d}$ relative to the axes of the $\underline{X}_1$ ellipsoid. Preferably, we want $\underline{d}'\underline{P}_1 = (c,0,\ldots,0)$. Then the larger $\lambda_{1,1}$ is, the less the least squares prediction risk will be.

6.3 BIASED ESTIMATORS AS PREDICTORS

Several studies have discussed the use of biased estimators to deal with the problem of out-of-sample prediction in the presence of multicollinearity. For example, Copas(1983), Friedman and Montgomery (1985) and Jones and Copas(1986).

Friedman and Montgomery(1985) showed that in the case of two dimensional space with severe multicollinearity, ridge regression and principal components estimator can have lower mean square error of prediction than that of least squares. They worked with the orthogonalized model

$$\underline{y} = Z\underline{\alpha} + \underline{e}$$

where $Z = XP$, $\underline{\alpha} = P'\underline{\beta}$ and P is the $(K \times K)$ orthogonal matrix of eigenvectors associated with the eigenvalues of $X'X$. Providing that α_1^2 is not too large relative to its standard error or the model error measured by σ^2, ridge regression and principal components estimator will allow reductions in mean square error over least squares. In general, ridge regression is a superior predictor if the prediction is in the direction of the less stable regressor. If the prediction is in the direction of the more stable regressor, principal components estimator is superior. The choice of the techniques, hence, depends of the region of the regressor space over which prediction will be made and the degree of multicollinearity.

Copas (1983) considered the out-of-sample prediction under the assumption that the future centered regressor, X_0, come from a distribution with the same mean and covariance as the in-sample centered regressor variables matrix X. The model considered is

$$y|\underline{x} \sim N(\alpha + \underline{\beta}'\underline{x}, \sigma^2)$$

where y is an observed response or dependent variable and $\underline{x}$ is a vector of p observable predictive factors or independent variables. Let $V = T^{-1}(X'X)$ where T is the sample size and the least squares predictor be denoted as

$$\hat{y} = \hat{\alpha} + \hat{\underline{\beta}}'\underline{x}$$

where $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$ and $\hat{\alpha} = \bar{y}$. The shrinkage predictor is of the form

$$\tilde{y} = \hat{\alpha} + K\hat{\underline{\beta}}'\underline{x}$$

where the index of shrinkage K is defined as

$$K = \hat{\underline{\beta}}'V\underline{\beta} \;/\; \hat{\underline{\beta}}'V\hat{\underline{\beta}}$$

The denominator of K is on the average larger than the numerator; thus, the expectation of K is less than one. K = 1 indicates no shrinkage and small K indicates substantial shrinkage. For a given number of dependent variables p, the distribution of K depends on a quantity δ given by

$$\delta^2 = \sigma^2 / \; T\underline{\beta}'V\underline{\beta}$$

The distribution of K becomes more concentrated about K = 1 as $\delta^2 \rightarrow 0$. K is likely to be small if p is not small relative to the sample size T and/or the signal to noise ratio measured by $\underline{\beta}'V\underline{\beta}/\sigma^2$ is small. The constant K is estimated by

$$\hat{K}(k) = 1 - \hat{\sigma}^2 k / T\hat{\underline{\beta}}'V\hat{\underline{\beta}}$$

where $\hat{\sigma}^2 = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})$. The shrinkage predictor considered is of the form

$$\tilde{y} = \hat{\alpha} + \hat{K}\hat{\underline{\beta}}'\underline{x}$$

The overall prediction mean square error is less than that of least squares if

$$0 < k < \frac{2(p-2)}{1 + 2/v}$$

where v = T-p-1. Copas showed the value of k = p-2 always gives lower prediction mean square error than that of least squares provided that v > 2 which yields the usual Stein-rule.

Jones and Copas(1986) examined a more general case of prediction. Using the same model as in Copas(1983), they assume that the future regressor, X_0, comes form a distribution with mean $\underline{\mu}_o$ and variance V_o and define $A_0 = V_0 + \underline{\mu}_0\underline{\mu}_0'$. They define a region in which the Stein-like shrinkage predictor dominates the least squares predictor for all values of $\underline{\beta}$ and σ^2. The region is defined as

$$w_{max} < 2\ tr\mathbf{V}^{-1}\mathbf{A}_0/(p + 2)$$

where w_{max} is the largest eigenvalue of $\mathbf{V}^{-1}\mathbf{A}_0$. The differences between $\mathbf{V}$ and $\mathbf{A}_0$ are measured by the ratio of the largest eigenvalue to the trace of $\mathbf{V}^{-1}\mathbf{A}_0$. The ratio reaches its minimum of 1/p when $\mathbf{V} = \mathbf{A}_0$ and $1/p < 2(p + 2)$ for all $p \geq 3$.

Fomby and Hill (1986) considered the traditional model as in the Equation (6.1). They showed that for the usual Stein rule estimator that shrinks all parameters toward the

origin

$$\underline{\delta} = (1 - \frac{as}{\underline{b}'X'X\underline{b}})\underline{b}$$

where $s = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})$ and $\underline{b}$ is the least squares estimator, the prediction mean square error of such a predictor is less than that of ordinary least squares for all $\underline{\beta}$ and σ^2 if

$$a \leq \frac{2}{T-K+2}\left[\text{tr}\ \frac{X_0'X_0(X'X)^{-1}}{\lambda_{max}} - 2\right]$$

where λ_{max} is the largest characteristic root of $X_0'X_0(X'X)^{-1}$. Thus, the necessary conditions

$$\lambda_{max} < \frac{\text{tr}\ X_0'X_0(X'X)^{-1}}{2} \quad .$$

All of the studies cited point out that the performances of the biased predictors are dependent on the unforeseen future values of the regressor. Friedman and Montgomery (1985) showed that ridge regression and principal components estimator may improve upon the out-of-sample mean square error of prediction of least squares for a certain range of parameter space. The choice of predictor depends of the direction of the prediction space.

Copas (1983), Jones and Copas(1986) and Fomby and Hill (1986) showed that the predictability of the Stein-like predictor is based on the similarity between the in-sample and out-of-sample data. Consequently, we can use the biased predictors as alternatives of least squares in the presence of multicollinearity.

6.4 BOOTSTRAP

6.4.1 BOOTSTRAP METHOD

Bootstrap is a computationally intensive method which attempts to determine the characteristics of the distribution of a random variable by using its observed values. Efron (1979) viewed the bootstrap method as a nonparametric method of estimating the bias and variance of a statistic of interest.

Let X be a random sample of size T with a complete unspecified probability distribution F, ie.

$$X_i = x_i \qquad X_i \sim_{ind} F \qquad i = 1,2,..,T$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_T)$ denotes the random sample and $x = (x_1, x_2, \ldots x_T)$ denotes the observed values. We are interested in the distribution of a specific parameter such as the mean or standard deviation of F. Let the parameter of interest be denoted as $\theta(F)$ and $t(X)$ be an estimator of $\theta(F)$. The sampling distribution that we are interested in is of the random sample

$$R(X,F) = t(X) - \theta(F)$$

The bootstrap method for the one-sample problem can be described in 3 steps.

1. Construct the sample probability distribution of x, namely $\hat{F}$.

2. Draw a sample of size T from $\hat{F}$ with replacement and calculate an estimate of $\theta(F)$. We call this a bootstrap sample which can be repeated.

3. Approximate the sampling distribution of R(X,F) by the distribution of the bootstrap sample.

Efron (1979) indicated that the accuracy of the approximation of the distribution of R(X,F) depends on the form of R. The Monte Carlo experiment sampling technique can also be applied to the bootstrap method by repeating step 2 for N times and use the resulting histogram as an approximation to the distribution of the bootstrap estimator.

Bickel and Freedman (1981) provided the proof of some asymptotic theories for the bootstrap method. They showed that the resampling by Efron (1979) from $\hat{F}$ converges in conditional probability to the true variance-covariance matrix of X_1. Freedman (1981) extended the asymptotic theories to be applied to regression equation.

Consider the model

$$\underline{y} = X\underline{\beta} + \underline{e}$$

where $\underline{y}$ is a (T×1) vector of observations on a dependent variable, X is a fixed (T×K) full rank matrix of observations on exogenous variables, $\underline{\beta}$ is a (K×1) vector of unknown parameters and $\underline{e}$ is a (T×1) vector of unobservable disturbance terms. Let $\underline{b}$ be the least squares estimator of $\underline{\beta}$, then the observed column vector of residuals $\hat{\underline{e}}$ is given by

$$\hat{\underline{e}} = \underline{y} - X\underline{b}$$

Let $\underline{e}^*$ be a (T×1) vector of resampling with replacement of

the elements of the vector of centered residuals $\hat{\underline{e}}$, $\underline{y}^*$ be a (T×1) vector of dependent variable generated by

$$\underline{y}^* = X\underline{b} + \underline{e}^* \tag{6.12}$$

We estimate Equation (6.12) by least squares and obtain an estimate for $\underline{b}$, $\hat{\underline{\beta}}^*$. We want to characterize the distribution of $(\underline{b} - \underline{\beta})T^{1/2}$ by using the distribution of $(\hat{\underline{\beta}}^* - \underline{b})T^{1/2}$ as an approximation. This approximation is likely to be good if T is large and $\sigma^2 tr(X'X)^{-1}$ is small. Notice that $\underline{e}^*$ is a vector of centered residuals, without centering the distribution of $(\hat{\underline{\beta}}^* - \underline{b})T^{1/2}$ incorporates a bias term which is random and has nondegenerating normal distribution and hence does not approximate the distribution $(\underline{b} - \underline{\beta})T^{1/2}$. If the matrix (X'X) is in the form of correlation matrix $\hat{\underline{e}}$ need not be centered for it is orthogonal to the matrix X.

Efron (1981) used the bootstrap method to calculate the standard deviations of the Pearson correlation coefficient. He discovered that the bootstrap performs best among the non-parametric methods such as Jackknife, half-sample and random subsampling. Efron (1982) extended his study of the comparison of the performances of the bootstrap and other non-parametric methods. The method that we are going to focus on is the percentile method for assigning approximate confidence intervals to any real valued parameter $\theta(F)$ based on the bootstrap distribution of $\hat{\theta} = \theta(\hat{F})$. Let

$$\hat{CDF}(t) = Prob_*(\hat{\theta}^* \leq t)$$

be the distribution function of the bootstrap distribution of $\hat{\theta}^*$ the estimated value of $\hat{\theta}$ obtained from the bootstrap sample. If the bootstrap distribution is obtained by the Monte Carlo method, the $\hat{CDF}(t)$ is approximated by $(\#\hat{\theta}^* \leq t)/N$; N is the number of the Monte Carlo samples. We define the lower and upper bound values of the confidence interval of $\hat{\theta}$ as

$$\hat{\theta}_L(\alpha) = \hat{CDF}^{-1}(\alpha/2), \ \hat{\theta}_U = \hat{CDF}^{-1}(1-\alpha/2)$$

where α is the level of significance. Consequently, an approximate $(1-\alpha)$ central confidence interval for $\hat{\theta}$ is $[\hat{\theta}_L(\alpha), \hat{\theta}_U(\alpha)]$.

Freedman and Peters (1984a) studied the performances of the bootstrap method in the generalized least squares context. They also outlined how the bootstrap method can be used to examine an estimator's forecasts distribution. Freedman and Peters (1984b) estimated the standard errors for regression coefficients obtained by constrained generalized least squares with an estimated asymptotic covariance matrix via the bootstrap method. They found that the standard deviations of the bootstrap estimates are larger than the estimated asymptotic standard errors. Moreover, the bootstrap estimates of standard errors are closer to the true values than the conventional asymptotic approach. Still, the bootstrap estimates of standard errors are biased downward.

Freedman and Peters (1985) demonstrated the use of the bootstrap method to find the out-of-sample forecasts standard errors and to select between alternative mode specifications in the context of a dynamic energy demand model. They had shown that the bootstrap standard errors are more reliable than those obtained through the asymptotic methods. Efron (1987) studied the setting of approximate confidence intervals for a real valued parameter. He considered a method called biased corrected method. This method corrects the biasedness in the percentile method to achieve second order correctness which makes the CDF of bootstrap distribution complete. For an application see DiCiccio and Tibshirani (1987).

