MULTIPLE COMPARISONS WITH THE BEST FOR INFERENCE IN STOCHASTIC FRONTIER MODELS

AND

SUBMODEL ESTIMATION FOR COUNTERFACTUAL POLICY ANALYSIS

By

William Clinton Horrace

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Economics

1996

ABSTRACT

MULTIPLE COMPARISONS WITH THE BEST FOR INFERENCE IN STOCHASTIC FRONTIER MODELS AND SUBMODEL ESTIMATION FOR COUNTERFACTUAL POLICY ANALYSIS

By

William Clinton Horrace

This is a dissertation in three chapters. In the first chapter we examine a statistical method for performing simultaneous inference on all distances from the "best" called *multiple comparisons with the best* or simply MCB. We find that MCB can be used on a fixed effects stochastic production frontier model for panel data to construct simultaneous confidence intervals for technical inefficiency and to perform inference on maximum efficiency measures, where previously no methods had been suggested.

In the second chapter we use the MCB analysis of chapter one to perform inference and point-estimation on some previously analyzed stochastic frontier data and compare these results to the inference and point estimates currently suggested but never exploited in the stochastic frontier literature.

The third chapter is a complete departure from the first two. In chapter three we construct a subset limited-information maximum likelihood estimator for a vector error correction model under the cointegration hypothesis for use in a counterfactual policy analysis model.

Dedicated to my children: Ava Deleon Horrace and Ian William Horrace

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Professor Peter Schmidt and the rest of my committee: Professor Ching-Fan Chung, Professor Robert Rasche and Professor Jeffrey Wooldridge. I would also like to thank the following individuals for their support: Ava, Ian, Sarah, Diana, Zoe, Brian, Bobby, Vaishali, Ren, Stimpy, Homer, Marge, Maggie, Lisa, Bart and the staff at the Peanut Barrel. Also special thanks to Professor Carl Davidson.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION TO THE DISSERTATION

This is a dissertation in three chapters. It is an attempt to address topics in econometric theory and applied econometrics in the cross sectional, panel data and time series contexts. Each chapter contains its own introductory remarks, so this introduction to the entire text is intended to briefly summarize the goals of this dissertation. The plan of the dissertation is as follows.

Chapters one and two are concerned with performing inference on technical efficiency estimates in stochastic frontiers models. These are cross sectional and panel data models that predict or estimate technical efficiency for a set of productive or decision-making units. Two different approaches dominate the efficiency measurement literature: the aforementioned stochastic frontiers approach and a deterministic approach. While the debate over which approach is "preferred" continues, a clear advantage of the stochastic frontiers approach is that it allows quantification of the uncertainty associated with the efficiency estimates while the deterministic approach does not. Therefore the import of the first two chapters of this dissertation is that they detail procedures to quantify this uncertainty in the stochastic frontiers model. This is accomplished through confidence intervals construction.

The first chapter details a new technique for confidence intervals construction in the fixed effects stochastic frontier model for panel data. To my knowledge this has heretofore never been accomplished. This technique, called Multiple Comparisons with the Best or MCB, allows simple confidence intervals construction that not only quantifies the uncertainty of the individual technical efficiency estimates of the productive units but also quantifies the uncertainty of which firm in the sample is the best in the population

and suggests point estimates for technical efficiency which have less positive bias than those currently exploited in the literature. Moreover, the fixed effects formulation of the stochastic frontier model requires weaker distributional assumptions than any other stochastic formulation, so the benefits of these MCB intervals are clear.

The second chapter discusses several other formulations of the stochastic frontier model. Unfortunately, all of these formulations require stronger distributional assumptions than the aforementioned fixed effects model, but the ability to construct confidence intervals for their technical efficiency estimates has been well documented. Strangely, this ability has never been systematically exploited in an empirical setting. Therefore, chapter two details the interval construction techniques of these formulations and, along with the MCB techniques of chapter one, presents a comprehensive empirical study in which confidence intervals on technical efficiency measures are constructed using various estimation and interval construction methods on three different data sets. In doing so, chapter two advances our understanding of the various sources of uncertainty inherent in econometric models for efficiency estimation.

Chapter three is a complete departure from the first two. This chapter presents a new method for estimating the parameters in a conditional submodel for counterfactual policy analysis under the cointegration hypothesis. Counterfactual analysis attempts to analyze an economy (system of equations) in which new (counterfactual) policy rules have been substituted for the historical policy rules which generated the data. The idea is to see how the economy would have behaved had a different policy regime been in effect. However, as we shall see, current counterfactual analysis techniques are limited

in their scope, so chapter three suggests a technique that can be more universally applicable in these types of analyses.

Finally, in the conclusion section of this dissertation I summarize the results of the research and suggest areas for additional work.

CHAPTER 1

MULTIPLE COMPARISONS WITH THE BEST AND THE FIXED EFFECTS MODEL

1 INTRODUCTION

It is often the case in empirical research that comparative studies are prescribed. For instance, one may wish to compare the effectiveness of several drugs in the treatment of a disease, or one may be interested in comparing the differences in crop yield for a variety of fertilizers. Such experiments might involve collecting a sample of some effectiveness or yield measure for each treatment or variety, calculating a summary statistic for each treatment (such as a sample mean) to estimate some population parameter, and then comparing these statistics using some inference techniques like an F-test or t-tests to test hypotheses on comparisons between the population parameters.

As an example consider a controlled experiment where 3 different fertilizers (A, B, and C) are each applied in the same quantity to 10 separate samples of the same soil, all receiving the same seed, sunlight and irrigation. At the end of the experiment crop output is measured for each of the 30 soil samples, and mean output is calculated for each of the 3 fertilizers. The three sample means are construed as estimates of population means, so that comparative hypotheses on the populations can be tested. For instance a typical comparative hypothesis would be that fertilizer A is better than fertilizer B or, perhaps, that the difference in crop yield between fertilizer A and fertilizer C is 20 units.

Very often in these studies it is advantageous to test a hypothesis about several comparisons between populations, *simultaneously* (i.e. a multiple hypothesis). To continue our crop example, one may be interested in testing the hypothesis that fertilizer A is better then fertilizer B *and* fertilizer A is better than fertilizer C. In these instances the advantages of *multiple comparison procedures* (or MCP) over single or per comparison testing have been well documented in the statistical literature. Basically, MCP precludes what has been deemed the *multiplicity effect*. Hochberg and Tamhane (1987) give an excellent exposition on the multiplicity effect and the justification for MCP.

This chapter deals with a specific case of MCP called *multiple comparisons with the best* (or MCB) in which simultaneous inference is performed on all differences from the *unknown* "best" population parameter. This procedure has been extensively exploited in various forms in the natural science and statistics literatures to allow ranking and selection of treatments. For example, see Becker (1961), Dalal and Srinivasan (1977), Gupta and Hsu (1977) and McDonald (1977). While generally ignored in the econometrics literature, MCB methods can be use to perform inference in stochastic production frontier models for panel data. These are econometric panel data models that produce estimates of technical efficiency for a set of firms or productive units. Our focus is the fixed effects estimate which produces a distinct intercept term for each firm. The firm with the highest intercept estimate is deemed "best", and differences between the best firm's estimate and those of the less efficient firms are technical *inefficiencies*. See Schmidt and Sickles (1984). These inefficiency estimates can be thought of as comparisons with the best. If we are interested in testing the hypothesis that the firm

with the highest efficiency estimate is, in fact, the true best, then we must test the simultaneous hypothesis that all other firms have positive technical inefficiencies. Therefore, our inference is necessarily simultaneous and constitutes multiple comparisons with the best or MCB. Once MCB intervals are calculated their midpoints suggest point-estimates of technical inefficiency and maximum efficiency which have less positive bias than those currently exploited in the stochastic frontier literature.

It is important to stress the fact that this chapter deals with inference on the maximal efficiency and on technical inefficiencies measured relative to this maximum. Therefore, confidence intervals and inference on the technical inefficiencies are necessarily complex due to the non-linearity associated with taking the maximum of the individual efficiencies. This chapter uses a pre-existing theoretical methodology to simplify construction of these intervals. It is also important to realize that determination of the most efficient firm should *not* be approached as simple selection of the firm with the largest technical efficiency estimate. This naive approach assumes that the estimation procedure reveals with certainty the true efficiency rankings of the firms. Conversely, *simultaneous* MCB inference on the technical inefficiency measures allows us to determine a confidence level with which we can say that the firm with the highest technical efficiency estimate is the true "best".

Therefore, the importance of these procedures is that they allow us to *make confidence statements about the efficiency estimates relative to an unknown standard and, subsequently, a statement about the maximum efficiency estimate's ability to serve as that standard.* For practical purposes these inefficiency estimates, their intervals and any inference on the efficient firm are measures of performance or success, and as such can

aid in technological decisions and the evaluation of managerial performance. Additionally, studying these intervals will provide insight into the sources of uncertainty associated with productivity estimation.

This chapter is concerned with performing MCB in the fixed effects stochastic frontier model. It is organized as follows. Section 2 briefly discusses estimation of the stochastic frontier model and summaries the main algebraic and statistical results of a fixed effects treatment. After a very brief historical account, section 3 explores developments in MCB theory. Section 4 discusses theoretical considerations in applying MCB theory to the stochastic frontier model. Finally, section 5 draws some conclusions and introduces areas for additional research.

2 STOCHASTIC FRONTIER MODELS

2.1 Introduction

Stochastic frontier models were originally due to Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). These models were based on cross sectional data and strong distributional assumptions. Similar models have also been developed for panel data. Pitt and Lee (1981) and Schmidt and Sickles (1984) were the first to exploit the advantages of a panel data over cross sectional data. Since this is not intended to be a comprehensive survey, the reader is referred to Cornwell and Schmidt (1995), Greene (1995), Lovell (1993), Lovell and Schmidt (1988) and Schmidt (1985) for further details.

This chapter deals with these models only in the context of panel data. Specifically, only the fixed effects formulation of the stochastic frontier model will be

detailed. However, we now present a more general discussion of the model to incorporate alternative formulations to be used in chapter two. The basic model that we will consider is as follows.

(1.1) $$y_{it} = \alpha + x_{it}\beta + v_{it} - u_i, \quad u_i \geq 0; \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

Here i indexes firms (or other productive units) and t indexes time periods. Typically y_{it} is the logarithm of output and x_{it} is a vector of inputs or functions of inputs. v_{it} is statistical noise and $u_i \geq 0$ represents technical inefficiency, assumed to be time invariant. More specifically, if y_{it} is the logarithm of output, technical efficiency of the i^{th} firm is $TE_i = \exp(-u_i)$ and technical inefficiency is $1-TE_i$. We will refer to the composite error as $\epsilon_{it} = v_{it} - u_i$. We will always assume the following:

(A.1) The v_{it} are iid $N(0, \sigma^2_v)$.

