TESTING OF AND ESTIMATION SUBJECT TO INEQUALITY RESTRICTIONS USING A MINIMUM DISTANCE ESTIMATOR

A Dissertation
for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Lawrence Capron Marsh
1976

This is to certify that the

thesis entitled

TESTING OF AND ESTIMATION SUBJECT TO
INEQUALITY RESTRICTIONS USING A
MINIMUM DISTANCE ESTIMATOR

presented by

Lawrence Capron Marsh

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Major professor
James B. Ramsey

Date 1-12-76

ABSTRACT

TESTING OF AND ESTIMATION SUBJECT TO
INEQUALITY RESTRICTIONS USING A
MINIMUM DISTANCE ESTIMATOR

By

Lawrence Capron Marsh

Economic theory and empirical evidence from previous research may suggest that certain inequality restrictions may be appropriate for particular parameters in an econometric model. This dissertation develops a test for the appropriateness of restrictions that constrain parameters to finite intervals. In addition, it considers the use of a minimum distance estimator that incorporates inequality restrictions into the estimation procedure itself.

By definition the minimum distance estimator selects that point in the constrained space which is nearest to the unconstrained maximum likelihood estimate. The distributional properties of the minimum distance estimator are compared to those of the unconstrained maximum likelihood estimator and the constrained maximum likelihood estimator in extensive Monte Carlo sampling experiments. It is shown that the minimum distance estimator is greatly superior to the unconstrained maximum likelihood estimator in terms of various definitions of mean squared error.

The equivalence of the constrained maximum likelihood estimator and the minimum distance estimator in a number of circumstances is demonstrated by noting that in general there is not a substantial difference between the estimated

distributional properties of the minimum distance estimator and those of the constrained maximum likelihood estimator.

A single sample approximation of the variance, absolute bias, and mean squared error of the minimum distance estimator is developed and many examples of applying the minimum distance estimator in regression analysis are presented indicating the effect of increasingly stringent inequality restrictions as imposed by the minimum distance estimator. Thus, in general, the usefulness and effectiveness of the minimum distance estimator are demonstrated.

TESTING OF AND ESTIMATION SUBJECT TO INEQUALITY RESTRICTIONS USING A MINIMUM DISTANCE ESTIMATOR

By

Lawrence Capron Marsh

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1976

ACKNOWLEDGMENTS

Professor James B. Ramsey provided the inspiration and guidance for this dissertation. Although the computer programming and testing procedures are my own work, the basic idea of the minimum distance estimator as defined herein and the alternative conceptions of mean squared error in the multivariate case belong to Dr. Ramsey. Professor Ramsey sacrificed a great deal of his own research time to read and reread the many variations of this manuscript including several parts that are not included in this final version. Professor Ramsey was exceptionally prompt in reviewing and returning chapters and providing detailed suggestions for revising both content and presentation.

Professor Robert H. Rasche, director of graduate studies, and professors William J. Haley, Robert L. Gustafson and Lester V. Manderscheid assisted me in the course of writing this dissertation and in learning the fundamentals of econometrics. Great appreciation in this regard is also reserved for Professor Jan Kmenta whose teaching served as the foundation upon which this dissertation is built. The advice of Dr. C. Shapiro in regard to hypothesis testing was also appreciated.

My greatest debt by far is to Janet K. Marsh for the countably infinite hours she spent typing and retyping the rough drafts and proofreading everything from computer output to the final document.

I retain credit for myself for any errors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Statement of the Problem

In order to understand fully how to utilize the behavioral and predictive capabilities of an econometric model, all available information, both *a priori* and empirical, must be used effectively in the estimation process. Such information could be in the form of inequality restrictions. Failure to use all available information results in models that tend to be less reliable and tend to have larger variances than necessary and are therefore inefficient. On the other hand, imposing restrictions that are invalid leads to estimators which may not only be biased but may be inconsistent as well. The need for testing for the appropriateness of restrictions is clear, especially in the case of inequality restrictions since they have been ignored most often both theoretically and practically. In particular, a test procedure for restrictions which constrains regression coefficients to finite intervals needs to be developed.

Research economists occasionally wish to restrict the values of one or more regression coefficients by inequality constraints. A coefficient might be desired which is positive, negative, or restricted to lie within a particular range such as zero to one. For example, the slope of a supply curve could be required to be positive while that of a demand curve could be required to be negative. A coefficient representing the marginal propensity to consume could be required to lie between zero and one.

A common ad hoc approach to this problem is first to run an unrestricted regression. If all the coefficients of interest have the right signs or lie within the right range, then the initial regression results are accepted. However, if any of the coefficients have the wrong sign or lie outside the restricted range, then a new combination of explanatory variables is tried or the form of the regression equation is altered. This procedure is continued until the desired results are obtained.

The above ad hoc procedure does not take into account the effect of such a procedure on the properties of the resulting estimators themselves. A more straightforward approach is to impose the desired restrictions on the estimation procedure itself. Such a one-step procedure simplifies the derivation of the properties of the resulting estimators.

Inequality restrictions may be incorporated into the estimation procedure by use of a minimum distance estimator such as that proposed by Ramsey (30, p.8). Ramsey's minimum distance estimator, $\hat{\underline{\theta}}_m$, selects that point in the constrained set which is nearest to the unconstrained maximum likelihood estimator, $\hat{\underline{\theta}}_u$. Ramsey defines his minimum distance estimator, $\hat{\underline{\theta}}_m$, by the expression

$$||\hat{\underline{\theta}}_u - \hat{\underline{\theta}}_m||^2 = \min_{\underline{\theta} \in C} ||\hat{\underline{\theta}}_u - \underline{\theta}||^2$$

which indicates that $\hat{\underline{\theta}}_m$ is defined as the point in the constrained space, C, which gives the minimum value for the

square of the Euclidean norm of the vector representing the difference between the unconstrained point, $\hat{\underline{\theta}}_u$, and every possible point, $\underline{\theta} \in C$, in the constrained space.

Obviously, by definition, the minimum distance estimator (MDE) will be at least as close to the unconstrained maximum likelihood estimator (UMLE) as the constrained maximum likelihood estimator (CMLE) will be. In some cases it may be possible for the MDE to be closer to the UMLE than the CMLE is. As will be demonstrated in a later chapter, this will often be the case for multivariate situations.

1.2 Outline of the Thesis

The next section of this chapter will briefly review some of the literature dealing with inequality restrictions which is appropriate for regression analysis. Chapter II will develop and explain a test procedure for deciding if inequality restrictions are appropriate for a particular situation where such restrictions limit the parameter of interest to a finite interval. Chapter III will discuss the minimum distance estimator as a particular method of dealing with inequality restrictions with reference to some of the work of Ramsey and Penneck. Chapter IV will determine estimates of small sample properties of the minimum distance estimator as compared to the constrained and unconstrained maximum likelihood estimators by means of Monte Carlo sampling experiments and present a method of obtaining an estimate of the variance of the minimum distance estimator given a particular sample.

Finally, Chapter V will provide some examples of using the minimum distance estimator and will demonstrate under what circumstances and to what extent in these examples use of the minimum distance estimator can result in a reduction in the estimated variance of the regression coefficients and how this reduction in estimated variance is not substantially offset by an increase in estimated bias in the calculation of estimated mean squared error.

1.3 Review of the Literature

Methods of testing for and incorporating exact linear restrictions in regression analysis have been well developed by Chow (6, p.591), Chipman and Rao (5, p.198), Toro and Wallace (37, p.558), Fisher (10, p.361), Wallace and Anderson (39, p.1), Wallace (38, p.689), and others. However, these techniques cannot be directly applied in the case of inequality restrictions. Dealing with inequality restrictions has proven to be a more difficult problem. Consequently, the literature on inequality restrictions has been much more limited and less productive.

Theil and Goldberger (35, p.65) proposed a method for approximating inequalities on coefficients so as to incorporate indirectly inequality restrictions into the estimation procedure. They combine a priori and sample information in the equation:

$$\begin{bmatrix} \underline{y} \\ \underline{r} \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \underline{b} + \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix}$$

where the a priori information is given by the expression $\underline{r} = R\underline{b} + \underline{v}$. $\underline{b}$ is the vector of coefficients; R is a matrix of weights; $\underline{r}$ is the resulting vector of values set by the restrictions except for a disturbance term vector, $\underline{v}$, which is assumed to have a zero mean and a nonsingular variance-covariance matrix. This a priori information tends to restrict the posterior coefficients to be within desired ranges where the a priori coefficients, $\underline{b}$, represent the midpoints of the desired ranges and the variance-covariance matrix of $\underline{v}$ is used to restrict the end points of the ranges to be a set number of standard deviations away from the midpoint so that the probability of a sample observation falling outside of a desired range is very small. The sample information is represented by the usual regression equation $\underline{y} = X\underline{b} + \underline{u}$, where $\underline{y}$ is a vector of observations on the dependent variable, X is a matrix of observations on the explanatory variables, and $\underline{u}$ is a disturbance term vector with zero mean and nonsingular variance-covariance matrix. The restricted parameter estimates can be represented by

$$\hat{\underline{b}} = (X^{*\prime}U^{*-1}X^{*})^{-1}X^{*\prime}U^{*-1}\underline{y}^{*}$$

where $\underline{y}^{*} = \begin{bmatrix} \underline{y} \\ \underline{r} \end{bmatrix}$, $X^{*} = \begin{bmatrix} X \\ R \end{bmatrix}$, and $U^{*} = \begin{bmatrix} \underline{uu}' & \underline{uv}' \\ \underline{vu}' & \underline{vv}' \end{bmatrix}$

However, their Bayesian method of mixing a priori and sample information does not guarantee that the inequality restrictions will hold. By adding a point estimate from near

the midpoint of the restricted range to the original sample data before estimating the coefficients and by setting a small standard deviation for the selected point estimate, a coefficient can be derived which has a high probability of being within the restricted range. Thus, it becomes highly likely using this method that an income elasticity will lie between zero and one. However, it cannot guarantee an estimate between zero and one. The smaller the standard deviation, the more concentrated the probability becomes around one point. In reducing the probability of getting an estimate outside the restricted range, the probability of getting an estimate near either of the end points of the range is substantially reduced. Furthermore, the larger the sample size, the more weight is given to the sample observations and the harder it is to guarantee an estimate within the restricted range. Also, the method is only relevant for interval inequality restrictions and is generally not very useful in cases of single one-sided inequality restrictions unless they are arbitrarily truncated at some point. Consequently, their approach leaves the way open for a technique that would constrain a coefficient by the desired inequality restrictions and at the same time would not result in excessive concentration of probability about a single point.

Judge and Takayama (19, p.166) combine prior and sample information in a regression model such that inequality restraints are placed on individual coefficients or combinations of coefficients by minimizing the quadratic form given

by the sum of squared residuals subject to the inequality constraints. The analysis is based on the standard regression model

$$\underline{y} = X\underline{b} + \underline{u}, \qquad E\underline{u} = 0, \qquad E\underline{u}\underline{u}' = \sigma^2 I$$

where $\underline{y}$ is a vector of observations on the dependent variable, X is the regressor matrix, $\underline{b}$ is the vector of regression coefficients, and $\underline{u}$ is the disturbance vector which has zero mean and a diagonal covariance matrix equal to the constant variance σ^2 times the identity matrix I. The estimator $\underline{b}^{**}$ is defined by that value of $\underline{b}$ which minimizes $\underline{u}'\underline{u} = (\underline{y}-X\underline{b})'(\underline{y}-X\underline{b})$ subject to inequality restrictions such as

1) $0 \leq \underline{r}_1^l \leq \underline{b}_1 \leq \underline{r}_1^u$

2) $\underline{r}_2^l \leq \underline{b}_2 \leq \underline{r}_2^u \leq 0$

3) $\underline{r}_3^l \leq \underline{b}_3 \leq \underline{r}_3^u$ and $\underline{r}_3 \leq 0 \leq \underline{r}_3^u$

4) $\underline{r}_4^l \leq \underline{s}\underline{b}_4 \leq \underline{r}_4^u$

where $\underline{r}_i^u$ and $\underline{r}_i^l$ for i = 1, . . . 4, are known vectors of upper and lower bound constraints for the unknown coefficients in the $i^{\underline{th}}$ set $\underline{b}_i$ and $\underline{s}$ is a row vector of weights with one s_j for each b_j.

This non-linear programming problem is solved by use of

the Kuhn-Tucker equivalence theorem. Their discussion of the sampling properties of the restricted estimators refers to Zellner (41, p.1). Zellner assumed a multivariate normally distributed disturbance term with zero means and known homogeneous variances. The inequality restricted estimators were indicated to be distributed as truncated normal. Properties such as bias, variance, and mean squared error can be determined for the single explanatory variable situation, but become quite difficult to evaluate for models with two or more explanatory variables. The latter situation must be dealt with by numerical integration procedures or simulation techniques. Zellner's work suggests that the mean squared error of the inequality restricted estimator is smaller than the mean squared error of the corresponding unrestricted estimator. However, the work in this area is incomplete since small sample properties of the inequality restricted estimator have not been comprehensively evaluated for a multiparameter situation. Moreover, neither Zellner nor Judge and Takayama have dealt with the problem of hypothesis testing.

Lovell and Prescott (25, p.913) presented an analysis of a related hypothesis testing problem with inequality constraints for a two-step procedure which is quite similar to a special case of the minimum distance estimator. Their two-step procedure involves imposing a single one-sided inequality constraint on one coefficient in a regression equation. In the example, the coefficient of the first explanatory variable is restricted to be non-negative. If in the

initial regression that coefficient is non-negative, then the initial results are accepted. Otherwise, that coefficient is set equal to zero and the regression is rerun without the first explanatory variable. Minimum distance estimation would also set the first coefficient equal to zero in this event, but would accept the other coefficients as initially estimated without rerunning the regression. Lovell and Prescott indicate that the two-step procedure results in parameter estimates that are biased and inefficient. The main thrust of the article, however, is to show that the final student-t statistics are upward biased in absolute value and in effect exaggerate the significance levels of the corresponding t-tests. This result follows from the work of Bancroft (1, p.190) and others. Lovell and Prescott do not, however, deal with the problem of testing for the inequality constraint in the beginning to decide if it is a valid restriction. Nor is the problem of testing within the restriction dealt with because the concern is with testing the unrestricted coefficients after the first coefficient has been restricted (i.e., dropped from the equation in this case) and the regression has been rerun on the remaining variables. However, Lovell and Prescott demonstrate that the two-step procedure results in parameter estimates with smaller mean squared error than the initial parameter estimates if the disturbance term is normally distributed. It is noted that both the exaggeration of significance levels and the reduction in mean squared error are directly related to the

absolute value of the degree of correlation between the restricted parameter variable and the other explanatory variables in the model. Thus, the Lovell and Prescott analysis is most appropriate for highly ill-conditioned data as is often found in economic time series.

Since Zellner's (41, p.1) analysis was performed for only one explanatory variable and one inequality constraint under the classical normal linear regression model assumptions, Hussain (17, p.1) extended his analysis to simultaneous equation models and in particular to two-stage least squares constrained and unconstrained estimators by means of Monte Carlo simulated sampling. Hussain used mean absolute error and root mean squared error to compare the estimators. His results indicated that constrained two-stage least squares was superior to unrestricted two-stage least squares by these criteria. Where appropriate, two-stage least squares was preferable to constrained ordinary least squares, and constrained ordinary least squares was preferable to unrestricted ordinary least squares.

A more general approach which deals with finitely bounded compact sets of which an inequality constrained parameter space may be viewed as a special case has been developed by Ramsey (30, p.1) who suggested his minimum distance estimator as an alternative to the maximum likelihood approach. Ramsey has shown the consistency of the minimum distance estimator and its superiority over the unconstrained maximum likelihood estimator in terms of mean squared error

analytically and in limited sampling experiments under various circumstances. In particular, he has shown that the MDE has a mean squared error smaller than, or at most equal to, the mean squared error of the UMLE in the single parameter case when the constrained parameter space is either a convex set or consists of a finite number of points. Also in the single parameter case Ramsey has shown that the CMLE is equivalent to the MDE when the rate of change of the logarithm of the likelihood function with respect to the parameter, θ, is a nowhere increasing and somewhere decreasing function of θ. This result implies that $\hat{\theta}_c = \hat{\theta}_m$ and that the mean squared errors of the CMLE and the MDE are equal and less than or equal to that of the UMLE:

$$\text{MSE } \hat{\theta}_c = \text{MSE } \hat{\theta}_m \leq \text{MSE } \hat{\theta}_u.$$

Under more general conditions specified by Ramsey for the single parameter case, he has shown that the above results hold at least asymptotically. In the multiparameter case Ramsey found that the MDE has a mean squared error at least as small as that of the UMLE in the sense that

$$E(\underline{\hat{\theta}}_m - \underline{\theta}_o)'(\underline{\hat{\theta}}_m - \underline{\theta}_o) \leq E(\underline{\hat{\theta}}_u - \underline{\theta}_o)'(\underline{\hat{\theta}}_u - \underline{\theta}_o)$$

under the assumption that the constrained parameter space consists of a finite number of points or contains a non-countably infinite number of points and is a convex set.

In addition to offering a substantial reduction in mean squared error, the MDE is particularly useful because it is considerably easier to calculate than the CMLE for regression coefficients restricted by inequality constraints. Since the distribution of the MDE can easily be derived from the distribution of the UMLE, the analytical form of the finite sample distribution of the MDE is at least known whenever the probability density function of the UMLE is known. Furthermore, although the CMLE has the same asymptotic distribution as the MDE, the derivation of the CMLE from the UMLE, as well as the proof of CMLE induced reduction in mean squared error over the UMLE, requires a number of conditions that may be more or less demanding and difficult to satisfy as indicated by Ramsey (31, p.8). Consequently, these advantages of the MDE suggest that it may be preferred to the CMLE.

The literature as reviewed in this chapter has led to the conclusion that a more extensive examination of the MDE and its distributional properties as compared with the CMLE and the UMLE would be useful and interesting for the purpose of deciding how best to incorporate inequality restrictions into estimation procedures.

However, before proceeding with a more detailed analysis of the minimum distance estimator, testing procedures need to be developed to determine if inequality restrictions are appropriate or not in a given situation in the first place. Such procedures will be developed in Chapter II.

CHAPTER II

TESTING FOR INEQUALITY RESTRICTIONS

Although test procedures for equality restrictions are well-known and extensively used, such procedures for inequality restrictions that limit a parameter to a finite interval have not been adequately developed, in particular where the variance is unknown (20, p.213) except in very general abstract terms for the exponential family (24, p.315). This chapter will develop the needed test procedures related to the normal distribution before proceeding to an analysis of the incorporation of such inequality restrictions into the estimation process by means of the MDE in Chapter III. In particular, this chapter will deal with a composite null hypothesis where the mean, θ, of a normally distributed random variable, X, is restricted to values in a closed interval. At first it will be assumed that the variance of X, σ^2, is a known constant. Later this assumption will be changed in order to deal with the case where the variance is unknown but takes on the same constant value under both the null and alternative hypotheses. This analysis will then be applied to a regression problem under the classical normal linear regression model assumptions.

The null hypothesis consists of an interval in that any point in the interval is an acceptable value for the true population mean under the null hypothesis. Chernoff and Moses (4, p.257) have referred to such an interval in hypothesis testing as an indifference zone. The null hypothesis

may be specified as:

$$H_o: \quad \theta_1 \leq \theta \leq \theta_2 \tag{1}$$

The corresponding alternative hypothesis is:

$$H_a: \quad \theta < \theta_1 \quad \text{or} \quad \theta > \theta_2 \tag{2}$$

as shown in Figure 1.

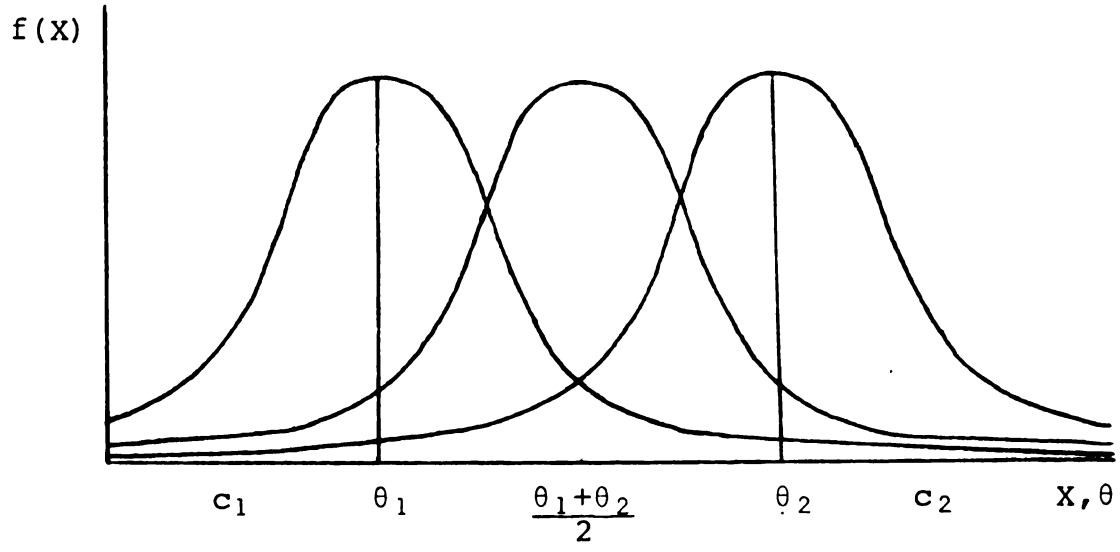


Figure 1

Normal Distribution with Mean
Restricted to Finite Interval

Note: Graph of normal distribution with mean between or at points θ_1 and θ_2. H_0: $\theta_1 < \theta < \theta_2$. H_a: $\theta < \theta_1$ or $\theta > \theta_2$.

θ: mean of normally distributed random variable, X.
f(X): probability density function of X.
c_1: critical point for left-hand critical region.
c_2: critical point for right-hand critical region.

Before proceeding, it is desirable to establish agreement on the following selected definitions. A critical region, C, will be defined as a subset of the sample space, S, such that if an observed value of a statistic, t, falls in the critical region C the null hypothesis will be rejected. If the observed statistic does not fall in the critical region, the alternative hypothesis will be rejected in favor of the null hypothesis.

A statistic is a real-valued function of the observed sample values alone and is not dependent on any unknown parameter values. In this initial case, the mean of the observed sample values will serve as the statistic.

The power of a test is the probability for a given parameter value that the sample point will fall in the critical region of the test, i.e., $\Pr \{t \in C|\theta\}$. The power function expresses power as a function of the admissible parameter values, which have been defined by Cramer (7, p.528) as all parametric points which are regarded as a priori possible under an admissible hypothesis. The parameter space is denoted by Ω where $\theta \in \Omega$, while the acceptance region is denoted Ω_o and the rejection region is denoted Ω_1 such that $\Omega = \Omega_o \cup \Omega_1$.