6.4.2 CONFIDENCE INTERVALS FOR FORECASTS

We are interested in constructing confidence intervals for the out-of-sample forecasts obtained from the biased estimators described in chapter 4. Consider the Equation (6.1)

$$\underline{y} = X\underline{\beta} + \underline{e}$$

Let $\underline{\delta}$ be an estimator of $\underline{\beta}$. The residuals are defined as

$$\hat{\underline{e}} = \underline{y} - X\underline{\delta}$$

Let $\underline{e}^*$ be a (T×1) vector of disturbance terms obtained by the resampling with replacement the elements of $\underline{e}$. We construct a bootstrap sample by the relationship

$$\underline{y}^* = X\underline{\delta} + \underline{e}^* \qquad (6.13)$$

From equation (6.13), we reestimate $\underline{\delta}$ to obtain $\underline{\delta}^*$. Then, we use $\underline{\delta}^*$ as a predictor by

$$\hat{\underline{y}}_0 = X_0 \underline{\delta}^*$$

where X_0 is previously defined. For simplicity, we assume that X_0 is a (1×K) dimension matrix. Repeating the procedure N number of times to get N values of forecasts. We use the N forecast values to construct a histogram for the distribution of the bootstrap estimates. The confidence interval is constructed by the percentile method. Suppose that we wish to construct a confidence interval with the level of significance α. The lower boundary $\hat{y}_{0L}$ is defined as

$$\frac{(\#\hat{y}_0 \leq \hat{y}_{0L})}{N} = (\alpha/2) \qquad (6.14a)$$

where $(\#\hat{y}_0 \leq \hat{y}_{0L})$ is the number of forecasts that are less than or equal to $\hat{y}_{0L}$. Similarly the upper boundary $\hat{y}_{0U}$ is defined as

$$\frac{(\#\hat{y}_0 \leq \hat{y}_{0U})}{N} = (1-\alpha/2) \qquad (6.14b)$$

The approximated confidence interval is $[\hat{y}_{0L}, \hat{y}_{0U}]$.

CHAPTER 7

THE BOOTSTRAP CONFIDENCE INTERVALS FOR OUT-OF-SAMPLE PREDICTION IN THE PRESENCE OF MULTICOLLINEARITY USING BIASED PREDICTORS

7.1 INTRODUCTION.

In this chapter, we use the bootstrap sampling technique to construcy the confidence intervals for out-of-sample predictions when multicollinearity is present and biased predictors are . The biased predictors of interest are the traditional biased estimators outlined in Chapters 4 and 5. They are simple ridge regression and two Stein-like estimators. As we have indicated in Chapter 6, the confidence intervals for the forecast values of these biased predictors cannot be acquired in the usual way. Consequently, we use the bootstrap re-sampling method to construct reasonable confidence intervals of the forecasts.

The outline of the chapter is as follows: Section 7.2, we portray the in-sample and out-of-sample data used in our study and describe the nature of multicollinearity in the data. In Section 7.3, we use our predictors obtained from the in-sample data to make out-of-sample forecasts. Then we perform a bootstrap re-sampling algorithm on each of the predictors to construct reasonable confidence intervals at various levels of confidence for their forecasts values. In addition, we construct confidence intervals for the least squares predictor to be used as a benchmark.

7.2 DATA DESCRIPTION.

The data that we use in this chapter is the price promotion data used in Chapter 5. Again, we use the price promotion model in Equation (5.18) to explain the unit sales of a major brand canned tuna as a function of its own price, the prices of its competitors, discounts and types of advertising being used. The data contains 52 store weeks of data for a chain store.

In order to perform out-of-sample prediction, we divide the data into two groups, in-sample data and out-of-sample data. The in-sample data is composed of the first 48 store weeks of data. We use the remaining 4 store weeks of data as the basis for out-of-sample predictions.

We apply three multicollinearity diagnostics to the in-sample data and list the findings below.

1. Simple Correlations among Regressors.

Severe collinearity is said to be present when a correlation coefficient exceeds 0.8. By examining pair-wise correlation coefficients between the explanatory variables (omitting the intercept term), we detect correlation between the variable $I_{DISMAD3}$ (the dummy variable indicating display and major ad campaign for brand #3) and d_3 (the price discount variable for brand #3) having the correlation coefficient equal to 0.84.

2. Determinant of $(X'X)$.

We center and normalize to unit length the matrix of exogenous variables X before computing the determinant. If the determinant is 0, then one or more exact linear dependencies exist among the columns of X. If the determinant is 1, then the columns of X are orthogonal. The determinant of the centered and normalized to unit length matrix X'X is 1.636E-8 which indicates that there exist certain linear dependencies among the exogenous variables.

3. Matrix Decompositions.

We calculate the condition numbers for the matrix X'X when the matrix X is un-conditioned, mean centered and centered and normalized to unit length. For the un-conditioned X matrix, we also include a column of ones for the intercept parameter. The condition number associated with the eigenvalue λ_i is defined as the square root of the ratio λ_1/λ_i, $i = 1,2,\ldots,K$. The condition numbers are presented in Table 7.1. Pertaining to the reported eigenvalues, we can deduce that by mean centering and centering and normalizing to unit length the exogenous variables we can eliminate the multicollinearity problem to a certain extent. The reasons for the decrease in the degree of multicollinearity by centering and centering and normalizing to unit length are given in Chapter 5.

As indicated in Section 6.2, if the in-sample data and out-of-sample data are similar enough or have the same pattern, so that they are rotationally and variationally equivalent, least squares prediction seems to have no

TABLE 7.1

The condition numbers for the untransformed and transformed data.

actual	centered	centerd and normalized
1.000	1.000	1.000
3.011	1.049	1.219
3.067	1.073	1.252
3.110	1.364	1.308
4.004	1.653	1.420
4.764	1.734	1.541
4.998	1.853	1.641
5.341	1.907	1.653
5.495	1.952	1.737
5.630	1.961	1.794
5.664	1.985	1.812
5.722	2.033	1.858
5.859	2.351	2.135
6.778	2.635	2.263
7.559	2.837	2.455
8.280	3.248	2.669
9.375	3.924	3.415
11.312	4.605	3.447
13.273	4.881	4.206
14.324	6.605	4.627
19.911	8.916	5.171
26.801	13.496	7.110
38.903	14.823	8.486
42.771	34.624	11.097
100.103	66.713	26.185
255.285		

serious prediction risk. The data ellipsoids of the in-sample and out-of-sample data are rotationally equivalent if $P_1 \cdot P_2 = I_K$ and are variationally equivalent if $\Lambda_1 = \Lambda_2$, the matrices P_1, P_2, Λ_1 and Λ_2 are defined in Chapter 6. We find little rotational equivalence or variational equivalence between the in-sample and out-of-sample data. Consequently, there appears to be little similarity between the in-sample and out-of-sample data that we use. Therefore, the least squares predictor may suffer large forecast variability.

The rotational equivalence between the in-sample and out-of-sample data are evaluated by the determinant of $(I-P_1'P_2)$. If the in-sample and out-of-sample data are rotationally equivalent, $\det(I-P_1'P_2)$ is equal to zero. For this study, the determinant has the value of -8.12e-18 which suggests near rotational equivalence between the data. However, by observation, the product $P_1'P_2$ is far from resembling an identity matrix. Furthermore, the eigenvalues of $P_1'P_2$ indicate severe collinearity among the column vectors. It appears that the determinant criteria used is too rough a measure of estimating rotational equivalence. Moreover, the measure relies on the matrix $(I-P_1'P_2)$ to be a null matrix when there is rotational equivalence and yield zero determinant. A matrix need not be a null matrix to have zero determinant; for instance, a singular matrix. The determinant of $(I-P_1'P_2)$ ought to be used along with other evaluation criteria to determine rotational equivalence.

The variational equivalence between the in-sample and

out-of-sample data are evaluated by the trace of $(I-\Lambda_2\Lambda_1^{-1})$. If the data are variationally equivalent, then $tr(I-\Lambda_2\Lambda_1^{-1})$ is equal to zero. The calculated trace for this study is 25.748 indicating no evidence of variational equivalence between the in-sample and out-of-sample data.

7.3 BOOTSTRAP CONFIDENCE INTERVALS.

7.3.1 THE CONFIDENCE INTERVALS.

When biased predictors are used to make out-of-sample forecasts, we cannot construct the confidence intervals for the forecast values in the usual way. As a consequence, we use the bootstrap re-sampling technique to assist in constructing reasonable confidence intervals.

As outlined in Section 6.4.2, in order to perform bootstrap re-sampling, we have to create N samples of the vector of disturbance term, namely $\underline{e}^*$. Suppose that N is equal to 1,000. Consider the equation

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{e} \qquad (7.1)$$

where $\underline{y}$ is a $(T_1\times K)$ vector of observations on a dependent variable, $\mathbf{X}$ is a fixed $(T_1\times K)$ full rank matrix of observations on the exogenous variables, $\underline{\beta}$ is a $(K\times 1)$ vector of unknown parameters and $\underline{e}$ is a $(T_1\times 1)$ vector of disturbance terms which are identically and independently distributed as $N(0,\sigma^2)$. In this study, the matrix $\mathbf{X}$ and the vector $\underline{y}$ represent the in-sample data set; hence, $T_1 = 48$.

Presume that we perform bootstrap re-sampling technique

on the least squares predictor. Initially, we apply least squares to Equation 7.1. Let $\underline{b}$ be the least squares estimator of $\underline{\beta}$

$$\underline{b} = (X'X)^{-1}X'\underline{y} \quad . \tag{7.2}$$

The vector of residuals $\hat{\underline{e}}$ is defined as

$$\hat{\underline{e}} = \underline{y} - X\underline{b} \quad . \tag{7.3}$$

We obtain a thousand samples of disturbance vector $\underline{e}^*$ by sampling with replacement the elements of the vector of residuals $\hat{\underline{e}}$. In selecting the elements of $\hat{\underline{e}}$, we have to ensure that each elements has equal opportunity of being selected to assure randomness. Therefore, we select the the elements of $\hat{\underline{e}}$ by using a uniform random number generator with range [0.5,48.5]. The generated random numbers are then rounded off to the nearest integer. The rows of $\hat{\underline{e}}$ are chosen in correspondence with the integers. We generate a thousand samples of the vector $\underline{e}^*(T\times 1)$.

With the samples of $\underline{e}^*$, we construct samples of vector of observations on the dependent variable by

$$\underline{y}^* = X\underline{b} + \underline{e}^* \quad . \tag{7.4}$$

We estimate Equation (7.4) via least squares and obtain a thousand vectors of $\underline{b}^*$ where

$$\underline{b}^* = (X'X)^{-1}X'\underline{y}^* \quad . \tag{7.5}$$

Note that the estimator applied in this stage must be the same as the one applied to Equation (7.2). Let X_0 be a $(T_2\times K)$ matrix containing the out-of-sample observations on

the endogenous variables; hence, $T_2 = 4$. We use the estimator $\underline{b}^*$ to make forecasts based on the X_0 matrix. Let $\underline{y}_P$ denote a $(T_2 \times 1)$ vector of forecast values; thus

$$\underline{y}_P = X_0 \underline{b}^* \quad . \tag{7.6}$$

From Equation (7.4), we have a thousand predictions for each of the four periods under study. The resulting forecasts are sorted from minimum to maximum. Suppose that we wish to construct a confidence interval with 99% level of confidence for the forecasts of week 49 of the least squares predictor, we use the 6-th lowest forecast value as the lower boundary and the 6-th highest forecast value as the upper boundary.

For the confidence intervals with 95% and 90% level of confidence, we use the 26-th and 51-th lowest and highest forecast values as the lower and upper boundary, respectively. We follow similar process in constructing confidence intervals for the ridge regression, Stein (D=I), Stein ($D=X_D'X_D$), PC-Stein (D=I) and PC-Stein ($D=X_D'X_D$) predictor.