(A.2) x_{it} and v_{js} are independent for $t, s = 1, \ldots, T, \quad i, j = 1, \ldots, N.$

We will sometimes but not always make the additional assumptions:

(A.3) The u_i are independent of x and v.

(A.4) $u_i = | U_i |$, where the U_i are iid $N(0, \sigma^2_u)$

Assumption (A.4) implies that the u_i are half-normal, but this assumption could be replaced by other specific distributional assumptions, as in Stevenson (1980) or Greene (1990).

Now define $\alpha_i = \alpha - u_i$, so that $\alpha_i \leq \alpha$ for all i. Then we can rewrite (1.1) as the usual panel data model

$$(1.2) \qquad y_{it} = \alpha_i + x_{it}\beta + v_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T.$$

We regard zero as the absolute minimal value of u_i, and hence α as the absolute maximal value of α_i, over any possible sample (essentially, as $N \to \infty$). This can be distinguished from the minimal value of u_i and the maximal value of α_i in a given sample of size N, and this distinction is relevant when N is small and the u_i (hence α_i) are treated as fixed. Let $\alpha_{[1]} \leq \alpha_{[2]} \leq \ldots \leq \alpha_{[N]}$ be the population rankings of the α_i, so $\alpha_{[N]} = \max_{i=1}^{N} \alpha_i$, and $\alpha_{[N]} \leq \alpha$. Similarly, let $u_{[N]} \leq u_{[N-1]} \leq \ldots \leq u_{[1]}$ be the population rankings of the u_i, so $u_{[N]} = \min_{i=1}^{N} u_i$, and $u_{[N]} \geq 0$. Then $u_{[i]} = \alpha - \alpha_{[i]}$. In this case the technical efficiency measures u_i are defined by comparing α_i to the absolute standard α. We can consider the alternative of comparing α_i to the within sample standard $\alpha_{[N]}$. Define $u^*_i = \alpha_{[N]} - \alpha_i = u_i - u_{[N]}$, so that $0 \leq u^*_i \leq u_i$. Then equation (1.2) can be rewritten as:

$$(1.3) \qquad y_{it} = \alpha_{[N]} + x_{it}\beta + v_{it} - u^*_i, \quad i = 1, \ldots, N, \qquad t = 1, \ldots, T.$$

The difference between the two definitions of u is substantive and will be considered further in chapter two.

Equation (1.2) is useful primarily as a basis for estimation that treats the α_i (or u_i) as fixed. A fixed effects treatment may be useful because it relies only on assumptions (A.1) and (A.2), not (A.3) and (A.4), and because it is applicable when N is small and T is large (as well as when N is large). Suppose we estimate (1.2) by the usual fixed effects estimation involving the within transformation (or, equivalently, dummy variables for firms), yielding estimates of $\alpha_1, \ldots, \alpha_N$, β and σ^2_v. Define

$$\hat{\alpha} = \max_{j=1}^{N} \hat{\alpha}_j \tag{1.4}$$

$$\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i \qquad i = 1, \ldots, N$$

Then, as T→∞ with N fixed, $\hat{\alpha}_i \to \alpha_i$, $\hat{\alpha} \to \alpha_{[N]}$ and $\hat{u}_i \to u^*_i = \alpha_{[N]} - \alpha_i$, so that $\hat{u}_i$ measures inefficiency relative to the standard of the best firm in the sample. Now consider what happens as N→∞. Under the assumption (A.4) of half-normality, or in fact under any mechanism for the generation of u_i that allows u arbitrarily close to zero with positive probability (density), $u_{[N]} \to 0$ and $\alpha_{[N]} \to \alpha$ as N→∞. Thus, $\hat{\alpha} \to \alpha$ and $\hat{u}_i \to u_i$ as both N and T→∞, so that inefficiency is measured relative to its absolute (not just within-sample) standard. This distinction will be important in the empirical analyses in chapter two.

The statistical properties of the estimated u_i are complicated because of the "max" operation involved in the definition of $\hat{\alpha}$ and therefore of $\hat{u}_i$. Consistency as both N and T → ∞ was argued heuristically (as above) by Schmidt and Sickles (1984). Park and

Simar (1994) and Kneip and Simar (1995) established the rate of convergence of the estimates. However, the asymptotic distributions of the estimates of α and the u_i are unknown, so that standard methods of construction of asymptotically-valid confidence intervals based on these asymptotic distributions are currently not possible. Additionally, the estimate $\hat{\alpha}$ is essentially based on the presumption that we are certain that our estimates of the α_i reveal the true ranking of the corresponding population parameters. Therefore any statistical inference or interval construction must be conditional on that certainty, which seems dubious. Despite these problems, MCB methods allow construction of confidence intervals and will be discussed in the next section. To this end, we examine the covariance structure of the $\hat{\alpha}_i$ conditional on x_i.

$$(1.5) \qquad \begin{aligned} \mathrm{Var}(\hat{\alpha}_i) &= \sigma^2/T + \bar{x}_i \mathrm{Var}(\hat{\beta}) \bar{x}_i' \qquad i = 1, \ldots, N \\ \mathrm{Cov}(\hat{\alpha}_i, \hat{\alpha}_j) &= \bar{x}_i \mathrm{Var}(\hat{\beta}) \bar{x}_j' \qquad i \neq j \end{aligned}$$

It is important to note in equation (1.5) that the consistency of $\hat{\beta}$ implies that as either $N \to \infty$ or $T \to \infty$, $\mathrm{Cov}(\hat{\alpha}_i, \hat{\alpha}_j) \to 0$, and $\mathrm{Var}(\hat{\alpha}_i) \to \sigma^2/T$ or 0 respectively, implying that asymptotically our estimates of α_i are orthogonal. This orthogonality greatly simplifies MCB analysis, so, as we will see, *MCB methods are most readily applied when N or T is large.*

2.2 Conclusions and Extensions

The previous fixed effects analysis is for a *balanced* panel where the number of periods of observation for each individual or firm is the same, T. Generalization to the

unbalanced case is straightforward and will not be discussed here. However, for the purposes of this chapter the generalization changes the variance of $\hat{\alpha}_i$ to:

$$(1.6) \qquad \mathrm{Var}(\hat{\alpha}_i) = \sigma^2/T_i + \bar{x}_i \mathrm{Var}(\hat{\beta}) \bar{x}_i' \qquad i = 1, \ldots, N$$

where: T_i is the number of periods for the i^{th} firm.

Of course the $\mathrm{Var}(\hat{\beta})$ is slightly different in the unbalanced case, and all the previous results hold with some minor algebraic modifications as well. Implications of the unbalanced panel to MCB will be discussed in subsequent sections. The following section is devoted to MCB procedures.

3 MULTIPLE COMPARISONS WITH THE BEST

3.1 Introduction

Multiple Comparisons with the Best (MCB) is a specific case of Multiple Comparison Procedures (MCP) which is performance of simultaneous inferences or construction of simultaneous confidence intervals in comparative analysis. MCB is a procedure for constructing simultaneous confidence intervals of the form:

$$(1.7) \qquad u^*_i = \alpha_{[N]} - \alpha_i \qquad i = 1, \ldots, N$$

where the α_i are *unknown* "goodness" parameters for N populations, and the population with the largest α_i is considered "best". It is important to realize that the best population is unknown. If this were not the case MCB procedures would not be required; some

multiple comparison or single comparison technique would probably suffice, depending on the experimenter's requirements. Even then they would typically be based on the estimate $\hat{\alpha}_{[N]} = \max_{j=1}^{N} \hat{\alpha}_j$ which assumes that estimation reveals the true ranking of α_i. The MCB intervals may also suggest consistent point-estimates with nice small sample properties.

MCB is similar to and can be adapted from Multiple Comparisons with a Control (MCC) procedures which construct simultaneous confidence intervals of the form:

$$(1.8) \qquad \alpha_N - \alpha_i \qquad i = 1, \ldots, (N-1)$$

where the N^{th} population is regarded as a control. Any population may be chosen for the control, but the choice must be independent of the data.

This section is intended to serve as in introduction to procedures for performing inference on equations (1.7) and (1.8). What follows is a brief historical survey of MCB procedures. Sections 3.3 - 3.5 outline the main MCB results. Finally, section 3.6 extends sections 3.3 - 3.5 to a more general case and draws some conclusions on the state of MCB literature as it pertains to stochastic frontier models.

3.2 Historical Perspective

MCP theory evolved during the late 1940s and early 1950s primarily due to David Duncan, S. N. Roy, Henry Scheffé and John Tukey. Harter (1980) gives a complete historical account. Shortly thereafter a related body of literature called ranking and selection surfaced with the work of Bechhofer (1954). Additional ranking procedures

followed due to Gupta (1956, 1965), Fabian (1962) and Desu (1970). MCC procedures were primarily due to Dunnett (1955, 1964). MCB evolved in the early 1980's with the work of Jason Hsu. The primary justification for MCB is that it not only allows significance testing and interval estimation for differences between populations, but it allows the experimenter to simultaneously determine which population is "best".

Hsu (1981) constructed parametric and non-parametric simultaneous one-sided upper confidence intervals for equation (1.7) under a location model. The parametric confidence intervals were stronger than those suggested by Bechhofer, Gupta, Fabian and Desu, while the non-parametric intervals were new to the literature.[1] Later Hsu (1984) constructed simultaneous two-sided MCB confidence intervals for equation (1.7) which implied his 1981 results. Additionally, Edwards and Hsu (1983) provide a general technique for adapting MCC intervals in equation (1.8) to MCB intervals in equation (1.7). Hochberg and Tamhane (1987) nicely summarize the main results of these three papers. The next section recaps those results that are germane to this chapter and provides a few additional insights.