The significance level of a test is the supremum of the power function of the test when the null hypothesis is true, i.e., $\alpha = \sup_{\theta \in \Omega_o} \Pr \{t \in C|\theta\}$ (see Hogg and Craig (16, p.269) or Rao (33, p.375).) As Rao indicates, although under a simple null hypothesis the power function yields a single value for

the significance level, under a composite null hypothesis the power function yields a set of values and the significance level is taken to be the supremum of that set. In the case of an interval null hypothesis with a single completely specified nuisance parameter, the supremum of the power function under the null hypothesis is the maximum value of the power function under the null hypothesis. A typical power curve drawn for values of the power function under the null hypothesis is shown in Figure 2

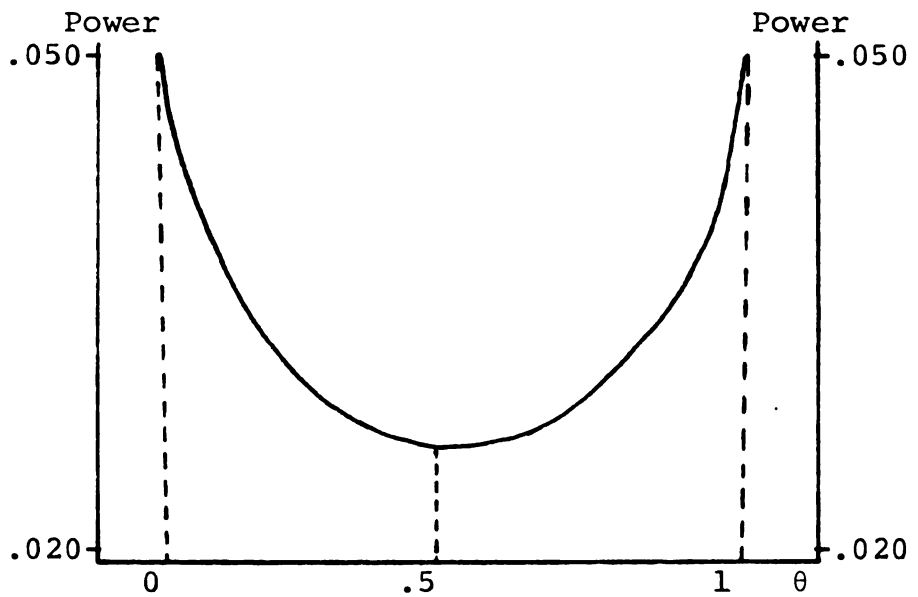


Figure 2

Power Curve Under Null Hypothesis
H_o: $0 \leq \theta \leq 1$ $\alpha = .05$

Ferguson (9, p.224) defines an unbiased test as a test which has a power function which takes on values less than or equal to the significance level under the null hypothesis and

values greater than or equal to the significance level under the alternative hypothesis. Given a particular level of significance, α, a uniformly most powerful unbiased test of size α is defined by Ferguson to be a test of size α which has power equal to or greater than the power of any other unbiased test of size α for all admissible parameter values under the alternative hypothesis. In other words, it is an unbiased test which has maximum power under the alternative hypothesis.

The following testing procedure may be used for testing the interval hypothesis (1) against the alternative hypothesis (2) when the parameter of interest is the mean of a normally distributed random variable with a known constant variance. The proposed test statistic is the mean of the observed sample values, which is distributed as $N(\theta, \sigma^2/n)$, where n is the sample size. Select two critical points c_1 and c_2, where $c_1 < \theta_1$ and $c_2 > \theta_2$, such that the significance level is equal to some desired value, α, that is $\alpha = \sup \Pr \{\bar{x} \in C\}$, where $C = \{(-\infty, c_1), (c_2, \infty)\}$. For reasons shortly to become apparent, the critical points must be located symmetrically relative to the interval hypothesized in (1) in order to have an unbiased test. This symmetry may be expressed by the condition:

$$\theta_1 - c_1 = c_2 - \theta_2 \tag{3}$$

This condition may be rewritten as:

$$(c_1 + c_2) = (\theta_1 + \theta_2) \tag{4}$$

Equivalently, c_1 and c_2 may be thought of as symmetrical about the midpoint of the interval, $\frac{\theta_1+\theta_2}{2}$. The power function of this test in terms of standardized units is:

$$\text{Power} = \Phi\left(\frac{c_1-\theta}{\sigma/\sqrt{n}}\right) + \left(1 - \Phi\left(\frac{c_2-\theta}{\sigma/\sqrt{n}}\right)\right) \qquad (5)$$

where $\Phi(\cdot)$ refers to the standard normal cumulative distribution function, and n is the number of observed sample values. If $\bar{x}$ is greater than c_2 or less than c_1, reject the null hypothesis (1). If $\bar{x}$ is contained in the interval $[c_1, c_2]$, reject the alternative hypothesis (2) in favor of the null hypothesis (1).

It is important to keep in mind that although all of the points in the indifference zone are acceptable under the null hypothesis, only one point is actually the true value of the population mean θ_o. For example, if the parameter of interest is the mean of the distribution generating the observations, only one point in the indifference zone is actually the true mean. However, the interval hypothesis may be accepted regardless of which particular point in the indifference zone is the mean of the generating distribution.

For example, suppose that $\theta_1 = 0$, $\theta_2 = 1$, $\sigma^2 = 10$, $n = 10$ and the desired significance level is .05. The appropriate critical points are $c_1 = -1.68$ and $c_2 = 2.68$. The power function for this example is:

$$\text{Power} = \Phi(-1.68 - \theta) + (1 - \Phi(2.68 - \theta)) \qquad (6)$$

the po

I

espon

eterm

ypoth

est is

he nu

ower f

he pow

Fi

ireter

$_0$ = 0,

unctio

as bee

The power function has been graphed in Figure 3.

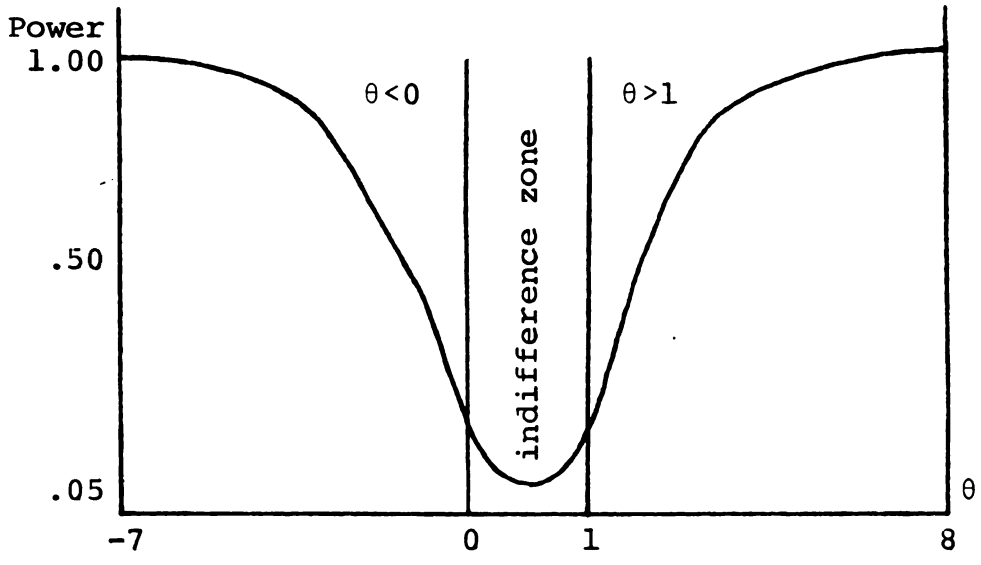
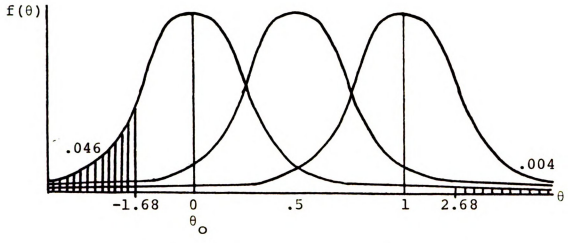


Figure 3

Power Function

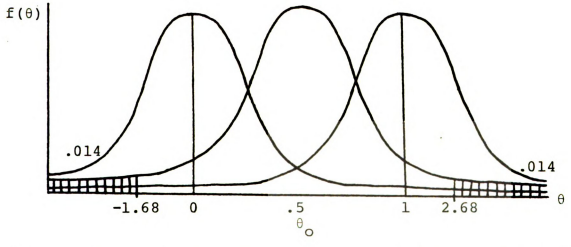
In order to illustrate that these critical points correspond to a significance level of .05, it is necessary to determine the supremum of the power function under the null hypothesis. As already noted, the significance level of a test is equal to the supremum of the power function under the null hypothesis, and in this case the supremum of the power function under the null hypothesis is the maximum of the power function under the null hypothesis.

Figure 4 shows what happens if the true value of the parameter of interest is assumed to be at the lower end, i.e., $\theta_o = 0$, of the hypothesized interval. The value of the power function at this end point is .046 + .004 = .05. This value has been plotted in Figure 2 which showed the values of the



$P(\theta \leq -1.68 \mid \theta_o=0) = .046$ $P(\theta \geq 2.68 \mid \theta_o=0) = .004$

Figure 4 - Population Mean at 0[1]



$P(\theta \leq -1.68 \mid \theta_o=.5) = .014$ $P(\theta \geq 2.68 \mid \theta_o=.5) = .014$

Figure 5 - Population Mean at .5[1]

---

1. Graph of normal distribution with mean between points 0 and 1 inclusive. H_o: $0 \leq \theta \leq 1$. H_a: $\theta < 0$ or $\theta > 1$. $c_1 = -1.68$ and $c_2 = 2.68$

power

the s

value

In th

.014

out t

Figur

value

(

the m

power

and po

than a

the al

define

a

the sy

Interv

If the

the in

favore

differ

site s

the po

Attach

Note t

for sy

power function under the null hypothesis. Figure 5 indicates the situation under the assumption that the true parameter value lies at the midpoint of the range, i.e., $\theta_o = .05$. In this case, the value of the power function is .014 + .014 = .028. As seen in Figure 2, this point turns out to be the minimum point of the power function. Reversed, Figure 4 would show the end point solution $\theta_o = 1$ yielding a value for the power function of .004 + .046 = .05.

Consequently, the supremum of the power function under the null hypothesis, which in this case is the maximum of the power function under the null hypothesis, occurs at either end point, 0 or 1, and is equal to .05. Since .05 is smaller than any of the values that the power function takes on under the alternative hypothesis, this test is an unbiased test as defined by Ferguson (9, p.224).

By examining Figure 3 it becomes intuitively clear that the symmetry of the critical points about the hypothesized interval is essential in order for the test to be unbiased. If the critical points were not placed symmetrically about the interval and, therefore, one side of the interval was favored over the other, the power curve would enter the indifference zone in Figure 3 at a lower power value on one side than on the other. Such a situation would result in the power function having at least one point under the alternative hypothesis with a smaller value than at least one point under the null hypothesis. Consequently, by definition such a test would be biased.

In addition to designating a test statistic, a testing procedure must indicate how the appropriate critical points are derived given a particular desired significance level. The four equations numbered (8) through (11) which will be discussed below can be solved simultaneously to obtain the required critical points, c_1 and c_2, given a desired significance level, α.

The likelihood function for the set of n observations, $x_1, \ldots, x_n$, given the mean, θ, of a normal distribution with variance, σ^2, is:

$$L(\underline{x}|\theta) = (2\pi\sigma^2)^{\frac{-n}{2}} \exp \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\theta)^2 \qquad (7)$$

Define α_1 as the probability of being in the left-hand critical region of the test under the normal distribution with mean, θ_1. Since the test is symmetrical and the normal distribution is also symmetrical, this is equivalent to the probability of being in the right-hand critical region when the true mean is θ_2.

$$\alpha_1 = \int_{-\infty}^{c_1} L(\underline{x}|\theta_1)\,d\underline{x} \quad \text{or} \quad \int_{c_2}^{\infty} L(\underline{x}|\theta_2)\,d\underline{x} \qquad (8)$$

Next define α_2 as the probability of being in the right-hand critical region of the test under the normal distribution with mean θ_1. Again, since the test is symmetrical, this is equivalent to the probability of being in the left-hand critical region when the true mean is θ_2.

$$\alpha_2 = \int_{c_2}^{\infty} L(\underline{x}|\theta_1)d\underline{x} \quad \text{or} \quad \int_{-\infty}^{c_1} L(\underline{x}|\theta_2)d\underline{x} \tag{9}$$

Recall the symmetry condition:

$$c_1 + c_2 = \theta_1 + \theta_2 \tag{10}$$

Since under the null hypothesis the power of this symmetrical test is maximized at θ_1, or, equivalently, θ_2, the significance level, α, of this test may be expressed as:

$$\alpha = \alpha_1 + \alpha_2 \tag{11}$$

Given a particular desired significance level α, the above four equations (8) to (11) may be solved for the four unknowns: α_1, α_2, c_1 and c_2.

The above test of size α for a normal distribution with unknown mean, θ_o, but known variance, σ^2, which has the test statistic, $\bar{x}$, and critical points c_1 and c_2 is a special case of the one-parameter exponential family and provides a test of the hypothesis H_o: $\theta_1 \leq \theta \leq \theta_2$ against the alternative H_a: $\theta < \theta_1$ or $\theta > \theta_2$ and is therefore a uniformly most powerful unbiased test as indicated by Lehmann (24, p.126) and Fisz (11, p.581).

Next consider the case where the variance, σ^2, is unknown. The usual student-t statistic may be used in this case where the normally distributed statistic, $\bar{x}$, is centered

about either of the end points of the interval (θ_1 or θ_2) and divided by the estimated standard deviation of $\bar{x}$, $s/n^{\frac{1}{2}}$ where n is the sample size and

$$s = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}) .$$

Thus,

$$t = \frac{\bar{x}-\theta_1}{s/n^{\frac{1}{2}}} \quad \text{or} \quad t = \frac{\bar{x}-\theta_2}{s/n^{\frac{1}{2}}}$$

The density function of the student-t distribution is

$$f(t) = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\ \Gamma(\frac{n-1}{2})\ (1+\frac{t^2}{n-1})^{\frac{1}{2}n}}\, dt \qquad -\infty < t < \infty$$

where $\Gamma(a) = \int_0^{\infty} y^{a-1} e^{-y}\, dy$ for $a > 0$ is the well-known gamma function. Define α_1' as the probability of being in the left-hand critical region of the test under the student-t distribution with n-1 degrees of freedom. This may be written as

$$\alpha_1' = \int_{-\infty}^{c_1'} f(t)\, dt \tag{12}$$

where c_1' is the left-hand critical point. Define α_2' as the probability of being in the right-hand critical region of the test under the same student-t distribution. Thus,

$$\alpha_2' = \int_{c_2'}^{\infty} f(t)\, dt \tag{13}$$

The symmetry condition for the critical points of the student-t distribution test may be derived as follows:

$$c_2'' - \theta_2 = \theta_1 - c_1''$$

$$c_2' - \frac{\theta_2-\theta_1}{s/n^{\frac{1}{2}}} = - c_1' \text{ where } c_2' = \frac{c_2''-\theta_1}{s/n^{\frac{1}{2}}} \text{ and } c' = \frac{c_1''-\theta_1}{s/n^{\frac{1}{2}}}$$

$$c_1' + c_2' = \frac{\theta_2-\theta_1}{s/n^{\frac{1}{2}}} \tag{14}$$

The level of significance, α', is

$$\alpha' = \alpha_1' + \alpha_2' \tag{15}$$

Given a particular level of significance α', the above four equations (12) to (15) may be solved for the four unknowns: α_1', α_2', c_1', and c_2'.

The student-t test described above for the case of unknown variance, σ^2, is not a uniformly most powerful unbiased test, but instead is a uniformly most powerful unbiased scale invariant test as follows from general results for the exponential family by Hodges and Lehmann (14, p.261).

The testing procedures described above may be applied to a linear regression problem when the dependent variable is normally distributed. An estimated regression coefficient of interest, $\hat{b}_k$, in the vector of coefficients (k is 0, . . ., or K) $\underline{\hat{b}}' = \hat{b}_0, \hat{b}_1, \ldots, \hat{b}_K$ defined by $\underline{\hat{b}} = (X'X)^{-1}X'\underline{y}$ for the regression model $\underline{y} = X\underline{b} + \underline{\varepsilon}$ is normally distributed when the dependent variable, $\underline{y}$, is normally distributed given a fixed regressor matrix, X, and an error term, $\underline{\varepsilon}$.

Gi

if the

$t_K < t$

four ec

Where t

student

explana

statis

diagona

and s^2

Ha

level

If s^2_K

To

simplic

approxi

Person

t test

Given the usual regression situation where the variance of the dependent variable σ^2 is unknown, the restriction $\theta_{1k} \leq b_k \leq \theta_{2k}$ may be tested by solving simultaneously the four equations

$$\alpha'_{1k} = \int_{-\infty}^{c'_{1k}} f(t_{\hat{b}_k})\, dt_{\hat{b}_k} \tag{16}$$

$$\alpha'_{2k} = \int_{c'_{2k}}^{\infty} f(t_{\hat{b}_k})\, dt_{\hat{b}_k} \tag{17}$$

$$c'_{1k} + c'_{2k} = \frac{\theta_{2k}-\theta_{1k}}{s_{\hat{b}_k}} \tag{18}$$

$$\alpha' = \alpha'_{1k} + \alpha'_{2k} \tag{19}$$

where $f(\cdot)$ is the probability density function of the student-t distribution which for a linear regression with K-1 explanatory variables will have n-K degrees of freedom. The statistic $t_{\hat{b}_k}$ is equal to $(\hat{b}_k-\theta_{1k})/s_{\hat{b}_k}$ where $s_{\hat{b}_k}$ is the k^{th} diagonal element of the variance-covariance matrix $s^2(X'X)^{-1}$ and $s^2 = \underline{e}'\underline{e}/(n-K)$ based on the vector of residuals, e.

Having thus determined c'_{1k} and c'_{2k} for the significance level α', the null hypothesis H_o: $\theta_{1k} \leq b_k \leq \theta_{2k}$ is rejected if $t_{\hat{b}_k}$ is less than c'_{1k} or greater than c'_{2k}.

To illustrate this finite interval test consider a consumption function of the form $C_i = b_o+b_1Y_i+\varepsilon_i$ where C_i is aggregate consumption of goods and services, Y_i is aggregate personal disposable income, and ε_i is the disturbance term. A test of the hypothesis that the marginal propensity to

consume lies between zero and one may be formulated as a test of the null hypothesis H_o: $0 \leq b_1 \leq 1$. If $s_{\hat{b}_1} = 1$, $\hat{b}_1 = 1.1$, and 20 annual observations provide 18 degrees of freedom, the critical points obtained by solving equations (16) through (19) at the 5% level of significance are -1.802 and 2.802. In this case the test statistic $t_{\hat{b}_1}$ is equal to 1.1 and the null hypothesis is not rejected.

In the multiparameter case, the hypothesis $\underline{\theta}_1 \leq \underline{b} \leq \underline{\theta}_2$ may be tested by obtaining the solution to the following equations:

$$\theta_1^* = \frac{\underline{\theta}_1'\underline{X}'\underline{X}\underline{\theta}_1/(K-1)}{\underline{e}'\underline{e}/(n-K)} \tag{20}$$

$$\theta_2^* = \frac{\underline{\theta}_2'\underline{X}'\underline{X}\underline{\theta}_2/(K-1)}{\underline{e}'\underline{e}/(n-K)} \tag{21}$$

$$c_1^* - \theta_1^* = \theta_2^* - c_2^* \tag{22}$$

$$\alpha_1^* = \int_0^{c_1^*} h(F_{\underline{\hat{b}}})\, dF_{\underline{\hat{b}}} \tag{23}$$

$$\alpha_2^* = \int_{c_2^*}^{\infty} h(F_{\underline{\hat{b}}})\, dF_{\underline{\hat{b}}} \tag{24}$$

$$\alpha^* = \alpha_1^* + \alpha_2^* \tag{25}$$

where $h(\cdot)$ is the probability density function of the Snedecor's F distribution with (K-1) and (n-K) degrees of freedom. The statistic $F_{\underline{b}}$ is equal to $\frac{\underline{b}'\underline{X}'\underline{X}\underline{b}/(K-1)}{\underline{e}'\underline{e}/(n-K)}$. With c_1^* and c_2^* obtained for significance level α^*, reject the

null hypothesis H_o: $\underline{\theta}_1 \leq \underline{b} \leq \underline{\theta}_2$ if $F_{\hat{\underline{b}}}$ is less than c_1^* or greater than c_2^*.

To demonstrate this test procedure the following standard linear multiple regression equation is used: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \varepsilon_i$ where Y_i is the dependent variable, X_{1i} and X_{2i} are the explanatory variables, and ε_i is the error term. The coefficients b_1 and b_2 might be restricted to non-negative values such that b_1 and b_2 are no greater than one. These restrictions and an X'X matrix based on 20 hypothetical observations result in a two-sided F-test with critical values $c_1^* = .05$ and $c_2^* = 62.37$ at the 5% level of significance. The estimated regression coefficients $\hat{b}_1 = .89$ and $\hat{b}_2 = .44$ yield a test statistic $F_{\hat{\underline{b}}} = 50.38$ which lies within the critical values and, therefore, indicates that the null hypothesis of restricting the coefficients to the finite intervals should not be rejected.

This chapter has developed a test procedure for deciding if inequality restrictions are appropriate for a particular situation where such restrictions limit the parameter of interest to a finite interval where the parameter may be a regression coefficient. Having determined the appropriateness of such a proposed inequality restriction, one may wish to proceed with the incorporation of such an inequality restriction into the estimation process by means of the minimum distance estimator whose distribution will now be discussed in Chapter III.

CHAPTER III

THE MINIMUM DISTANCE ESTIMATOR AND ITS DISTRIBUTION

Once an inequality restriction has been deemed appropriate as by the testing procedures developed in Chapter II, alternative means of dealing with the inequality restriction in estimation need be considered. In order to compare such estimators in terms of their distributional properties, it is first of all necessary to have some idea of the nature of their distributional characteristics. Since the minimum distance estimator is a new estimator in relation to the well-known constrained maximum likelihood and unconstrained maximum likelihood estimators, the distribution of the minimum distance estimator will be discussed in this chapter relative to those of the maximum likelihood estimators to serve as a basis for understanding the comparison of the estimated distributional properties by means of Monte Carlo experiment in Chapter IV.

As already noted, the minimum distance estimator is defined in terms of the unconstrained maximum likelihood estimator. Intuitively, this relationship is best expressed graphically. Figure 6 gives an example of a constrained space with an unconstrained maximum likelihood estimate $\hat{\underline{\theta}}_u$ lying outside of the constrained region.

The point labeled $\hat{\underline{\theta}}_c$ is the constrained maximum likelihood point which represents the point in the constrained set with the greatest likelihood. When the constrained space is defined by inequality restrictions on $\underline{\theta}$, $\hat{\underline{\theta}}_c$ can be obtained

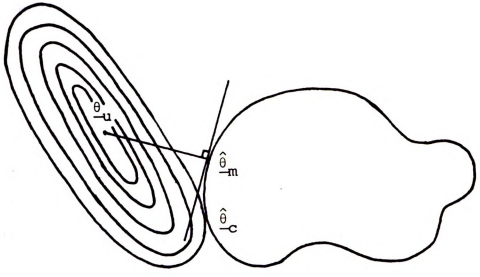


Figure 6

Likelihood Ellipse Contours
and the Constrained Space

by the constrained likelihood function by use of a Lagrangian function subject to the Kuhn-Tucker conditions which require the solution to lie in the constrained space. The point in the constrained space that is closest to the unconstrained maximum likelihood estimate is the minimum distance estimate which is labeled $\hat{\theta}_m$ in Figure 6.

Figure 6 emphasizes the importance of requiring that the constrained parameter space be convex in order to obtain unique minimum distance estimates $\hat{\theta}_m$ since multiple solutions might occur for estimates in non-convex regions. In addition to the convexity assumption, and in order to relate the discussion more easily to classical normal linear regression analysis, the underlying random variable $\underline{y}$ is assumed to be normally distributed with mean θ_o and variance-covariance

matrix $\sigma^2 I$ where I is an appropriately dimensioned identity matrix.