Let $\underline{\delta}$ denote a biased estimator; then, the vector of residuals $\underline{\tilde{e}}$ is defined as

$$\underline{\tilde{e}} = \underline{y} - X\underline{\delta} \quad .$$

Similar to the case of least squares predictor, a thousand samples of disturbance vectors of $\underline{\tilde{e}}^*$ are obtained by sampling with replacement the elements of using uniform random number generator. With the vectors $\underline{\tilde{e}}^*$, we construct

samples of vector of observations on the dependent variable by

$$\tilde{\underline{y}}^{*} = X\underline{\delta} + \tilde{\underline{e}}^{*} \quad .$$

Then we estimate $\tilde{\underline{y}}^{*}$ via the same estimation technique as $\underline{\delta}$ to obtain a thousand vectors of $\underline{\delta}^{*}$. We use $\underline{\delta}^{*}$ to make out-of-sample forecasts based on the matrix X_0; ie.

$$\tilde{\underline{y}}_P = X_0\underline{\delta}^{*} \quad .$$

The confidence intervals for the forecast values are constructed in the same manner as those of the least squares predictor described earlier.

7.3.2 BOOTSTRAP SAMPLING RESULTS.

In Table 7.2, we present the parameter estimates obtained from applying the least squares, ridge regression, Stein (D=I), Stein ($D=X_D'X_D$), PC-Stein (D=I) and PC-Stein ($D=X_D'X_D$) estimator to the in-sample data. The derivation of each of the estimators and the abbreviations used are outlined in Chapter 5. For the PC-Stein estimators, we use the same elimination rule as in Chapter 5. We disregard the n smallest eigenvalues of the matrix $X_D'X_D$ that contribute in total merely about 5% of the total variation. The number of restrictions is 10.

From Table 7.2, we find that only the ridge regression estimator gives the correct signs for most of the parameter estimates associated with the own price variable (negative sign), the own discount variable (positive sign) and the

TABLE 7.2

The estimated parameter values of the equity estimator and its alternatives.

Parameter	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)	OLS
α_0	9.449	1.828	2.421	0.728
α_1	-1.241	6.927	6.319	8.177
α_2	1.461	1.166	1.164	1.376
α_3	0.381	0.283	0.293	0.334
α_4	0.267	0.199	0.199	0.235
α_5	1.403	1.238	1.434	1.462
β_{22}	0.480	0.739	0.629	0.872
β_{24}	-0.340	-0.292	-0.361	-0.344
β_{25}	-0.550	-0.626	-0.700	-0.739
β_{32}	-3.102	-3.634	-3.377	-4.289
β_{34}	0.189	0.138	0.100	0.163
β_{35}	0.330	0.635	0.421	0.750
β_{42}	0.710	2.872	2.611	3.390
β_{44}	-0.395	-0.448	-0.399	-0.529
β_{45}	-1.080	-1.349	-1.327	-1.592
μ_2	-0.122	-0.262	-0.148	-0.309
μ_3	-0.117	-0.467	-0.390	-0.551
μ_4	-0.419	-1.109	-0.993	-1.309
μ_5	-0.347	-1.042	-0.855	-1.230
μ_6	-0.057	-1.051	-0.918	-1.240
μ_7	-0.550	-1.620	-1.601	-1.912
μ_8	-0.899	-2.033	-2.063	-2.400
μ_9	-0.236	-1.410	-1.318	-1.665
μ_{10}	-0.558	-1.888	-1.744	-2.228
μ_{11}	-0.056	-1.483	-1.362	-1.750
μ_{12}	-0.464	-1.730	-1.642	-2.042

competitors' promotion variables (negative sign). The only exceptions are the parameters estimate associated with the variables I_{DIS3} and $I_{DISMAD3}$, I_{DIS3} is the dummy variable indicating the use of display only promotion campaign for brand #3 and $I_{DISMAD3}$ is the dummy variable indicating the display and major ad promotion campaign for brand #3, which have positive signs. Furthermore, we find that the estimates of the Stein (D=I) and PC-Stein (D=I) estimator are identical to those obtained via the least squares estimator. We also find that the parameter estimates obtained from the Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) closely resemble those obtained from the least squares estimator. The reason why the PC-Stein (D=I) and Stein (D=I) estimator yield identical results to the least squares estimator is the same as in Chapter 5, a_{max} being zero.

In Table 7.3, we present the values of the exogenous variables used in our projections. In Table 7.4, we report the bootstrap confidence intervals of the variable ln s_1 for week 49 obtained via various predictors. The use of ridge regression predictor produces the least variability while the PC-Stein predictor produces the least predicted mean square error. However, we note that the 99% confidence interval of the ridge regression predictor is the only confidence interval of the ridge regression predictor that contains the actual value. From Figure 7.1, we observe that the predicted values obtained from the ridge regression predictor are compactly distributed around the peak. The

TABLE 7.3

The out-of-sample exogenous variables

Parameter	variables	week 49	week 50	week 51	week 52
α_1	R_1	1.060	1.050	1.040	1.060
α_2	d_1	0.170	0.124	0.106	0.208
α_3	I_{MAD1}	0.000	0.000	0.000	0.000
α_4	I_{DIS1}	0.000	0.000	0.000	0.000
α_5	$I_{DISMAD1}$	0.000	0.000	0.000	0.000
β_{22}	d_2	0.302	0.255	0.283	0.264
β_{24}	I_{DIS2}	1.000	1.000	0.000	0.000
β_{25}	$I_{DISMAD2}$	0.000	0.000	1.000	0.000
β_{32}	d_3	0.178	0.149	0.149	0.149
β_{34}	I_{DIS3}	0.000	0.000	0.000	0.000
β_{35}	$I_{DISMAD3}$	0.000	0.000	0.000	0.000
β_{42}	d_4	0.230	0.113	0.113	0.103
β_{44}	I_{DIS4}	0.000	0.000	0.000	0.000
β_{45}	$I_{DISMAD4}$	1.000	0.000	0.000	0.000

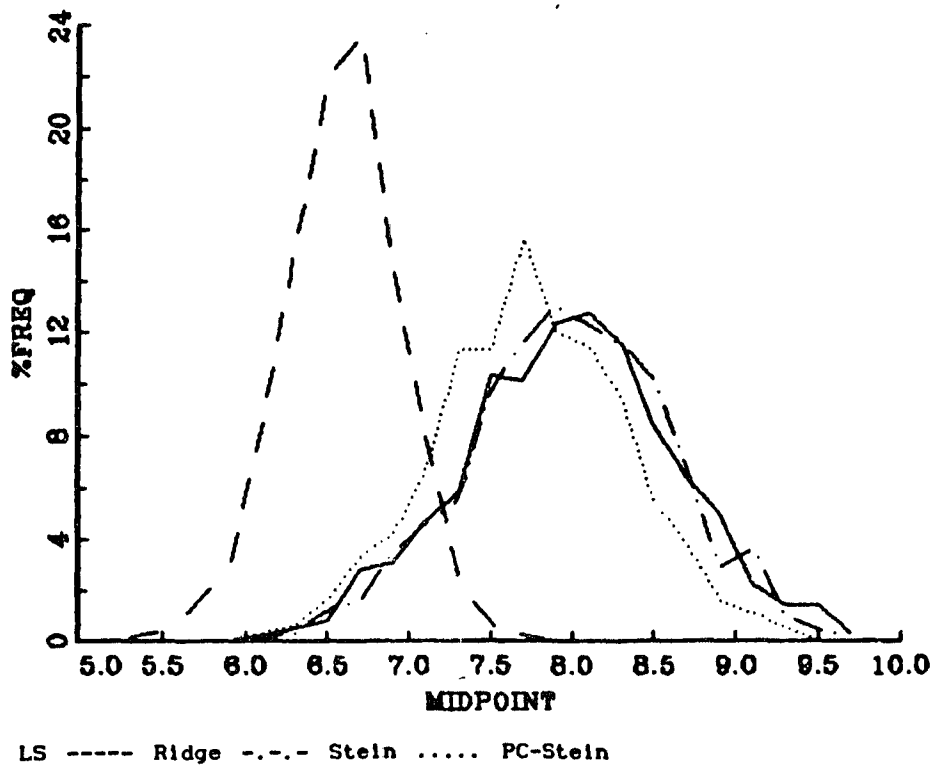
TABLE 7.4

The $\ln(s_1)$ forecasts of week 49 obtained from the predictors. (actual value = 7.336)

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	7.974	6.583	7.969	7.741
Std.	0.647	0.349	0.616	0.573
Max	9.717	7.626	9.703	9.387
Min	6.068	5.385	6.084	6.082
Med.	7.990	6.593	7.981	7.725
Pmse	0.825	0.689	0.780	0.492
99% confidence	[6.314, 9.499]	[5.636, 7.543]	[6.406, 9.422]	[6.322, 9.185]
95% confidence	[6.706, 9.254]	[5.879, 7.224]	[6.756, 9.148]	[6.622, 8.926]
90% confidence	[6.835, 9.022]	[6.001, 7.141]	[6.885, 9.010]	[6.765, 8.694]

FIGURE 7.1

The distribution of the forecast of $\ln(s_1)$ (week 49).



predicted values obtained from the least squares, Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) predictor are dispersed with similar characteristics.

From Table 7.5, we find that the ridge regression predictor offer the predicted values of ln s_1 for week 50 with the most precision. The PC-Stein ($D=X_D'X_D$), Stein ($D=X_D'X_D$) and least squares predictor yield the predicted values with almost identical distributions , Figure 7.2. However, the PC-Stein ($D=X_D'X_D$) and Stein ($D=X_D'X_D$) predictor produce predicted values that are more accurate than those of the least squares predictor and ,hence, yield confidence intervals with smaller ranges.

From Tables 7.6-7.7 and Figures 7.3-7.4, we observe that the ridge regression predictor gives the predicted values with the least variability for week 51-52 and, as a consequence, produces the confidence intervals with the smallest ranges. Once again, the PC-Stein ($D=X_D'X_D$) and Stein ($D=X_D'X_D$) predictor produce the predicted values with similar characteristics to those of the least squares predictor. However, the PC-Stein ($D=X_D'X_D$) and Stein ($D=X_D'X_D$) predictor still improve upon the least squares predictor by providing smaller prediction mean square error and variability. Overall, the ridge regression predictor yields the predicted values with the most accuracy which becomes increasingly evident as we further forecast into the future. The PC-Stein ($D=X_D'X_D$), Stein ($D=X_D'X_D$) predictor offer predictions with less variability than those of the least sqares estimator.

TABLE 7.5

The $\ln(s_1)$ forecasts of week 50 obtained from the predictors. (actual value = 7.286)

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	9.116	7.484	8.847	8.637
Std.	0.617	0.240	0.586	0.557
Max	11.085	8.562	10.632	10.302
Min	7.139	6.478	6.569	6.625
Med.	9.116	7.469	8.834	8.612
Pmse.	3.729	0.097	2.781	2.136
99% confidence	[7.615, 10.522]	[6.876, 8.214]	[7.314, 10.316]	[7.289, 10.059]
95% confidence	[7.956, 10.316]	[7.047, 8.038]	[7.696, 9.993]	[7.634, 9.742]
90% confidence	[8.129, 10.168]	[7.131, 7.886]	[7.908, 9.778]	[7.766, 9.622]

FIGURE 7.2

The distribution of the forecast of $\ln(s_1)$ (week 50).