3.3 Hsu (1981) - Multiple Comparisons with the Best

Let $\pi_1, \ldots, \pi_N$ be N *independent* populations or treatments. Let the distributions of the N populations differ only by location, so for $i = 1, \ldots, N$, $F(x - \alpha_i)$ is the distribution of X in π. Let $X_{i1}, \ldots, X_{iT}$ be an i.i.d. random sample of size T from

---

[1] Hsu's intervals are stronger in several senses. See Hsu (1981) for specific comparisons with Bechhofer (1954), Gupta (1956, 1965), Fabian (1962) and Desu (1970).

population π_i. Let the "best" population be that with the largest location parameter, α_i. Let

$$\alpha_{[1]} \leq \ldots \leq \alpha_{[N]}$$

be the ordered location parameters. Let $F(x) = \Phi(x/\sigma)$, a normal distribution with mean 0 and standard deviation σ, unknown. Let

$$\hat{\alpha}_1, \ldots, \hat{\alpha}_N$$

be the sample means of the N populations, so that $E[\hat{\alpha}_i] = \alpha_i$ for all i. Let s be the usual pooled estimate of σ with $\upsilon = N(T-1)$ degrees of freedom. Then $var(\hat{\alpha}_i) = \sigma^2/T$ and the independence of the populations implies $cov(\hat{\alpha}_i,\hat{\alpha}_i) = 0$ for all $i \neq j$. Simple calculations show that $var(\hat{\alpha}_k - \hat{\alpha}_i) = 2\sigma^2/T$ and $cov(\hat{\alpha}_k - \hat{\alpha}_i, \hat{\alpha}_k - \hat{\alpha}_j) = \sigma^2/T$ for all $i \neq j \neq k$, so that $corr(\hat{\alpha}_k - \hat{\alpha}_i, \hat{\alpha}_k - \hat{\alpha}_j) = ½$, $i \neq j \neq k$. To perform inference on equation (1.7) select

$$(1.9) \qquad \hat{\gamma}_i = \hat{\alpha}_{[N]} - \hat{\alpha}_i \qquad i = 1, \ldots,N; \quad i \neq [N],$$

where $\hat{\alpha}_{[N]}$ is any of the $\hat{\alpha}_i$. Then

$$corr(\hat{\gamma}_i, \hat{\gamma}_j) = \rho_{ij} = \rho = ½ \qquad i \neq j.$$

Notice that this correlation structure will emerge *regardless of which of the N sample means we select as* $\hat{\alpha}_{[N]}$, provided that the choice does not depend on the data. That is,

the variance and covariances of the $\hat{\alpha}_i$ are not conditional on the rankings. This equicorrelated structure facilitates calculation and tabulation of the necessary multivariate critical values. Based on this structure, one-sided intervals are given by Theorem 3.1.

Theorem 3.1 (Hsu 1981).

Simultaneous $(1 - \lambda)$ confidence intervals for $\alpha_{[N]} - \alpha_i$ are given by:

$$\alpha_{[N]} - \alpha_i \in \{0, \max(\max_{j \neq i} \hat{\alpha}_j - \hat{\alpha}_i + d, 0)\}$$

where $d = T^{(\lambda)}_{N-1, \nu, \frac{1}{2}} s(2/T)^{\frac{1}{2}}$

and $T^{(\lambda)}_{N-1,\nu,\rho}$ is the solution in t for

$$\int_0^\infty \int_{-\infty}^\infty \Phi^{N-1}[(z\rho^{\frac{1}{2}} + ts)(1-\rho)^{-\frac{1}{2}}] d\Phi(z) dQ_\nu(s) = 1 - \lambda$$

where Q_ν is the distribution of a $\chi_\nu \nu^{-\frac{1}{2}}$ random variable.

For a proof see Hsu (1981). Clearly, the critical point and hence the inference hinges on the $\hat{\gamma}_i$'s being equicorrelated with correlation coefficient $\rho = \frac{1}{2}$. This means that for more general correlation structures these intervals are less useful, but for the purposes of this chapter they are instructive. Tables for the critical value $T^{(\lambda)}_{N-1,\nu,\rho}$, can be found in Hochberg and Tamhane (1987), Bechhofer and Dunnett (1986) and Gupta, Panchapakesan and Sohn (1985). $T^{(\lambda)}_{N-1,\nu,\rho}$ will be positive for values of $\lambda < 0.50$, so

for reasonable confidence levels, d will be positive. Since the lower bound is constrained positive, these intervals are not as informative as the two-sided intervals discussed later. However, they are of particular interest to those concerned with ranking and selection of the treatments. Therefore, we now discuss this theorem in terms of the inference implied by the intervals and defer any discussion of the uncertainty associated with our estimates until we discuss two-sided intervals in section 3.5.

The constructed intervals imply statements about the ranking of the treatment means and subsequently about the most efficient treatment mean. Notice that the upper bound is constrained non-negative, so for any treatment to be in contention for the "best" it is necessary that its upper bound be zero. We can make this point clear by examining specific cases. If the difference between the largest and second largest sample means is large relative to d, then the treatment with the largest sample mean will have an upper bound of zero and can be considered "best" at an $1-\lambda$ confidence level. Conversely, if this difference is small relative to d, then the upper bound will be greater than zero and the treatment is not "best". In fact, it is easy to show that the upper bounds for the remaining treatments (i = [1],..., [N-1]) will *always* be greater than zero, so, in this case, none of the treatments is "best".

If we are only concerned with the hypothesis that the treatment with the largest sample mean is best, then the notion of the remaining N-1 treatments always being less than best implies that we only need test the treatment with the largest sample mean to perform this inference. What distinguishes this inference from a single t-test is the selection of the critical value, which is drawn from a multivariate-t distribution to

account for the uncertainty associated with the best treatment mean being unknown and to account for the simultaneity associated with the ranking.

If our inference fails to produce a single "best" treatment, then we may also be able to test additional hypotheses about the population by constructing two-sided confidence intervals, the lower bounds of which will tell us which of the treatments cannot be best at the 1-λ level. This way if we cannot pick a single "best" treatment, at least we can determine the treatments that cannot be best, leaving us with a subset of the treatments which *might* be best. Also, the two-sided intervals may suggest consistent estimates with nice small sample properties. This particular inference technique is explored in section 3.5, but first we examine construction of intervals for equation (1.8) using MCC procedures.

3.4 Two-Sided MCC Intervals

Multiple comparisons with a control or MCC is primarily due to Dunnett (1955). If one of the N treatments, say N, can be regarded as a control, then we can construct intervals for equation (1.8). We use the same notation as section 3.3. We consider an balanced layout where for each treatment i = 1...N the number of observations is T.

Theorem 3.2

Let $| T |^{(\lambda)}_{N-1,\nu,\rho}$ be the solution in t for

$$\int_0^\infty \int_{-\infty}^{\infty} \{\Phi^{N-1}[(z\rho^{1/2} + ts)(1-\rho)^{-1/2}] - \Phi^{N-1}[(z\rho^{1/2} - ts)(1-\rho)^{-1/2}]\} d\Phi(z) dQ_\nu(s) = 1 - \lambda$$

A set of 1 - λ simultaneous confidence intervals for

$$\alpha_N - \alpha_1, \ldots, \alpha_N - \alpha_{N-1}$$

is given by

$$[\hat{\alpha}_N - \hat{\alpha}_i - d', \hat{\alpha}_N - \hat{\alpha}_i + d'] \qquad i = 1, \ldots, N-1$$

where

$$d' = | T |^{(\lambda)}_{N-1,\nu,\rho} S(2/T)^{1/2}$$

Tables for $| T |^{(\lambda)}_{N-1,\nu,\rho}$ can be found in Hochberg and Tamhane (1987), Dunnett (1964), Hahn and Hendrickson (1971) and Dunn and Massey (1965). A few observations on these intervals are in order. First, the critical value and hence the intervals are based on the equicorrelated structure of the $\hat{\alpha}_N - \hat{\alpha}_i$, but there are several approximation technique for dealing with a more general correlation structure. We defer discussion of these techniques until later. Second, if we restrict the bounds non-negative, these intervals can be regarded as MCB intervals conditional on α_N being the known best. Third, the two-sided critical value $| T |^{(\lambda)}_{N-1,\nu,\rho}$ is necessarily larger than the one-sided critical value $T^{(\lambda)}_{N-1,\nu,\rho}$ for equal values of N-1, ν and ρ, meaning that if we want to make statements about both upper and lower bounds, the upper bound is necessarily larger. These intervals are important insofar as they lay the foundation for adapting MCC intervals to MCB intervals which we address in the next section.

3.5 Edwards and Hsu (1983) - MCB Intervals from MCC Intervals

To perform two-sided inference on equation (1.7) we examine the 1983 paper of Edwards and Hsu. If we can regard *any* of the N populations as a control and construct

MCC intervals, they can be adapted to the MCB intervals of equation (1.7). Computation of these MCB intervals only requires that the MCC intervals exist (i.e. can be constructed); it does not require independence or equal variances for the N populations, π_i. As mentioned earlier MCC intervals can be thought of as MCB intervals conditional on the knowledge that the control is the best. Here Edwards and Hsu adapt these to *unconditional* MCB intervals to incorporate uncertainty about the best treatment. We use the same notation as section 3.3.

Let I represent the set of population indices, so $I = \{1, \ldots, N\}$. If there is any reference group or control $j \in I$ such that a random confidence interval $[L_i^j, U_i^j]$ exists for $i \in I - \{j\}$ satisfying

$$P\{L_i^j \leq \alpha_j - \alpha_i \leq U_i^j \text{ for every } i \in I - \{j\}\} \geq 1 - \lambda,$$

and each joint distribution, P, is an element of some family of distributions at least partially indexed by the α_i, then MCC intervals on j exist and can be adapted to MCB intervals using the following theorem.

Theorem (Edwards and Hsu 1983)

When MCC intervals on $j \in I$ exist as defined above,

$$P\{[L_i \leq u^*_i \leq U_i \; \forall i \in I] \cap [[N] \in \zeta]\} \geq 1 - \lambda$$

where

$$\zeta = \{j: U_i^j \geq 0 \text{ for } i \in I - \{j\}\},$$

and for each $i \in I$

$$L_i = 0 \qquad \zeta = \{i\}$$

$$L_i = \max(\min_{j \in \zeta} L_i^j, 0) \qquad \zeta \neq \{i\}$$

$$U_i = \max(\max_{j \neq i} U_i^j, 0).$$

For a proof see Edwards and Hsu (1983). A few observations are in order. First, ζ is the set of all treatments that have *all* non-negative MCC upper bounds. If we construct a set of simultaneous MCC intervals for each $j \in I$, then those j that conform to the selection criteria of ζ must have all non-negative MCC upper bounds. Second, if a treatment i is not in the set ζ, then its lower bound is $\max(\min_{j \in \zeta} L_i^j, 0)$. If the treatment i is contained in ζ, then its lower bound is 0. So, a treatment's lower bound is 0 if it is in ζ or if it is close to being in ζ. Third, the lower bound is constrained non-negative. In their paper Edwards and Hsu give a normal distribution example of the above theorem. While not an explicit theorem in their paper, the following theorem is adapted from their normal example. It assumes that the $\hat{\gamma}_i$'s possess the aforementioned equicorrelated structure.