3.1 Distributional Characteristics

In the single parameter case, $\hat{\theta}_u$ lies in a one dimensional parameter space. The dependent variable $\underline{y}$ takes on values y_i with the error term $\underline{\varepsilon}$ defined to be the difference between the $\underline{y}$ and the population parameter θ_o such that $\varepsilon_i = y_i - \theta_o$. In this case, θ_o is assumed to be the expected value of y_i and therefore the expected value of ε_i is zero. The variance of y_i is given by σ^2, and since θ_o is a constant, the variance of ε_i is also given by σ^2. Consequently, the probability density function of y_i is

$$f(y_i) = (2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}} \exp \frac{-(y_i-\theta_o)^2}{2\sigma^2} \tag{26}$$

The unconstrained maximum likelihood estimator, $\hat{\theta}_u$, is equal to the arithmetic mean of the sample $\bar{y} = \sum_{i=1}^{n} y_i$ where n is the sample size. Therefore, $\hat{\theta}_u$ is distributed as normal with mean θ_o and variance $\frac{\sigma^2}{n}$. This may be summarized as follows:

$$\hat{\theta}_u \sim N(\theta_o, \frac{\sigma^2}{n}) \tag{27}$$

$$f(\hat{\theta}_u) = (2\pi)^{-\frac{1}{2}} (\frac{\sigma^2}{n})^{-\frac{1}{2}} \exp \frac{-n(\hat{\theta}_u-\theta_o)^2}{2\sigma^2} \tag{28}$$

The constrained maximum likelihood estimators $\hat{\theta}_c$ and $\hat{\sigma}_c^2$, are derived by maximizing the likelihood function, $L(\theta|\underline{x})$, subject to the general inequality constraints, $g(\theta)$, where $g_j(\theta) \leq 0$ for $j = 1,2, \ldots, r$. The natural logarithm of the likelihood function $l(\theta|\underline{x}) = \ln L(\theta|\underline{x})$ may be substituted into the Lagrangian expression:

$$L = l(\theta|\underline{x}) - \lambda\, g(\theta) \tag{29}$$

subject to the following Kuhn-Tucker conditions:

$$\partial L/\partial\theta - \lambda\, \partial g(\theta)/\partial\theta = 0 \tag{30}$$

$$(\partial L/\partial\theta - \lambda\, \partial g(\theta)/\partial\theta)\ \theta = 0 \tag{31}$$

$$\theta \geq 0 \tag{32}$$

$$g(\theta) \leq 0 \tag{33}$$

$$\lambda\, g(\theta) = 0 \tag{34}$$

$$\lambda \geq 0 \tag{35}$$

Note that solving the Kuhn-Tucker conditions for the desired estimates is a considerably more difficult problem than solving the usual Lagrangian partial derivatives based on equality restrictions.

The minimum distance estimator, $\hat{\theta}_m$, is specified in the single parameter case by finding the θ that minimizes the squared distances such that

$$(\hat{\theta}_u - \hat{\theta}_m)^2 = \min_{\theta \in C} (\hat{\theta}_u - \theta)^2 \tag{36}$$

Given a lower boundary constraint θ_1 and an upper boundary constraint θ_2, $\hat{\theta}_m$ is restricted to take on values as follows:

$$\hat{\theta}_m = \begin{cases} \theta_1, \text{ if } \hat{\theta}_u \leq \theta_1 \\ \theta_2, \text{ if } \hat{\theta}_u \geq \theta_2 \\ \hat{\theta}_u, \text{ otherwise (i.e., } \theta_1 < \hat{\theta}_u < \theta_2) \end{cases} \tag{37}$$

Thus, a regression coefficient that falls outside of a restricted interval is set at the nearest boundary end-point of the restricted interval.

The probability density function of the minimum distance estimator is of the mixed discrete and continuous type:

$$g(\hat{\theta}_m) = \begin{cases} p_1 \text{ for } \hat{\theta}_m = \theta_1 \\ p_2 \text{ for } \hat{\theta}_m = \theta_2 \\ f(\hat{\theta}_u) \text{ for } \theta_1 < \hat{\theta}_m < \theta_2 \end{cases} \tag{38}$$

where

$$p_1 = \int_{\hat{\theta}_u < \theta_1} f(\theta_u) d\theta_u \quad \text{and} \quad p_2 = \int_{\hat{\theta}_u > \theta_2} f(\hat{\theta}_u) d\hat{\theta}_u.$$

The probability density function of the minimum distance estimator can be graphed as shown in Figure 7.

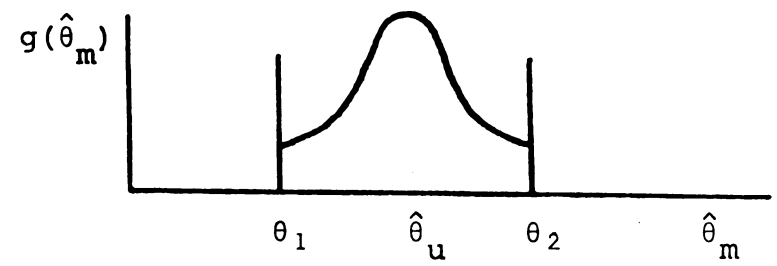


Figure 7

p.d.f. of MDE

Ramsey (30, p.11) shows that the minimum distance estimator is a consistent estimator of θ_o. Referring to Wald's 1949 proof of the consistency of the unconstrained maximum likelihood estimator, $\hat{\theta}_u$, Ramsey points out that the definition of the minimum distance estimator given in equation (36) implies that:

$$(\hat{\theta}_u-\hat{\theta}_m)^2 \leq (\hat{\theta}_u-\theta_o)^2 \tag{39}$$

The consistency of the minimum distance estimator then follows from the consistency of the maximum likelihood estimator, since the consistency of $\hat{\theta}_u$ implies that:

$$\lim \Pr\ (|\hat{\theta}_u-\theta_o| = 0) = 1 \tag{40}$$

Thus, $\hat{\theta}_m$ is also a consistent estimator of θ_o.

The expected value of the minimum distance estimator in this case can be expressed as follows:

$$E\hat{\theta}_m = p_1 \cdot \theta_1 + p_2 \cdot \theta_2 + \int_{\theta_1}^{\theta_2} \hat{\theta}_u \cdot f(\hat{\theta}_u) \cdot d\hat{\theta}_u \qquad (41)$$

Although the unconstrained maximum likelihood estimator, $\hat{\theta}_u$, is an unbiased estimator of θ_o, the minimum distance estimator, $\hat{\theta}_m$, is biased except in the special case where θ_1 and θ_2 are equi-distant from θ_o to form a symmetrical distribution as was shown by Penneck (29, p.15). However, θ_1 and θ_2 are determined by economic theory, and in general, θ_o cannot be expected to fall midway between them. In the case of single one-sided inequality constraints, the minimum distance estimator will generally be biased. In that event the direction of the bias will be in the unconstrained direction.

Since $\hat{\theta}_m$ is generally a biased estimator of θ_o, it would not be appropriate to compare its efficiency relative to $\hat{\theta}_u$ in terms of variance alone. Mean squared error offers a better basis for comparison in this case. The mean squared error of the minimum distance estimator may be written in this case as:

$$MSE\ \hat{\theta}_m = E\ (\hat{\theta}_m - \theta_o)^2 \qquad (42)$$

This expression may be restated algebraically as the sum of the variance plus the bias squared:

$$\text{MSE } \hat{\theta}_m = E\ (\hat{\theta}_m - E\hat{\theta}_m)^2 + (E\hat{\theta}_m - \theta_o)^2 \tag{43}$$

Similarly, the mean squared error of the unconstrained maximum likelihood estimator is:

$$\text{MSE } \hat{\theta}_u = E\ (\hat{\theta}_u - \theta_o)^2 \tag{44}$$

Since in this case the unconstrained maximum likelihood estimator is unbiased, its mean squared error is equal to its variance:

$$\text{MSE } \hat{\theta}_u = E\ (\hat{\theta}_u - E\hat{\theta}_u)^2 = \sigma_u^2 \tag{45}$$

Clearly, when calculating the mean squared error under the assumption that the inequality restrictions are valid, i.e., that $\theta_1 \leq \theta_o \leq \theta_2$, the squared distances will be less for the minimum distance estimator than for the unconstrained maximum likelihood estimator for those points that fall outside the restricted range. For points within the restricted range, the squared distances will be the same. Consequently, the minimum distance estimator will have a mean squared error as small or smaller than that of the unconstrained maximum likelihood estimator when the inequality restrictions hold true.

the M

be co

M to

Having discussed the distributional characteristics of the MDE, the distributional properties of the MDE will now be compared by means of a Monte Carlo experiment in Chapter IV to those of the UMLE and the CMLE.

CHAPTER IV

MONTE CARLO EXPERIMENTS

In this chapter the bias, variance, and mean squared error of the UMLE, CMLE, and MDE will be compared for small samples to determine which estimator is preferable for finitely bounded compact sets of which inequality restrictions serve as a special case.

In his recent paper, Ramsey (30, p.36) has shown that under certain assumptions the constrained maximum likelihood estimator has the same asymptotic distribution as the minimum distance estimator. His preliminary sampling experiments have indicated that both estimators lead to a considerable reduction in mean squared error.

However, a more extensive empirical analysis is needed to compare the small sample characteristics of the unconstrained maximum likelihood (UML) estimator, the constrained maximum likelihood (CML) estimator, and the minimum distance (MD) estimator.

4.1 Criteria for Comparing the UML, CML, and MD Estimators

The following criteria will be used for comparing and evaluating the UML, CML, and MD estimators.

Estimated Expected Value

The estimated expected value of the $j^{\underline{th}}$ parameter may be expressed by:

$$\bar{\hat{\theta}}_j = \frac{1}{L} \sum_{i=1}^{L} \hat{\theta}_{ij} \tag{46}$$

where L is the number of replications of the Monte Carlo sampling experiment.

Estimated Variance

The estimated variance can be calculated for the $j^{\underline{th}}$ parameter by the estimated expected value of the square of an estimator minus the square of its estimated expected value:

$$\text{Est. Var } \hat{\theta}_j = \bar{\hat{\theta}}_j^2 - (\bar{\hat{\theta}}_j)^2 \tag{47}$$

where

$$\bar{\hat{\theta}}_j^2 = \frac{1}{L} \sum_{i=1}^{L} \hat{\theta}_{ij}^2 \quad .$$

Although this estimator of the variance is appropriate for Monte Carlo sampling experiments, it cannot be used to estimate the variance of the minimum distance estimator in the usual single sample situation since one sample only yields one minimum distance estimate and thus the variation of the MDE cannot be observed in this manner when only one sample is available. However, an estimate of the variance of the minimum distance estimator can be obtained for the one sample case by first considering the true population variance of the minimum distance estimator.

The true population variance of the minimum distance estimator may be expressed as:

$$\text{Var } \hat{\theta}_m = E(\hat{\theta}_m - E\hat{\theta}_m)^2 = E\hat{\theta}_m^2 - (E\hat{\theta}_m)^2 \tag{48}$$

where

$$E\hat{\theta}_m = p_1 \cdot \theta_1 + p_2 \cdot \theta_2 + \int_{\theta_1}^{\theta_2} \hat{\theta}_u \cdot f(\hat{\theta}_u | \theta_o) \cdot d\hat{\theta}_u$$

and

$$E\hat{\theta}_m^2 = p_1 \cdot \theta_1^2 + p_2 \cdot \theta_2^2 + \int_{\theta_1}^{\theta_2} \hat{\theta}_u^2 \cdot f(\hat{\theta}_u | \theta_o) \cdot d\hat{\theta}_u$$

Therefore,

$$\text{Var } \hat{\theta}_m = p_1 \cdot \theta_1^2 + p_2 \cdot \theta_2^2 + I_2 - (p_1 \cdot \theta_1 + p_2 \cdot \theta_2 + I_1)^2 \quad (49)$$

where

$$I_1 = \int_{\theta_1}^{\theta_2} \hat{\theta}_u \cdot f(\hat{\theta}_u | \theta_o) \cdot d\hat{\theta}_u$$

and

$$I_2 = \int_{\theta_1}^{\theta_2} \hat{\theta}_u^2 \cdot f(\hat{\theta}_u | \theta_o) \cdot d\hat{\theta}_u$$

p_1 and p_2 may be calculated from the cumulative distribution function of $\hat{\theta}_u$ as:

$$p_1 = F\left(\frac{\theta_1 - \theta_o}{\sigma_u}\right)$$

and

$$p_2 = F\left(\frac{\theta_o - \theta_2}{\sigma_u}\right) = 1 - F\left(\frac{\theta_2 - \theta_o}{\sigma_u}\right)$$

where σ_u^2 is the variance of $\hat{\theta}_u$ and $F(\cdot)$ is the cumulative distribution function of a standardized normally distributed random variable with mean zero and variance one. I_1 is then determined as in Penneck (29, p.14) by:

$$I_1 = (1-p_1-p_2) \cdot \theta_o + \frac{\sigma_u}{(2\pi)^{\frac{1}{2}}} \left[\exp\left\{\frac{-(\theta_1-\theta_o)^2}{2\sigma_u^2}\right\} - \exp\left\{\frac{-(\theta_2-\theta_o)^2}{2\sigma_u^2}\right\}\right]$$

Penneck (29, p.19) also derives I_2 as:

$$I_2 = (\sigma_u^2-\theta_o^2)\cdot(1-p_1-p_2) + 2\cdot\theta_o\cdot I_1$$

For a one-sided left-hand constraint $\hat{\theta}_m \geq \theta_1$ the above two-sided expressions reduce to:

$$E\hat{\theta}_m = p_1\theta_1 + I_1^*$$

$$E\hat{\theta}_m^2 = p_1\theta_1^2 + I_2^*$$

$$\text{Var } \hat{\theta}_m = p_1\theta_1^2 + I_2^* - (p_1\theta_1+I^*)^2$$

where

$$I_1^* = \int_{\theta_1}^{\infty} \hat{\theta}_u\cdot f(\hat{\theta}_u|\theta_o)\cdot d\hat{\theta}_u = (1-p_1)\theta_o + \frac{\sigma_u}{(2\pi)^{\frac{1}{2}}}\left[\exp\{\frac{-(\theta_1-\theta_o)^2}{2\sigma_u^2}\}\right]$$

and

$$I_2^* = \int_{\theta_1}^{\infty} \hat{\theta}_u^2\cdot f(\hat{\theta}_u|\theta_o)\cdot d\hat{\theta}_u = (\sigma_u^2-\theta_o^2)(1-p_1) + 2\theta_o\cdot I_1^*$$

For application to classical regression analysis, the appropriate diagonal element of $s^2(X'X)^{-1}$ serves as a consistent estimator of the variance, σ_u^2, of an ordinary least squares coefficient as indicated by Kendall and Stuart (20, p.82). In addition the unconstrained maximum likelihood estimator, $\hat{\theta}_u$, serves as a consistent estimator of θ_o as follows from Theil (34, p.392) in obtaining an estimator of the variance of the minimum distance estimator by substituting in for σ_u^2 and θ_o in the above equations.

In support of these results ten Monte Carlo sampling experiments based on a sample size of ten and an experiment size of 500 replications compared the true population variance, σ_m^2, of the minimum distance estimator to its sample estimator, $\hat{\sigma}_m^2$. The unconstrained population variance was set at 1.0 while the corresponding minimum distance population variance was .341 based on the constraint that the estimates be positive with population mean at zero. The results indicate that the estimator, $\hat{\sigma}_m^2$, of the variance of the minimum distance estimator did almost as well in estimating the true population variance, $\sigma_m^2 = .341$, of the minimum distance estimator as the standard unbiased estimator, $\hat{\sigma}_u^2$, of the variance of the unconstrained estimator did in estimating its true population value, $\sigma_u^2 = 1.0$. These results appear to support the use of $\hat{\sigma}_m^2$ as a reasonably good approximation of σ_m^2.

Estimated Bias

The estimated bias of each of the three estimators for the $j^{\underline{th}}$ parameter can be calculated as the estimated expected value minus the true population parameter value:

$$\text{Estimated Bias of } \hat{\theta}_j = \bar{\hat{\theta}}_j - \theta_{oj} \tag{50}$$

where θ_{oj} is the true population parameter value.

In the one sample situation the bias of the minimum distance estimator can be expressed for a single constraint

as $\theta_1 p_1 + I_1^* - \theta_o$ where the maximum bias occurs when the true population parameter is at the boundary of the constrained region such that $\theta_o = \theta_1$. With θ_1 substituted for θ_o the bias becomes $\theta_1 p_1 + I_1^* - \theta_1$ where $p_1 = F(\frac{\theta_1 - \theta_1}{\sigma_u}) = F(0) = .5$ so that the maximum bias is $\theta_1(.5) + (1-.5)\theta_1 + \frac{\sigma_u}{(2\pi)^{\frac{1}{2}}} - \theta_1 = \frac{\sigma_u}{(2\pi)^{\frac{1}{2}}}$.

Estimated Mean Squared Error

Although estimated bias and estimated variance are of interest separately their joint importance might be considered by different possible measures of estimated mean squared error. Each different definition of mean squared error reflects a loss function for the trade-off between the larger bias but smaller variance of the minimum distance estimator. Although a multitude of arbitrary definitions may be conceived of to reflect the trade-off between bias and variance, the following definitions are multiparameter extensions of the usual univariate definition of mean squared error and, therefore, seem to be more likely to be generally acceptable than many possible alternatives.

Estimated MSE A

The standard definition of mean squared error is equivalent to variance plus squared bias for each parameter separately in the multiparameter situation. The estimated mean squared error (MSE A) for the $j^{\underline{th}}$ parameter could then be calculated by the estimated variance plus squared estimated

bias:

$$\text{Est. MSE A of } \hat{\theta}_j = \frac{1}{L} \sum_{i=1}^{L} (\hat{\theta}_{ij} - \theta_{oj})^2$$

$$= \text{Est. Var } \hat{\theta}_j + (\text{Est. Bias of } \hat{\theta}_j)^2 \tag{51}$$

for the j^{th} parameter where $j = 1, \ldots, J$ for the J parameters.

Estimated MSE B

A summary measure of mean squared error (MSE B) in the multiparameter situation might be obtained by summing MSE A over all parameters for each estimator and dividing by the number of parameters. Estimated MSE B summarizes the importance of estimated variance and estimated bias for all parameters simultaneously just as estimated MSE A does for each parameter individually. Since MSE B is defined by Ramsey (30, p.20) by $E(\underline{\hat{\theta}}-\underline{\theta}_o)'(\underline{\hat{\theta}}-\underline{\theta}_o)$, estimated MSE B may be defined by:

$$\text{Est. MSE B of } \underline{\hat{\theta}} = \frac{1}{J \cdot L} \sum_{j=1}^{J} \sum_{i=1}^{L} (\hat{\theta}_{ij} - \theta_{oj})^2 \tag{52}$$

Estimated MSE D[1]

Another formulation of mean squared error (MSE D) considers the mean cross-product of errors between parameters.

---

1. Since MSE C is defined by Ramsey (30, p.20) for all positive semi-definite matrices A in the expression $E(\hat{\theta}-\theta_o)'A(\hat{\theta}-\theta_o)$ nothing conclusive can be shown by considering only one such matrix or even a limited number of such matrices in a sampling experiment. Moreover, Ramsey has indicated that MSE C and MSE D are equivalent (30, p.20).

This c

in a v

produc

the di

parame

from i

elemen

cross-

terms

ments

errors

MSE D

A matr

is the

mean e

estima

differ

first

square

this d

This cross-product concept is similar to that of a covariance in a variance-covariance matrix. The estimated mean cross-product is calculated as the average value of the product of the difference of one estimator from its true population parameter value times the difference of another estimator from its true population parameter value. The diagonal elements in the matrix of estimated mean error squares and cross-products are the estimated mean squared error (MSE A) terms themselves for each parameter. The off-diagonal elements in this matrix are the estimated mean cross-product errors between parameters. Ramsey (30, p.20) expresses MSE D algebraically by $E(\underline{\hat{\theta}}-\underline{\theta}_o)(\underline{\hat{\theta}}-\underline{\theta}_o)'$.

$$\text{Est. MSE D of } \underline{\hat{\theta}} = \frac{1}{L}\sum_{i=1}^{L}(\underline{\hat{\theta}}_i-\underline{\theta}_o)(\underline{\hat{\theta}}_i-\underline{\theta}_o)' \qquad (53)$$

A matrix of estimated mean error squares and cross-products is then calculated for each estimator. If the estimated mean error matrix of one estimator is subtracted from the estimated mean error matrix of another estimator and the difference is a positive semi-definite matrix, then the first estimator is said to have a smaller estimated mean squared error than the second estimator in the sense of this definition of estimated MSE D.

Estimated MSE E

A final definition of estimated mean squared error (MSE E) might be considered which also uses the estimated

mean

fines

repre

mated

Under

finan

is sm

matri

said

lowin

cons-

as sh

mean error matrix described above. Ramsey (30, p.20) defines MSE E algebraically by $|E(\underline{\hat{\theta}}-\underline{\theta}_o)(\underline{\hat{\theta}}-\underline{\theta}_o)'|$ where $|M|$ represents the determinant of any square matrix M. Estimated MSE E may be expressed:

$$\text{Est. MSE E of } \underline{\hat{\theta}} = \left|\frac{1}{L}\sum_{i=1}^{L}(\underline{\hat{\theta}}_i-\underline{\theta}_o)(\underline{\hat{\theta}}_i-\underline{\theta}_o)'\right| \qquad (54)$$

Under this definition of estimated MSE E, if the determinant of the estimated mean error matrix of one estimator is smaller than the determinant of the estimated mean error matrix of another estimator, then the first estimator is said to have a smaller mean squared error.

4.2 Three Monte Carlo Models

Monte Carlo experiments will be performed on the following three models: (1) a discrete points model where the constrained space consists of a set of nine discrete points as shown in Figure 8;

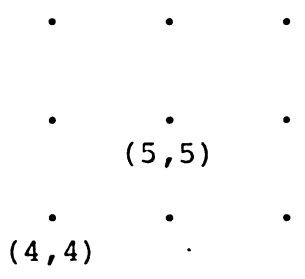


Figure 8

Discrete Points Model

(2) a square model where the constrained space is a square as shown in Figure 9;

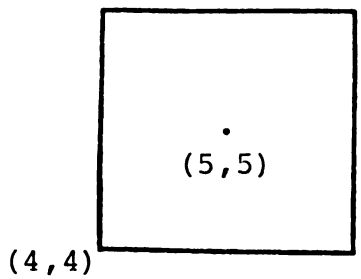


Figure 9

Square Model

(3) an elliptical model where the constrained space is an ellipse with center at (5,5) as shown in Figure 10.

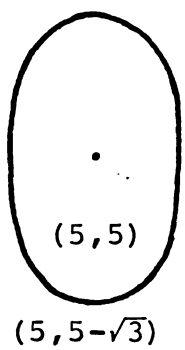


Figure 10

Elliptical Model

Samples of size 30, 60, and 100 will be used for each of the three models. Unconstrained maximum likelihood (UML), minimum distance (MD), and constrained maximum likelihood (CML) estimates will be calculated for each sample and a Monte Carlo sampling experiment size of 1,000 replications will be used.

UML Estimator

Since the three models are the same except for the nature of their constrained space, the unconstrained maximum likelihood estimator is calculated in the same manner for each model. Various specifications of the variance-covariance matrix are tried in order to examine the effect of altering the shape and direction of the major axis of the set of ellipses representing the likelihood function. In the sampling experiments, this set of likelihood ellipses is centered at the unconstrained maximum likelihood estimate point $(\bar{x}, \bar{y})$ where for the bivariate pair of random variables (X, Y), the sample means are defined by $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ where (x_i, y_i) represents an individual sample observation and n is the number of observations in the sample. Thus, the unconstrained maximum likelihood estimator is given as $\underset{\sim}{\hat{\theta}}_u' = (\hat{\theta}_{u1}, \hat{\theta}_{u2}) = (\bar{x}, \bar{y})$ for the first three models. Since calculation of the MD and CML estimates depends upon the nature of the constrained space, their calculations will be discussed separately as each model is examined.

Population Mean Specification

The mean of the bivariate normal distribution will initially be set at the point (5,5) which represents the center point of the constrained space for all three models. Penneck (29, p.15) has indicated that the minimum distance estimator is an unbiased estimator of the population mean

in the special case where the population mean happens to lie at the center of the constrained space. In order to compare this situation with the other extreme possibility of having the mean as far away from the center point as possible, the entire set of sampling experiments will be repeated with the mean at the corner point of the constrained space which is at the point (4,4) for both the discrete points model and the square model. Penneck (29, p.17) indicates that the largest values for bias for the MD and CML estimators occur when the true population mean is as far away from the center point as possible. Consequently, the sampling experiment for the elliptical model will be rerun with the mean at the bottom point of the constrained space ellipse which happens to be the point $(5,5-\sqrt{3})$ for the particular ellipse chosen for this experiment and is as far away from the center point as possible.

Before proceeding with the analysis of the individual models it might be worth noting at this point that in order to get a different estimate for constrained maximum likelihood than for the minimum distance approach, it is necessary that the two variances, σ_x^2 and σ_y^2, on the diagonal of the variance-covariance matrix be significantly different from one another. If the variances are the same, the likelihood "ellipse" will approximate a circle and the CML and MD estimates will tend to be the same.