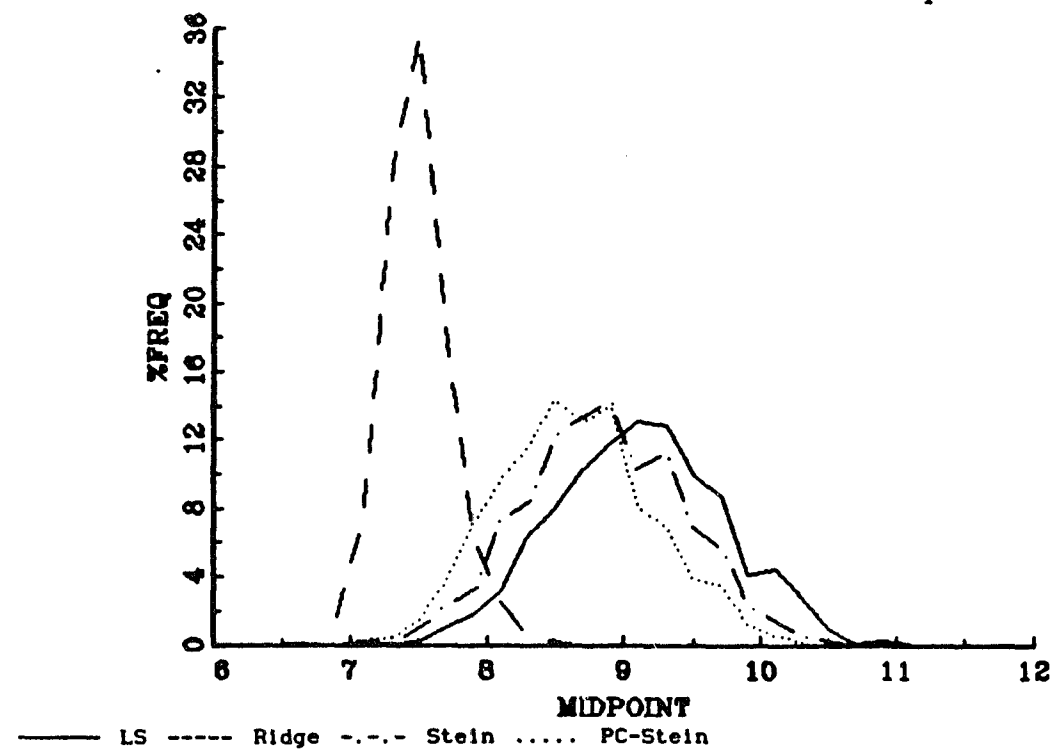


TABLE 7.6

The $\ln(s_1)$ forecasts of week 51 obtained from the predictors. (actual value = 7.263)

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	8.638	7.356	8.483	8.266
Std.	0.541	0.220	0.517	0.487
Max	10.485	8.410	10.195	10.130
Min	6.856	6.619	6.676	6.690
Med.	8.647	7.340	8.483	8.242
Pmse.	2.182	0.057	1.755	1.244
99% confidence	[7.368, 9.939]	[6.740, 7.999]	[7.232, 9.732]	[7.212, 9.468]
95% confidence	[7.645, 9.723]	[6.951, 7.781]	[7.511, 9.428]	[7.408, 9.207]
90% confidence	[7.756, 9.543]	[7.016, 7.712]	[7.645, 9.345]	[7.484, 9.092]

FIGURE 7.3

The distribution of the forecast of $\ln(s_1)$ (week 51).

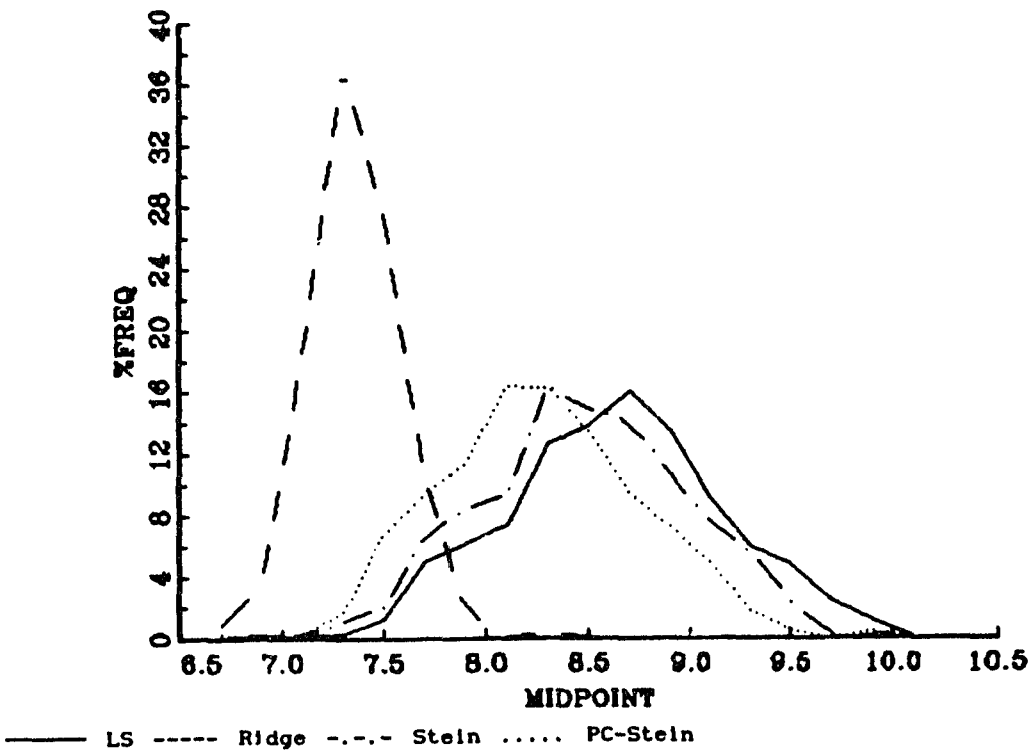


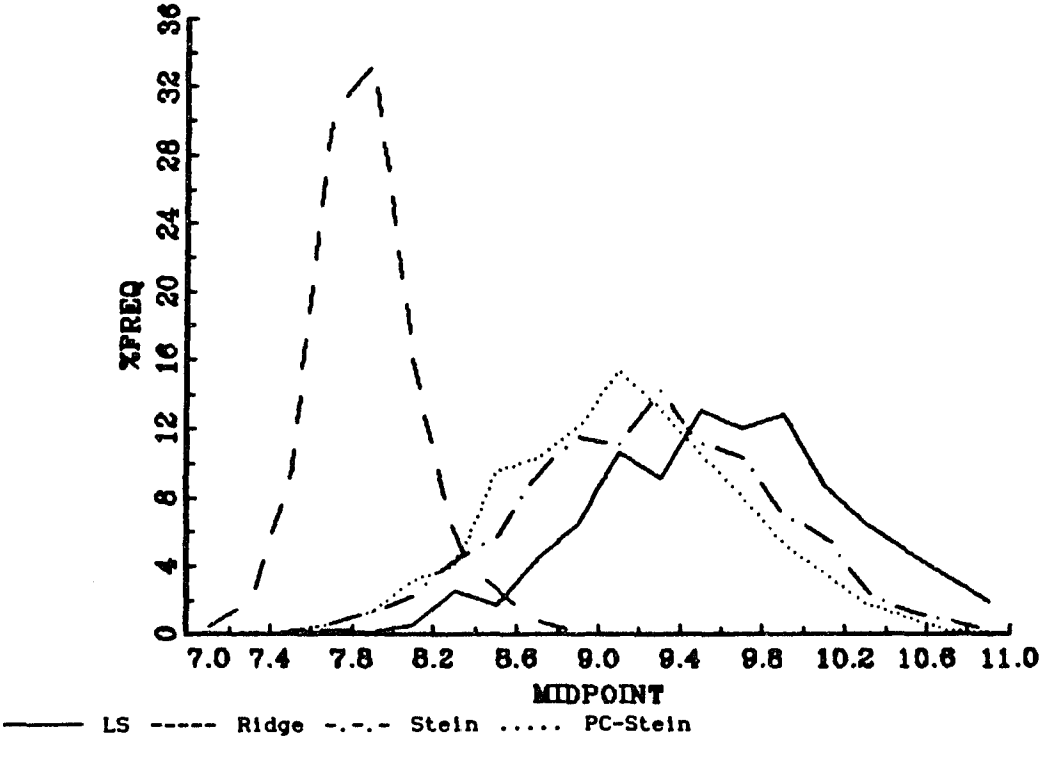
TABLE 7.7

The $\ln(s_1)$ forecasts of week 52 obtained from the predictors. (actual value = 7.672)

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	9.620	7.864	9.234	9.124
Std.	0.629	0.249	0.601	0.566
Max	11.528	8.860	11.013	10.910
Min	7.798	7.044	7.077	7.389
Med.	9.629	7.840	9.231	9.122
Pmse.	4.189	0.099	2.801	2.429
99% confidence	[8.083, 11.092]	[7.245, 8.673]	[7.716, 10.647]	[7.771, 10.531]
95% confidence	[8.332, 10.859]	[7.409, 8.431]	[8.045, 10.391]	[8.051, 10.297]
90% confidence	[8.614, 10.691]	[7.496, 8.308]	[8.223, 10.189]	[8.195, 10.066]

FIGURE 7.4

The distribution of the forecast of $\ln(s_1)$ (week 52).



The next issue to be addressed is how well do the bootstrap confidence intervals represent the actual confidence intervals. Unlike the biased predictors that we considered, the covariance matrix for the prediction error of the least squares predictor can be calculated. The covariance matrix of the least squares predictions is defined as

$$E\left[(\hat{\underline{y}}_0 - \underline{y}_0)'(\hat{\underline{y}}_0 - \underline{y}_0)\right] = E\left[(\mathbf{X}_0\underline{b} - \mathbf{X}_0\underline{\beta} - \underline{e}_0)(\ .\ )'\right]$$

$$= E\left[(\mathbf{X}_0(\underline{b} - \underline{\beta}) - \underline{e}_0)(\ .\ )'\right]$$

$$= \mathbf{X}_0 E\left[(\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})'\right]\mathbf{X}_0 + E\left[\underline{e}_0\underline{e}_0'\right]$$

$$= \sigma^2\left[\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0' + \mathbf{I}_{T2}\right] . \quad (7.7)$$

This relationship indicates that the prediction variability of the least squares predictor is composed of (a) the equation error $\underline{e}_0$ and (b) the error in predicting the parameters $\underline{\beta}$. From Equation (6.2), the random variable $\underline{y}_0$ is expressed as

$$\underline{y}_0 = \mathbf{X}_0\underline{\beta} + \underline{e}_0$$

where $\underline{e}_0 \sim N(\underline{0}, \sigma^2\mathbf{I}_{T2})$ and $E[\underline{e}\underline{e}_0'] = 0$. Let $\underline{x}_0'$ be a row of the matrix $\mathbf{X}_0$. Then, the forecast value associated with $\underline{x}_0'$ is $y_0 = \underline{x}_0'\underline{b}$ which has its variance equal to $\sigma^2(\underline{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\underline{x}_0 + 1)$. Consequently, the random variable

$$\frac{\underline{x}_0'\underline{b} - y_0}{\sigma(\underline{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\underline{x}_0 + 1)^{1/2}}$$

is distributed as a standard normal random variable with mean 0 and variance 1. It follows that the random variable

$$\frac{\underline{x}_0'\underline{b} - y_0}{\hat{\sigma}(\underline{x}_0'(X'X)^{-1}\underline{x}_0+1)^{1/2}}$$

is distributed as a t random variable with T-K degree of freedom; $\hat{\sigma}^2 = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})/(T-K)$. The confidence interval is, therefore, established by

$$Pr\left[-t_{(T-K,\alpha/2)} \leq t_{(T-K)} \leq t_{(T-K,\alpha/2)}\right] = 1-\alpha$$

$$\text{or } Pr\Big[\underline{x}_0'\underline{b} - t_{(T-K,\alpha/2)}\hat{\sigma}(\underline{x}_0'(X'X)^{-1}\underline{x}_0 + 1)^{1/2} \leq$$

$$y_0 \leq \underline{x}_0'\underline{b} + t_{(T-K,\alpha/2)}\hat{\sigma}(\underline{x}_0'(X'X)^{-1}\underline{x}_0 +1)^{1/2}\Big] = 1-\alpha \ . \quad (7.8)$$

We use the relationship in Equation (7.8) to construct the prediction confidence intervals for the least squares predictor at various levels of confidence. The results are reported in Table 7.8a.