Theorem 3.3 (Edwards and Hsu 1983)

A set of $1 - \lambda$ simultaneous confidence intervals for

$$\alpha_{[N]} - \alpha_1, \ldots, \alpha_{[N]} - \alpha_N$$

is given by

$$P\{[L_i \leq u^*_i \leq U_i \; \forall i \in I] \cap [[N] \in \zeta]\} \geq 1 - \lambda$$

where

$$\zeta = \{j: \hat{\alpha}_j \geq \max_{i=1}^{N} \hat{\alpha}_i - d'\}.$$

and for each $i \in I$

$$L_i = \max(\min_{j \in \zeta}[\hat{\alpha}_j - \hat{\alpha}_i - d'], 0)$$

$$U_i = \max(\max_{j \neq i}[\hat{\alpha}_j - \hat{\alpha}_i + d'], 0)$$

where

$$d' = | T |^{(\lambda)}_{N-1,\nu,\rho} s(2/T)^{1/2}$$

Edwards and Hsu refer to these as *adaptive intervals*. Again, tables for $| T |^{(\lambda)}_{N-1,\nu,\rho}$ can be found in Hochberg and Tamhane (1987), Dunnett (1964), Hahn and Hendrickson (1971) and Dunn and Massey (1965). Notice that the upper bound in this case is the same as the one-sided upper bound derived by Hsu (1981), save for the critical value, so its interpretation is the same. In terms of inference on which treatment is best, the interpretation of these intervals is straightforward. If the lower bound for a group of treatments is not zero (positive), these treatments can be eliminated as being best at a 1-λ confidence level. If the upper bound for any single treatment is zero, then it is best at the 1-λ confidence level. We now discuss the interpretation of the lower bound.

The lower bound is based on the smallest treatment mean of those that satisfy the selection criterion, the set ζ. ζ is the set of all treatments in contention for the best where selection is based on the MCC intervals for each treatment. A treatment is eliminated from contention (excluded from ζ), if *any one* of its MCC upper bounds are negative. Clearly, this means we need only test the interval that could result in the largest negative upper bound for each treatment, namely the interval for $\alpha_i - \max_{j=1}^{N} \alpha_j$

for each treatment i. This leads to the definition of ζ in Theorem 3.3 which is the basis for determining the MCB lower bounds.

Two polar examples help to illustrate the purpose of ζ. If ζ consists of one element, say {m}, then α_m is the best treatment with confidence level 1-λ. In this case the MCB intervals reduce to MCC intervals with the lower bound constrained positive, and the MCC intervals provide us with all the information we need to determine that m is best, so the MCB intervals are narrow. If ζ consists of all N treatments, then all the treatments will have an MCB lower bound of zero, meaning that none of the treatments can be eliminated from being the best. In this case the MCC intervals have provided us with little information and the MCB intervals are wide.

The lower bound is constructed so that it equals zero, when the treatment in question is in the set ζ or when it is within d' of the lowest value in this set. This means that even if a treatment is not an element of ζ, it may still have zero as a lower bound and might not be eliminated as being in contention for the best. This is due to the uncertainty associated with the best treatment being unknown.

Since we have two-sided intervals a discussion of the factors which determine their width is more instructive than in the case of one-sided intervals. The following factors affect the bounds and hence the width of the intervals. T and s effect both bounds in the usual way though d'. The critical value, $| T |^{(\lambda)}_{N-1,\rho,\rho}$, merits some further examination. When N is large, the critical value is larger making the intervals wider to account for the added uncertainty associated with estimating more parameters. This widening also compensates for additional multiplicity in the probability integral lest the confidence level be diminished.

The lower bound involves ζ and, as we have seen, is based on the smallest sample mean with a positive MCC upper bound. As T gets large the probability that ζ contains the single element [N] approaches 1. That is: as $T \to \infty$, $P[\zeta = \{[N]\}] \to 1$ (provided that there are no ties for the best in the population). Since this single element is the largest in the population it will tend to be the largest in the sample as the accuracy of the estimates improves; this forces the lower bound up, causing the intervals to get narrower. So, increasing T narrows the lower bound in two ways: through d' and through ζ. The former is due to improved accuracy of our estimates and the latter is due to improved accuracy in the ranking of the sample means.

Figure 1 (in Tables and Figures Section) illustrates the relationship between MCC intervals and MCB intervals. The width of the MCC interval (top) is determined solely by the allowance, d'. The width of the MCB interval (bottom) is determined by the allowance *and* the uncertainty of the ranking, which is determined by the number of elements contained in ζ. So, when T is large enough that ζ consists of a single element then the uncertainty of the ranking vanishes, and the MCB interval is identical to the MCC interval with the maximum population parameter known.

As previously stated the size of T has a direct effect on the number of elements that satisfy the selection statement, ζ. As T gets large the probability that ζ contains the single element [N] approaches 1, meaning that the firm with the maximum sample mean will ultimately be determined best by the confidence interval. If T is small, ζ will probably consist of many elements, and we may not be able to say with any reasonable degree of confidence that one treatment is significantly better than the rest.

One must keep in mind that these intervals are *not* based on the estimator $\max_{j=1}^{N} \hat{\alpha}_j - \hat{\alpha}_i$, for this would imply that we are certain that our estimation procedure reveals the true ranking of the population parameters. If we *do* regard these intervals as based on the estimator $\max_{j=1}^{N} \hat{\alpha}_j - \hat{\alpha}_i$, then (ignoring the max(., 0) operators in Theorem 3.3) these intervals can be rewritten as follows:

$$L_i = \max_{j=1}^{N} \hat{\alpha}_j - \hat{\alpha}_i - d' - (\max_{j=1}^{N} \hat{\alpha}_j - \min_{j \in \zeta} \hat{\alpha}_j)$$

$$U_i = \max_{j=1}^{N} \hat{\alpha}_j - \hat{\alpha}_i + d'$$

If we consider the mid-point of the interval a point estimate for $\max_{j=1}^{N} \alpha_j - \alpha_i$, then an estimate for inefficiency, u_i, based on the intervals is:

$$\tilde{u}_i = (\max_{j=1}^{N} \hat{\alpha}_j + \min_{j \in \zeta} \hat{\alpha}_j)/2 - \hat{\alpha}_i \tag{1.10}$$

suggesting the point estimate for $\alpha_{[N]}$:

$$\tilde{\alpha}_{[N]} = (\max_{j=1}^{N} \hat{\alpha}_j + \min_{j \in \zeta} \hat{\alpha}_j)/2 \tag{1.11}$$

which has been shown to be consistent and have less positive bias than $\hat{\alpha}$ (Edwards and Hsu, 1983, p967). The bias of the estimate $\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i$ is positive, is of magnitude $(\max_{j=1}^{N} \hat{\alpha}_j - \min_{j \in \zeta} \hat{\alpha}_j)/2$ and is caused by the over estimation of the largest $\hat{\alpha}_i$ of the sample. Notice that this bias is a function of the critical value which depends on our confidence level. As the confidence level rises so does the critical value, potentially making $\min_{j \in \zeta} \hat{\alpha}_j$ smaller and the bias larger. This bias is eliminated when the difference

between the largest $\hat{\alpha}_i$ and the second largest $\hat{\alpha}_i$ is large, the probability of which approaches 1 as T gets large. When T is small the probability that there are several $\hat{\alpha}_i$ near the maximum is large, so the bias is large.

3.6 Extensions and Conclusions

Clearly, the inference of the preceding theorems hinges on a specific form for the correlation of the $\hat{\gamma}_i$: the correlation matrix must have identical off-diagonal elements, so the equicorrelated structure can emerge. If the correlation matrix is of a more general form, one must appeal to a more robust inference technique. This section briefly summarizes a few of these techniques.

If the data are unbalanced, things become more complicated. In general, unbalanced data lead to a lack of the equicorrelated structure, making determination of the critical value more cumbersome. In fact, we would have to calculate a more general critical value, $T^{(\lambda)}_{N-1,\nu,\rho(ij)}$ or $|T|^{(\lambda)}_{N-1,\nu,\rho(ij)}$, where $\rho(ij)$ is the correlation between the $\hat{\gamma}_i$ and $\hat{\gamma}_i$ in the unbalanced case and is <u>not</u> constant over all $i \neq j$. (This notation, $\rho(ij)$, is to distinguish this correlation from that of the balanced case, ρ_{ij}). Tabulation of such a critical value would be impractical, and even computation with a computer would be costly, since it would require solving N-1 dimensional integrals. All is not lost however, for if the correlation matrix of the $\hat{\gamma}_i$ takes on a *product structure* where:

$$\rho(ij) = \delta_i\delta_j \qquad i, j \neq N \tag{1.12}$$

where $\delta_i = [T_i/(T_i + T_N)]^{1/2}$ $i \neq N$

then the N-1 dimensional probability integral reduces to a lower dimensional iterated integral, so $T^{(\lambda)}_{N-1,\nu,\delta}$ and $|T|^{(\lambda)}_{N-1,\nu,\delta}$ are the solutions in t of:

$$\int_0^{\infty}\int_{-\infty}^{\infty} \Pi_i^{N-1}\Phi[(\delta_i z + ts)(1-\delta_i^2)^{-1/2}]d\Phi(z)dQ_\nu(s) = 1 - \lambda$$

and

$$\int_0^{\infty}\int_{-\infty}^{\infty} \Pi_i^{N-1}\{\Phi[(\delta_i z + ts)(1-\delta_i^2)^{-1/2}] - \Phi[(\delta_i z - ts)(1-\delta_i^2)^{-1/2}]\}d\Phi(z)dQ_\nu(s) = 1 - \lambda$$

respectively. These can be solved using an iterative search technique or simulated. If this is the case, then the respective allowances for the MCC and MCB intervals become:

$$(1.13) \qquad d_{ij} = T^{(\lambda)}_{N-1,\nu,\delta} S(1/T_i + 1/T_j)^{1/2} \qquad d_{ij}' = |T|^{(\lambda)}_{N-1,\nu,\delta} S(1/T_i + 1/T_j)^{1/2}$$

If we cannot appeal to this special structure in the unbalanced case (or for that matter, in the balanced case) than we must use a conservative approximation technique. Matejcik (1992) suggests techniques for adaptive MCB intervals that are robust to a generalization of the correlation matrix and compares their performance using computer simulation. These techniques are based on several MCC methods that are themselves robust and include: an MCC method based on Banjeree's Inequality due to Tamhane (1977), a procedure using a moment-based approximation to the Behrens-Fisher problem due to Tamhane (1977), a method using all-pairwise procedures due to Dunnett (1980)

and lastly his own technique based on a heteroscedastic selection procedure. The techniques will not be discussed here; the reader is referred to the above citations. An obvious line for further research is to examine the applicability of these techniques to stochastic frontier models.