In addition, in the case of the square, the covariance, σ_{xy}, should be nonzero or the CML and MD estimators will

tend to be the same even when the variances, σ_x^2 and σ_y^2, differ significantly from one another. This tendency occurs because the square was constructed with sides parallel to the horizontal and vertical axes. Since zero covariance implies that the major axis of the likelihood ellipse is also parallel to one of these axes, the major or minor axis of the likelihood ellipse will tend to form a 90° angle with the side of the square and thus result in equal MD and CML estimates. Also note that the correlation coefficient, ρ, for the likelihood function is defined by $\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$, or in other words, the covariance divided by the square root of the variances. Maximum difference between the MD and CML estimates might be expected when in addition to unequal variances, the correlation coefficient is set at .5 or -.5. This corresponds to Ramsey's contention (30, p.8) that "in two-dimensional space the values taken by the two estimators will differ most markedly when the angle between the major 'axes' of the likelihood contour and of the parameter space is $\frac{\pi}{4}$."

Variance-Covariance Specifications

The square roots of the variances and covariance of the bivariate normal random number generator will be designated V1, V2, and CV, respectively, and the correlation coefficient will be expressed as ρ. The experimental design for the sampling experiments for each model and each of the three sample sizes consists of four different specifications for

TABLE 1

VARIANCE AND COVARIANCE SPECIFICATIONS

a. Unequal Variances and Nonzero Covariance

V1	V2	CV	ρ
5	20	-7.07	-.5
10	15	+8.66	+.5
15	10	-8.66	-.5
20	5	+7.07	+.5

b. Equal Variances and Nonzero Covariance

V1	V2	CV	ρ
5	5	-3.54	-.5
10	10	+7.07	+.5
15	15	-10.61	-.5
20	20	+14.14	+.5

c. Unequal Variances and Zero Covariance

V1	V2	CV	ρ
5	20	0.00	.0
10	15	0.00	.0
15	10	0.00	.0
20	5	0.00	.0

d. Equal Variances and Zero Covariance

V1	V2	CV	ρ
5	5	0.00	.0
10	10	0.00	.0
15	15	0.00	.0
20	20	0.00	.0

the variances and covariances each of which was run at four different settings with the results combined. The variance-covariance specifications are given in Table 1.

In order to summarize the resulting estimates of bias, variance, and the various definitions of mean squared error, two sets of tables are presented. One set of tables is designed to indicate how often one estimator is better than another in terms of smaller bias, variance and mean squared error. The three estimators were ranked one, two, or three with a rank of one corresponding to the smallest value for bias, variance or mean squared error. These ranks were then averaged for each of the variance-covariance specifications. A second set of tables shows the average values of bias, variance, and mean squared error for each of the three estimators under the various variance-covariance specifications. This set of tables is designed to give some idea as to the size of the bias, variance, and mean squared error of each of the estimators.

4.3 Discrete Points Model

The first model to be considered will be the discrete points model in which the constrained space consists of a set of discrete points. The rationale behind this model is that it is designed to fill the gap cited by Ramsey (30, p.25) for multiparameter situations that "with respect to the situation in which θ contains only a finite number of points, no finite sample size results have yet been obtained

on mean squared error properties." This model will consider not only the mean squared errors, but also the other characteristics of the estimators which were discussed in detail above.

More specifically, the following nine points have been chosen to represent the constrained space for the discrete points model: (4,4), (4,5), (4,6), (5,4), (5,5), (5,6), (6,4), (6,5), and (6,6). The midpoint (5,5) serves as the true population mean of the bivariate normal distribution.

MD Estimator

The minimum distance estimator, $\underline{\hat{\theta}}_m' = (\hat{\theta}_{m1}, \hat{\theta}_{m2})$, can be calculated for the discrete points model by substituting the nine possible values for $\underline{\hat{\theta}}_m' = (\hat{\theta}_{m1}, \hat{\theta}_{m2})$, into the expression $D^2 = (\hat{\theta}_{m1}-\bar{x})^2 + (\hat{\theta}_{m2}-\bar{y})^2$ where D^2 represents the square of the distance between the UML estimate $\underline{\hat{\theta}}_u$ and the MD estimate $\underline{\hat{\theta}}_m$, and by choosing the value for $\underline{\hat{\theta}}_m' = (\hat{\theta}_{m1}, \hat{\theta}_{m2})$ that corresponds to the smallest value for D^2.

CML Estimator

The constrained maximum likelihood estimator, $\underline{\hat{\theta}}_c' = (\hat{\theta}_{c1}, \hat{\theta}_{c2})$, is calculated as follows. The paired random variables (X, Y) have the joint bivariate normal probability density function given by

$$f(x,y) = \frac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi\sigma_x\sigma_y} \exp \frac{-1}{2(1-\rho^2)} \cdot \left[\left(\frac{x-\theta_{o1}}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\theta_{o1}}{\sigma_x}\right)\left(\frac{y-\theta_{o2}}{\sigma_y}\right) + \left(\frac{y-\theta_{o2}}{\sigma_y}\right)^2\right]$$

where ρ, σ_x^2, and σ_y^2 are the known correlation coefficient and variances as described above and where $\underline{\theta}_o' = (\theta_{o1}, \theta_{o2})$ is the unknown mean of the bivariate normal distribution which will be estimated in the constrained case by the constrained maximum likelihood estimator $\hat{\underline{\theta}}_c' = (\hat{\theta}_{c1}, \hat{\theta}_{c2})$.

The product of n values of the above probability density function where (x,y) is replaced by each of the n sample values (x_i, y_i) and where $\underline{\theta}_o' = (\theta_{o1}, \theta_{o2})$ is replaced by each of nine possible values, $(\theta_{c1}, \theta_{c2})$, for the constrained maximum likelihood estimator $\hat{\underline{\theta}}_c' = (\hat{\theta}_{c1}, \hat{\theta}_{c2})$ forms the appropriate likelihood function which may be maximized by minimizing its exponent as follows:

Minimize,

$$C = \sum_{i=1}^{n} \left[\left(\frac{x_i - \theta_{c1}}{\sigma_x}\right)^2 - 2\rho \left(\frac{x_i - \theta_{c1}}{\sigma_x}\right)\left(\frac{y_i - \theta_{c2}}{\sigma_y}\right) + \left(\frac{y_i - \theta_{c2}}{\sigma_y}\right)^2 \right]$$

by trying each of the nine possible values, $(\theta_{c1}, \theta_{c2})$, and picking the CML value, $\hat{\underline{\theta}}_c = (\hat{\theta}_{c1}, \hat{\theta}_{c2})$, which corresponds to the smallest C value for the nine points. The point thus chosen will be the constrained maximum likelihood estimate for the discrete points case.

Sampling Results

Estimated Bias

The sampling results corresponding to the three sample sizes, two means, and various general specifications of the variances and covariance are summarized in Table 2 and

TABLE 2

AVERAGE RANKS
OF SMALLER ABSOLUTE ESTIMATED BIAS
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	2.75	1.63	1.63	a.	1.00	2.50	2.50
	b.	2.75	1.63	1.63	b.	1.00	2.50	2.50
	c.	2.63	1.75	1.63	c.	1.00	2.50	2.50
	c.	2.75	1.81	1.44	d.	1.00	2.63	2.38
Total		2.72	1.70	1.58		1.00	2.53	2.47
N=60	a.	2.63	1.50	1.88	a.	1.00	2.38	2.63
	b.	2.50	1.81	1.69	b.	1.00	2.50	2.50
	c.	2.75	1.63	1.63	c.	1.00	2.81	2.19
	d.	2.75	1.63	1.63	d.	1.00	2.75	2.25
Total		2.65	1.70	1.71		1.00	2.61	2.39
N=100	a.	2.75	1.50	1.75	a.	1.00	2.38	2.63
	b.	2.13	2.25	1.63	b.	1.00	2.50	2.50
	c.	2.63	2.00	1.38	c.	1.00	2.25	2.75
	d.	2.88	1.75	1.38	d.	1.00	2.25	2.75
Total		2.60	1.87	1.53		1.00	2.34	2.66
TOTAL		2.66	1.74	1.60		1.00	2.49	2.51

Table 3. The notations a,b,c, and d refer to the variance-covariance specifications previously referred to in Table 1.

As expected when the true population mean happens to be at the center point (5,5) of the constrained space, Table 3 shows that the estimated bias is quite small for all three estimators which tends to support the hypothesis that they may be unbiased estimators in this circumstance. In addition to being small, the estimated bias for the three estimators tends to be positive almost as often as negative and

TABLE 3

AVERAGE VALUES OF ESTIMATED BIAS
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	-.006	-.008	.001	a.	-.006	.283	.340
	b.	-.031	-.021	.000	b.	-.031	.568	.740
	c.	-.187	-.057	-.043	c.	-.187	.766	.759
	d.	-.075	-.046	-.044	d.	-.075	.571	.575
N=60	a.	.030	.028	.032	a.	.030	.217	.262
	b.	.006	-.004	.002	b.	.006	.497	.668
	c.	-.019	.002	.002	c.	-.019	.718	.712
	d.	-.034	-.027	-.027	d.	-.034	.467	.461
N=100	a.	.006	.001	-.004	a.	.006	.128	.161
	b.	-.058	-.034	-.021	b.	-.058	.390	.576
	c.	.079	.018	.016	c.	.079	.653	.658
	d.	-.015	.000	.000	d.	-.015	.354	.357

since estimated skewness in this case is approximately zero for all three estimators, this further suggests that the true value of the bias for all three estimators in this special situation is zero. It is interesting to note that while the estimated bias is quite small for all three estimators when the mean is at the center point, the UMLE tends to have a larger average estimated bias in Table 3 than the other two estimators. Furthermore, Tabel 2 indicates that the UMLE tends to have a larger estimated bias more frequently in addition to having a larger average value. This is particularly interesting because the UMLE is known to be unbiased even for small samples.

At the center point, the MDE tends to have a slightly smaller average estimated bias than the CMLE as can be seen

in Table 3. Table 2 indicates that the MDE has the smallest estimated bias more frequently than the CMLE when the mean is at (5,5).

The estimated bias of the CMLE is almost equal to that of the MDE for variance-covariance specifications c and d which suggests that the MDE and the CMLE may be giving nearly equal estimates when the covariance is zero. Increasing the sample size when the population mean is at the (5,5) center point does not appear to influence the size of the estimated bias of any of the three estimators which is to be expected because it is essentially zero initially and therefore not subject to further reduction.

When the true population mean of the random number generator is shifted to the (4,4) corner point, the average estimated bias of the CMLE and MDE becomes larger than that of the UMLE and strictly positive as it tends toward the center point of the constrained space as predicted by Penneck (29, p.17). Table 2 indicates that the UMLE has the smallest bias every time for all sample sizes and specifications when the mean is at the (4,4) corner point. The average estimated bias of the MDE tends to be slightly larger than that of the CMLE especially for specifications a and b which require non-zero covariances. However, the estimated bias of the MDE is smaller than that of the CMLE more frequently for small sample sizes as shown in Table 2. For sample size 100 the CMLE tends to have a smaller estimated bias more frequently than the MDE. The relative performance of the CML and MD

estimators in this situation is primarily dependent upon the slope of the major axis of the likelihood ellipse. The CMLE does better than the MDE when the slope of the major axis of the likelihood ellipse is negative in the (4,4) case. A negative slope means that the CMLE is more likely to select the (4,4) corner point or the nearest side points when the (4,4) corner point is the true population mean.

Table 3 indicates that the estimated bias for the MDE and the CMLE decreases for all cases as sample size increases when the mean is at the (4,4) corner point. The UMLE gives the same values for estimated bias at the (4,4) corner point as it gave for the (5,5) center point since it is not influenced in any way by the constrained space. Sample size does not appear to influence the estimated bias of the UMLE which is shown in Table 3 to be practically zero anyway.

Estimated Variance

In all cases the average estimated variance of the UMLE is substantially larger than that of either the CMLE or the MDE as shown in Table 5. When the mean is at the (5,5) center point the estimated variance of the UMLE is almost always larger than that of the other estimators as indicated in Table 4. For sample size 30 the UMLE's estimated variance is always the largest, but as the sample size increases a few exceptions occur but the UMLE still has the largest estimated variance in most cases. This result seems intuitively plausible since the UMLE can choose points anywhere in

TABLE 4

AVERAGE RANKS
OF SMALLER ESTIMATED VARIANCE
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.13	1.88
	b.	3.00	1.00	2.00	b.	3.00	1.38	1.63
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.25	1.75		3.00	1.50	1.50
N-60	a.	3.00	1.13	1.88	a.	3.00	1.25	1.75
	b.	3.00	1.13	1.88	b.	3.00	1.50	1.50
	c.	2.88	1.88	1.25	c.	3.00	1.75	1.25
	d.	2.75	1.94	1.31	d.	3.00	1.75	1.25
Total		2.90	1.52	1.58		3.00	1.56	1.44
N=100	a.	2.75	1.25	2.00	a.	3.00	1.25	1.75
	b.	2.50	1.38	2.13	b.	3.00	1.38	1.63
	c.	2.50	1.50	2.00	c.	3.00	1.25	1.75
	d.	2.50	1.88	1.63	d.	3.00	1.25	1.75
Total		2.56	1.50	1.94		3.00	1.28	1.72
TOTAL		2.82	1.42	1.76		3.00	1.45	1.55

two-dimensional space whereas the constrained estimators are limited to choosing from among nine particular points. However, as sample size increases the variation of the unconstrained sample estimates should be expected to decrease. If this decrease is substantial enough, most if not all of this variation may take place within the grid of the nine constrained space points. This situation might cause the UMLE to have a smaller variance than either of the constrained estimators in some cases especially when the sample

N=30

N=60

N=100

size is

Table

mean is

which

the MDP

of the

estimat

though

average

nately

select

When th

spectiv

TABLE 5

AVERAGE VALUES OF ESTIMATED VARIANCE
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	10.415	.879	.880	a.	10.415	.859	.820
	b.	10.693	.851	.868	b.	10.693	.691	.804
	c.	14.342	.890	.897	c.	14.342	.847	.840
	d.	14.267	.885	.881	d.	14.267	.855	.846
N=60	a.	5.454	.812	.829	a.	5.454	.781	.750
	b.	5.526	.802	.832	b.	5.526	.592	.724
	c.	6.752	.852	.854	c.	6.752	.779	.777
	d.	6.470	.838	.836	d.	6.470	.777	.774
N=100	a.	2.967	.754	.782	a.	2.967	.655	.636
	b.	3.278	.771	.799	b.	3.278	.468	.618
	c.	4.093	.804	.806	c.	4.093	.710	.698
	d.	4.051	.789	.788	d.	4.051	.678	.671

size is large. The average estimated variance of the MDE in Table 5 is slightly larger than that of the CMLE when the mean is at the (5,5) center point except for specification d which requires equal variances with zero covariance. There the MDE's estimated variance is slightly smaller than that of the CMLE. The table of ranks indicates that the MDE's estimated variance is often larger than that of the CMLE although specification d is again an exception. Since the average estimated sample correlation coefficient was approximately -.003, the CMLE may have had a slight tendency to select the upper left-hand and lower right-hand corner points when the MDE chose the left or right-hand side points, respectively. The ranks of the CMLE and the MDE are equal for

specification c at N=30 and the MDE has a smaller rank than the CMLE at N=60 for specification c. Since specification c requires zero covariance, it is likely that this results in CML and MD estimates that are nearly identical. In any event the estimated variances of the MDE and the CMLE are very close to one another. As sample size increases when the mean is at the (5,5) center point, the estimated variances of all three estimators decrease with that of the UMLE decreasing most substantially as indicated in Table 5.

When the population mean is at the (4,4) corner point the estimated variance of the UMLE is always larger than that of the CMLE or the MDE as can be seen in Table 4. The average value of the estimated variance of the MDE is smaller than that of the CMLE when the population mean is at the (4,4) corner point except for specification b. The average estimated sample correlation coefficient for specification b was substantially negative resulting in a negative slope for the major axis of the likelihood ellipse. The unequal variances may have accentuated the elongation of the likelihood ellipse in this case. These developments would tend to cause the CMLE to be more likely to choose the (4,4) corner point or a nearby side point than it otherwise would.

The ranks of the estimated variance of the MDE are smaller than those of the CMLE when the sample size is 60, equal when the sample size is 30, and larger when the sample size is 100. In general the estimated variance of the MDE and that of the CMLE are so close to one another, no

clear cut pattern emerges. However, the estimated variances of both of the constrained estimators are clearly smaller for the (4,4) corner point case than for the (5,5) center point case. This reduction in estimated variance may help offset the increase in estimated bias that occurs when shifting from the (5,5) center point to the (4,4) corner point. The average estimated variances of all three estimators substantially decrease as sample size increases as shown in Table 5 for both the (5,5) center point case and the (4,4) corner point case.

MSE A

The sampling results for estimated mean squared error A are summarized in Table 6. In the case where the constrained space consists of a finite number of points Ramsey (30, p.17) has indicated that "it appears that nothing can be said in general about the mean squared error gain for the minimum distance estimator. A similar tentative conclusion holds for the constrained maximum likelihood estimator for finite sample sizes." In the absence of an analytical proof of mean squared error gain for the MDE or the CMLE, sampling experiments provide a means of determining the relative values of mean squared error for the alternative estimators. In this instance the sampling experiments indicate that in almost all cases the estimated MSE A of the UMLE is larger than that of either the CMLE or the MDE. In particular, Table 7 reveals that when the mean is at the (5,5) center point, the

TABLE 6

AVERAGE RANKS
OF SMALLER ESTIMATED MSE A FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.25	1.75		3.00	1.62	1.38
N=60	a.	3.00	1.13	1.88	a.	3.00	1.38	1.63
	b.	3.00	1.13	1.88	b.	3.00	1.50	1.50
	c.	2.63	2.00	1.38	c.	3.00	1.75	1.25
	d.	2.75	2.00	1.25	d.	3.00	1.75	1.25
Total		2.84	1.56	1.60		3.00	1.59	1.41
N=100	a.	2.75	1.25	2.00	a.	3.00	1.38	1.63
	b.	2.50	1.38	2.13	b.	3.00	1.50	1.50
	c.	2.50	1.50	2.00	c.	3.00	1.38	1.63
	d.	2.50	1.81	1.69	d.	3.00	1.38	1.63
Total		2.56	1.48	1.96		3.00	1.41	1.59
TOTAL		2.80	1.43	1.77		3.00	1.54	1.46

UMLE's average estimate of MSE A is larger than that of either the CMLE or the MDE. As sample size increases the UML estimate of the MSE A decreases substantially while the estimates of the CMLE and MDE decrease moderately. This decrease is due entirely to the decrease in variance as sample size increases since the bias remains essentially unchanged.

Table 6 shows that the UMLE estimate of MSE A is almost always larger than the CMLE and MDE estimates. The frequency with which the UMLE is larger decreases as sample size increases.

TABLE 7

AVERAGE VALUES OF ESTIMATED MSE A
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	10.415	.879	.880	a.	10.415	1.639	1.386
	b.	10.694	.851	.868	b.	10.694	1.014	1.352
	c.	14.401	.892	.900	c.	14.401	1.404	1.402
	d.	14.269	.885	.881	d.	14.269	1.471	1.465
N=60	a.	5.457	.812	.830	a.	5.457	1.386	1.187
	b.	5.526	.802	.832	b.	5.526	.839	1.170
	c.	6.758	.854	.856	c.	6.758	1.324	1.329
	d.	6.485	.840	.838	d.	6.485	1.316	1.303
N=100	a.	2.969	.754	.783	a.	2.969	1.080	.968
	b.	3.281	.772	.799	b.	3.281	.620	.950
	c.	4.094	.805	.806	c.	4.094	1.140	1.117
	d.	4.058	.790	.789	d.	4.058	1.037	1.026

When the mean is at the (4,4) corner point, the average value of the UMLE estimate of the MSE A is still larger than that of either of the other estimators. Table 6 indicates that with the mean at (4,4) the UMLE estimate is always the largest. However, the CMLE and MDE average estimates are larger than they were in the (5,5) case. All the average estimates decrease as the sample size increases but again the UMLE estimate of MSE A decreases faster than that of either the CMLE or the MDE. Ramsey (30, p.25) has indicated that for the situation where the constrained space consists of only a finite number of points mean squared error gains can be shown to hold asymptotically under suitable regularity conditions in the multivariate case. However, the above sampling results suggest that for small samples the mean squared error gain is even greater than for large samples.

In comparing the CML and MD average values of the estimates of MSE A, Table 7 shows very little difference between their average values in the (5,5) case. When the mean is shifted to the (4,4) point there is only slightly more difference. When comparing the average ranks of the CML and MD estimates, Table 6 shows a difference between the (5,5) and the (4,4) case. When the mean is at the center point, the CMLE rank is more frequently smaller than the MDE, while the reverse is true at the corner point at least most of the time for specifications a, c, and d. Since the estimated variance is dominant over the estimated bias in the calculation of estimated MSE A even in the (4,4) case, the explanation relating to the negative estimated sample correlation coefficient for the estimated variance would also apply to estimated MSE A. In particular the CMLE may have a greater tendency to select the (4,4) corner point or near by points than the MDE does in this special circumstance.

MSE B

The sampling results for estimated mean squared error B are summarized in Table 8 and Table 9. When the mean is at the (5,5) center point the UMLE almost always gives the largest estimate of MSE B, but doing so less frequently as sample size increases as can be seen in Table 8. Table 9 shows that the average value of the estimated MSE B of the UMLE is considerably larger than those of the CMLE and MDE. As the sample size increases the average values given by all

TABLE 8

AVERAGE RANKS
OF SMALLER ESTIMATED MSE B FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.63	1.38	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	1.63	1.38
Total		3.00	1.28	1.72		3.00	1.59	1.41
N=60	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.88	1.13	c.	3.00	2.00	1.00
	d.	2.50	2.13	1.38	d.	3.00	1.75	1.25
Total		2.87	1.50	1.63		3.00	1.69	1.31
N=100	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	2.50	1.50	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.25	1.75	c.	3.00	1.25	1.75
	d.	2.50	1.88	1.63	d.	3.00	1.25	1.75
Total		2.75	1.41	1.84		3.00	1.37	1.63
TOTAL		2.87	1.40	1.73		3.00	1.55	1.45

three estimators decrease, with the average value of the UMLE estimate decreasing most substantially. When the mean is moved to the (4,4) corner point the UMLE always has the largest estimate of MSE B. Table 9 shows that the average values of estimated variance for the CMLE and MDE have increased overall, but continue to decrease as sample size increases.

When the average ranks of the CMLE and MDE estimates of the MSE B are compared in Table 8 it can be seen that the CMLE on the average gives smaller estimates more frequently

TABLE 9

AVERAGE VALUES OF ESTIMATED MSE B
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	6.901	.821	.838	a.	6.901	1.321	1.180
	b.	9.555	.840	.862	b.	9.555	.967	1.289
	c.	7.615	.743	.742	c.	7.615	.927	.918
	d.	13.889	.897	.892	d.	13.889	1.457	1.441
N=60	a.	3.566	.728	.768	a.	3.566	1.072	.946
	b.	4.696	.778	.822	b.	4.696	.784	1.151
	c.	3.608	.657	.657	c.	3.608	.815	.813
	d.	6.884	.850	.847	d.	6.884	1.294	1.291
N=100	a.	1.962	.655	.696	a.	1.962	.793	.716
	b.	2.773	.742	.772	b.	2.773	.572	.908
	c.	2.174	.565	.566	c.	2.174	.649	.639
	d.	3.952	.805	.803	d.	3.952	1.090	1.084

when the mean is at (5,5). However closer examination reveals that when the covariance is zero the MDE more frequently gives a smaller estimate of MSE B. When the mean is shifted to the (4,4) point, the average rank of the MDE is smaller than that of the CMLE. Table 8 also shows that when the covariance is nonzero (specifications a and b) the CMLE and MDE have the smallest estimate of MSE B equally often. Table 9 shows that the average values of the CMLE and MDE estimates of MSE B are very close, particularly when the covariance is zero. When the mean is shifted to (4,4) the difference between the CMLE and MDE estimates increases, but is still small. In particular, the MDE has a slightly smaller estimated MSE B when the variances are equal while the CMLE has a slightly smaller MSE B when the variances are unequal.

Thus the MDE tends to have a slight edge over the CMLE when the likelihood ellipse approximates a circle. Only specification b in the (4,4) case showed the estimated MSE B of the CMLE to be slightly smaller than that of the MDE. Since the average estimated sample correlation coefficient tended to be negative in this case, the slope of the major axis of the likelihood ellipse tended to be negative which tended to keep the CML estimates closer to the (4,4) point than those of the MDE. The opposite result would be expected if the slope had been positive.