In order to evaluate the appropriateness of the bootstrap confidence intervals of the ridge, Stein and PC-Stein predictors, we utilize the Monte Carlo Experiment. Using the parameter values from the last column of Table 5.2, 400 Monte Carlo samples are generated. The disturbance terms are generated from a standard normal distribution. For each Monte Carlo sample, we construct 95% bootstrap confidence intervals for periods 49-52 forecasts; the number of bootstrap samples used is 200. This procedure is designed to test the method that is used to construct the confidence intervals. If the constructed confidence intervals

TABLE 7.8a

The calculated true confidence intervals of the least squres predictors.

period	90% confidence	95% confidence	99% confidence
49	[6.200,9.714]	[5.832,10.110]	[5.064,10.878]
50	[7.420,10.799]	[7.069,10.151]	[6.336,11.884]
51	[7.130,10.136]	[6.818,10.449]	[6.166,11.101]
52	[7.891,11.351]	[7.531,11.711]	[6.780,12.461]

TABLE 7.8b

The percentage of time that the bootstrap confindence intervals contain the actual forcast values (level of confidence is 95%).

period	ridge	Stein($D=X_D'X_D$)	PC-Stein($D=X_D'X_D$)	LS
49	99.75	99.50	100.00	94.25
50	100.00	99.25	100.00	94.50
51	100.00	99.75	100.00	95.00
52	100.00	99.25	100.00	96.25

reasonably represent the true confidence intervals, we expect roughly 95% of the times the constructed intervals contain the actual forecast values.

From Table 7.8b, we discover that the bootstrap confidence intervals contain the 'true' forecast values more often than expected for the biased predictors. This seems to be the result of the overestimation of the forecast variability of the bootstrap. The larger the estimated the forecast variability is, the larger the confidence intervals become. As a consequence, the constructed confidence intervals often contain the actual forecast values more than they should. However, the bootstrap is still a useful tool to construct confidence intervals for biased predictors which previously are unknown.

Now, we forecast the values of ln s_1 along the own discount variable, d_1. We are presuming the role of the management of the target brand emphasizing price discount promotion campaign. The assumption is that the competitive brands maintain the discount rates and promotion campaign as in week 49 which are assumed to be known throughout the analysis. By this assumption, the out-of-sample exogenous data used is non-stochastic. The rates of discount used are 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5. The examination of the in-sample data reveals that the target brand (brand #1) discount its product at the rate between 0.0 and 0.51 which suggests that any discount rate beyond 0.50 is very unlikely and may not be feasible. We construct confidence intervals

for alternative values of discount rate by the bootstrap sampling technique.

Tables 7.9-7.14 present the confidence intervals obtained from each of the predictors considered together with the descriptive statistics of the distributions of the predicted values. Figures 7.5-7.10 illustrate the frequency distributions of the acquired predicted values. In general, the ridge regression predictor provides the confidence intervals with the smallest ranges and yields predicted values that are similarly distributed for all values of d_1 considered; futhermore, it also have central tendencies, measured by the means and medians that are lower than those of the other predictors considered. The least squares predictor gives confidence intervals that have the largest ranges. The Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) produce confidence intervals that are very comparable in width.

Figures 7.11-7.16, we make pair-wise comparison between the 95% confidence intervals, which has moderate level of confidence, obtained from each of the predictors considered. The confidence intervals constructed from the ridge regression predictor evidently have smaller ranges than those constructed from the other predictors. Moreover, we discover that the ranges of the confidence intervals of the ridge regression predictor are quite stable throughout the analysis. The means of the predicted values of the ridge regression predictor are lower than those of the Stein ($D=X_D'X_D$), PC-Stein ($D=X_D'X_D$) and the least squares predictor.

TABLE 7.9

The ln (s_1) forecasts obtained from the predictors when $d_1 = 0$.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	7.738	6.331	7.788	7.559
Std.	0.639	0.342	0.609	0.565
Max	9.512	7.373	9.486	9.086
Min	5.906	5.119	6.065	5.917
Med.	7.735	6.335	7.804	7.554
99% confidence	[6.099, 9.293]	[5.305, 7.230]	[6.225, 9.251]	[6.151, 8.965]
95% confidence	[6.486, 9.033]	[5.666, 6.985]	[6.547, 8.956]	[6.447, 8.679]
90% confidence	[6.624, 8.782]	[5.758, 6.880]	[6.729, 8.768]	[6.581, 8.480]

FIGURE 7.5

The distribution of the forecast of ln (s_1) $d_1 = 0$.

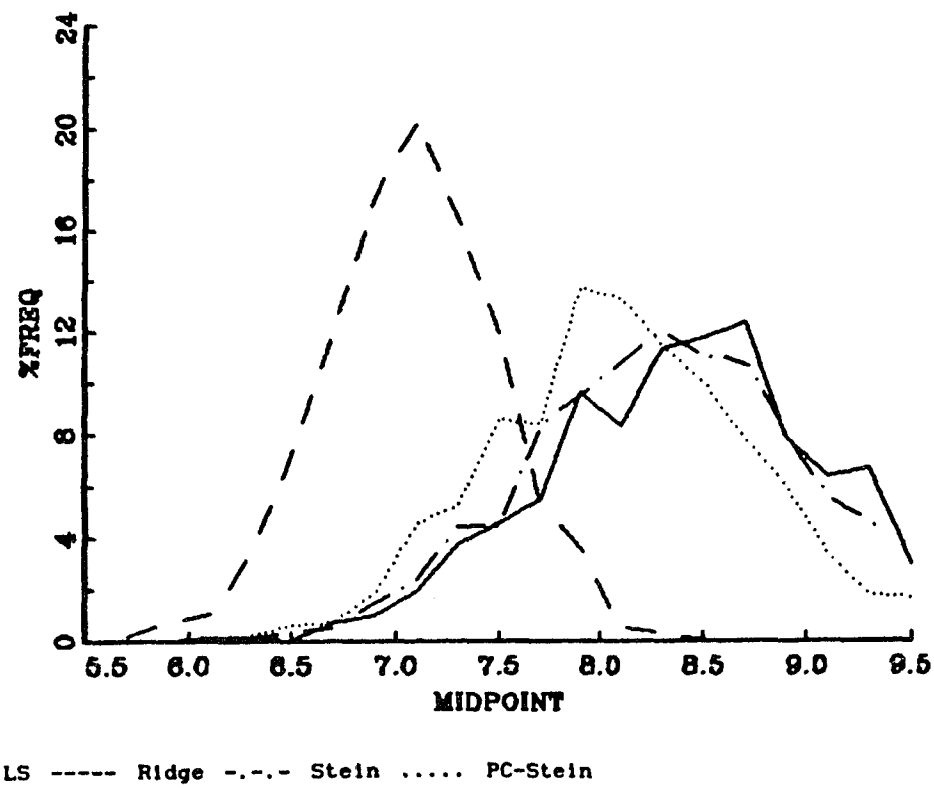


TABLE 7.10

The ln (s_1) forecasts obtained from the predictors when d_1= 0.1.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	7.877	6.480	7.895	7.666
Std.	0.642	0.344	0.612	0.568
Max	9.633	7.510	9.614	9.263
Min	6.027	5.310	6.084	6.076
Med.	7.883	6.487	7.907	7.653
99%	[6.240,	[5.511,	[6.310,	[6.291,
confidence	9.413]	7.368]	9.373]	9.109]
95%	[6.618,	[5.785,	[6.688.	[6.527,
confidence	9.154]	7.118]	9.050]	8.857]
90%	[6.730,	[5.897,	[6.789,	[6.683,
confidence	8.917]	7.035]	8.926]	8.587]

FIGURE 7.6

The distribution of the forecast of ln(s_1) d_1= 0.1.

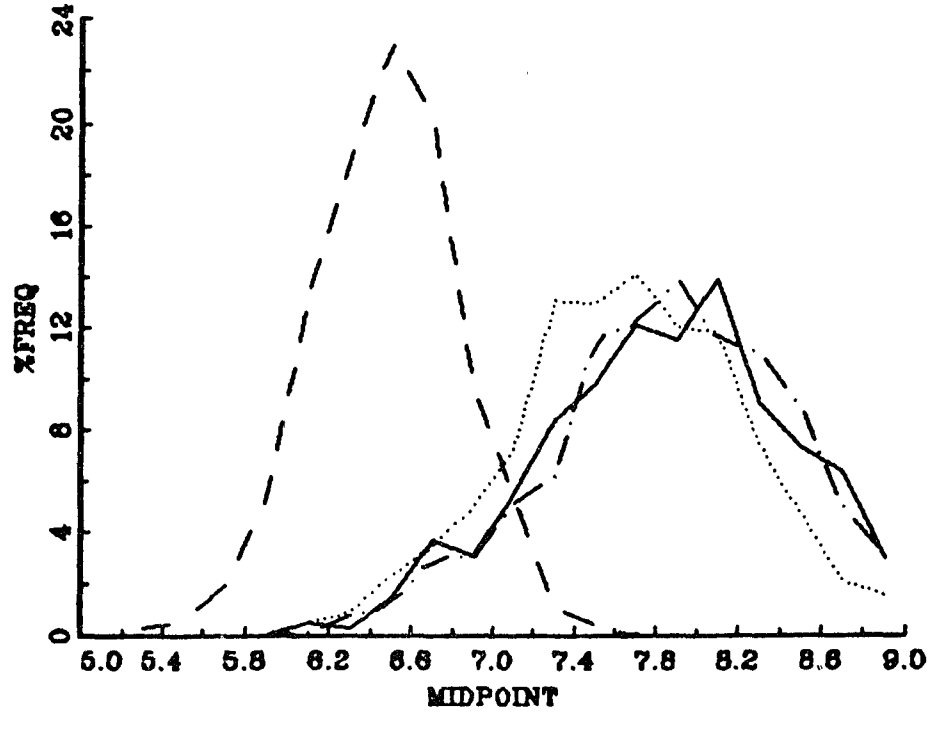


TABLE 7.11

The ln (s_1) forecasts obtained from the predictors when d_1= 0.2.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	8.016	6.628	8.001	7.773
Std.	0.650	0.352	0.619	0.576
Max	9.764	7.676	9.741	9.440
Min	6.085	5.417	6.083	6.078
Med.	8.034	6.636	8.011	7.764
99% confidence	[6.346, 9.549]	[5.687, 7.503]	[6.448, 9.434]	[6.347, 9.216]
95% confidence	[6.738, 9.287]	[5.905, 7.279]	[6.791, 9.194]	[6.651, 8.951]
90% confidence	[6.877, 9.049]	[6.045, 7.198]	[6.912, 9.033]	[6.800, 8.724]

FIGURE 7.7

The distribution of the forecast of ln (s_1) d_1= 0.2.