The adaptive MCB intervals of Theorem 3.3 (Edwards and Hsu 1983) imply the one-sided results of Theorem 3.1 (Hsu 1981), so the rest of this chapter will be in terms of Theorem 3.3. To construct adaptive MCB intervals one need only be able to construct the MCC intervals. So the questions remains, can we construct MCC intervals in the fixed effects stochastic production frontiers context? The answer to this question is yes, and the following section seeks to explain why.

4 APPLYING MCB METHODS TO STOCHASTIC FRONTIER MODELS

4.1 Motivation and Considerations

The motivation for MCB stems from the problems specified in section 2.2. MCB methodology eliminates these problems. First, we can perform comparative inference on the best population mean with only a normality assumption (no half normals, exponentials or extreme value distributions). Second, MCB allows construction of intervals for an *unknown* best treatment, so we don't have to assume that the sample rankings coincide with the population rankings. Additionally, the point estimates of equations (1.10) and (1.11) are consistent as $T \to \infty$ and may have less positive bias than those of equation (1.4) when T is small.

To use Theorem 3.3 we must be able to construct MCC intervals for the α_i, so we would like to identify the circumstances under which MCC is the most readily

applicable to constructing intervals for stochastic frontiers. First, it is important to point out that the MCB intervals are not intervals constructed on the $\hat{u}_i$ but instead on $\hat{\gamma}_i$, for we are not certain which firm is the true best. However, the MCC intervals (from which the MCB intervals are adapted) are for $\alpha_N - \alpha_i$, where N is not necessarily the index of the best firm in the population, but merely a control. Therefore, a priori knowledge of the best firm is not necessary, and MCC intervals constructed on $\hat{u}_i$ are perfectly acceptable. This is the strength of the adaptive intervals: the experimenter first assumes that the estimation reveals the true best firm and constructs MCC intervals. Then she drops this assumption and adapts the intervals to accommodate this uncertainty. Consequently one would expect MCC intervals with the lower bound constrained non-negative to be narrower than the subsequent MCB intervals.

Construction of the MCC and MCB intervals is simple, however determination of the appropriate critical value is non-trivial. When we can appeal to the equicorrelated structure the critical value can be drawn from the aforementioned references, but even then the existing tables are for small values of N only ($N \leq 20$, typically). For large values of N the critical value would have to be calculated using a iterative search technique or simulated. Failing the equicorrelated case, if the product structure of equation (1.12) emerges, then we can again calculate a critical value using a iterative search. When we cannot appeal to the equicorrelated or product structures, approximation techniques can be employed as noted in section 3.6, but even then one runs into problem of no critical value tables for large N.

To construct MCB intervals we must examine the correlation structure of the $\hat{\gamma}_i$, which entails calculating the variance and the covariances of the $\hat{\alpha}_i$ in equation (1.5). If

the firm efficiencies are orthogonal (e.g when β is known and $\text{Var}(\hat{\beta}) = 0$), and we have a balanced panel, then $\text{var}(\hat{\gamma}_i) = 2\sigma^2/T$ and $\rho_{ij} = \rho = \frac{1}{2}$. The equicorrelated structure emerges; the critical value can be looked up or calculated, and the MCB intervals can be constructed directly from Theorem 3.3 without appealing to an approximation technique. The problem is that the firm efficiency estimates are, in general, not orthogonal as evidenced by equation (1.5), so the equicorrelated structure is unlikely to emerge in small samples. All is not lost however, for we can appeal to the large sample properties of the $\hat{\alpha}_i$. Another potential problem arises when the panel is unbalanced. If the panel is unbalanced then the $\hat{\gamma}_i$ do not possess the equicorrelated structure. In this case we would like to appeal to product structure of equation (1.12). Again, unfortunately, the lack of orthogonality within the firm efficiencies, as evidenced in equation (1.6), precludes us from appealing to this simpler structure. However, again we may be rescued by the large sample properties of $\hat{\alpha}_i$. What follows are summaries of the properties of the MCB intervals under different configurations of the panel data set in both the balanced and unbalanced contexts.

4.2 Large N, Small T

For a balanced panel, when N gets large the covariance matrix of $\hat{\beta}$ approaches zero, causing the off-diagonal elements of $\text{Var}(\hat{\alpha})$ to approach zero and the diagonal elements to approach σ^2/T, hence $\text{var}(\hat{\gamma}_i) \to 2\sigma^2/T$, $\text{cov}(\hat{\gamma}_i, \hat{\gamma}_j) \to \sigma^2/T$ and $\rho_{ij} \to \rho = \frac{1}{2}$, giving us the equicorrelated structure. In this case MCB procedures can be directly applied using Theorem 3.3 (Edwards and Hsu). If the critical constant $| T |^{(\lambda)}_{N-1,r,\rho}$ cannot be found in the aforementioned references, because N is too large, then it can be

calculated or simulated. If the panel is unbalanced, as N gets large the product structure of equation (1.12) will emerge, and we can apply Theorem 3.3 as modified by equation (1.13) with critical value $|T|^{(\lambda)}_{N-1,r,\delta}$ calculated or simulated. In either case, balanced or unbalanced, point-estimates based on these intervals follow directly; since T is small the point estimates will have smaller positive bias than $\hat{u}_i$ and $\hat{\alpha}$.

4.3 Large T

For a balanced panel, if T is large, the variance of our estimate of β is small, so the of diagonal elements elements of the covariance matrix of the $\hat{\alpha}_i$ become negligable. Additionally, the bias of $\hat{u}_i$ and $\hat{\alpha}$ will be minimal. In the unbalanced case, as T_i gets large for all i similar results arise. In either case MCB procedures can be directly applied using Theorem 3.3 (or as modified by equation (1.13)) and should result in narrow intervals for the u_i and a strong statement about which treatment is best.

4.4 Small N and T

If N is less than large, the covariance matrix of $\hat{\alpha}$ is non-spherical, and Theorem 3.3 can no longer be used to perform MCB inference on the estimates. However, if the variance structure of the $\hat{\alpha}_i$ takes on the special form;

$$(1.14) \qquad \mathrm{Var}(\hat{\alpha}_i) = \sigma^2/T + C \qquad \mathrm{Cov}(\hat{\alpha}_i,\hat{\alpha}_j) = C \qquad C = \text{constant}$$

then again $\mathrm{Var}(\hat{\gamma}_i) = 2\sigma^2/T$, $\mathrm{Cov}(\hat{\gamma}_i, \hat{\gamma}_j) = \sigma^2/T$, $\rho_{ij} = \rho = \frac{1}{2}$ and the correlation structure of the $\hat{\gamma}_i$'s is the same as when N is large. Again Theorem 3.3 can be directly

applied. Since T is small the bias of $\hat{u}_i$ and $\hat{\alpha}$ will be large, so we will need to rely on the point estimates suggested by the intervals. Similarly for the unbalanced case, if

$$(1.15) \qquad \mathrm{Var}(\hat{\alpha}_i) = \sigma^2/T_i + C \qquad \mathrm{Cov}(\hat{\alpha}_i,\hat{\alpha}_j) = C,$$

then we can use Theorem 3.3 as modified by equation (1.13). If the special forms of equations (1.14) or (1.15) cannot be assumed we must appeal to an approximation technique as suggested in section 3.6.

5 CONCLUSIONS AND EXTENSIONS

In the broadest sense this chapter has attempted to provide some insight into the sources of uncertainty in the estimation of inefficiency measures in the fixed effects model. Specifically, we find that MCB methodology is a useful tool for understanding this uncertainty in the panel data context. For the econometric theorist the brief MCB survey should shed some light on the issues surrounding this uncertainty. We conclude that the assumption that the sample rankings reveal the true population rankings may be dubious, particularly when T is small.

Consideration of this ranking uncertainty has strong implications for the overall productivity modelling approach. This uncertainty uncovers additional inaccuracy inherent in the efficiency estimates of these models, which heretofore has gone unnoticed. If we are interested in preserving the accuracy of our estimates and reducing this uncertainty, then we would typically have to sacrifice model flexibility to achieve it.

Therefore, this MCB methodology may argue for more accurate and less flexible productivity models. Of course this remains to be seen.

The next chapter is concerned with applying this MCB methodology to real data sets and comparing the resulting intervals with those generated under other techniques that are fairly well-established in the stochastic frontiers literature.

CHAPTER TWO

CONFIDENCE STATEMENTS FOR EFFICIENCY ESTIMATES

1 INTRODUCTION

This chapter is a comprehensive study of methods of inference associated with technical efficiency estimates in the stochastic frontier model. We seek to characterize the nature and the empirical magnitude of the uncertainty associated with the usual estimates of efficiency levels.

From our perspective, deterministic approaches (e.g., DEA) produce efficiency *measures*, while statistical approaches (stochastic frontier models) produce efficiency *estimates*. The relative strengths and weaknesses of these approaches have been vigorously debated, and will continue to be. However, the strongest argument in favor of a statistical approach has always been that it provides a straightforward basis for inference, not just for point estimates. Thus, for example, one can construct standard errors and confidence intervals for estimates of technical efficiency. A statistical approach recognizes that uncertainty exists and is capable of quantifying it. In our view, uncertainty also exists within the deterministic approach, but methods of characterizing and quantifying it are still not well developed. Consistency of the DEA estimates has been established by Banker (1993) and by Korostelev, Simar and Tsybakov (1992, 1995). Korostelev, Simar and Tsybakov also establish the rate of convergence of the estimates, and Banker (1995) considers certain types of hypothesis tests. These results are important but they do not lead to confidence intervals. Confidence intervals can be

constructed by bootstrapping the DEA estimates; for example, Simar and Wilson (1995) give some theoretical results and an empirical example. However, in our view bootstrapping procedures are an imperfect substitute for an adequately developed distributional theory.

Ironically, the ability to conduct inference on efficiency estimates in stochastic frontier models has previously been noted approvingly, but has never been systematically exploited in an empirical setting. This chapter seeks to fill this void and, in doing so, advances our understanding of the various sources of uncertainty inherent in econometric models for efficiency estimation.