MSE D

The sampling results for estimated mean squared error D are summarized in Table 10 and Table 11. Since estimated MSE D is a matrix, an estimator is said to have a smaller estimated MSE D when the difference matrix is positive semi-definite. The difference matrix is obtained by subtracting the MSE D matrix of that first estimator from the MSE D matrix of a second estimator. The average values given in Table 11 are the values of the determinants of the MSE D difference matrices.

Table 10 shows in the (5,5) center point case and the (4,4) corner point case that the CMLE frequently has the smallest estimated MSE D, followed by the MDE, and then by the UMLE with the largest estimate. As sample size increases the prominence of this pattern diminishes. Table 11 shows a large range in the size of the determinants of the

TABLE 10

AVERAGE RANKS
OF SMALLER ESTIMATED MSE D FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.00	2.00
	b.	3.00	1.00	2.00	b.	2.75	1.00	2.25
	c.	3.00	1.00	2.00	c.	3.00	1.75	1.25
	d.	3.00	1.00	2.00	d.	3.00	1.50	1.50
Total		3.00	1.00	2.00		2.94	1.31	1.75
N=60	a.	2.00	1.50	2.50	a.	2.75	1.25	2.00
	b.	2.50	1.25	2.25	b.	2.75	1.25	2.00
	c.	2.25	1.75	2.00	c.	3.00	1.25	1.75
	d.	2.50	1.50	2.00	d.	3.00	1.50	1.50
Total		2.31	1.50	2.19		2.88	1.31	1.81
N=100	a.	2.00	1.50	2.50	a.	2.75	1.25	2.00
	b.	2.50	1.25	2.25	b.	2.75	1.25	2.00
	c.	2.00	1.50	2.50	c.	3.00	1.00	2.00
	d.	2.50	1.50	2.00	d.	3.00	1.00	2.00
Total		2.25	1.44	2.31		2.88	1.12	2.00
TOTAL		2.52	1.31	2.17		2.90	1.25	1.85

MSE D difference matrix for U-C and U-M. When one of the constrained estimators has a larger MSE D matrix than the UMLE it is only slightly larger. As sample size increases the average values of the determinants decrease for U-C and U-M in both the (5,5) and (4,4) case. These two columns also show that the determinants are usually smaller when the variances are unequal (specifications a and c) than when they are equal (specifications b and d). The values of specification c are especially small. The values of bias and variance

N=30

N=60

N=100

in th

estim

avera

stand

CVLI

point

well

as th

varia

tion

point

rate

the C

USP D

TABLE 11

AVERAGE VALUES OF ESTIMATED MSE D
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		U-C	U-M	C-M		U-C	U-M	C-M
N=30	a.	16.549	18.132	-.120	a.	15.277	17.084	-.036
	b.	38.466	42.530	-.187	b.	33.351	30.105	.075
	c.	3.186	3.305	-.000	c.	4.864	5.071	.000
	d.	168.535	168.662	-.000	d.	153.552	153.945	.000
N=60	a.	2.684	3.009	-.055	a.	2.555	3.121	.004
	b.	7.253	8.765	-.134	b.	29.919	34.323	-.080
	c.	-.011	.001	-.000	c.	.779	.833	-.000
	d.	36.253	36.281	-.000	d.	30.704	30.731	-.000
N=100	a.	.332	.390	-.018	a.	.553	.737	.003
	b.	1.697	2.208	-.053	b.	1.643	.344	.112
	c.	-.231	-.234	-.000	c.	.281	.275	-.000
	d.	9.872	9.888	.000	d.	7.861	7.090	-.000

in this case are small and very nearly equal for all three estimators. The second variate has a considerably larger average estimated variance than the first variate in this instance and since the covariance is zero, the MDE and the CMLE may have an unusually strong tendency to choose the points immediately above and below the (5,5) center point as well as the center point itself when the mean is at (5,5). As the sample size increases, the variance of the second variate declines sufficiently to allow a sufficient proportion of the UML estimates to fall between the (5,4) and (5,6) points so that the estimated MSE D of the UMLE actually became smaller than that of the constrained estimators. While the CML MSE D matrix is almost always smaller than the MD MSE D matrix in the (5,5) case, it is by only a small amount.

When the mean is at (4,4) each of the two constrained estimators has the smallest MSE D matrix half of the time.

MSE E

The sampling results for the estimated mean squared error E are summarized in Table 12 and Table 13. As in the other mean squared error cases, the UMLE usually gives the largest estimated MSE E when compared to the CMLE and MDE. Table 12 shows that when the mean is at the (5,5) center point, the UMLE almost always has the largest estimated MSE, although the frequency of this decreases as the sample size increases. When the mean is moved to the (4,4) corner point, the UMLE always has the largest estimated MSE E. Table 13 shows that the average values of estimated MSE E given by all three estimators decrease as sample size increases for both the (5,5) and (4,4) cases. Again, as in previous measures of mean squared error, the estimated values given by the UMLE decrease much more rapidly than do those of the CMLE and MDE. It can also be seen in Table 13 that each estimator usually gives smaller average estimates of MSE E when the variances are not equal (specifications a and c) than when the variances are equal (specifications b and d). This is not the case, however, for the CMLE when the mean is at (4,4). The larger estimated MSE E for the equal variance cases may be explained by noting that the sum of squares of the variances listed in Table 1 is greater than the sum of their cross-products. The exception for the CMLE in the (4,4)

TABLE 12

AVERAGE RANKS
OF SMALLER ESTIMATED MSE E FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.75	1.25	a.	3.00	1.50	1.50
	b.	3.00	1.75	1.25	b.	3.00	1.50	1.50
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	2.00	1.00
Total		3.00	1.62	1.38		3.00	1.69	1.31
N=60	a.	3.00	1.75	1.25	a.	3.00	1.50	1.50
	b.	3.00	1.75	1.25	b.	3.00	1.50	1.50
	c.	3.00	2.00	1.00	c.	3.00	2.00	1.00
	d.	2.50	2.25	1.25	d.	3.00	1.75	1.25
Total		2.87	1.94	1.19		3.00	1.69	1.31
N=100	a.	3.00	1.25	1.75	a.	3.00	1.50	1.50
	b.	2.50	2.00	1.50	b.	3.00	1.50	1.50
	c.	3.00	1.25	1.75	c.	3.00	1.25	1.75
	d.	2.50	2.00	1.50	d.	3.00	1.25	1.75
Total		2.75	1.625	1.625		3.00	1.37	1.63
TOTAL		2.87	1.73	1.40		3.00	1.58	1.42

case is again due to a negative average estimated correlation coefficient. In all cases the average value of the UML estimated MSE E is considerably larger than the estimates of the other two estimators.

When the performance of the CMLE and MDE are compared in Table 12, it can be seen that the MDE more frequently gives the smallest estimate of MSE in both the (5,5) and the (4,4) case, especially for smaller sample sizes. However Table 13 shows that there is little actual difference between the size of the CML and MD estimates of MSE E. When the mean is

TABLE 13

AVERAGE VALUES OF ESTIMATED MSE E
FOR POINTS MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	26.919	.670	.594	a.	26.919	1.379	.830
	b.	53.723	.699	.616	b.	53.723	.792	1.615
	c.	11.950	.530	.526	c.	11.950	.567	.542
	d.	192.630	.803	.794	d.	192.630	1.792	1.746
N=60	a.	6.699	.510	.465	a.	6.699	.824	.532
	b.	13.617	.602	.497	b.	13.617	.568	1.300
	c.	3.097	.393	.392	c.	3.097	.367	.356
	d.	47.219	.722	.717	d.	47.219	1.428	1.408
N=100	a.	2.055	.399	.399	a.	2.055	.431	.299
	b.	4.775	.528	.452	b.	4.775	.315	.817
	c.	1.043	.262	.263	c.	1.043	.170	.170
	d.	15.585	.647	.643	d.	15.585	1.028	1.025

shifted to (4,4) both the CML and MD average estimates usually increase in size, but still remain much smaller than the UMLE average estimate.

Summary of Discrete Points Model Sampling Results

In summarizing the discrete points model, all three estimators appeared to be unbiased when the mean was at the (5,5) center point. The CMLE and the MDE had small to moderate estimated positive bias when the mean was shifted to the (4,4) corner point and the size of their biases was nearly the same. Also at the (4,4) point, the estimated biases of the CMLE and MDE decreased as sample size increased, but remained unchanged at close to 0 in the (5,5) case.

Calculations of the average estimated variance of each of the three estimators showed it to be much larger for the UMLE than for either the CMLE or the MDE in both the (5,5) and the (4,4) situations. As in the case of bias, the two constrained estimators had average estimated variances of nearly the same size. As sample size increased, the estimated variances of all three estimators decreased. Although that of the UMLE declined most dramatically, it still remained the largest.

For all four of the mean squared errors which were examined, the UMLE gave substantially larger average estimates than did the CMLE or the MDE. The estimates of MSE A, B and E always decreased as sample size increased for all three estimators. Because MSE D was presented in matrix form, it was not determined if the estimates were also decreasing there. In particular, the MSE A average estimate of all three estimators was dominated by the effect of the variance, especially in the (5,5) case. In the (4,4) case, the bias affected the CMLE and the MDE about equally. For both the MSE A and MSE B situations, the CMLE tended to give smaller estimates than the MDE when the mean was at (5,5), and the MDE tended to give smaller estimates when the mean was at (4,4). In the case of estimated MSE D the CMLE performed slightly better than the MDE in many cases. In the case of estimated MSE E the MDE usually did slightly better than the CMLE.

In the discrete points case, the estimated skewness of all three estimators was nearly always equal and very small relative to the standard deviation when the mean was at (5,5). There was also about an equal number of positive and negative values. This suggests that all three estimators had symmetric distributions in the (5,5) case. In the (4,4) corner point case, the small estimated skewness of the UMLE remained unchanged. The average estimated skewness of the CMLE and the MDE increased, but remained close to one another. The ratio of average estimated skewness to standard deviation for the constrained estimators ranged from about 0.2 to 2.0. While the UMLE had about an equal number of negative and positive estimates of skewness, the constrained estimators always gave positive estimates in the (4,4) case.

In both the (5,5) and (4,4) cases, the estimated kurtosis of the UMLE, which was standardized relative to the normal distribution, was quite small and fluctuated between negative and positive values. The estimated kurtosis of the CMLE and MDE was almost always negative, and the ratio of their kurtosis to the standard deviations ranged from about -.04 to -1.4.

Also in the discrete points case, all four estimated moments of the CMLE and the MDE were close to one another. The first moments of all three estimators tended to be very close in the (5,5) case, while the CMLE and MDE estimated first moments were slightly smaller than that of the UMLE in the (4,4) case. The second estimated moment of the UMLE

was somewhat larger than those of the constrained estimators in many cases. The third estimated moment of the UMLE was usually larger than the estimated third moments of the constrained estimators in the (5,5) case. However, the reverse was true in the (4,4) case. The fourth estimated moment of the UMLE was considerably larger than those of the constrained estimators in almost all cases when the mean was at (5,5) and was often larger in the (4,4) case.

4.4 Square Model

In the second model the constrained space is defined as the points in and on a square which is located with corner points at (4,4), (4,6), (6,4), and (6,6). As in the first model, the midpoint of the constrained space is at (5,5). The constrained space defined as a square effectively restricts both variates of the bivariate normal distribution to finite intervals. Such restrictions may occur in regression analysis as double inequality restrictions on each of two regression coefficients. As in the previous case, the characteristics of the estimators and their distributions will be analyzed. As before, the UML estimate for this model is $\hat{\underline{\theta}}_u' = (\hat{\theta}_{u1}, \hat{\theta}_{u2}) = (\bar{x},\bar{y})$.

MD and CML Estimates

The minimum distance estimate and the constrained maximum likelihood estimate may be calculated as follows. If the UML estimate lies within the square or on its edge, then the

UML estimate, the CML estimate, and the MD estimate are all equal.

If the UML estimate lies outside the square, then the calculation of the MD estimate, $\underline{\hat{\theta}}_m' = (\hat{\theta}_{m1}, \hat{\theta}_{m2})$, can be calculated by projecting directly onto the nearest side of the square in a horizontal or vertical direction as appropriate. If the UML estimate lies in one of the corner quadrants formed by extending the sides of the square outward from the corner points, then the corner point itself corresponds to the MD estimate.

The CML estimate, $\underline{\hat{\theta}}_c' = (\hat{\theta}_{c1}, \hat{\theta}_{c2})$, is calculated by a search routine that involves minimizing the exponent of the likelihood function which in this case is equivalent to maximizing the likelihood function itself. If the UML estimate lies directly to one side of the square, then that side is searched for the point that corresponds to the maximum of the likelihood function. If the UML estimate lies in one of the corner quadrants, then the two nearest sides of the square are searched. Therefore, as in the discrete points model, the CML estimate, $\underline{\hat{\theta}}_c$, is obtained by minimizing the exponent:

$$C = \sum_{i=1}^{n} \left[\left(\frac{xi-\theta c1}{\sigma x}\right)^2 - 2\rho \left(\frac{xi-\theta c1}{\sigma x}\right)\left(\frac{yi-\theta c2}{\sigma y}\right) + \left(\frac{yi-\theta c2}{\sigma y}\right)^2 \right]$$

subject to the condition that the solution lies on the appropriate boundary of the constrained space.

TABLE 14

AVERAGE RANKS
OF SMALLER ABSOLUTE ESTIMATED BIAS
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	2.63	1.63	1.75	a.	1.00	2.50	2.50
	b.	3.00	1.50	1.50	b.	1.00	2.50	2.50
	c.	3.00	1.63	1.38	c.	1.00	2.50	2.50
	d.	3.00	1.63	1.38	d.	1.00	2.63	2.38
Total		2.91	1.59	1.50		1.00	2.53	2.47
N=60	a.	2.88	1.88	1.25	a.	1.00	2.38	2.63
	b.	2.88	1.63	1.50	b.	1.00	2.50	2.50
	c.	3.00	1.63	1.38	c.	1.00	3.00	2.00
	d.	3.00	1.50	1.50	d.	1.00	2.75	2.25
Total		2.94	1.66	1.40		1.00	2.66	2.34
N=100	a.	2.50	2.00	1.50	a.	1.00	2.50	2.50
	b.	2.63	1.88	1.50	b.	1.00	2.50	2.50
	c.	2.50	1.63	1.88	c.	1.00	2.38	2.63
	d.	2.50	1.63	1.88	d.	1.00	2.38	2.63
Total		2.53	1.78	1.69		1.00	2.44	2.56
TOTAL		2.79	1.68	1.53		1.00	2.54	2.46

Sampling Results

Bias

The sampling results for estimated bias for the square model are summarized in Table 14 and Table 15. As expected when the true population mean happens to be at the (5,5) center point of the constrained space, the estimated bias is quite small for all three estimators which tends to support the hypothesis that they may be unbiased estimators in this circumstance. In addition to being small, the estimated

TABLE 15

AVERAGE VALUES OF ESTIMATED BIAS
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	-.006	-.006	.001	a.	-.006	.299	.352
	b.	-.031	-.018	-.003	b.	-.031	.573	.742
	c.	-.187	-.049	-.041	c.	-.187	.762	.759
	d.	-.075	-.044	-.045	d.	-.075	.572	.574
N=60	a.	.030	.030	.029	a.	.030	.244	.281
	b.	.006	.001	.005	b.	.006	.494	.672
	c.	-.019	.005	.005	c.	-.019	.717	.716
	d.	-.034	-.023	-.023	d.	-.034	.478	.474
N=100	a.	.006	.006	.007	a.	.006	.171	.205
	b.	-.058	-.025	-.020	b.	-.058	.379	.578
	c.	.079	.011	.011	c.	.079	.655	.655
	d.	-.015	-.010	-.010	d.	-.015	.369	.371

bias in the (5,5) case tends to be positive almost as much as negative at least for sample sizes 60 and 100. Since the estimated skewness for these three estimators is quite small and therefore suggests that their distributions are symmetric, the frequency of positive and negative values further suggests that the true value of bias for all three estimators in this special situation is zero. While the estimated bias is quite small for all three estimators, the UMLE tends to have a larger estimated bias than the other two estimators. The estimated bias of the CMLE is almost identical to that of the MDE in the (5,5) case. The table of ranks indicates that the MDE had a smaller estimated bias slightly more frequently than the CMLE although both of these constrained estimators did somewhat better than the UMLE in terms of

frequency of smaller bias. The estimated bias of all three estimators remains essentially unchanged as sample size increases in the (5,5) case.

When the true population mean of the random number generator is shifted to the corner point (4,4), the estimated bias of the CMLE and MDE becomes larger than that of the UMLE and strictly positive so that it tends toward the center point of the constrained space. The rank table indicates that the UMLE always has the smallest estimated bias in the (4,4) case. The estimated bias of the CMLE and the MDE is essentially the same except for specification b. Since specification b requires equal variances with nonzero covariance, the likelihood ellipse should approximate a circle. However, in this case the average estimated standard deviations are approximately 15 and 18 with a substantially negative estimated correlation coefficient. Consequently, the CMLE is able to choose points closer to the (4,4) corner point than those points chosen by the MDE on the average under specification b. When sample size increases in the (4,4) case, the estimated bias of the UMLE remains virtually unchanged. However, the estimated bias of the CMLE and the MDE declines in all cases as sample size increases when the mean is at the (4,4) corner point. Overall there is not much difference between the MDE's estimated bias and that of the CMLE although both constrained estimators do have a larger estimated bias than the UMLE in the (4,4) case.

TABLE 16

AVERAGE RANKS
OF SMALLER ESTIMATED VARIANCE
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.25	1.75
	b.	3.00	1.00	2.00	b.	3.00	1.38	1.63
	c.	3.00	1.63	1.38	c.	3.00	1.88	1.13
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.28	1.72		3.00	1.56	1.44
N=60	a.	3.00	1.13	1.88	a.	3.00	1.13	1.88
	b.	3.00	1.00	2.00	b.	3.00	1.38	1.63
	c.	3.00	1.88	1.13	c.	3.00	2.00	1.00
	d.	3.00	2.00	1.00	d.	3.00	2.00	1.00
Total		3.00	1.50	1.50		3.00	1.62	1.38
N=100	a.	3.00	1.00	2.00	a.	3.00	1.25	1.75
	b.	3.00	1.00	2.00	b.	3.00	1.25	1.75
	c.	3.00	1.38	1.63	c.	3.00	1.75	1.25
	d.	3.00	1.63	1.38	d.	3.00	1.63	1.38
Total		3.00	1.25	1.75		3.00	1.47	1.53
TOTAL		3.00	1.34	1.66		3.00	1.55	1.45

Estimated Variance

The sampling results for estimated variance for the square model are summarized in Table 16 and Table 17. The UMLE always gives the largest estimate of variance both when the mean is at the (5,5) center point and at the (4,4) corner point. The estimated variance of the UMLE must necessarily be equal to or greater than that of the constrained estimators because the constrained space square is a convex set. Table 17 of the average values of the estimates shows that

TABLE 17

AVERAGE VALUES OF ESTIMATED VARIANCE
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	10.415	.829	.835	a.	10.415	.811	.783
	b.	10.693	.804	.825	b.	10.693	.654	.773
	c.	14.342	.863	.862	c.	14.342	.817	.809
	d.	14.267	.853	.851	d.	14.267	.823	.808
N=60	a.	5.454	.751	.775	a.	5.454	.717	.691
	b.	5.526	.737	.766	b.	5.526	.542	.682
	c.	6.752	.804	.805	c.	6.752	.741	.738
	d.	6.470	.794	.793	d.	6.470	.731	.727
N=100	a.	2.967	.693	.710	a.	2.967	.605	.572
	b.	3.278	.664	.725	b.	3.278	.396	.564
	c.	4.093	.743	.743	c.	4.093	.643	.646
	d.	4.051	.727	.726	d.	4.051	.620	.619

the UML estimate of variance is much larger than either of the two constrained estimates in both the (5,5) and (4,4) case. This difference is greatest when the true population mean is on the boundary but it is also very large when the mean is in the center of the constrained space. All of the estimates decrease as sample size increases, with the UML estimate showing the most substantial reduction. Consequently, the advantage in terms of smaller estimated variance of the MDE and CMLE over the UMLE is especially apparent in small samples but was substantial even for the largest sample size.

A comparison of the CML and MD estimates in Table 16 indicates that the CMLE more frequently gives the smallest estimate when the mean is at (5,5). The situation is reversed

in the (4,4) case, but the difference between the average ranks is not as great. Table 17 shows that the average values of the two constrained estimates are very close to one another particularly when the covariance is zero (specifications c and d). A zero covariance implies that the major axis of the likelihood ellipse tends to be either vertical or horizontal and therefore its point of tangency with the square (which has sides that are also either vertical or horizontal) represents not only the CML estimate, but the MD estimate as well. When the mean is moved from the center to the corner point, both of the constrained estimators give smaller average estimates of variance as expected.

Estimated MSE A

The sampling results for estimated MSE A for the square model are summarized in Table 18 and Table 19. As in the case of estimated variance, the UMLE always gives the largest estimate of MSE A in both the case where the mean is at the (5,5) center point and when it is at the (4,4) corner point. The similarities between the estimated variance and the estimated MSE A are also evident in Table 19 where the average values of the estimates of MSE A of all three estimators are nearly the same as their estimates of variance, particularly in the (5,5) case where the bias is essentially zero and for the UMLE in the (4,4) case. In terms of MSE A which considers each parameter separately, Ramsey (30, p.12) has proven analytically that the mean

TABLE 18

AVERAGE RANKS
OF SMALLER ESTIMATED MSE A FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.38	1.63
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.63	1.38	c.	3.00	1.88	1.13
	d.	3.00	1.50	1.50	d.	3.00	1.63	1.38
Total		3.00	1.28	1.72		3.00	1.59	1.41
N=60	a.	3.00	1.13	1.88	a.	3.00	1.38	1.63
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.88	1.13	c.	3.00	2.00	1.00
	d.	3.00	2.00	1.00	d.	3.00	2.00	1.00
Total		3.00	1.50	1.50		3.00	1.72	1.28
N=100	a.	3.00	1.00	2.00	a.	3.00	1.38	1.63
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.38	1.63	c.	3.00	1.75	1.25
	d.	3.00	1.63	1.38	d.	3.00	1.63	1.38
Total		3.00	1.25	1.75		3.00	1.56	1.44
TOTAL		3.00	1.34	1.66		3.00	1.62	1.38

squared error of the MDE is less than that of the UMLE. The sampling experiments for the square model completely support Ramsey's theorem since in all cases the estimated MSE A of the UMLE is larger than that of either the CMLE or the MDE. This held true for the experiments performed with the true population mean at the (5,5) center point as well as for those carried out with the mean at the (4,4) corner point. As sample size increases, the UMLE, CMLE, and MDE all give smaller estimates of MSE A, with those of the UMLE decreasing most rapidly. This result suggests that all three estimators

TABLE 19

AVERAGE VALUES OF ESTIMATED MSE A
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	10.415	.829	.835	a.	10.415	1.607	1.346
	b.	10.694	.804	.825	b.	10.694	.983	1.324
	c.	14.401	.866	.865	c.	14.401	1.375	1.368
	d.	14.269	.853	.851	d.	14.269	1.452	1.424
N=60	a.	5.457	.751	.776	a.	5.457	1.317	1.132
	b.	5.526	.737	.766	b.	5.526	.785	1.133
	c.	6.758	.806	.806	c.	6.758	1.293	1.289
	d.	6.485	.795	.795	d.	6.485	1.268	1.266
N=100	a.	2.969	.693	.710	a.	2.969	1.052	.901
	b.	3.281	.665	.726	b.	3.281	.540	.898
	c.	4.094	.743	.744	c.	4.094	1.064	1.070
	d.	4.058	.728	.727	d.	4.058	.981	.978

may be consistent estimators and that they may converge in terms of MSE A as sample size increases.