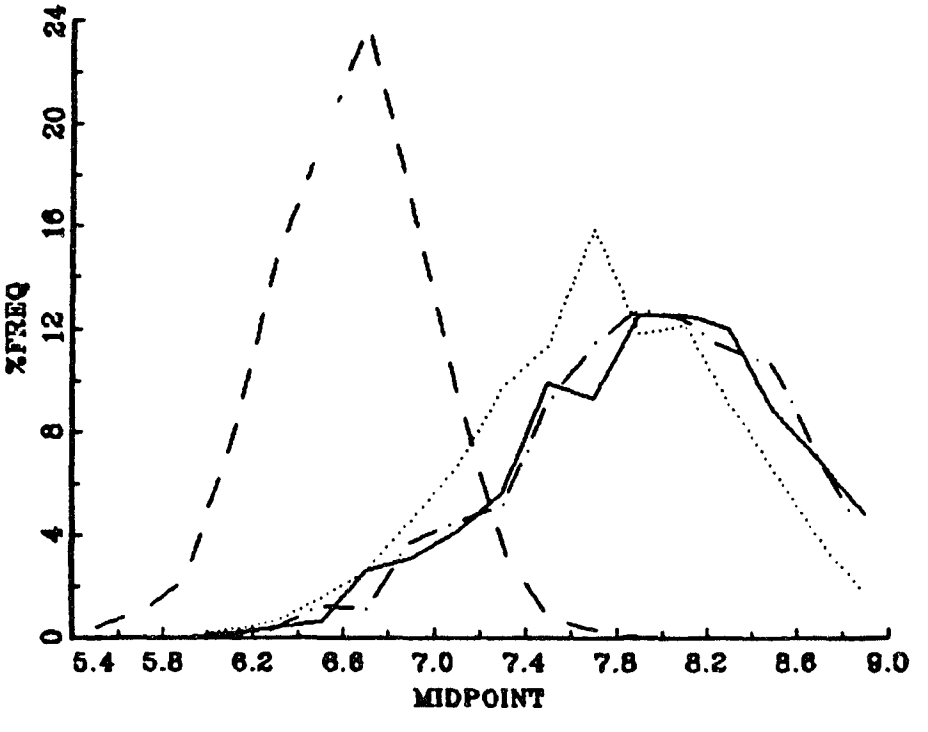


TABLE 7.12

The ln (s_1) forecasts obtained from the predictors when d_1= 0.3.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	8.155	6.776	8.107	7.881
Std.	0.662	0.367	0.630	0.588
Max	9.968	7.872	9.869	9.617
Min	6.123	5.526	6.083	6.067
Med.	8.177	6.785	8.107	7.868
99% confidence	[6.485, 9.712]	[5.787, 7.734]	[6.508, 9.474]	[6.400, 9.317]
95% confidence	[6.850, 9.468]	[6.015, 7.450]	[6.855, 9.309]	[6.736, 9.042]
90% confidence	[7.022, 9.246]	[6.176, 7.379]	[7.012, 9.158]	[6.905, 8.864]

FIGURE 7.8

The distribution of the forecast of ln (s_1) d_1= 0.3.

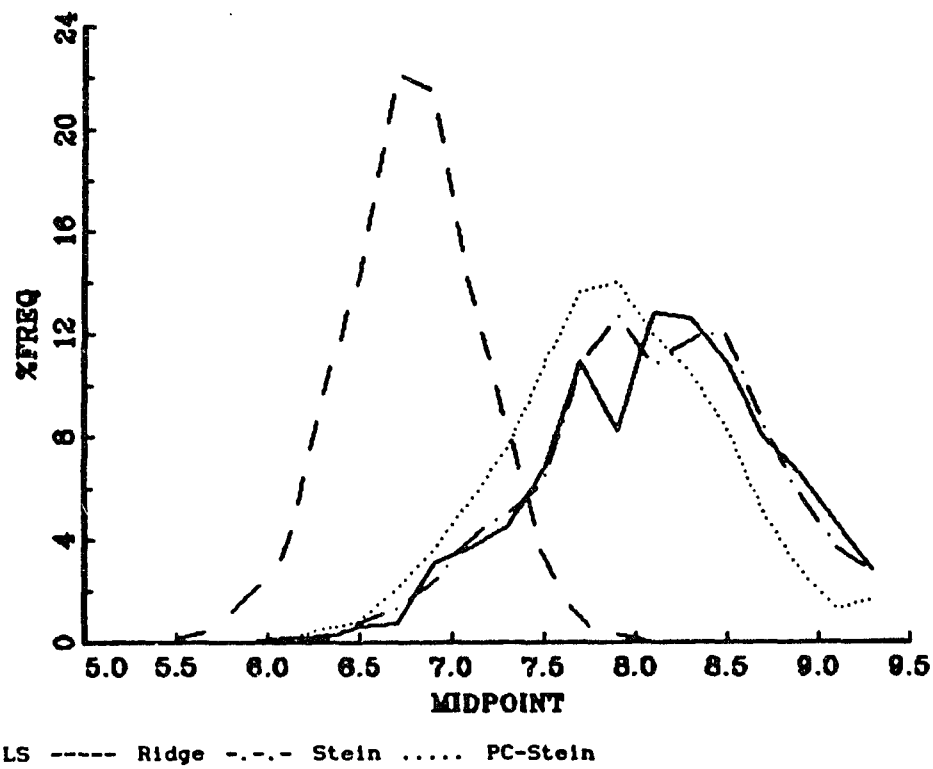


TABLE 7.13

The ln (s_1) forecasts obtained from the predictors when $d_1 = 0.4$.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	8.293	6.924	8.213	7.988
Std.	0.678	0.388	0.646	0.603
Max	10.171	8.068	9.996	9.794
Min	6.130	5.634	6.083	6.057
Med.	8.314	6.925	8.209	7.979
99% confidence	[6.652, 9.890]	[5.842, 7.914]	[6.567, 9.641]	[6.423, 9.436]
95% confidence	[6.963, 9.632]	[6.159, 7.645]	[6.914, 9.432]	[6.841, 9.157]
90% confidence	[7.142, 9.417]	[6.283, 7.570]	[7.123, 9.293]	[6.987, 9.003]

FIGURE 7.9

The distribution of the forecast of ln(s_1) $d_1 = 0.4$.

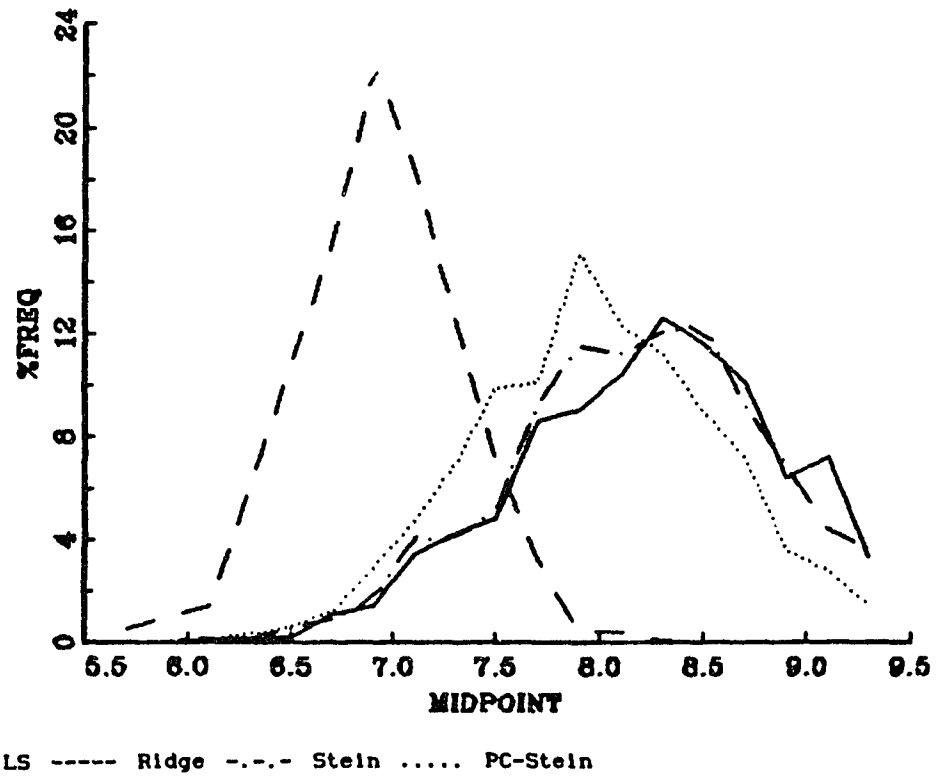


TABLE 7.14

The ln (s_1) forecasts obtained from the predictors when d_1= 0.5.

	LS	Ridge	Stein ($D=X_D'X_D$)	PC-Stein ($D=X_D'X_D$)
Average	8.432	7.072	8.320	8.095
Std.	0.699	0.413	0.666	0.623
Max	10.375	8.309	10.137	9.970
Min	6.138	5.742	6.083	6.046
Med.	8.444	7.078	8.326	8.083
99% confidence	[6.715, 10.039]	[5.844, 8.102]	[6.611, 9.843]	[6.467, 9.565]
95% confidence	[7.077, 9.825]	[6.279, 7.862]	[7.007, 9.554]	[6.912, 9.312]
90% confidence	[7.251, 9.579]	[6.395, 7.763]	[7.221, 9.426]	[7.072, 9.139]

FIGURE 7.10

The distribution of the forecast of ln(s_1) d_1= 0.5.

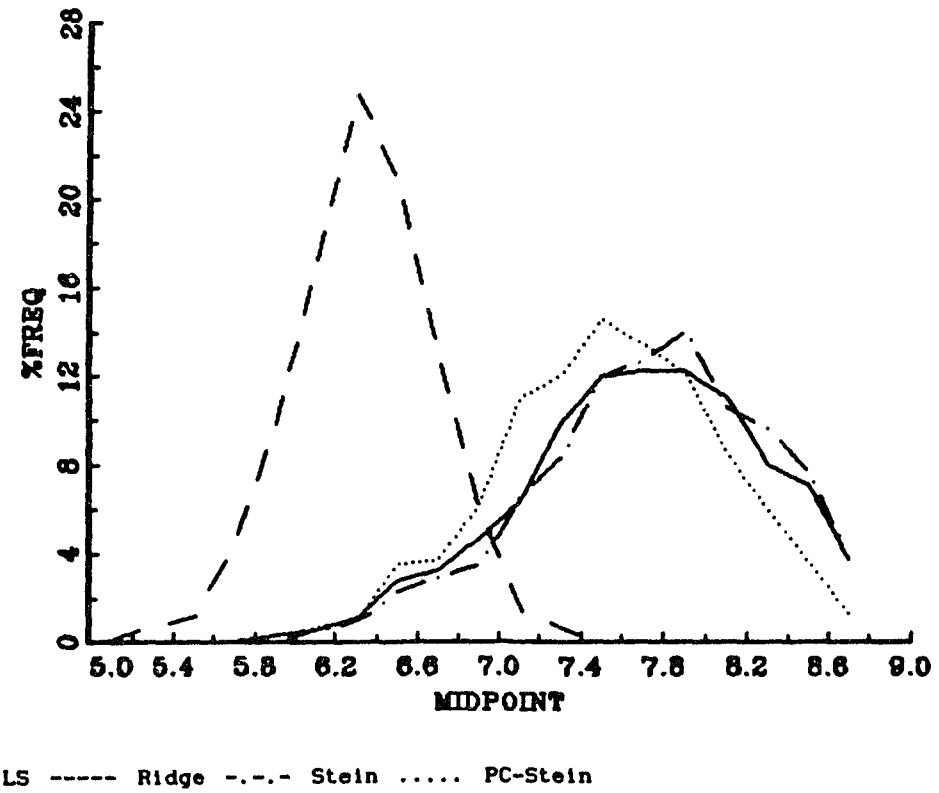
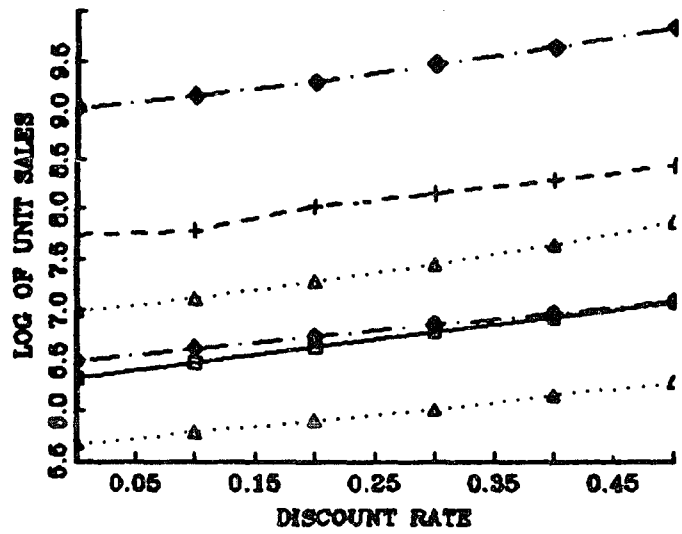