Of course, the strength of the econometric approach comes at a cost: strong and often arbitrary distributional assumptions are necessary to extract technical efficiency estimates and ultimately to construct confidence intervals. Therefore, a major aim of this chapter will be to show how to perform inference on efficiency estimates under different sets of assumptions that range from the very strong to the relatively weak, and to see how the degree of uncertainty associated with these estimates relates to the strength of the assumptions made. Some of the methods we discuss require panel data. Most make specific distributional assumptions for statistical noise and technical inefficiency. However, we also make use of the methodology of Multiple Comparisons with the Best (MCB), developed by Edwards and Hsu (1983) and described in chapter one, which uses panel data to construct confidence intervals without the need for strong distributional assumptions.

In this chapter, technical efficiency estimates and their confidence intervals are generated for three different panel data sets with different dimensional characteristics,

using several formulations of the stochastic frontier model. We analyze these panel data as complete data sets and also in some cases broken down into their component cross sections to construct confidence intervals for technical efficiency estimates using different interval construction techniques. The results highlight the relevant strengths and weaknesses of the various techniques and data configurations. The chapter addresses practical aspects of interval construction that may present problems for the data analyst.

The plan of the chapter is as follows. Section 2 briefly reviews the stochastic frontier model as it relates to this chapter. Section 3 reviews two interval construction techniques: the Jondrow, Lovell, Materov and Schmidt (JLMS) (1982) method and the Battese-Coelli (BC) (1988) method. Section 4 is an empirical analysis of three panel data sets for which we construct confidence intervals for technical efficiency estimates. Section 5 summarizes and concludes.

2 STOCHASTIC FRONTIER MODELS

In this section we detail various estimation techniques for the stochastic frontier model of equation (1.1). Since chapter one detailed fixed effects estimation in the panel data context, this section concentrates on other estimation techniques in both the cross-sectional and panel data contexts.

2.1 Cross-sectional Data

In the case of a single cross-section, $T = 1$ and t is irrelevant and can be suppressed. Under assumptions (A.1)-(A.4), the model as given in equation (1.1) can be estimated by maximum likelihood (MLE). Details of this estimation, including the

likelihood function, can be found in Aigner et al. (1977) and will not be addressed here. MLE of equation (1.1) yields $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_v$ and $\hat{\sigma}_u$, which are consistent as $N \to \infty$.

Define $\mu = E(u_i) \geq 0$. Under assumption (A.4), $\mu = (\pi/2)^{1/2}\sigma_u$. Ordinary Least Squares (OLS) applied to equation (1.1) yields consistent estimates of $(\alpha-\mu)$ and β. The Corrected Ordinary Least Squares (COLS) method constructs a consistent estimate of α by adding a consistent estimate of μ to the OLS intercept. This requires a consistent estimate of σ_u, say $\hat{\sigma}_u$, which can be derived from the third moment of the OLS residuals. Also a consistent estimate of σ_v can be derived from the second moment of the OLS residuals. See Olson et al. (1980) for details.

So, in summary, both COLS and MLE yield consistent estimates of α, β σ_u and σ_v. COLS is less efficient than MLE. In either case, point estimates for u_i and $TE_i = \exp(-u_i)$ can be obtained, as described in section 3.1.

2.2 Panel Data

We now turn to the case of panel data with $T > 1$. Under assumptions (A.1)-(A.4) equation (1.1) can be estimated by MLE. See Pitt and Lee (1981) for the likelihood function and other details. MLE yields estimates of the same parameters as in the cross-sectional case: α, β, σ_u and σ_v. These estimates are consistent as $N \to \infty$; therefore MLE is appropriate when N is large. Large T is not a substitute for large N.

Equation (1.1) can also be estimated by Generalized Least Squares (GLS). This requires assumptions (A.1)-(A.4), except that it does not rely on specific distributional assumptions (normality of v, or half-normality of u). The standard panel data GLS procedure yields estimates of $(\alpha-\mu)$, β, σ^2_v and $\text{var}(u_i)$ that are consistent as $N \to \infty$. Care

must be taken to distinguish $var(u_i)$ and σ^2_u; the usual GLS procedure uses $var(u_i)$, not σ^2_u. Under the half-normal distributional assumption, $\sigma^2_u = var(u_i)\pi/(\pi-2)$, so that the estimate of $var(u_i)$ is easily converted to an estimate of σ^2_u. This is required to estimate $\mu = E(u_i)$ and to convert the intercept, exactly as in the discussion of COLS above. We will refer to GLS with this intercept correction as the CGLS method. Point estimates for u_i and TE_i can be obtained, as described in section 3.2.

3 TECHNIQUES FOR CONSTRUCTION OF CONFIDENCE INTERVALS

We use two different techniques to construct confidence intervals for technical efficiency estimates in stochastic frontier models. The first technique is based on the (conditional) distribution $u_i \mid \epsilon_i$, where $\epsilon_i = [\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iT}]$. It was developed for the cross-sectional case by Jondrow, Lovell, Materov and Schmidt (JLMS) (1982) and later generalized to the panel data case by Battese and Coelli (BC) (1988). The second technique is based on the MCB procedures described on chapter one. The MCB method will be based on fixed effects estimates described in chapter one, while the JLMS and BC methods will be applied to the results of the other estimation techniques described above; this choice is primarily driven by the difference in distributional assumptions of the models.

3.1 Cross-sectional Data: JLMS Method

For either cross-sectional estimation method, MLE or COLS, we use the JLMS method for interval construction. The JLMS technique follows from the distribution of u_i conditional on ϵ_i (which is a scalar, since $T=1$ for a cross section). JLMS show that

given distributional assumptions (A.1) and (A.4), the distribution of $u_i \mid \epsilon_i$ is that of a $N(\mu^*_i, \sigma^2_*)$ random variable truncated (from the left) at zero, where $\mu^*_i = \sigma^2_u\epsilon_i(\sigma^2_u + \sigma^2_v)^{-1}$ and $\sigma^2_* = \sigma^2_u\sigma^2_v(\sigma^2_v + \sigma^2_u)^{-1}$. They evaluate $E(u_i \mid \epsilon_i)$, which is regarded as a point estimate for u_i. A point estimate for TE_i, due to Battese and Coelli (1988), is given by:

$$(2.1) \quad TE_i = E[\exp(-u_i) \mid \epsilon_i] = \exp\{-\mu^*_i + \tfrac{1}{2}\sigma^2_*\}\{1 - \Phi(\sigma_* - \mu^*_i/\sigma_*)\}\{1 - \Phi(-\mu^*_i/\sigma_*)\}^{-1},$$

where Φ is the standard normal cdf. Implementing this procedure requires estimates of μ^*_i and σ^2_*; this in turn requires estimates of σ^2_u and σ^2_v, and the use of $e_i = y_i - \hat{\alpha} - x_i\hat{\beta}$.

Empirical implementations of the JLMS technique have focused on the point estimate $E(u_i \mid \epsilon_i)$. However, confidence intervals for u_i or TE_i are easily constructed from the density of $u_i \mid \epsilon_i$. Critical values can be extracted from a standard normal density to place lower and upper bounds on $u_i \mid \epsilon_i$. Because TE_i is a monotonic transformation of u_i, the lower and upper bounds for $u_i \mid \epsilon_i$ translate directly into upper and lower bounds on $TE_i \mid \epsilon_i = \exp(-u_i) \mid \epsilon_i$. Specifically, a $(1-\lambda)100\%$ confidence interval (L_i, U_i) for $TE_i \mid \epsilon_i$ is given by:

$$(2.2) \qquad L_i = \exp(-\mu^*_i - z_L\sigma_*),$$

$$U_i = \exp(-\mu^*_i - z_U\sigma_*),$$

where: $\Pr(Z > z_L) = (\lambda/2)[1 - \Phi(-\mu^*_i/\sigma_*)]$,

$\Pr(Z > z_U) = (1 - \lambda/2)[1 - \Phi(-\mu^*_i/\sigma_*)]$,

with Z distributed as N(0, 1); so that

$$z_L = \Phi^{-1}\{1 - (\lambda/2)[1 - \Phi(-\mu^*_i/\sigma_*)]\} \quad \text{and}$$

$$z_U = \Phi^{-1}\{1 - (1 - \lambda/2)[1 - \Phi(-\mu^*_i/\sigma_*)]\}.$$

As a semantic point, we will refer to the implementation of equation (2.2) in the cross-sectional context as the *JLMS method*, since it relies on the JLMS result for the distribution of $u_i \mid \epsilon_i$, even though equation (2.1) is due to BC. The *BC method* will refer to the corresponding calculations in the panel-data case, described in the next section. It should be noted that both the JLMS and the BC methods treat α, β, σ^2_u and σ^2_v as known, so that the confidence intervals do <u>not</u> reflect uncertainty about these parameters. For large N, this is probably unimportant, since the variability in the parameter estimates is small compared to the variability intrinsic to the distribution of the $u_i \mid \epsilon_i$ (and due to the presence of the statistical noise v_{it}).

3.2 <u>Panel Data: Battese-Coelli Method</u>

The BC method for construction of confidence intervals is a generalization of the JLMS method and also follows from the distribution of $u_i \mid \epsilon_i$. The BC technique can be based on the MLE or CGLS estimates of α, β, σ^2_u and σ^2_v. It extends the JLMS method to accommodate the case of panel data ($T > 1$), so that now $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT}) = (v_{i1} - u_i, \ldots, v_{iT} - u_i)$. Define $\bar{\epsilon}_i = (1/T)\Sigma_t \epsilon_{it}$, $\mu^*_i = \sigma^2_u \bar{\epsilon}_i(\sigma^2_u + \sigma^2_v/T)^{-1}$ and $\sigma^2_* = \sigma^2_u \sigma^2_v(\sigma^2_v + T\sigma^2_u)^{-1}$. The latter expressions are essentially the same as in JLMS, with $\bar{\epsilon}_i$ replacing ϵ_i and σ^2_v/T replacing σ^2_v. Then the distribution of $u_i \mid \epsilon_i$ is that of a $N(\mu^*_i, \sigma^2_*)$ random

variable truncated at zero, a point estimate for TE_i is given by equation (2.1) above, and confidence intervals are constructed as in equation (2.2) above.