While the size of the average estimates of the CMLE and MDE are very close to one another in all cases, the CMLE more frequently has the smallest estimate when the mean is at (5,5). However the MDE most frequently has the smallest estimate when the mean is (4,4). The average values table shows a departure from the similarities with the variance estimates. Instead of decreasing, the CML and MD estimates of MSE increase when the mean is shifted to the (4,4) corner point. This is due entirely to the increase in estimated bias since estimated variance decreased slightly when the mean shifted to (4,4). This result differs from the mean

TABLE 20

AVERAGE RANKS
OF SMALLER ESTIMATED MSE B FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.25	1.75		3.00	1.62	1.38
N=60	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	2.00	1.00	c.	3.00	2.00	1.00
	d.	3.00	2.00	1.00	d.	3.00	2.00	1.00
Total		3.00	1.50	1.50		3.00	1.75	1.25
N=100	a.	3.00	1.00	2.00	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.50	1.50
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.00	2.00	d.	3.00	1.75	1.25
Total		3.00	1.12	1.88		3.00	1.62	1.38
TOTAL		3.00	1.29	1.71		3.00	1.66	1.34

squared error graph drawn by Penneck (29, p.22) which shows the maximum mean squared error of the MDE at the midpoint of the restricted interval at least in the univariate case. The CML estimate usually increases more than the MD estimate except for specification b where the variances are equal and the covariance is nonzero. Here the CML average estimate remains smaller than the MD average estimate.

Estimated MSE B

The sampling results for estimated MSE B are summarized in Table 20 and Table 21.

TABLE 21

AVERAGE VALUES OF ESTIMATED MSE B
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	6.901	.756	.780	a.	6.901	1.286	1.131
	b.	9.555	.784	.813	b.	9.555	.930	1.255
	c.	7.615	.669	.667	c.	7.615	.882	.874
	d.	13.889	.865	.861	d.	13.889	1.425	1.402
N=60	a.	3.566	.655	.700	a.	3.566	1.007	.895
	b.	4.696	.709	.757	b.	4.696	.742	1.106
	c.	3.608	.576	.575	c.	3.608	.768	.765
	d.	6.884	.800	.799	d.	6.884	1.256	1.255
N=100	a.	1.962	.577	.608	a.	1.962	.760	.660
	b.	2.773	.644	.697	b.	2.773	.513	.860
	c.	2.174	.490	.491	c.	2.174	.596	.599
	d.	3.952	.741	.741	d.	3.952	1.032	1.030

Ramsey (30, p.22) has proven in his theorem 3 that even in the multiparameter situation mean squared error B of the UMLE is larger than that of the MDE when the parameter space is a convex set. Ramsey's analytical proof is further verified by the sampling experiments for the square model in that the estimated MSE B of the UMLE was always larger than that of the MDE for all sample sizes and for both the (5,5) centered population mean and the (4,4) corner mean. Table 21 shows that the average estimated MSE B of the UMLE is considerably larger than either of the constrained estimates. As sample size increases, the estimates of MSE B of all of the estimators decrease in both the (5,5) and (4,4) case. As with variance and MSE A, the UML estimates decrease most rapidly, but remain larger than the constrained estimates.

Table 21 also shows that almost all the average estimates are smaller when the variances are unequal (specifications a and c) than when the variances are equal (specifications b and d). Here again this is probably due to the fact that the sum of the squares of the variances is greater than the sum of their cross-products as given in Table 1.

The average ranks table for the CML and MD estimates of MSE B shows that the CML estimate is more frequently smaller when the mean is at (5,5) while the MD estimate is more frequently smaller when the mean is at (4,4). Table 21 indicates that the CML and MD average estimates of MSE B are very close to one another, especially when the covariance is zero (specifications c and d). This result relates to the previous explanation concerning the major (and minor) axis of the likelihood ellipse being parallel to a side of the square. Shifting the mean to (4,4) almost always increases the size of the CML and MD average estimates which again is a result of the increased estimated bias of the constrained estimators as the true population mean moves toward the boundary of the constrained space.

Estimated MSE D

The sampling results of estimated MSE D for the square model are summarized in Table 22 and Table 23. As in the points model, MSE D is a matrix and the average values given in Table 23 are the values of the determinants of the MSE D difference matrices.

TABLE 22

AVERAGE RANKS
OF SMALLER ESTIMATED MSE D FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.00	2.00	a.	3.00	1.00	2.00
	b.	3.00	1.00	2.00	b.	2.75	1.00	2.25
	c.	3.00	1.00	2.00	c.	3.00	1.50	1.50
	d.	3.00	1.25	1.75	d.	3.00	1.50	1.50
Total		3.00	1.06	1.94		2.94	1.25	1.81
N=60	a.	3.00	1.00	2.00	a.	3.00	1.00	2.00
	b.	3.00	1.00	2.00	b.	2.75	1.25	2.00
	c.	3.00	1.50	1.50	c.	3.00	2.00	1.00
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.25	1.75		2.94	1.50	1.56
N=100	a.	3.00	1.00	2.00	a.	2.75	1.25	2.00
	b.	3.00	1.00	2.00	b.	2.75	1.25	2.00
	c.	3.00	1.00	2.00	c.	3.00	1.00	2.00
	d.	3.00	1.00	2.00	d.	3.00	1.25	1.75
Total		3.00	1.00	2.00		2.87	1.19	1.94
TOTAL		3.00	1.10	1.90		2.92	1.31	1.77

Table 22 shows that the UMLE always gives the largest estimate of MSE D when the mean is at (5,5). When the mean is moved to the (4,4) corner point, it still gives the largest estimate most frequently, but not as often as the sample size increases. The CMLE gives a smaller estimate of MSE D more frequently than does the MDE in both the (5,5) and (4,4) cases.

Table 23 indicates that the difference matrices for U-C and U-M are much larger than the matrix of C-M in both the (5,5) and (4,4) cases. Since the determinants of U-C and

TABLE 23

AVERAGE VALUES OF ESTIMATED MSE D
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		U-C	U-M	C-M		U-C	U-M	C-M
N=30	a.	17.353	18.852	-.121	a.	15.695	17.643	-.023
	b.	39.251	43.089	-.193	b.	33.842	30.537	.077
	c.	4.829	4.866	-.000	c.	5.673	5.794	-.000
	d.	169.349	169.460	.000	d.	154.337	154.337	.001
N=60	a.	3.056	3.410	-.061	a.	2.823	3.388	-.004
	b.	7.696	9.161	-.132	b.	6.009	3.610	.128
	c.	.665	.682	-.000	c.	1.116	1.131	.000
	d.	36.854	36.862	.000	d.	31.115	31.126	-.000
N=100	a.	.537	.642	-.024	a.	.628	.857	.005
	b.	1.959	2.475	-.065	b.	1.828	.469	.120
	c.	.057	.058	-.000	c.	.376	.376	-.000
	d.	10.277	10.275	-.000	d.	8.197	8.211	-.000

U-M are strictly positive here as are the diagonal elements of the corresponding difference matrices, the MSE D of the UMLE is clearly larger than that of the MDE and the CMLE. Also in both the (5,5) and (4,4) cases the U-C and U-M determinants always decrease as sample size increases which suggests the possible convergence of the unconstrained and constrained estimators in terms of MSE D as sample size increases. When the mean is shifted from (5,5) to (4,4) the U-C and U-M determinants almost always decrease which probably reflects the increased estimated bias of the CMLE and the MDE in the (4,4) case since the UMLE is not affected by the constraints. When the mean is at (5,5), the C-M determinant is almost always negative, while in the (4,4) situation there is an equal number of positive and negative

TABLE 24

AVERAGE RANKS
OF SMALLER ESTIMATED MSE E FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.75	1.25	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	1.50	1.50
	c.	3.00	1.50	1.50	c.	3.00	1.75	1.25
	d.	3.00	1.50	1.50	d.	3.00	1.75	1.25
Total		3.00	1.69	1.31		3.00	1.62	1.38
N=60	a.	3.00	2.00	1.00	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	2.00	1.00
	c.	3.00	2.00	1.00	c.	3.00	2.00	1.00
	d.	3.00	2.00	1.00	d.	3.00	2.00	1.00
Total		3.00	2.00	1.00		3.00	1.87	1.13
N=100	a.	3.00	1.75	1.25	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	1.50	1.50
	c.	3.00	1.25	1.75	c.	3.00	1.50	1.50
	d.	3.00	1.25	1.75	d.	3.00	1.75	1.25
Total		3.00	1.56	1.44		3.00	1.56	1.44
TOTAL		3.00	1.75	1.25		3.00	1.68	1.32

values. Table 23 also shows the C-M determinant is almost zero whenever the covariance is zero. This reflects the closeness of the CMLE to the MDE in terms of MSE D.

Estimated MSE E

The sampling results for estimated MSE E for the square model are summarized in Table 24 and 25. The average ranks table indicates that the UMLE always gives the largest estimate of MSE E in both the (5,5) and (4,4) cases. Table 25 also shows its estimates to be considerably larger than

TABLE 25

AVERAGE VALUES OF ESTIMATED MSE E
FOR SQUARE MODEL

		Mean at (5,5)				Mean at (4,4)		
		UML	CML	MD		UML	CML	MD
N=30	a.	26.919	.564	.508	a.	26.919	1.284	.733
	b.	53.723	.605	.540	b.	53.723	.711	1.523
	c.	11.950	.409	.407	c.	11.950	.465	.454
	d.	192.630	.748	.741	d.	192.630	1.691	1.626
N=60	a.	6.699	.414	.373	a.	6.699	.711	.435
	b.	13.617	.500	.413	b.	13.617	.501	1.198
	c.	3.097	.279	.277	c.	3.097	.272	.268
	d.	47.219	.639	.638	d.	47.219	1.320	1.309
N=100	a.	2.055	.306	.285	a.	2.055	.383	.231
	b.	4.775	.404	.353	b.	4.775	.249	.730
	c.	1.043	.177	.177	c.	1.043	.119	.120
	d.	15.585	.549	.549	d.	15.585	.906	.905

those of the CMLE and MDE. As sample size increases, the average estimates of all three estimators decrease with those of the UMLE decreasing most rapidly. This result suggests the possible consistency and convergence of the three estimators.

Comparing the CML and MD estimate average ranks, Table 24 shows that the MDE most frequently gave the smallest estimate of MSE E, regardless of the location of the mean. The average values of the CML and MD estimates are very close to one another, particularly when the covariance is zero (specifications c and d). Zero covariance implies the parallel relationship between the likelihood ellipse and the square that often results in equal estimates for the CMLE and the MDE as previously discussed. When the mean is

shifted from (5,5) to (4,4) almost all of the constrained estimates of MSE E increase, especially for small sample sizes which probably reflects the increased estimated bias observed as the true mean shifts to the boundary of the constrained space.

Summary of Square Model Sampling Results

All three estimators appeared to be unbiased when the population mean was in the center of the constrained space. When the true population mean shifted to the boundary of the constrained space, the UMLE appeared to remain unbiased but the CMLE and the MDE appeared to have small to moderate estimated biases which were nearly the same size. The estimated bias of the constrained estimators when the mean was at the boundary decreased as sample size increased.

The estimated variances of the constrained estimators were considerably smaller than that of the UMLE at both the center and on the boundary of the constrained space. The estimated variances of all the estimators tended to decline as sample size increased.

The UMLE always had the largest estimated mean squared error for MSE A, B, and E, and almost always for MSE D. The estimated MSE A, B, and E of all three estimators always decreased as sample size increased. It was not determined if the estimated MSE D was decreasing for all the estimators because it was in matrix form. For the estimates of MSE A and B the CMLE tended to do slightly better than the MDE when

the true population mean was in the center of the constrained space, but the reverse was true when the mean was at the boundary of the constrained space. For MSE D the CML estimate was generally slightly smaller than the MD estimate. For MSE E the MD estimate was generally slightly smaller than the CML estimate.

When the true population mean was at the center of the constrained space, the estimates of the skewness of the UMLE, the CMLE, and the MDE were very small relative to their standard deviations and the estimates were nearly always equal. There were about an equal number of positive and negative values which suggests that these estimators have symmetric distributions when the population mean is at the center of the constrained space. The UMLE continued to have small estimated skewness when the mean shifted to the boundary of the constrained space, however the estimates of the skewness of the CMLE and the MDE, while remaining nearly equal to one another, increased. The ratio of the average estimated skewness to the standard deviations for the constrained estimators ranged from about 0.5 to 6.5 when the population mean was at the boundary, while the ratio for the UMLE remained essentially zero.

The estimated kurtosis of the UMLE was quite small in almost all cases and fluctuated between positive and negative values. The estimates of the kurtosis of both of the constrained estimators were nearly the same and were almost always negative. When the mean was at the center point,

the ratio of the average estimates of the kurtosis of the constrained estimators to their standard deviations ranged from about -1 to -2. When the mean was at the boundary, the average ratio of the estimated kurtosis of the constrained estimators to their standard deviations ranged from about -.5 to -7.0.

The CML and MD estimates of the four moments were quite close to one another in each case. When the mean was in the center of the constrained space, the estimated first moments of all three estimators were nearly the same. When the mean shifted to the boundary of the constrained space the estimated first moments of the CMLE and the MDE remained close to one another but were somewhat larger than that of the UMLE. The estimated second moment of the UMLE was somewhat larger than those of the constrained estimators when the mean was in the center of the constrained space, but the estimated second moments of all three estimators were about the same when the mean shifted to the boundary of the constrained space. The estimated third moment of the UMLE was greater than that of the constrained estimators when the mean was in the center of the constrained space, but became about equal to those of the constrained estimators when the mean was shifted to the boundary. The estimated fourth moment of the UMLE was considerably larger than that of the constrained estimators when the mean was at the center point of the constrained space, and was about equal to the constrained estimates when the mean was at the boundary. The

estimated third and fourth moments of the UMLE exhibited considerably greater variation than those of the constrained estimators when the mean was at the boundary of the constrained space.

4.5 Elliptical Model

An alternative formulation of the constrained space in the continuous case is an ellipse. As in the case of the discrete points and the continuous square, the ellipse may also be centered around the point (5,5). In particular, the ellipse may be given by the equation:

$$\frac{(y-5)^2}{3} + \frac{(x-5)^2}{2} = 1 \qquad (55)$$

The UML estimate is the average of the observed sample values, and is designated for this bivariate normal situation as $\underline{\hat{\theta}}_u' = (\hat{\theta}_{u1}, \hat{\theta}_{u2}) = (\bar{x}, \bar{y})$ as described above. If the UML estimate is in or on the ellipse then the CML and MD estimates are equal to the UML estimate. Otherwise the MD estimate can be obtained by finding the values of θ_1 and θ_2 that minimize the expression $D^2 = (\theta_1 - \bar{x})^2 + (\theta_2 - \bar{y})^2$ where D^2 represents the square of the distance between the UML estimate and the MD estimate subject to the condition that the solution be on the ellipse. A boundary search procedure is used to obtain the MD estimates. If the UML estimate falls outside of the ellipse, the CML estimate is obtained by maximizing the likelihood function (or minimizing the exponent of

the likelihood function) subject to the condition that the solution be a point on the constrained space ellipse. The CML estimates are also obtained by a boundary search procedure.

Sampling Results

Since in the discrete points model and in the square model the shift from the (5,5) center point to the (4,4) corner point was accomplished by shifting both coordinates of the mean of the bivariate normal in the same direction and by the same amount, the two variates of the bivariate normal were considered together in the presentation of tables and exposition for the discrete points model and the square model. However, in the elliptical model the shift from the (5,5) center point to the $(5,5-\sqrt{3})$ bottom point on the ellipse is accomplished by changing only the second coordinate of the true population mean. Consequently the two variates of the bivariate normal will be treated separately in the tables and exposition for the elliptical model.

Estimated Bias

The sampling results for estimated bias for the elliptical model are summarized in Table 26 and Table 27. When the true population mean is at the (5,5) center point both variates of the UMLE give the largest estimates of bias most frequently. When the mean is moved to the $(5,5-\sqrt{3})$ bottom point on the ellipse, however, the first variate of the UMLE

gives the largest estimate of bias less often than the CMLE. Meanwhile the second variate always gives the smallest estimate of bias. The average values table shows that the estimated bias of all three estimators is quite small and about equal for all three estimators when the mean is at the center of the constrained space and also for the first variate when the mean is shifted to the bottom point on the constrained space ellipse. This result suggests that all

TABLE 26

AVERAGE RANKS
OF SMALLER ABSOLUTE ESTIMATED BIAS
FOR ELLIPTICAL MODEL

		Mean at (5,5)					
		V1			V2		
		UML	CML	MD	UML	CML	MD
N=30	a.	2.75	1.75	1.50	2.50	1.75	1.75
	b.	2.50	1.50	2.00	3.00	1.25	1.75
	c.	3.00	1.00	2.00	2.75	1.25	2.00
	d.	3.00	1.00	2.00	3.00	1.25	1.75
Total		2.81	1.31	1.88	2.81	1.38	1.81
N=60	a.	3.00	1.50	1.50	2.00	1.75	2.25
	b.	3.00	1.75	1.25	3.00	1.50	1.50
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	1.75	1.25	2.75	1.75	1.50
Total		3.00	1.62	1.38	2.69	1.62	1.69
N=100	a.	2.75	1.75	1.50	2.50	2.00	1.50
	b.	2.50	2.00	1.50	2.00	2.00	2.00
	c.	2.25	1.50	2.25	2.50	2.00	1.50
	d.	2.25	2.00	1.75	2.50	1.75	1.75
Total		2.44	1.81	1.75	2.37	1.94	1.69
TOTAL		2.75	1.58	1.67	2.62	1.65	1.73

TABLE 26 - Continued

		Mean at $(5,5-\sqrt{3})$						
		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	1.50	3.00	1.50	a.	1.00	2.50	2.50
	b.	1.75	2.75	1.50	b.	1.00	2.50	2.50
	c.	3.00	1.25	1.75	c.	1.00	2.50	2.50
	d.	3.00	1.25	1.75	d.	1.00	2.50	2.50
Total		2.31	2.06	1.63		1.00	2.50	2.50
N=60	a.	1.25	3.00	1.75	a.	1.00	2.50	2.50
	b.	1.00	3.00	2.00	b.	1.00	2.25	2.75
	c.	3.00	1.75	1.25	c.	1.00	2.50	2.50
	d.	3.00	1.75	1.25	d.	1.00	3.00	2.00
Total		2.06	2.38	1.56		1.00	2.56	2.44
N=100	a.	1.00	3.00	2.00	a.	1.00	2.50	2.50
	b.	1.00	3.00	2.00	b.	1.00	2.50	2.50
	c.	2.75	2.00	1.25	c.	1.00	2.50	2.50
	d.	2.50	2.00	1.50	d.	1.00	3.00	2.00
Total		1.81	2.50	1.69		1.00	2.62	2.38
TOTAL		2.06	2.31	1.63		1.00	2.56	2.44

three estimators may be unbiased in this special circumstance. The exception is when the mean is shifted to $(5,5-\sqrt{3})$. While this obviously does not affect the UMLE, it greatly increases the average estimated bias of the second variates of the constrained estimators which become strictly positive. For both variates and for both means, the estimates of the UMLE are about equally negative and positive. Increasing the sample size does not appear to affect the size of the estimated bias of the UMLE which is practically zero anyway.

TABLE 27

AVERAGE VALUES OF ESTIMATED BIAS
FOR ELLIPTICAL MODEL

Mean at (5,5)

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	-.006	-.009	-.006	a.	-.234	-.072	-.090
	b.	-.045	-.025	-.018	b.	-.031	-.029	-.022
	c.	-.006	-.000	.001	c.	-.243	-.066	-.085
	d.	-.075	-.051	-.056	d.	.035	.028	.028
N=60	a.	.030	.029	.018	a.	.032	.015	.020
	b.	.052	.033	.029	b.	.006	.003	-.002
	c.	.030	.025	.011	c.	.076	.040	.044
	d.	-.034	-.019	-.021	d.	-.018	-.015	-.015
N=100	a.	.006	.006	.008	a.	.005	.005	.004
	b.	.034	.013	.009	b.	-.058	-.027	-.032
	c.	.006	.007	.007	c.	.014	.021	.021
	d.	-.015	-.014	-.015	d.	-.016	-.013	-.012

Mean at $(5,5-\sqrt{3})$

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	-.006	-.083	-.035	a.	-.234	1.177	1.100
	b.	-.045	-.190	-.094	b.	-.031	1.079	1.097
	c.	-.006	.001	.002	c.	-.243	1.167	1.084
	d.	-.075	-.048	-.050	d.	.035	.863	.863
N=60	a.	.030	-.045	-.020	a.	.032	1.055	1.012
	b.	.052	-.173	-.087	b.	.006	.927	.937
	c.	.030	.028	.012	c.	.076	1.042	.992
	d.	-.034	-.006	-.009	d.	-.018	.632	.630
N=100	a.	.006	-.058	-.033	a.	.005	.854	.824
	b.	.034	-.195	-.105	b.	-.058	.744	.752
	c.	.006	.007	.005	c.	.014	.828	.795
	d.	-.015	-.007	-.008	d.	-.016	.474	.473

When the CMLE and MDE average ranks are compared it can be seen that the two estimators give the smallest estimates almost equally often in the (5,5) case with the CMLE doing slightly better for both variates. However when the mean is moved to the bottom point at $(5,5-\sqrt{3})$, the variates perform differently. The first variate of the CMLE more frequently gives the largest estimate while that of the MDE more often gives the smallest. The second variate of the CMLE also gives the largest estimate most often, with the MDE average rank not far behind. Table 27 shows that the average estimates of bias of both estimators are usually fairly close to one another. Also, the average estimates do not usually seem to be affected by sample size. The exception to this statement is the case of the second variate when the mean is at $(5,5-\sqrt{3})$. Increasing the sample size here reduced the average estimates of the CMLE and the MDE. This suggests the possibility that both constrained estimators may be asymptotically unbiased. The second variate here is also the only place where all of the estimates of bias of the constrained estimators are positive. Elsewhere the number of positive and negative values are about equal.

Estimated Variance

The sampling results for estimated variance for the elliptical model are summarized in Table 28 and Table 29. The UMLE always gives the largest estimate of variance regardless of which variate or which mean is being examined.

Table 29 shows that the UMLE's average estimate of variance is usually considerably larger than the CML and MD estimates except for the first variate under specification c where in this case the average estimated variance of the first variate itself was quite small relative to its value for specifications a, b, and d in this instance. As a result all three estimators had smaller estimates for variance. The UMLE estimated variance is closer to that of the constrained

TABLE 28

AVERAGE RANKS
OF SMALLER ESTIMATED VARIANCE
FOR ELLIPTICAL MODEL

		Mean at (5,5)					
		V1			V2		
		UML	CML	MD	UML	CML	MD
N=30	a.	3.00	1.50	1.50	3.00	1.50	1.50
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	1.50	1.50	3.00	1.50	1.50
Total		3.00	1.62	1.38	3.00	1.37	1.63
N=60	a.	3.00	1.50	1.50	3.00	1.50	1.50
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	1.50	1.50	3.00	1.50	1.50
Total		3.00	1.63	1.37	3.00	1.37	1.63
N=100	a.	3.00	1.75	1.25	3.00	1.25	1.75
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	2.00	1.00	3.00	1.00	2.00
Total		3.00	1.81	1.19	3.00	1.19	1.81
TOTAL		3.00	1.69	1.31	3.00	1.31	1.69

TABLE 28 - Continued

		Mean at $(5,5-\sqrt{3})$						
		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	1.75	1.25	b.	3.00	1.00	2.00
	c.	3.00	1.50	1.50	c.	3.00	1.50	1.50
	d.	3.00	1.75	1.25	d.	3.00	1.50	1.50
Total		3.00	1.62	1.38		3.00	1.25	1.75
N=60	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	1.75	1.25	b.	3.00	1.00	2.00
	c.	3.00	1.50	1.50	c.	3.00	1.25	1.75
	d.	3.00	1.75	1.25	d.	3.00	1.50	1.50
Total		3.00	1.62	1.38		3.00	1.19	1.81
N=100	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	1.50	1.50	b.	3.00	1.00	2.00
	c.	3.00	1.50	1.50	c.	3.00	1.25	1.75
	d.	3.00	2.00	1.00	d.	3.00	1.50	1.50
Total		3.00	1.62	1.38		3.00	1.19	1.81
TOTAL		3.00	1.62	1.38		3.00	1.21	1.79

estimators were small to begin with since they were restricted to lie in the constrained space, while that of the UMLE tends to decrease rapidly as a larger and larger proportion of its estimates fall within the constrained space. As sample size increases the UMLE average estimate always decreases fairly substantially.