FIGURE 7.11

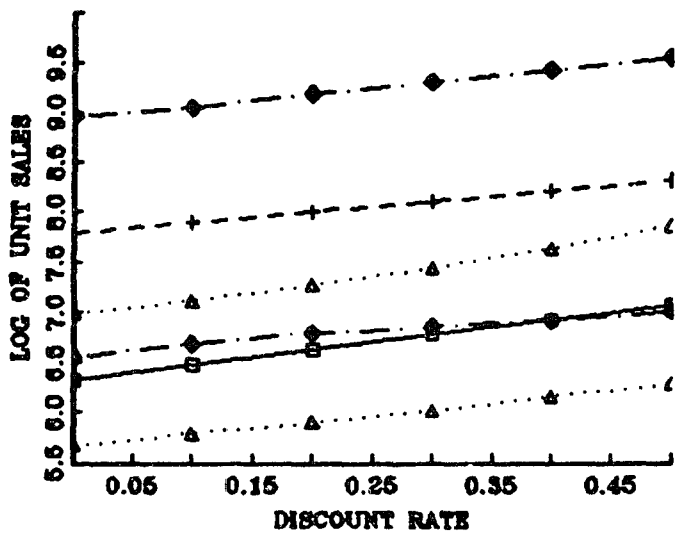
The 95% confidence intervals of the ridge regression and the least squares predictor



——— Ridge mean ..... Ridge boundaries
---- OLS mean .-.-.- OLS boundaries

FIGURE 7.12

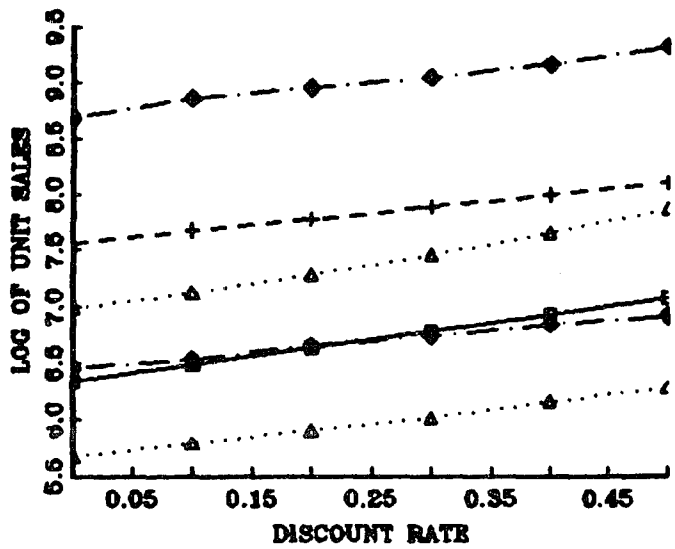
The 95% confidence intervals of the ridge regression and Stein ($D=X_D'X_D$) predictor.



——— Ridge mean .... Ridge boundaries
----- Stein mean .-.-. Stein boundaries

FIGURE 7.13

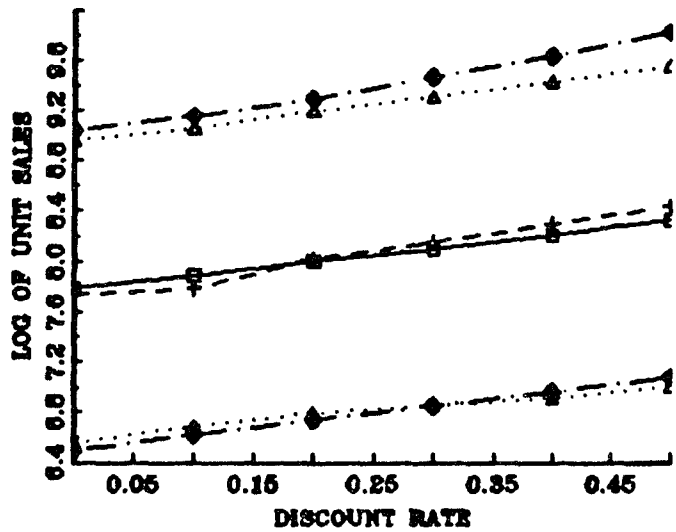
The 95% confidence intervals of the ridge regression and the PC-Stein ($D=X_D'X_D$) predictor.



——— Ridge mean ..... Ridge boundaries
---- PC-Stein mean .-.-.- PC-Stein boundaries

FIGURE 7.14

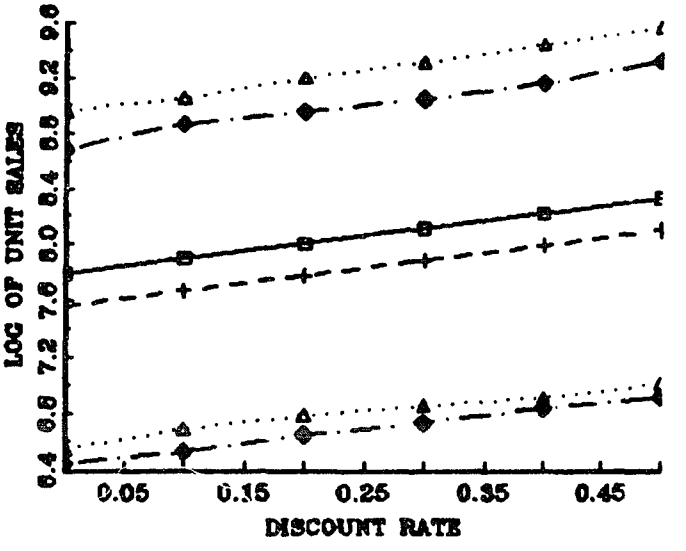
The 95% confidence intervals of the Stein ($D=X_D'X_D$) and the least squares predictor.



——— Stein mean .... Stein boundaries
----- OLS mean .-.-. OLS boundaries

FIGURE 7.15

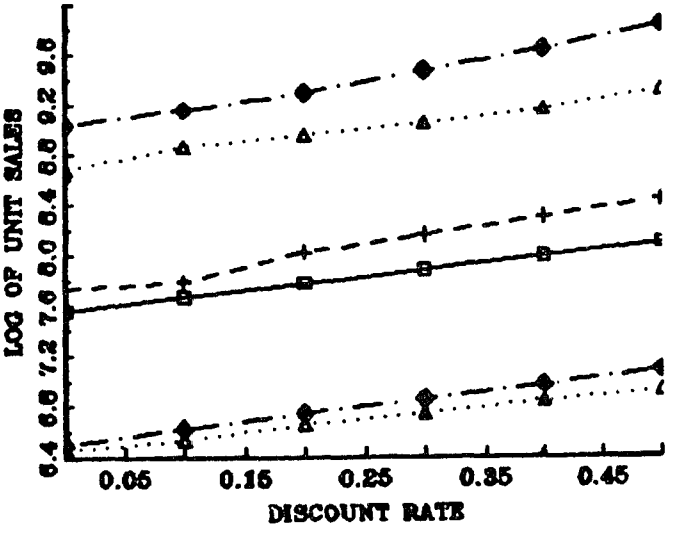
The 95% confidence intervals of the Stein ($D=X_D'X_D$) and the PC-Stein ($D=X_D'X_D$) predictor.



——— Stein mean ..... Stein boundaries
---- PC-Stein mean .-.-.- PC-Stein boundaries

FIGURE 7.16

The 95% confidence intervals of the PC-Stein ($D=X_D'X_D$) and the least squares predictor.



——— PC-Stein mean .... PC-Stein boundaries
----- OLS mean .-.-. OLS boundaries

The confidence intervals obtained through the Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) have smaller ranges to those of the least squares predictor. The PC-Stein ($D=X_D'X_D$) produces confidence intervals than those of the Stein ($D=X_D'X_D$) predictor.

7.4 CONCLUSIONS.

We have demonstrated that reasonable confidence intervals can be established via the bootstrap resampling technique. From our study, we make several deductions.

1. For the case of forecasting ln s_1 for week 49-52, the confidence intervals obtained from the ridge regression predictor have the smallest ranges as well as the least predicted mean square errors; nevertheless, not all of the confidence intervals contain the true forecast value. The PC-Stein ($D=X_D'X_D$) and Stein ($D=X_D'X_D$) predictor offer some reduction in the variability of the predicted values and prediction mean square errors over the least squares predictor.

2. The bootstrap confidence intervals for the biased predictors considered appear to be larger than the unknown true confidence intervals which can be a result of the overestimation of the forecast variability.

3. For the case of forecasting along the values of the discount rate, the confidence intervals obtained from the ridge regression predictor still yield the confidence

constructed from the Stein ($D=X_D'X_D$) and PC-Stein ($D=X_D'X_D$) predictor do not completely overlap those of the least squares predictor.

REFERENCES

Adkins, L.C. and R.C. Hill "The RLS Positive-part Stein Estimator," forthcoming American Journal of Agricultural Economics.

Amemiya, T. "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica, 46(1978): 1193-1205.

Armstrong, J.S., R.J. Brodie and S.H. McIntyre "Forecasting Methods for Marketing: Review of Empirical Research," International Journal of Forecasting, 3(1987):355-376.

Belsley, D.A., E. Kuh and R.E. Welsh Regression Diagnostics. (John Wiley & Sons, New York),1980.

Belsley, D.A. "Demeaning Conditioning Diagnostic through Centering," The American Statistician, 38(1984):73-77.

Bickel, P.J. and K.A. Doksum Mathematical Statistics. (Holden-Day, San Francisco),1977.

Bickel, P.J. and Freedman, D.A. "Some Asymptotic Theory for the Bootstrap," Annals of Statistics, 9(1981): 1196-1217.

Bigham, C. and K. Larntz "A Simulation Study of Alternatives to Ordinary Least Squares: Comment," Journal of the American Statistical Association, 72(1977):97-103.

Blattberg, R.C. and K.J. Wisniewski "Price-induced Patterns of Competition," Marketing Science, 8(1989):291-309.

Calzolari, G. and L. Panattoni "Alternative Estimator for FIML Covariance Matrix: A Monte Carlo Study (Notes and Comments)," Econometrica, 56(1988):701-714.

Casella, G. Minimax Ridge Estimator, unpublished Ph.D. dissertation,(Purdue University),1977.

Canavos, C.G. Applied Probability and Statistical Mothods, (Little, Brown and Company, Boston),1984.

Chow, C.G. Econometrics,(McGraw-Hill, New York),1983.

Copas, J.B. "Regression, Prediction and Shrinkage," Journal of the Royal Statistic Society, 45B(1983): 311-354.

Cramer, J.S. Econometric Application of Maximum Likelihood Methods,(Cambridge University Press,New York),1986.

Deegan, J. "Optimal k in Ordinary Ridge Regression," Report, (Dept. of Political Sciences, University of Rochester), 1975.

Demspeter, A.P., M. Schatzoff and N. Wermuth "A Simulation Study of Alternatives to Ordinary Least Squares," Journal of the American Statistical Association, 72(1977):79-91.

Dhrymes, P.J. Econometrics, Statistical Foundations and Applications,(Springer-Verlag,New York),1974.

Dhrymes, P.J. Mathematic for Econometrics, (Springer-Verlag,New York),1978.

Efron, B. "Bootstrap Methods: Another Look at Jackknife," Annals of Statistics, 7(1979):1-26.

Efron, B. "Nonparametric Estimates of Standard Errors: The Jackknife, the Bootstrap and other Methods," Biometrika, 68(1981):589-599.

Efron, B. The Jackknife, the Bootstrap, and Other Resampling Plans,(SIAM,Philadelphia),1982.

Efron, B. "Better Bootstrap Confidence Intervals," Journal of the American Statistical Association, 82(1987):171-185.

Farber, H.S. "The Evolution of Public Sector Bargaining Laws," in When Public Sector Workers Unionize, edited by R. Freeman and C Ichiniowski, (University of Chicago Press,Chicago),1988.

Farebrother, R.W. "Principal Component Estimators and Minimum Mean Square Error Criteria in Regression Analysis," Review of Economics and Statistics, 54(1972):332-336.