The Battese-Coelli method can also accommodate the case of an unbalanced panel, in which there are different numbers of time-series observations per firm. Suppose that for firm i there are T_i observations, where the notation reflects the fact that T_i varies over i. We simply have to replace T by T_i in the definition of μ^*_i and σ^2_* above, so that $\mu^*_i = \sigma^2_u\bar{\epsilon}_i(\sigma^2_u + \sigma^2_v/T_i)^{-1}$ and $\sigma^2_{*i} = \sigma^2_u\sigma^2_v(\sigma^2_v + T_i\sigma^2_u)^{-1}$; note that now σ^2_{*i} varies over i. Then equations (2.1) and (2.2) hold exactly as before, except that σ^2_{*i} replaces σ^2_*. Thus an unbalanced panel causes no real problems for the BC method.

3.3 Panel Data: Multiple Comparisons with the Best

MCB intervals are naturally based on the fixed effects estimates of chapter one and use only assumptions (A.1) and (A.2) above; they do not require a distributional assumption for the u_i. In this study, when we cannot appeal to asymptotics, the condition for the special covariance structure of equations (1.14) and (1.15) is met or nearly met, so MCB is at least approximately applicable.

We note in passing that several recent models in the frontiers literature have featured time-varying technical inefficiency. For example, see Cornwell, Schmidt and Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), and Lee and Schmidt (1993). These models imply intercepts α_{it} that vary over i and t. For a given value of t, it is natural to proceed as before to consider comparisons relative to the maximum (over i) of these intercepts, so that we essentially have a separate MCB problem for each t. However, there is no apparent reason to expect the equicorrelatedness condition to

hold for the estimated α_{it} from any of these models, and if it does not hold the methods surveyed above would not apply.

3.4 Comparison of Different Techniques

A discussion of the differences between the interval construction techniques is in order. First, it should be noted that MCB provides joint confidence intervals for $u^*_i = \alpha_{[N]} - \alpha_i$ of equation (1.3), whereas JLMS and BC provide marginal intervals for $u_i = \alpha - \alpha_i$ of equation (1.2). The difference between u^*_i and u_i is $u_{[N]} = \min_{i=1}^{N} u_i$ which may be non-trivial when N is small. Conversely, the difference between joint and marginal intervals may be substantial when N is large. For example, one of our data sets has N = 171. Although independence would be a poor assumption, it is instructive to note that a set of 171 independent intervals, each holding with a marginal probability of 0.95, would hold jointly with a probability of only $(.95)^{171} = 0.000116$. Conversely, joint confidence intervals that hold with a probability of 0.95 would correspond to marginal intervals with a confidence level far in excess of 0.95. Other things equal, we would certainly expect joint confidence intervals to be wider than corresponding marginal intervals, for a given level of confidence like 0.95.

The MCB and JLMS/BC methods also differ substantially in the way they handle estimation error. One sense in which this is true is that, assuming that the equicorrelated structure emerges for the $\hat{\alpha}_i$, the MCB intervals reflect the variability of $\hat{\beta}$, which the JLMS and the BC intervals ignore. This is probably not an important difference, since uncertainty about β is not the only source, or in most cases the major source, of

uncertainty about u_i. To be more specific, consider the following expression for the fixed effects estimate of α_i:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta} = \alpha + (\bar{v}_i - u_i) - \bar{x}_i(\hat{\beta} - \beta) = \alpha_i + \bar{v}_i - \bar{x}_i(\hat{\beta} - \beta)$$

The term $\bar{x}_i(\hat{\beta} - \beta)$ reflects estimation error in β. As noted above, BC ignores this source of uncertainty while MCB does not. This term disappears as either N or $T\to\infty$, and is probably not important empirically for most data sets. More fundamentally, $\hat{\alpha}_i$ contains the error $\bar{v}_i = (1/T)\Sigma_t v_{it}$; the fixed effects procedure separates u_i from v_{it} by averaging away the v_{it}. The significance of $\bar{v}_i$ depends on T and on the relative sizes of σ^2_u and σ^2_v; it is most troublesome when T is small and/or σ^2_v is large relative to σ^2_u. It is important to realize that the fixed effects estimate of u^*_i, namely $\hat{\alpha}_{[N]} - \hat{\alpha}_i$, is generally biased upward (inefficiency is overstated), because the larger $\hat{\alpha}_i$, such as $\hat{\alpha}_{[N]}$, will on average contain positive estimation error $\bar{v}_i$, while the smaller $\hat{\alpha}_i$ will on average contain negative estimation error. (That is, the $\hat{\alpha}_i$ will obviously be more variable than the α_i.) MCB recognizes this variability by including the sample equivalent of $\sigma_v(2/T)^{1/2}$ in the formula for the allowance, d, above. Also, the MCB intervals can be thought of as removing the bias just described; they are not centered on the value $\hat{\alpha}_{[N]} - \hat{\alpha}_i$.

The BC method uses distributional assumptions to remove estimation error more effectively. The first step in the BC procedure is to calculate $\bar{\epsilon}_i = \bar{v}_i - u_i$ (ignoring estimation error in β), so that the v_{it} are averaged away, as in the fixed effects procedure. The second step is to construct μ^*_i, which equals $\bar{\epsilon}_i$ times the shrinkage factor $\sigma^2_u(\sigma^2_u + \sigma^2_v/T_i)^{-1} < 1$. This corresponds to the "best linear predictor" in the random-effects panel

data literature; see Schmidt and Sickles (1984). It reflects the relative variability of u_i and $\bar{v}_i$. Finally, the distributional assumptions are used to imply the further shrinkage factor $\{1 - \Phi(\sigma_* - \mu^*_i/\sigma_*)\}\{1 - \Phi(-\mu^*_i/\sigma_*)\}^{-1} < 1$ in the calculation of the expectation of $TE_i \mid \epsilon_i$.

4 EMPIRICAL ANALYSES

4.1 Indonesian Rice Farms - Erwidodo (1990)

We analyze data previously analyzed by Erwidodo (1990), Lee (1991) and Schmidt and Lee (1993). For a complete discussion of the data see Erwidodo (1990). 171 rice farms in Indonesia were observed for six growing seasons. The data were collected by the Agro Economic Survey, as part of the Rural Dynamic Study in the rice production area of the Chimanuk River Basin, West Java and obtained from the Center for Agro Economic Research, Ministry of Agriculture, Indonesia. The 171 farms were located in six different villages and the six growing seasons consisted of three wet and three dry seasons. Thus the data configuration features large N and small T.

Inputs to the production of rice included in the data set are seed (kg), urea (kg), trisodium phosphate (TSP) (kg), labor (labor-hours) and land (hectares). Output is measured in kilograms of rice. The data also include dummy variables. DP equals 1 if pesticides were used and 0 otherwise. DV1 equals 1 if high yield varieties of rice were planted and DV2 equals 1 if mixed varieties were planted; the omitted category represents that traditional varieties were planted. DSS equals 1 if it was a wet season.

The are also 5 region dummy variables, DR1, DR2, DR3, DR4 and DR5, for the six different villages in the survey.

COLS and MLE were performed on each of the six different periods (cross-sections) in the panel. DSS, the dummy for wet season, had to be excluded for the cross section models, because it was constant across farms for a single period. Results are in Table 1. Unfortunately, periods 2, 3, 4 and 5 produced a positive third-order moment of the residuals, causing the MLE estimate to coincide with the OLS estimate as discussed in Waldman (1982). Additionally, this problem precludes COLS estimation since $\hat{\sigma}_u$ is negative. Therefore only periods 1 and 6 are analyzed as cross-sections for this data set. Since the results for the two periods were similar only the period 1 results are reported. Technical efficiencies and confidence intervals were produced using the JLMS technique; i.e., equations (2.1) and (2.2) above. Confidence levels are 95%, 90% and 75%. These results are contained in Tables 2A and 2B. Due to the large number of firms in the sample (171), only nine firms are reported here and subsequently: the three firms with the highest $\hat{\alpha}_i$, the three firms with the lowest $\hat{\alpha}_i$, and the three firms with the median $\hat{\alpha}_i$.

The choice of the estimation procedure (COLS versus MLE) made very little difference, so we will discuss only the MLE results in Table 2B. Efficiency levels are not estimated as precisely as one might hope. The firm with the highest estimated efficiency level had estimated efficiency of 0.9452, but a 95% confidence interval ranged from 0.8322 to 0.9982. The median firm had estimated efficiency of 0.9053, with a 95% confidence interval of (0.7576, 0.9957); and the worst firm in the sample had estimated efficiency of 0.8040 with a 95% confidence interval of (0.6415, 0.9694).

These are fairly wide confidence intervals. In fact the uncertainty about the inefficiency level of a given firm is definitely not small relative to the within-sample variability of the efficiency measures, and we would have little reason to have much faith in our efficiency rankings. The reason for this lack of precision is straightforward - most of the variation in $\epsilon_i = v_i - u_i$ is due to v_i, not u_i. We have (for MLE, t=1) $var(v_i) = \sigma^2_v = 0.0579$ and $var(u_i) = \sigma^2_u(\pi-2)/\pi = 0.00633$, so the variance of v is over nine times as large as the variance of u. This makes it very difficult to estimate u_i precisely.

Next, CGLS and MLE were performed on the entire panel. The variable DSS could now be included. Results are in Table 3. Technical efficiencies and confidence intervals were produced using the BC technique. These results are contained in Tables 4A and 4B. Efficiency levels based on the CGLS and MLE estimates are again similar. Not surprisingly, the panel data confidence intervals are tighter than their cross sectional counterparts, because $var(u_i \mid \epsilon_i)$ is smaller with six observations than with one. Nevertheless, the confidence intervals do not shrink as much as one might hope - compare a 95% confidence interval for the median firm of (0.7638, 0.9945) in Table 4B to (0.7576, 0.9957) in Table 2B. This is partly due to having only six observations per firm, and partly to getting a larger value of σ^2_v for the panel than for the t=1 cross section, which diminishes the value of the panel.

The fixed effects estimates were calculated for the panel, with time invariant regressors excluded to preclude multicolinearity. These results are also in Table 3. The covariance matrix for the $\hat{\alpha}_i$ very nearly exhibited the equicorrelated structure necessary to justify the MCB procedure:

Mean of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .04572$

Standard deviation of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .002211$

Maximum of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .05523$

Minimum of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .03918$

95%, 90% and 75% MCB intervals were constructed for technical inefficiencies using critical values of $|T|^{(\lambda)}_{170,846,½} = 3.42$, 3.18 and 2.71, respectively, and are given in Table 5A. The intervals are too wide to be of much use. For example, the firm with the highest $\hat{\alpha}_i$ (and hence with estimated efficiency of 100% by the usual calculation) has a confidence interval ranging from 0.5613 to 1. Every firm in the sample has a confidence interval with upper limit equal to one; that is, at the 95% confidence level, no firm is revealed to be inefficient. In fact, this is still true at the 75% confidence level.