The average ranks of the two constrained estimators show that the first variate of the MDE most frequently gives the smallest estimate of variance while the second variate of the CMLE most often gives the smallest estimate, regardless of

TABLE 29

AVERAGE VALUES OF ESTIMATED VARIANCE
FOR ELLIPTICAL MODEL

Mean at (5,5)

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.382	.836	.635	a.	10.415	1.431	1.731
	b.	7.415	.918	.814	b.	10.693	1.355	1.512
	c.	.830	.535	.271	c.	14.342	1.750	2.142
	d.	13.474	.957	.956	d.	14.267	1.442	1.444
N=60	a.	1.674	.682	.555	a.	5.454	1.399	1.589
	b.	3.863	.873	.805	b.	5.526	1.266	1.368
	c.	.457	.375	.239	c.	6.752	1.700	1.904
	d.	7.283	.960	.962	d.	6.470	1.291	1.286
N=100	a.	.954	.538	.468	a.	2.967	1.265	1.370
	b.	2.264	.793	.731	b.	3.278	1.128	1.221
	c.	.255	.237	.181	c.	4.093	1.610	1.696
	d.	3.840	.918	.916	d.	4.051	1.231	1.233

Mean at $(5,5-\sqrt{3})$

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.382	.788	.580	a.	10.415	1.174	1.443
	b.	7.415	.822	.746	b.	10.693	1.102	1.281
	c.	.830	.520	.244	c.	14.342	1.501	1.814
	d.	13.474	.907	.903	d.	14.267	1.286	1.283
N=60	a.	1.674	.606	.489	a.	5.454	.997	1.147
	b.	3.863	.744	.698	b.	5.526	.902	1.020
	c.	.457	.368	.223	c.	6.752	1.364	1.511
	d.	7.283	.871	.873	d.	6.470	1.069	1.062
N=100	a.	.954	.443	.371	a.	2.967	.737	.811
	b.	2.264	.611	.587	b.	3.278	.674	.766
	c.	.255	.226	.141	c.	4.093	1.024	1.089
	d.	3.840	.765	.763	d.	4.051	.830	.831

the location of the mean. The average values table shows that the average estimates of the CMLE and MDE are relatively close to one another compared to the much larger size of the UMLE. In addition the constrained average estimates always decrease as sample size increases. The estimated variance of the constrained estimators for the second variate is always larger than that for the first variate because the major axis of the constrained space ellipse is parallel to the axis of the second variate which allows a greater range for variation for the second variate than for the first variate.

Estimated MSE A

The sampling results for estimated MSE A for the elliptical model are summarized in Table 30 and Table 31. The UMLE always gives the largest estimate of MSE A when the mean is at the (5,5) center point. When the mean is moved to the $(5,5-\sqrt{3})$ bottom point the UMLE still always gives the largest estimate for the first variate, however it occasionally does not give the largest estimate for the second variate. The estimated MSE A of the UMLE is occasionally smaller than that of the constrained estimators for the second variate partly because the estimated variance of the UMLE is generally smaller than usual in this case and partly because the estimated bias of the constrained estimators is considerably larger for the second variate when the population mean is set at the bottom point on the ellipse. Table 31 shows that the

UMLE average estimate of MSE A is usually considerably larger than the average estimates of the CMLE and MDE. All of the UMLE average estimates decrease as sample size increases.

Comparing the CMLE and the MDE in the average ranks table shows that the MD estimate of the first variate is most frequently the smallest in both the center point and bottom point mean situations. On the other hand, the CML

TABLE 30

AVERAGE RANKS
OF SMALLER ESTIMATED MSE A
FOR ELLIPTICAL MODEL

		Mean at (5,5)					
		V1			V2		
		UML	CML	MD	UML	CML	MD
N=30	a.	3.00	1.50	1.50	3.00	1.50	1.50
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	1.50	1.50	3.00	1.50	1.50
Total		3.00	1.62	1.38	3.00	1.38	1.62
N=60	a.	3.00	1.50	1.50	3.00	1.50	1.50
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	1.50	1.50	3.00	1.50	1.50
Total		3.00	1.62	1.38	3.00	1.38	1.62
N=100	a.	3.00	1.75	1.25	3.00	1.25	1.75
	b.	3.00	2.00	1.00	3.00	1.00	2.00
	c.	3.00	1.50	1.50	3.00	1.50	1.50
	d.	3.00	2.00	1.00	3.00	1.00	2.00
Total		3.00	1.81	1.19	3.00	1.19	1.81
TOTAL		3.00	1.68	1.32	3.00	1.32	1.68

TABLE 30 - Continued

Mean at $(5,5-\sqrt{3})$								
		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	2.00	1.00	b.	3.00	1.00	2.00
	c.	3.00	1.50	1.50	c.	2.75	1.00	2.25
	d.	3.00	1.75	1.25	d.	3.00	1.50	1.50
Total		3.00	1.69	1.31		2.94	1.12	1.94
N=60	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	2.00	1.00	b.	3.00	1.00	2.00
	c.	3.00	1.50	1.50	c.	2.75	1.25	2.00
	d.	3.00	1.75	1.25	d.	3.00	2.00	1.00
Total		3.00	1.69	1.31		2.94	1.31	1.75
N=100	a.	3.00	1.50	1.50	a.	3.00	1.00	2.00
	b.	3.00	2.00	1.00	b.	3.00	1.25	1.75
	c.	3.00	1.50	1.50	c.	2.75	1.25	2.00
	d.	3.00	2.00	1.00	d.	3.00	1.75	1.25
Total		3.00	1.75	1.25		2.94	1.31	1.75
TOTAL		3.00	1.71	1.29		2.94	1.25	1.81

estimate for the second variate is most frequently the smallest regardless of where the mean is. Table 31 shows that the average MSE A estimates of the CMLE and the MDE are relatively close to one another. The average constrained estimates for the second variate are larger than those for the first variate. Shifting the mean to the $(5,5-\sqrt{3})$ bottom point of the ellipse slightly reduces the size of the average constrainted estimates for the first variate but increases the average estimates for the second variate which reflects the positive bias for the second variate at the bottom

TABLE 31

AVERAGE VALUES OF ESTIMATED MSE A
FOR ELLIPTICAL MODEL

Mean at (5,5)

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.387	.839	.639	a.	10.415	1.431	1.731
	b.	7.417	.919	.814	b.	10.694	1.356	1.513
	c.	.830	.535	.271	c.	14.401	1.755	2.149
	d.	13.509	.959	.958	d.	14.269	1.442	1.444
N=60	a.	1.675	.683	.556	a.	5.457	1.400	1.590
	b.	3.866	.874	.806	b.	5.526	1.266	1.368
	c.	.458	.376	.239	c.	6.758	1.701	1.906
	d.	7.283	.960	.962	d.	6.485	1.293	1.289
N=100	a.	.954	.539	.469	a.	2.969	1.266	1.371
	b.	2.265	.793	.731	b.	3.281	1.129	1.222
	c.	.255	.237	.181	c.	4.094	1.611	1.696
	d.	3.846	.918	.916	d.	4.058	1.232	1.234

Mean at $(5,5-\sqrt{3})$

		V1				V2		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.387	.798	.580	a.	10.415	2.455	2.622
	b.	7.417	.859	.754	b.	10.694	2.267	2.485
	c.	.830	.520	.244	c.	14.401	2.862	2.989
	d.	13.509	.910	.906	d.	14.269	2.793	2.769
N=60	a.	1.675	.632	.495	a.	5.457	1.844	1.934
	b.	3.866	.774	.705	b.	5.526	1.762	1.897
	c.	.458	.368	.224	c.	6.758	2.449	2.495
	d.	7.283	.871	.873	d.	6.485	2.251	2.243
N=100	a.	.954	.468	.378	a.	2.969	1.274	1.306
	b.	2.265	.649	.599	b.	3.281	1.228	1.333
	c.	.255	.226	.141	c.	4.094	1.710	1.721
	d.	3.846	.765	.763	d.	4.058	1.582	1.583

TABLE 32

AVERAGE RANKS
OF SMALLER ESTIMATED MSE B
FOR ELLIPTICAL MODEL

		Mean at (5,5)				Mean at (5,5-$\sqrt{3}$)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	1.50	1.50	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.25	1.75
	c.	3.00	1.25	1.75	c.	3.00	1.50	1.50
	d.	3.00	1.50	1.50	d.	3.00	1.50	1.50
Total		3.00	1.31	1.69		3.00	1.44	1.56
N=60	a.	3.00	1.75	1.25	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.25	1.75
	c.	3.00	1.50	1.50	c.	3.00	1.50	1.50
	d.	3.00	1.50	1.50	d.	3.00	2.00	1.00
Total		3.00	1.44	1.56		3.00	1.56	1.44
N=100	a.	3.00	1.25	1.75	a.	3.00	1.50	1.50
	b.	3.00	1.00	2.00	b.	3.00	1.25	1.75
	c.	3.00	1.50	1.50	c.	3.00	1.50	1.50
	d.	3.00	1.00	2.00	d.	3.00	2.00	1.00
Total		3.00	1.19	1.81		3.00	1.56	1.44
TOTAL		3.00	1.31	1.69		3.00	1.52	1.48

point on the constrained space ellipse. In all cases the average constrained estimate of MSE A almost always declines as sample size increases which suggests that the CMLE and the MDE may be consistent estimators of the population mean.

Estimated MSE B

The sampling results for estimated MSE B for the elliptical model are summarized in Table 32 and Table 33. The variates are combined under the definition of estimated

TABLE 33

AVERAGE VALUES OF ESTIMATED MSE B
FOR ELLIPTICAL MODEL

		Mean at (5,5)				Mean at (5,5-$\sqrt{3}$)		
		UML	CML	MD		UML	CML	MD
N=30	a.	6.901	1.135	1.185	a.	6.901	1.627	1.601
	b.	9.555	1.137	1.164	b.	9.555	1.563	1.620
	c.	7.615	1.145	1.210	c.	7.615	1.691	1.617
	d.	13.889	1.201	1.201	d.	13.889	1.851	1.837
N=60	a.	3.566	1.041	1.073	a.	3.566	1.238	1.215
	b.	4.696	1.070	1.087	b.	4.696	1.268	1.301
	c.	3.608	1.038	1.073	c.	3.608	1.409	1.359
	d.	6.884	1.127	1.126	d.	6.884	1.561	1.558
N=100	a.	1.962	.902	.920	a.	1.962	.871	.842
	b.	2.773	.961	.976	b.	2.773	.938	.966
	c.	2.174	.924	.938	c.	2.174	.968	.931
	d.	3.952	1.075	1.075	d.	3.952	1.174	1.173

MSE B. The UMLE always gives the largest estimate of MSE B regardless of the location of the mean. Table 33 shows that the average UML estimate is always considerably larger than those of the CMLE and the MDE, which as in the case of the square tends to reconfirm Ramsey's theorem (30, p.22) for convex sets indicating that the MSE B of the UMLE should be larger than that of the MDE. As sample size increases, the UML average estimates decline.

The average ranks table shows that the CML estimate of MSE B is most frequently smallest when the mean is at (5,5). The MDE gives smaller estimates slightly more often than the CMLE when the mean is moved to (5,5-$\sqrt{3}$). Table 33 shows that the CML and MD average estimates are relatively close to one another particularly when the variances are equal and the

TABLE 34

AVERAGE RANKS
OF SMALLER ESTIMATED MSE D
FOR ELLIPTICAL MODEL

		Mean at (5,5)			Mean at (5,5-$\sqrt{3}$)		
		UML	CML	MD	UML	CML	MD
N=30	a.	3.00	1.00	2.00	2.75	1.00	2.25
	b.	3.00	1.00	2.00	3.00	1.00	2.00
	c.	3.00	1.00	2.00	2.75	1.00	2.25
	d.	3.00	1.00	2.00	3.00	1.50	1.50
Total		3.00	1.00	2.00	2.87	1.13	2.00
N=60	a.	3.00	1.00	2.00	2.75	1.00	2.25
	b.	3.00	1.00	2.00	3.00	1.00	2.00
	c.	3.00	1.00	2.00	2.75	1.25	2.00
	d.	3.00	1.00	2.00	3.00	1.50	1.50
Total		3.00	1.00	2.00	2.87	1.19	1.94
N=100	a.	3.00	1.00	2.00	2.75	1.00	2.25
	b.	3.00	1.00	2.00	3.00	1.00	2.00
	c.	3.00	1.00	2.00	2.75	1.25	2.00
	d.	3.00	1.00	2.00	3.00	1.50	1.50
Total		3.00	1.00	2.00	2.87	1.19	1.94
TOTAL		3.00	1.00	2.00	2.87	1.17	1.96

covariance is nonzero. Shifting the mean to (5,5-$\sqrt{3}$) usually increases the size of the average estimates as a result of the increase in bias. However, the larger the sample size the smaller the increase. As the sample size is increased, all of the average estimates of the constrained estimators decline.

Estimated MSE D

The sampling results of estimated MSE D for the elliptical model are summarized in Table 34 and Table 35. As for

TABLE 35

AVERAGE VALUES OF ESTIMATED MSE D
FOR ELLIPTICAL MODEL

		Mean at (5,5)				Mean at (5,5-$\sqrt{3}$)		
		U-C	U-M	C-M		U-C	U-M	C-M
N=30	a.	15.048	17.103	-.100	a.	13.529	15.300	-.046
	b.	35.476	38.466	-.115	b.	32.244	33.514	-.061
	c.	3.734	6.841	-.104	c.	3.571	6.681	-.035
	d.	160.863	160.854	-.000	d.	144.515	144.868	.000
N=60	a.	2.226	2.816	-.037	a.	2.232	2.699	-.013
	b.	6.077	7.096	-.051	b.	5.623	6.010	-.023
	c.	.418	1.064	-.028	c.	.383	.998	-.007
	d.	32.819	32.830	-.000	d.	27.140	27.185	-.000
N=100	a.	.297	.423	-.010	a.	.468	.604	-.003
	b.	1.285	1.618	-.023	b.	1.629	1.710	-.009
	c.	.043	.178	-.005	c.	.068	.269	-.001
	d.	8.254	8.252	-.000	d.	7.595	7.599	-.000

the two previous models, MSE D is a matrix and the average values given in Table 35 are the values of the determinants of the MSE D difference matrices.

Table 34 shows that when the mean is at (5,5) the CMLE always gives the smallest estimate of MSE D, the MDE gives the second smallest estimate, and the UMLE always gives the largest estimate. When the mean is moved to (5,5-$\sqrt{3}$) there are a few exceptions, but the general pattern remains the same. The average values table shows that the estimated MSE D matrices of the CMLE and the MDE are always smaller that that of the UMLE. Table 35 also shows that shifting the mean to (5,5-$\sqrt{3}$) usually decreases the average difference between the UMLE and the constrained estimators' MSE D

matrices, but less so as sample size increases. It can also be seen that the CML average estimate of the MSE D matrix is always slightly smaller or the same as the average MD estimate. The difference is always zero when the variances are equal and the covariance is zero (specification d) where the likelihood "ellipse" approximates a circle. Shifting the mean to $(5,5-\sqrt{3})$ decreases the difference between the CML and MD estimated MSE D matrices. Increasing the sample size also decreases the difference which suggests that the constrained estimators may converge asymptotically.

Estimated MSE E

The sampling results for estimated MSE E for the elliptical model are summarized in Table 36 and Table 37. The UMLE always gives the largest estimate of MSE E regardless of the location of the mean. Table 37 shows that its average estimates are considerably larger than those of the CMLE and the MDE. As sample size increases, the estimates of all three estimators decrease, with the UMLE estimate declining most rapidly suggesting its convergence asymptotically to the constrained estimators.

The MDE more frequently gives the smallest estimate of MSE than does the CMLE. This is true in both the (5,5) and the $(5,5-\sqrt{3})$ cases. Table 37 shows that the constrained average estimates are relatively close to one another particularly when the variances are equal and the covariance is zero (specification d). Shifting the mean to $(5,5-\sqrt{3})$ tends

TABLE 36

AVERAGE RANKS
OF SMALLER ESTIMATED MSE E
FOR ELLIPTICAL MODEL

		Mean at (5,5)				Mean at (5,5-$\sqrt{3}$)		
		UML	CML	MD		UML	CML	MD
N=30	a.	3.00	2.00	1.00	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	2.00	1.00
	c.	3.00	2.00	1.00	c.	3.00	1.50	1.50
	d.	3.00	1.50	1.50	d.	3.00	1.50	1.50
Total		3.00	1.87	1.13		3.00	1.62	1.38
N=60	a.	3.00	2.00	1.00	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	2.00	1.00
	c.	3.00	2.00	1.00	c.	3.00	1.50	1.50
	d.	3.00	1.50	1.50	d.	3.00	2.00	1.00
Total		3.00	1.87	1.13		3.00	1.75	1.25
N=100	a.	3.00	2.00	1.00	a.	3.00	1.50	1.50
	b.	3.00	2.00	1.00	b.	3.00	1.75	1.25
	c.	3.00	2.00	1.00	c.	3.00	1.50	1.50
	d.	3.00	1.00	2.00	d.	3.00	2.00	1.00
Total		3.00	1.75	1.25		3.00	1.69	1.31
TOTAL		3.00	1.83	1.17		3.00	1.69	1.31

to increase the average estimates of the CMLE and the MDE, especially for small sample sizes which reflects the increased estimated bias for the constrained estimators. When the sample size is 100, the estimates are actually decreased in some cases which results from the compensating decrease in estimated variance especially for the second variate.

TABLE 37

AVERAGE VALUES OF ESTIMATED MSE E
FOR ELLIPTICAL MODEL

		Mean at (5,5)				Mean at (5,5-$\sqrt{3}$)		
		UML	CML	MD		UML	CML	MD
N=30	a.	26.919	1.192	1.021	a.	26.919	1.907	1.414
	b.	53.723	1.245	1.108	b.	53.723	1.878	1.666
	c.	11.950	.938	.583	c.	11.950	1.489	.730
	d.	192.630	1.383	1.383	d.	192.630	2.532	2.500
N=60	a.	6.699	.907	.773	a.	6.699	1.061	.831
	b.	13.617	1.066	.934	b.	13.617	1.254	1.139
	c.	3.097	.639	.455	c.	3.097	.901	.557
	d.	47.219	1.241	1.241	d.	47.219	1.960	1.958
N=100	a.	2.055	.623	.559	a.	2.055	.514	.411
	b.	4.775	.829	.742	b.	4.775	.688	.645
	c.	1.043	.382	.306	c.	1.043	.387	.243
	d.	15.585	1.130	1.131	d.	15.585	1.209	1.207

Summary of Elliptical Model Sampling Results

In the elliptical model the estimated bias was essentially zero for both variates of all three estimators when the mean was in the center of the constrained space and for the first variate when the mean was at the bottom of the constrained space ellipse. For the second variate when the mean was at the bottom of the ellipse, the UML estimate continued to appear to be unbiased but both of the constrained estimators had moderate positive estimated bias that declined as sample size increased.

The estimated variance of the UMLE was always greater than that of the constrained estimators. However, the estimated variance of the UMLE tended to converge toward that of

the constrained estimators as sample size increased. The estimated variance of the CMLE was quite close to that of the MDE in most cases. The estimated variance of each of the three estimators always decreased as sample size increased.

The UMLE always had the largest estimated mean squared error for MSE A, B, and E. It also always had the largest estimate of MSE D when the mean was at the center of the constrained space, and almost always had the largest estimate of MSE D when the mean was at the bottom of the ellipse. The MDE tended to have the smallest estimate of MSE A for the first variate while the CMLE tended to have the smallest estimate of MSE A for the second variate. For estimated MSE B, the CMLE tended to have the smallest estimate when the mean was at the center of the ellipse, while the MDE tended to have the smallest estimate when the mean was at the bottom of the ellipse. The CMLE always had the smallest estimate of MSE D when the mean was in the center of the constrained space and almost always had the smallest estimate when the mean was at the bottom. The MDE almost always had the smallest estimate of MSE E when the mean was at the center of the constrained space and usually had the smallest estimate when the mean was at the bottom. In all of the estimates of mean squared error as variously defined, the CMLE's estimate of MSE was quite close to that of the MDE.

When the true population mean was at the center of the constrained space, the estimates of the skewness of the UMLE, the CMLE, and the MDE were very small relative to their

standard deviations, and the estimates were nearly always equal. There were about an equal number of positive and negative values which suggests that these estimators have symmetric distributions when the population mean is at the center of the constrained space. The UMLE as well as the first variate of the constrained estimators continued to have small estimated skewness when the mean was shifted to the bottom of the elliptical constrained space. However for the second variate the estimates of skewness of the CMLE and the MDE increased and were almost always positive when the mean shifted to the boundary. The ratio of the average estimates of skewness to the standard deviations for the second variates of the constrained estimators when the mean was at the boundary ranged from about 0.5 to 5.0.

The estimated kurtosis of the UMLE was quite small in almost all cases and fluctuated between positive and negative values. When the mean was at the center point the estimates of kurtosis of both the constrained estimators were nearly the same and were usually negative. The ratio of kurtosis to their standard deviations ranged from about -0.1 to -1.0. When the mean was shifted to the boundary of the constrained space the ratios of kurtosis to the standard deviations for the first variates ranged from -0.5 to -1.5 while the ratios of the second variate ranged from -1.0 to +6.0.

The CML and MD estimates of the four moments were quite close to one another in each case. When the mean was in the

center of the constrained space, the estimated first moments of all three estimators were nearly the same. When the mean shifted to the boundary the first moments for the first variate of the three estimators were about the same, but the UMLE's first moment for the second variate tended to be smaller than those of the constrained estimators. The estimated second moment of the UMLE was somewhat larger than those of the constrained estimators when the mean was in the center of the constrained space. When the mean was at the bottom of the ellipse, the second moments of the three estimators for the first variate were about the same, but the second moment of the UMLE for the second variate tended to be smaller than those of the constrained estimators. The estimated third moment of the UMLE was greater than those of the constrained estimators when the mean was in the center of the constrained space and was greater for the first variate when the mean was at the boundary. The estimated third moment of the UMLE for the second variate was often smaller but occasionally somewhat larger than those of the constrained estimators, and in general exhibited greater variation than those of the constrained estimators. The estimated fourth moment of the UMLE was generally quite a bit larger than those of the constrained estimators, although occasionally its estimated fourth moment for the second variate was smaller than those of the constrained estimators when the mean was at the boundary.

4.6 Conclusion

The above Monte Carlo sampling experiments have clearly demonstrated the superiority of the MDE and the CMLE over the UMLE in terms of the various definitions of mean squared error under different specifications and conditions as suggested by Ramsey (30, p.20). In addition, the virtual equivalence of the MDE and the CMLE in a number of situations has been demonstrated. In general, there has not been a great deal of difference between the results obtained for the MDE and those obtained for the CMLE. In those circumstances where the CMLE gave a smaller mean squared error than the MDE, the MDE may be preferred because of its ease of calculation relative to the difficulties encountered in calculating the CMLE for situations involving finitely bounded compact sets.

Now that the MDE has been shown to lead to considerable reduction in mean squared error, examples of its use in single sample situations will be presented in Chapter V.

CHAPTER V

APPLYING THE MINIMUM DISTANCE ESTIMATOR

Having discussed the testing procedures of Chapter II and the minimum distance estimating procedures of Chapters III and IV, Chapter V will now demonstrate the use of the tests and the minimum distance estimator in single sample situations where restrictions are imposed on various production functions. This chapter will demonstrate under what circumstances and to what extent in these particular examples use of the minimum distance estimator can result in a reduction in the estimated variance of the regression coefficients and how this reduction in estimated variance is not substantially offset by an increase in estimated bias in the calculation of estimated mean squared error.

Restrictions may be imposed for any number of reasons including a priori theoretical considerations, restrictions suggested by previous research and published studies, or by testing procedures such as those described in Chapter III. In the heuristic examples given below of applying the minimum distance estimator, no attempt has been made to justify each restriction demonstrated although some such reasons may occur to the reader for these or for additional restrictions in these or in other examples of interest.