Farebrother, R.W. "Further Results on the Mean Square Error of Ridge Regression," Journal of Royal Statistic Society, 3B(1976):248-250.

Farrar, D.E. and R.R. Glauber "Multicollinearity in Regression Analysis: The Problem Revisited," Review of Econometrics and Statistics, 49(1967): 92-107.

Finney,D. Probit Analysis,(Cambridge University Press, New York),1984.

Fomby, T.B., R.C. Hill and J.R. Johnson Advanced Econometrics Methods,(Springer-Verlag,New York),1984.

Fomby, T.B. and R.C. Hill "Deletion Criteria for Principal Components Regression Analysis," American Journal of Agricultural Economics, 60(1987):524-527.

Freedman, D.A. "Bootstrapping Regression Models," Annals of Statistics,9(1981):1218-1228.

Freedman, D.A. "On Bootstrapping Two-stage Least Squares in Stationary Linear Model," Annals of Statistics, 12(1984):827-842.

Freedman, D.A. and S.C. Peters " Bootstrapping an Econometric Model: Some Empirical Results," Journal of Business & Economic Statistics,2(1984a): 150-158.

Freedman, D.A. and S.C. Peters "Bootstrapping a Regression Equation," Journal of the American Statistical Association,46(1984b):97-106.

Freedman, D.A. and S.C. Peters "Using the Bootstrap to Evaluate Forecasting Equations," Journal of Forecasting,4(1985):251-262.

Freeman, R.B. and R. Valletta, "The Effects of Pubic Sector Labor Laws on Labor Market Institutions and Outcomes," in When Public Sector Workers Unionize, edited by R.B. Freeman and C. Ichniowsky, (University of Chicago Press, Chicago)1988.

Friedman, D.J. and D.C. Montergomery "Evaluation of the Predictive Performances of Biased Regression Estimators," Journal of Forecasting,4(1985):153-163.

Goldberger, A.S. Econometric Theory,(John Wiley & Sons, New York),1964.

Gorman, J.W. and R.J. Toman "Selection of Varibles for Fitting Equations to Data," Technometrics,8(1966): 27-51.

Greenberg, E. "Minimum Variance Properties of Principal Components Regression," Journal of the American Statistical Association,70(1975):194-197.

Granger, C.W.J. and P. Newbold Forecasting Economic Time Series,(Academic Press,Florida),1986.

Guerard, J.B. and R.T. Clemen "Collinearity and the Use of Latent Root Regression for Combining GNP Forecasts," Journal of Forecasting,8(1989):231-238.

Gunst, R.F. and R.L. Mason "Biased Estimation in Regression: An Evaluation Using Mean Square Error," Journal of the American Statistical Association,72(1987):616-628.

Heckman, J. "Dummy Endogenous Variables in Simultaneous Equation System," Econometrica,46(1978):931-951.

Hemmerle, D.M. "An Explicit Solution for Generalize Ride Regression," Technometrics,17(1975):309-314.

Hogg, R.V. and Craig A.T. Introduction to Mathematical Statistics,(MacMillan, New York),1978.

Hill, R.C., R.F. Ziemer and F.C. White "Mitigating the Effects of Multicollinearity Using Exact and Stochastic Restrictions: The Case of an Aggregate Agricultural Function in Thailand: Comment," American Journal of Agricultural Economics, 63(1981):298-300.

Hill, R.C. and R.F. Ziemer "The Application of Generalized Ridge and Stein-like General Minimax Rules to Multicollinearity Data," Communications in Statistics, 11(1982):623-638.

Hill, R.C. "Modeling Multicollinearity and Extrapolation in Monte Carlo Experiments on Regression," Advanced in Econometrics,6(1987):127-155.

Hill, R.C. and G.G. Judge "Improved Prediction in the Presence of Multicollinearity," Journal of Econometrics,35(1987):83-100.

Hill, R.C., P.A. Cartwright and J.F. Arbaugh " Using Aggregate Data to Estimate Micro-level Parameteres with Shrinkage Rules," Working paper,1991.

Hoerl, A.E. and R.W. Kennard "Ridge Regression: Applications to Nonorthonormal Problems," Technometrics,12(1979a):55-67.

Hoerl, A.E. and R.W. Kennard "Ridge Regression: Application to Nonorthonormal Problems," Technometrics,12(1979b):69-82.

Hoerl, A.E. "A Simulation Study of Alternatives to Ordinary Least Squares: Comment," Journal of the American Statistical Association,72(1977):94-95.

Hoerl, A.E. and R.W. Kennard "Ridge Regression -1980 Advances, Algorithm, and Applications," American Journal of Mathematical and Management Science,1(1981),5-83.

Hoerl, A.E., R.W. Kennard and K.F. Baldwin "Ridge Regression: Some Simulation," Communications in Statistics,4A(1975):105-123.

Hoffman, K. and R. Kunze Linear Algebra, (Prentice Hall,New Jersey),1975.

Hocking, R.R., F.M. Speed and M.J. Lynn "A Class of Biased Estimators in Linear Regressions," Technometrics,18(1976):425-438.

Hunt, J.C. and R.A. White "The Legal Status of Collective Bargaining by Public School Teachers," Journal of Labor Research,4(1983):213-224.

Hunt,J.C., J.V. Terza, R.A. White and T.A. Moore "Wages, Union Membership, and Public Sector Bargaining Legislation: Simultaneous Equations with an Ordinal Qualitative Variable," Journal of Labor Research, 7(1986):255-267.

Johnson, S.R., S.C. Reimer and T.P. Rothrock "Principal Components and the Problem of Multicollinearity," Metroeconmica,25(1973):306-317.

Judge, G.G. and M.E. Bock The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics, (North-Holland,Amsterdam),1978.

Judge, G.G., R.C. Hill, W. Griffiths, H. Lütkepohl and T. Lee Introduction to the Theory and Practice of Econometrics,(John Wiley & Sons,New York),1985.

Judge, G.G. and M.E. Bock " Biased Estimation," Handbook of Econometrics, vol.1,edited by Z. Griliches and M.D. Intriligator, (North-Holland,Amsterdam),1983.

Judge, G.G., W. Griffiths, R.C. Hill, H. Lütkepohl and T. Lee The Theory and Practice of Econometrics, (John Wiley & Sons,New York),1985.

Judge, G.G. and T. Yancey Improved Method of Inference in Econometrics,(North-Holland,Amsterdam),1986.

Kennedy, P. A Guidge to Econometrics, (Basil Bleackwel,Oxford),1986.

Khuri, A. and C. Ofir " Multicollinearity in Marketing Models: Diagnostics and Remedial Measures," International Journal of Research in Marketing, 3(1986):181-205.

Krishnamurati, L., A. Rangaswamy and A. Zoltners " Equity Estimator: A New Method for Sales Response Estimation in the Presence of Multicollinearity," working paper, The Wharton School, University of Pennsylvania,1986.

Krishnamurati, L. and A. Rangaswamy " The Equity Estimator for Marketing Research," Marketing Science,6(1987): 336-357.

Kuther, M.H., J. Neter and W. Wasserman Applied Linear Regression Models,(Richard D. Irwin,New York),1989.

Lawless, J.F. and P. Wang " A Simulation Study of Ridge and other Regression Estimator," Communications in Statistics,A5(1976):307-323.

Lowerre, J.M. " On the Means Square Error of Parameter Estimates for Some Biased Estimator," Technometrics, 16(1974):461-464.

Lee, L.F. " Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," International Economic Review,19(1978): 415-433.

Marquardt, D.W. " Generalized Inverses, Ridge Regression, Biase Linear Estimator and Nonlinear Estimation," Technometrics,12(1970):591-612.

McKelvey, R. and W. Zavoina "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," Journal of Mathematical Sociology,4(1975):103-120.

Mittelhammer, R.C. and J.L. Baritelle " On Two Strategies for Choosing Principal Components in Regression Analysis," American Journal of Agricultural Economics, 59(1977):336-343.

Montgomery, D.C. and L.A. Johnson Forecasting and Time Series Analysis,(McGraw-Hill,New York),1976.

Montgomery, D.C. and E.A. Peek Introduction to Linear Regression Analysis,(Wiley,New York),1982.

Mundlak, Y. " On the Concept of Non-significant Function and its Implications for Regression Analysis," Journal of Econometrics,16(1981):139-149.

McKelvey, R. and W. Zavolina " A Statistical Model for Ordinal Level Dependent Variable," Journal of Mathematical Sociology,4(1975):103-120.

Pagan, A. " On the Role of Simulation in the Statistical Evaluation of Econometric Model," Journal of Econometrics,40(1989):125-139.

Pidot, G.,Jr. " A Principal Component Analysis to the Determinants of Local Government Fiscal Patterns," Review of Economics and Statistics,51(1969):176-188.

Rangaswamy, A. and L. Krishnamurati "Response Function Estimation Using the Equity Estimator," Journal of Marketing Research,28(1991):72-83.

Rao, C.R. Linear Statistical Inference and its Application, (Wiley,New York),1973.

Saltzman, G.M. "Bargaining Laws as A Cause and Consequence of the Growth of Teacher Unionism," Industrial & Labor Relations Review,38(1985):335-351.

Schmidt, P. " The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," Econometrica,42(1974):303-309.

Schmidt, P. " Some Small Sample Evidence on the Distribution of Dynamic Simulation Forecasts," Econometrica,45(1977):991-1005.

Schmidt, P. " Estimation of a Simultaneous Equation Model with Jointly Dependent Continuous and Qualitative Variables: The Union-Earnings Question Revisited," International Economic Review,19(1978):453-466.

Sidik, S.M. " Comparison of Some Biased Estimation Methods (Including Ordinary Least Squares) in the Linear Model," NASA Technical Note,1975.

Silvey, S.D. " Multicollinearity and Imprecise Estimation," Journal of the Royal Statistical Society,3(1969):539-552.

Strawderman, W.E. " Minimax Adaptive Generalized Ridge Regression Equation Estimation," Journal of the American Statistical Association,73(1978):623-627.

Tate, R.F. " The Theory of Correlation between Two Continuous Variables when One is Dichotomous," Biometrika,42(1955):205-216.

Teekens, R. amd P.M.C. De Boer " The Exact MSE-Efficiency of the General Ridge Estimator Relative to the OLS," Report, Eramus University,summer 1977.

Theobald, C.M. " Generalization of Mean Square Error Applied to Ridge Regression," Journal of the Royal Statistic Society,36B(1974):103-106.

Thisted, R.A. Ridge Regression, Minimax Estimation and Empirical Bayes Methods, unpublised Ph.D. dissertation, Standford University,1977.

Thisted, R.A. " On Generalized Ridge Regressions," Technical report, Department of Statistics, University of Chicago,1978a.

Thisted, R.A. " Multicollinearity, Information and Ridge Regression," Technical report, Department of Statistics,University of Chicago,1978b.

Ullah, A. and H.D. Vinod " A Familt of Improved Ridge Estimator," Report, Econometric Institute,1978.

Ullah, A. and H.D. Vinod Recent Advances in Regression Methods,(Mercel Dekker,New York),1981.

Waters, M.S. Three Essays on the Determinants and Effects of Public Sector Bargaining Law, Unpubished Ph.D. dissertation, Department of Economics, Louisiana State University,1989.

Watts, D.G. and A.R. Willan " Meaningful Multicollinearity Measures," Technometrics,20(1978):407-412.

Zellner, A. Basic Issues in Econometrics, (University of Chicago Press,Chicago),1984.

VITA

Parisun Chantanahom was born on June 8, 1963 to Namsap and Prempree Chantanahom in Bangkok, Thailand. He was the only child of the family.

His educations include a Bachelor of Art degree in Economics from Thammasat University, a Master of Science in Economics from Louisiana State University and also complete course works in Economics from Illinois State University.

His areas of interest are Econometrics and Econometrics applications to monetary economics.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Parisun Chantanahom

Major Field: Economics

Title of Dissertation: Three Essays on Sampling Techniques: Small Sample Performances of Estimators and Predictors

Approved:

R. Carter Hill
Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

4/12/91