This technique may be applied to data from Kendrick (21, p.192) using a log-linear representation of a CES production function derived from a truncated Taylor series expansion of the form:

$$\log Q = b_0 + b_1 \log K + b_2 \log L + b_3 (\log K - \log L)^2 + \varepsilon \quad (56)$$

where Q = output, K = capital, L = labor, and ε = error term. b_0 is the constant intercept term, b_1 may be referred to as the capital coefficient, b_2 may be referred to as the labor coefficient, and b_3 is the coefficient of the substitution term. The above equation may be estimated by ordinary least squares. Tests may then be performed on various proposed restrictions which might be incorporated by use of the minimum distance estimator. Tests of these inequality restrictions are performed as special cases of the test procedures discussed in Chapter III and are in terms of the student-t distribution. Table 39 lists some possible restrictions on the coefficients of labor and capital where the test statistic derived from the regression for each proposed restriction must be greater than the critical value of -1.7 for the restriction to be accepted at the 5% level of significance. None of the coefficients violated the restriction that the coefficients be greater than zero. Consequently, this restriction was accepted at the 5% level for the coefficients of capital and labor at each of the three sample sizes. The restriction that the coefficients by greater than one-half was violated for the labor coefficient in both the sample size 60 and sample size 78 cases. However, these violations were not substantial enough to cause the one-half restriction to be rejected in any of the tests of these coefficients. All of the labor coefficients but none of the capital

coefficients violated the restriction requiring the coefficients to be greater than one, but this restriction was rejected only for the labor coefficient in the sample size 78 case. All of the coefficients violated the restriction that the coefficients be greater than two. Moreover, the restriction was rejected at the 5% level in every case except for the capital coefficient for sample size 60 where the estimated variance was noticeably larger than for the other two sample sizes. This larger variance might explain why the test statistic was not sufficiently large enough to result in the rejection of the restriction at the 5% level in this one case.

The estimated variance of the minimum distance estimator, the estimated absolute value of the bias of the minimum distance estimator, and the estimated mean squared error of the minimum distance estimator are also given in Table 38 along with the percent reduction in mean squared error resulting from the use of the minimum distance estimator instead of the unconstrained estimator. The maximum absolute bias was also calculated and happens to correspond to the estimated absolute bias under the restriction $\hat{\theta}_m \geq 2$ in each case. The reduction in estimated variance was generally not substantially different although usually slightly larger than that of the estimated mean squared error. The reduction in estimated variance tended to increase as progressively more stringent restrictions were placed on the regression coefficients. Thus in this sense the efficiency of the

TABLE 38

ESTIMATED VARIANCE, ABSOLUTE BIAS AND MEAN SQUARED ERROR OF THE MDE FOR THE KENDRICK REGRESSIONS

	$\hat{\theta}_u$	Est. Var $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m \geq$	Est. Var $\hat{\theta}_m$	Est. |Bias| $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
N=30								
Labor	.655	.108	2.0	0	.106	.001	.106	3%
			.5	½	.059	.068	.063	42%
			-1.0	1	.001	.131	.017	85%
			-4.1	2	.000	.131	.000	100%
Capital	1.246	.036	6.5	0	.036	.000	.036	0%
			3.9	½	.036	.000	.036	0%
			1.3	1	.031	.007	.031	16%
			-4.0	2	.000	.076	.000	100%
N=60								
Labor	.282	.382	.5	0	.204	.130	.221	42%
			- .4	½	.078	.247	.139	64%
			-1.2	1	.000	.247	.048	87%
			-2.8	2	.000	.247	.001	99%
Capital	1.658	.482	2.4	0	.479	.001	.479	1%
			1.7	½	.445	.010	.445	8%
			.9	1	.353	.063	.357	26%
			- .5	2	.074	.277	.150	69%
N=78								
Labor	.428	.059	1.8	0	.055	.003	.055	6%
			- .3	½	.013	.096	.023	62%
			-2.4	1	.000	.097	.001	99%
			-6.5	2	.000	.097	.000	100%
Capital	1.549	.053	6.7	0	.053	.000	.053	0%
			4.6	½	.053	.000	.053	0%
			2.4	1	.052	.000	.052	1%
			-2.0	2	.000	.092	.001	97%

estimates of the coefficients of labor and capital for this production function tended to improve substantially when the minimum distance estimator was used to impose effective a priori restrictions in the estimation process. The contribution of the minimum distance estimator to production function theory consists of the extent to which this improvement in efficiency is important to the production function theorist which in turn may depend upon the nature and stringency of the restrictions that such a theorist deems reasonable to impose.

In another example, this approach may be applied to the production functions estimated by Ramsey and Zarembka (32, p.5) in their study of alternative function forms of the aggregate production function. The stochastic formulations of the five functional forms used were the Cobb-Douglas (CD) production function:

$$\ln y_i = \ln \alpha_0 + \alpha_1 \ln k_i + \alpha_2 \ln L_i + u_i;$$

the constant elasticity of substitution (CES) production function:

$$y_i^{\rho}/V = \delta_1 k_i^{\rho} + \delta_2 L_i^{\rho} + u_{2i};$$

the variable elasticity of substitution (VES) production function:

$$\ln y_i = \ln \gamma + \alpha(1-\delta\rho) \ln k_i + \alpha\delta\rho \ln(L_i+(\rho-1)k_i) + u_{3i};$$

the generalized production function (GPF):

$$\ln y_i + \gamma y_i = \ln \alpha_0 + \alpha_1 \ln k_i + \alpha_2 \ln L_i + u_{4i};$$

and the quadratic production function (QP):

$$y_i = \alpha_0 + \alpha_1 L_i + \alpha_2 k_i + \alpha_3 L_i^2 + \alpha_4 k_i^2 + \alpha_5 L_i k_i + u_{5i}$$

where y_i is the value added of aggregate output of manufacturing industry in each state in 1957, k_i is the value of capital services, and L_i is the labor input in terms of employment in each state in 1957. The CD and QP production functions were estimated by ordinary least squares while the other three were estimated by maximum likelihood.

Table 39 indicates the reduction in mean squared error which is achieved by using the minimum distance estimator for the capital and labor coefficients. A percent reduction approaching 100% implies an estimated variance approaching zero and suggests that fewer and fewer observations fall in the constrained region. The sensitivity of the reduction in estimated variance to the restriction appears to depend upon the size of the unconstrained estimate relative to its estimated variance. In this case slightly tighter restrictions resulted in substantial reductions in estimated variance. The degrees of freedom for the Ramsey-Zarembka regressions

TABLE 39

ESTIMATED VARIANCE, ABSOLUTE BIAS
AND MEAN SQUARED ERROR OF THE MDE
FOR THE RAMSEY-ZAREMBKA REGRESSIONS

	$\hat{\theta}_u$	Est. Var. $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m \geq$	Est. Var $\hat{\theta}_m$	Est. |Bias| $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
I. C-D								
Labor	.689	.003	13.2	0	.003	.000	.003	0%
			3.6	½	.003	.000	.003	0%
			- 6.0	1	.000	.021	.000	100%
			-25.2	2	.000	.021	.000	100%
Capital	.313	.003	5.7	0	.003	.000	.003	0%
			- 3.4	½	.000	.022	.000	100%
			-12.4	1	.000	.022	.000	100%
			-30.6	2	.000	.022	.000	100%
II. CES								
Labor	.096	.000	16.0	0	.000	.000	.000	0%
			-67.3	½	.000	.002	.000	100%
			-150.7	1	.000	.002	.000	100%
			-317.3	2	.000	.002	.000	100%
Capital	.657	.986	.7	0	.613	.150	.635	36%
			.2	½	.400	.323	.504	49%
			- .3	1	.204	.396	.361	63%
			- 1.4	2	.067	.396	.090	91%
III. VES								
Labor	1.152	.007	13.6	0	.007	.000	.007	0%
			7.7	½	.007	.000	.007	0%
			1.8	1	.007	.001	.007	6%
			-10.0	2	.001	.034	.000	100%
Capital	-.138	.007	- 1.6	0	.003	.034	.004	50%
			- 7.4	½	.001	.034	.000	100%
			-13.2	1	.001	.034	.000	100%
			-24.9	2	.001	.034	.000	100%

W. G

Labor

Capita

W. QP

Labor

Capit

ran

the

reg

te

The

tio

Vari

all

Capit

The co

TABLE 39 - Continued

	$\hat{\theta}_u$	Est. Var. $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m \geq$	Est. Var $\hat{\theta}_m$	Est. |Bias| $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
IV. GPF								
Labor	.678	.002	13.8	0	.002	.000	.002	0%
			3.6	½	.002	.000	.002	0%
			- 6.6	1	.000	.020	.000	100%
			-27.0	2	.000	.020	.000	100%
Capital	.300	.003	5.8	0	.003	.000	.003	0%
			- 3.8	½	.000	.021	.000	100%
			-13.5	1	.000	.021	.000	100%
			-32.7	2	.000	.021	.000	100%
V. QP								
Labor	.004	.000	6.7	0	.000	.000	.000	0%
			- 826.7	½	.000	.000	.000	100%
			-1660.0	1	.000	.000	.000	100%
			-3326.7	2	.000	.000	.000	100%
Capital	3.318	.687	4.0	0	.687	.000	.687	0%
			3.4	½	.687	.001	.687	0%
			2.8	1	.687	.002	.687	0%
			1.6	2	.626	.015	.627	8%

range from 43 for the quadratic production function to 47 for the constant elasticity of substitution production function resulting in a critical value of approximately -1.7 for the tests of the restrictions at the 5% level of significance. The results range from almost all of the proposed restrictions being rejected for the capital coefficient of the variable elasticity of substitution production function to all of the proposed restrictions being accepted for the capital coefficient of the quadratic production function. The coefficients whose test statistics showed the greatest

sensitivity to the ever more stringent restrictions had the smallest estimated variances. In particular, the estimated variances of the coefficients of labor in both the CES and QP models were very small resulting in test statistics that went from positive to quite large negative values. On the other hand, the estimated variances of the coefficients of capital in both the CES and QP models were quite large resulting in test statistics that did not change radically as more stringent restrictions were imposed. Just as in the Monte Carlo experiments of Chapter IV, the estimated variance tended to dominate the estimated bias in the calculation of estimated mean squared error. The maximum absolute bias again happens to correspond to the estimated absolute bias under the restriction $\hat{\theta}_m \geq 2$ in each case except for that relating to the capital coefficient for the quadratic production function where the maximum absolute bias is .331. In general, tighter restrictions result in a considerable reduction in variance especially where the minimum distance estimate is at the boundary of the constrained space.

Ferguson's (8, p.135) time-series study of a CES production function provides a convenient example of the effect of the minimum distance estimator on the estimated variance of the regression coefficient representing the estimated elasticity of substitution. The elasticity of substitution between labor and capital was estimated in the lumber, furniture, paper, chemicals, petroleum, metal l, and machinery industries with the equation:

Ind

Foo

Tob

App

Lum

Fur

Pap

Pri

Che

TABLE 40

ESTIMATED VARIANCE, ABSOLUTE BIAS AND MEAN SQUARED ERROR OF THE MDE FOR THE FERGUSON REGRESSIONS

Industry	$\hat{\theta}_u$	Est. Var. $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m \geq$	Est. Var $\hat{\theta}_m$	Est. $\lvert$Bias$\rvert$ $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
Food	.241	.040	1.2	0	.033	.008	.033	18%
			- 1.3	½	.001	.080	.004	89%
			- 3.8	1	.000	.080	.000	99%
			- 8.8	2	.000	.080	.000	100%
Tobacco	1.183	.212	2.6	0	.212	.010	.212	0%
			1.5	½	.191	.003	.191	10%
			.4	1	.107	.106	.118	44%
			- 1.8	2	.000	.184	.011	95%
Apparel	1.084	.026	6.8	0	.026	.000	.026	0%
			3.7	½	.026	.001	.026	0%
			.5	1	.014	.030	.015	40%
			- 5.7	2	.000	.064	.000	100%
Lumber	.905	.004	13.5	0	.004	.000	.004	0%
			6.0	½	.004	.000	.004	0%
			- 1.4	1	.000	.027	.000	91%
			-16.3	2	.000	.027	.000	100%
Furniture	1.123	.002	25.0	0	.002	.000	.002	0%
			13.8	½	.002	.000	.002	0%
			2.7	1	.002	.001	.002	0%
			-19.5	2	.000	.018	.000	100%
Paper	1.016	.004	16.9	0	.004	.000	.004	0%
			8.6	½	.004	.000	.004	0%
			.3	1	.002	.017	.002	47%
			-16.4	2	.000	.024	.000	100%
Printing	1.147	.096	3.7	0	.096	.002	.096	0%
			2.1	½	.096	.006	.096	0%
			.5	1	.052	.063	.056	42%
			- 2.8	2	.000	.124	.001	99%
Chemicals	1.248	.005	17.3	0	.005	.000	.005	0%
			10.4	½	.005	.000	.005	0%
			3.4	1	.005	.001	.005	0%
			-10.4	2	.000	.029	.000	100%

Ind

Pet

Rub

Lea

Gla

Met

Met

Mac

Ele

Tra

TABLE 40 - Continued

Industry	$\hat{\theta}_u$	Est. Var. $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m \geq$	Est. Var $\hat{\theta}_m$	Est. |Bias| $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
Petroleum	1.300	.022	8.7	0	.022	.000	.022	0%
			5.4	½	.022	.000	.022	0%
			2.0	1	.022	.003	.022	0%
			- 4.7	2	.000	.059	.000	100%
Rubber	.759	.314	1.4	0	.274	.007	.274	12%
			.5	½	.167	.116	.181	42%
			- .4	1	.057	.223	.107	66%
			- 2.2	2	.000	.223	.010	97%
Leather	.865	.020	6.2	0	.020	.000	.020	0%
			2.6	½	.020	.003	.020	0%
			- 1.0	1	.000	.056	.004	82%
			- 8.1	2	.000	.056	.000	100%
Glass	.666	.221	1.4	0	.197	.002	.197	11%
			.4	½	.107	.115	.120	45%
			- .7	1	.020	.188	.055	75%
			- 2.8	2	.000	.188	.003	99%
Metal 1	1.200	.011	11.4	0	.011	.000	.011	0%
			6.7	½	.011	.000	.011	0%
			1.9	1	.011	.003	.011	0%
			- 7.6	2	.000	.042	.000	100%
Metal 2	.926	.068	3.6	0	.068	.002	.068	0%
			1.6	½	.064	.001	.064	6%
			- .3	1	.016	.104	.026	61%
			- 4.1	2	.000	.104	.000	100%
Machinery	1.041	.002	25.4	0	.002	.000	.002	0%
			13.2	½	.002	.000	.002	0%
			1.0	1	.001	.003	.001	24%
			-23.4	2	.000	.016	.000	100%
Electrical	.643	.130	1.8	0	.126	.004	.126	0%
			.4	½	.066	.083	.072	44%
			- 1.0	1	.002	.144	.022	83%
			- 3.8	2	.000	.144	.000	100%
Transport	.237	.314	.4	0	.162	.124	.178	43%
			- .5	½	.052	.223	.102	68%
			- 1.4	1	.000	.223	.032	90%
			- 3.1	2	.000	.223	.002	99%

TABLE 40 - Continued

Industry	$\hat{\theta}_u$	Est. Var. $\hat{\theta}_u$	Test Stat.	$\hat{\theta}_m$	Est. Var. $\hat{\theta}_m$	Est. |Bias| $\hat{\theta}_m$	Est. MSE $\hat{\theta}_m$	MSE Reduction
Instruments	.763	.084	2.6	0	.084	.006	.084	0%
			.9	½	.060	.026	.061	27%
			- .8	1	.005	.116	.018	78%
			- 4.3	2	.000	.116	.000	100%
Textiles	1.104	.194	2.5	0	.194	.010	.194	0%
			1.4	½	.169	.008	.169	13%
			.2	1	.085	.128	.101	48%
			- 2.0	2	.000	.176	.007	97%

$$\log v = a + b_1 \log w + u$$

where v is value added per man-year, w is the real wage rate, and b_1 is an estimate of the elasticity of substitution between labor and capital. The other 12 industries had estimates based on the equation:

$$\log v = a + b_1 \log w + b_2 t + u$$

where t is an index of time. The observations were for the years 1949 through 1961 except for rubber, glass, and metal 1 which only covered up through 1958 because of material changes in industry designations for those three industries. The equations for all 19 industries were estimated by ordinary least squares. The rubber, glass, and metal 1 equations have about seven degrees of freedom with a critical value for the tests of approximately -1.9 at the 5% level.

All the other equations have about ten degrees of freedom with a critical value of about -1.8. None of the coefficients in the Ferguson regressions violated the restriction that the coefficients be greater than zero. Consequently, this restriction was accepted at the 5% level in the tests for every industry. The restriction that the coefficients be greater than one-half was violated only by the food and transport industries but not by a sufficient amount for the restriction to be rejected. Nine industries had coefficients that violated the restriction that they be greater than one although only in the case of the food industry was this violation substantial enough to cause the restriction to be rejected at the 5% level. The coefficients of all 19 industries violated the restriction that they be greater than two and the restriction was rejected by all 19 industries at the 5% level although the rejection was marginal for the tobacco industry. Consequently, it is not surprising that the percentage reduction in mean squared error approached 100% for the restriction that the coefficients be greater than two. The estimated absolute bias under the restriction $\hat{\theta}_m \geq 2$ is the same as the maximum absolute bias in each case. In virtually every case the estimated bias was offset by the substantial reduction in estimated variance brought about by using the minimum distance estimation technique.

The above examples demonstrate that the minimum distance estimator can be used to impose a priori inequality restrictions that result in a substantial reduction in

estimated variance although, as expected, this effect is greatest when the restrictions are violated by the unrestricted estimates. This result confirms the usefulness and effectiveness of the minimum distance estimator in incorporating inequality restrictions in the estimation procedures as demonstrated by the Monte Carlo experiment in Chapter IV.

Conclusion

This dissertation has emphasized the importance of using all information both *a priori* and empirical in the estimation process, in particular when the *a priori* information is in the form of inequality restrictions. Procedures for testing for the appropriateness of inequality restrictions have been developed. The distributional properties of the minimum distance estimator have been compared to those of the unconstrained maximum likelihood estimator and the constrained maximum likelihood estimator in extensive Monte Carlo sampling experiments. It has been shown that the minimum distance estimator is greatly superior to the unconstrained maximum likelihood estimator in terms of various definitions of mean squared error. The equivalence of the constrained maximum likelihood estimator and the minimum distance estimator in a number of circumstances has been indicated by noting that in general there has not been a substantial difference between the distributional properties of the minimum distance estimator and those of the constrained maximum likelihood estimator. A single sample approximation of the variance, absolute bias, and mean squared error of the minimum distance estimator has been developed and many examples of applying the minimum distance estimator in regression analysis have been presented indicating the effect of increasingly stringent inequality restrictions. Thus, in general, the usefulness and effectiveness of the minimum distance estimator have been demonstrated.

LIST OF REFERENCES

LIST OF REFERENCES

1. Bancroft, T. A. "On Biases in Estimation Due to the Use of Preliminary Tests of Significance." The Annals of Mathematical Statistics 15 (1944): 190-204.

2. Berman, Simeon M. Mathematical Statistics. Scranton: International Textbook Company, 1971.

3. Brown, Murray, ed. The Theory and Empirical Analysis of Production. New York: National Bureau of Economic Research, 1967.

4. Chernoff, Herman and Moses, Lincoln E. Elementary Decision Theory. New York: John Wiley & Sons, 1959.

5. Chipman, J. S. and Rao, M. M. "The Treatment of Linear Restrictions in Regression Analysis." Econometrica 32 (1964): 198-209.

6. Chow, G. C. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." Econometrica 28 (1960): 591-605.

7. Cramer Harald. Mathematical Methods of Statistics. Princeton: Princeton University Press, 1946.

8. Ferguson, C. E. "Time Series Production Functions and Technological Progress in American Manufacturing Industry." J. of Political Economy, April 1965: 135-147.

9. Ferguson, Thomas S. Mathematical Statistics, A Decision Theoretic Approach. New York: Academic Press, Inc., 1967.

10. Fisher, F. M. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note." Econometrica 38 (1970): 361-66.

11. Fisz, Marek. Probability Theory and Mathematical Statistics. New York: John Wiley & Sons, Inc., 1963.

12. Freund, John E. Mathematical Statistics. 2d ed. Englewood Cliffs: Prentice-Hall, Inc., 1971.

13. Graybill, Franklin A. An Introduction to Linear Statistical Models. Vol. 1. New York: McGraw-Hill Book Co., Inc., 1961.

14. Hodges, J. L. and Lehmann, E. L. "Testing the Approximate Validity of Statistical Hypotheses." Journal of the Royal Statistical Society, Series B, Vol. 16 (1954): 261-268.

15. Hoel, Paul G.; Port, Sidney C.; and Stone, Charles J. Introduction to Statistical Theory. Boston: Houghton Mifflin Co., 1971.

16. Hogg, Robert V. and Craig, Allen T. Introduction to Mathematical Statistics. 3rd ed. New York: Macmillan Publishing Co., Inc., 1970.

17. Hussain, Muhammad. On Inequality Constraints in Regression Analysis. Ph.D. dissertation, George Washington University, 1972.

18. Johnson, Norman L. and Kotz, Samuel. Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley & Sons, Inc., 1972.

19. Judge, G. G. and Takayama, T. "Inequality Restrictions in Regression Analysis." Journal of the American Statistical Association 61 (1966): 166-181.

20. Kendall, Maurice G. and Stuart, Alan. The Advanced Theory of Statistics. Vol. 2. New York: Hafner Publishing Co., 1961.

21. Kendrick, John W. Long Term Economic Growth 1860-1970. Bureau of Economic Analysis, Social and Economic Statistics Administration, U. S. Department of Commerce, 1973.

22. Kmenta, Jan. Elements of Econometrics. New York: Macmillan Publishing Co., Inc., 1971.

23. Kmenta, Jan. "On Estimation of the CES Production Function," University of Wisconsin Social Systems Research Institute Series, no. 6410, October 1964.

24. Lehmann, E. L. Testing Statistical Hypotheses. New York: John Wiley & Sons, Inc., 1959.

25. Lovell, Michael, and Prescott, Edward. "Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency." Journal of the American Statistical Association 65 (1970): 913-25.

26. Meyer, Paul L. Introductory Probability and Statistical Applications. 2d ed. Reading, Mass.: Addison-Wesley Publishing Co., 1970.

27. Mood, Alexander M. and Graybill, Franklin A. Introduction to the Theory of Statistics. 2d ed. New York: McGraw-Hill Book Co., 1963.

28. Moran, P. A. P. An Introduction to Probability Theory. London: Oxford University Press, 1968.

29. Penneck, Stephen J. Maximum Likelihood Estimation Using Inequality Constraints. Master's thesis, University of Birmingham, 1974.

30. Ramsey, James B. "The Maximum Likelihood Estimation of Parameters Contained Within Finitely Bounded Compact Sets: Some Preliminary Results." Mimeographed. East Lansing, Mich.: Michigan State University, 1974.

31. Ramsey, James B. "Mixtures of Distributions and Maximum Likelihood Estimation of Parameters Contained in Finitely Bounded Compact Spaces." Mimeographed. East Lansing, Mich.: Michigan State University, 1975.

32. Ramsey, James B. and Zarembka, Paul. "Specification Error Tests and Alternative Functional Forms of the Aggregate Production Function." Journal of the American Statistical Association 66 (1971): 471-77.

33. Rao, C. Radhakrishna. Linear Statistical Inference and Its Applications. New York: John Wiley & Sons, Inc., 1965.

34. Theil, Henri. Principles of Econometrics. New York: John Wiley & Sons, Inc., 1971.

35. Theil, Henri and Goldberger, A. "On Pure and Mixed Statistical Estimation in Economics," International Economic Review 2: 65-78.

36. Theil, Henri, and Van de Panne, Management Science (1960): 1-20.

37. Toro, Carlos and Wallace, T. D. "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression." Journal of the American Statistical Association 63 (1968): 558-572.

38. Wallace, T. D. "Weaker Criteria and Tests for Linear Restrictions in Regression." Econometrica 40 (1972): 689-98.

39. Wallace, T. D. and Anderson, R. L. "Notes on Exact Linear Restrictions and the Generalized Linear Hypothesis." Unpublished paper.

40. Wilks, S. S. Mathematical Statistics. Princeton: Princeton University Press, 1947.

41. Zellner, Arnold. "Linear Regression with Inequality Constraints on the Coefficients." Mimeographed Report 6109 of the International Center for Management Science, 1961.