The MCB intervals are much wider than their BC counterparts based on CGLS and MLE. We next attempt to determine the relative importance of three sources of width: estimation error, uncertainty of the identity of the most efficient firm, and the multiplicity of the probability statement. The easiest of these factors to investigate is uncertainty about the identity of the most efficient firm. To do so we simply assume that firm 164, which is the firm with the largest $\hat{\alpha}_i$, is most efficient in the sense of having the largest α_i (equivalently, smallest u_i). Under this assumption we construct the MCC intervals with firm 164 as the control. 95%, 90% and 75% confidence intervals required critical values of $|T|^{(\lambda)}_{170,846,½} = 3.42$, 3.18 and 2.71, respectively. Results are in Table 5B. The MCC intervals are necessarily tighter than the MCB intervals, but not

tight enough to be useful. In other words, the width of the MCB intervals is not significantly decreased by knowing which firm is best. We can conclude that the width is primarily due to either estimation error or multiplicity or both.

To disentangle the effect of multiplicity on the interval width, we would like to be able to construct marginal intervals for each firm. In the case where MCB reveals a single firm as efficient, this can be accomplished with a simple application of the Bonferroni inequality. This will be demonstrated later. In the present case, where there is no single firm revealed as most efficient, the construction of marginal intervals is less clear, because it is necessary to make a simultaneous statement about the firms to determine a subset of firms that may be efficient, and then to reduce this joint statement to a marginal statement about a single firm. However, we can get some idea of the effect of the multiplicity of the intervals just by reducing the number of confidence intervals created, which we can do by considering a subset of the firms. We therefore redid MCB for only the nine firms for which we have reported results in Table 5A. (However, the parameter estimates are still from the whole sample of 171 firms.) 95%, 90% and 75% confidence intervals required critical values of $| T |^{(\lambda)}_{9,846,\frac{1}{2}} = 2.56, 2.38$ and 1.96, respectively. Results are in Table 6A. As was the case in the MCC experiment, controlling for multiplicity did not result in a significant tightening of the intervals. For example, for the median firm, compare the new interval (0.3354, 0.9837) with N=9 to the old interval (0.2899, 1.0000) with N=171. We conclude that the multiplicity component of the interval width is small, leaving only estimation error to account for the large width of the intervals.

Further evidence on this point is obtained by considering the smallest possible subset of firms (N=2) and assuming that it is known that one of them is the most efficient. Thus, as in our MCC calculations, we assert that firm 164 is most efficient and we simply construct confidence intervals for $\hat{\alpha}_{164} - \hat{\alpha}_i$ for a given value of i. This is a standard calculation based on the estimate $\hat{\alpha}_{164} - \hat{\alpha}_i$ and its standard error, $se_{164,i} = [var(\hat{\alpha}_{164}) + var(\hat{\alpha}_i) - 2cov(\hat{\alpha}_{164}, \hat{\alpha}_i)]^{1/2}$, and using critical values from the standard normal distribution. (Note that we have not imposed the equicorrelatedness assumption in this calculation, so our results will be slightly different from the results for MCC with N=2, which would impose this assumption.) These are called "per comparison" intervals; they are given in Table 6B.

The per comparison intervals are indeed narrower than the MCB and the MCC intervals, but they are still fairly wide. For example, for the median firm we still have a 95% confidence interval of (0.3788, 0.8102). This confirms our conclusion that, for this data set, the width of the confidence intervals is due primarily to the estimation error. As noted above, estimation error is important for this data set because T is small and σ^2_v is large relative to $var(u_i)$. There is simply too much noise to get a clear picture of the value of u_i. The BC method does significantly better because it makes strong distributional assumptions that allow a much better separation of v from u. For this data set there does not seem to be a substitute for these strong assumptions.

4.2 Texas Utilities - Kumbhakar (1994)

In this study we reanalyze data originally analyzed by Kumbhakar (1994). Kumbhakar estimated a cost function, whereas we will estimate the production function.

The data set consists of observations on 10 major privately-owned Texas electric utilities observed annually over 18 years from 1966 to 1985, and includes information on annual labor, capital and fuel (inputs) for electrical power generation (output). Due to the relatively small number of firms a cross sectional study of the data was precluded. However, with 18 periods of observation per firm we have T larger than N, the opposite of the case with the Erwidodo rice farm data.

The model was estimated by CGLS and MLE with results given in Table 7. Notice that now σ^2_v is small relative to σ^2_u, so our estimates of technical efficiency should be more reliable than for the previous data set. It is instructive to point out that numerical accuracy became a problem in calculating TE_i using equation (2.1). The small value for σ_* produced extremely large values of μ^*_i/σ_* which, when evaluated in the standard normal cdf $\Phi(.)$, produced technical efficiencies greater then 100%. This was due to rounding error in the software package we originally selected. Fortunately, another package was found that evaluated the normal cdf more accurately. Tables 8A and 8B give our results for all 10 firms.

As expected, the efficiency estimates are much more precise than for the previous data set. For example, for firm 8 (one of the two median firms) and using MLE, efficiency is estimated as 0.8472 with a 95% confidence interval of (0.8264, 0.8683). These are useful results in the sense that the uncertainty about a given firm's efficiency level is small relative to the between-firm variation in efficiencies; we can have some faith in our rankings.

constr

equico

The M

the mo

the M

In fac

essenti

firms a

Erwido

(0.780

MCB i

contrib

make th

MCB c

exp(-u*

The fixed effects estimator was also calculated (Table 7), and MCB intervals were constructed (Table 9A). The covariance matrix for the $\hat{\alpha}_i$ again exhibited an almost-equicorrelated structure, so that MCB was applicable.

Mean of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .06765$

Standard deviation of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .00289$

Maximum of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .07255$

Minimum of $\bar{x}_i Var(\hat{\beta})\bar{x}_j' = .06166$

The MCB intervals successfully determined at the 95% confidence level that firm 5 was the most efficient firm in the sample and that all others were inefficient. Consequently, the MCC intervals coincided with the MCB intervals and are not reported separately. In fact, firm 5 was identified as most efficient at the 99.9% confidence level, so essentially we were certain that it was the best. The confidence intervals for the other firms are wider than the corresponding BC intervals, but still not nearly as wide as the Erwidodo rice farm data. For example, for firm 8 compare the MCB intervals of (0.7809, 0.8603) to the BC interval of (0.8264, 0.8683).

It is interesting to note that there is very little overlap between the BC and the MCB intervals, with the MCB intervals being generally lower. Two opposing sources contribute to this difference. The difference between u^*_i and u_i when N is small should make the BC intervals lower, since BC constructs a confidence interval for $\exp(-u_i)$ while MCB constructs confidence intervals for $\exp(-u^*_i)$, and since $u^*_i < u_i$ implies $\exp(-u_i) < \exp(-u^*_i)$. However, this is apparently more than offset by the BC technique's more

successful reduction of the effects of the estimation error. As noted above, the BC technique can be viewed as a set of shrinkages of the fixed effects inefficiency measures, leading to generally higher efficiency measures.

Marginal intervals were easily constructed for each firm using the Bonferroni inequality. Since we knew the probability with which firm 5 could be identified as most efficient, we simply constructed a joint probability statement with this and a per comparison interval and selected the marginal confidence levels so that the Bonferroni inequality produced the desired joint confidence level. Here, since we were essentially certain that firm 5 was efficient, the joint probability statement essentially reduced to a single per comparison probability statement. In the rice farm data the per comparison intervals were conditional on firm 164 being efficient; we just assumed that this was the case. However, for the current data, we knew with "almost" certainty (99.9% certainty) that firm 5 was efficient, so our marginal statement essentially coincides with the per comparison statement, just as the MCB intervals coincide with the MCC intervals.

The marginal/per comparison intervals are contained in Table 9B. Again the actual standard errors were used, since we did not have to appeal to equicorrelatedness to get our critical values. As a general statement, the marginal (per comparison) intervals are comparable to the MCB intervals. Surprisingly, in many cases the marginal intervals are actually wider than the MCB intervals. This must reflect failure of the equicorrelatedness assumption underlying our MCB intervals, but it also is a reflection of the relative sizes of N and T in the data. To be more specific, consider the following expression for the standard error of the estimate $\hat{\alpha}_5 - \hat{\alpha}_j$:

$$se_{5,j} = [var(\hat{\alpha}_5) + var(\hat{\alpha}_j) - 2cov(\hat{\alpha}_5, \hat{\alpha}_j)]^{1/2}$$

$$= [2\sigma^2_v/T + \bar{x}_5 V(\hat{\beta})\bar{x}_5' + \bar{x}_j V(\hat{\beta})\bar{x}_j' - 2\bar{x}_5 V(\hat{\beta})\bar{x}_j']^{1/2}$$

When T is small and N is large (e.g. T = 6 as in the rice farm data), the term $2\sigma^2_v/T$ is large relative to the other three terms, so any differences between $\bar{x}_5 Var(\hat{\beta})\bar{x}_5'$, $\bar{x}_j Var(\hat{\beta})\bar{x}_j'$ and $\bar{x}_5 Var(\hat{\beta})\bar{x}_j'$ are unimportant. For MCB we assume $\bar{x}_5 Var(\hat{\beta})\bar{x}_5' = \bar{x}_j Var(\hat{\beta})\bar{x}_j' = \bar{x}_5 Var(\hat{\beta})\bar{x}_j'$, so these insignificant differences are ignored. When T is large, however, the term $2\sigma^2_v/T$ is small and the aforementioned differences may become significant. However, if we ignore this difference in MCB, then the standard error for MCB may be smaller than the standard error for the some of the marginal (per comparison) intervals. This is less of a problem when both N and T are large, because large N tends to shrink the $V(\hat{\beta})$ term so any differences in the $\bar{x}_i V(\hat{\beta})\bar{x}_j'$ term will become less pronounced.

In cases where equicorrelatedness of the $\hat{\alpha}_i$ cannot be assumed, there are some conservative MCB approximations available. Matejcik (1992) suggests techniques for adaptive MCB intervals that are robust to a generalization of the correlation matrix and compares their performance using computer simulation. These techniques are based on several MCC methods that are themselves robust and include: an MCC method based on Banjeree's Inequality due to Tamhane (1977), a procedure using a moment-based approximation to the Behrens-Fisher problem due to Tamhane (1977), a method using all-pairwise procedures due to Dunnett (1980), and his own technique based on a heteroscedastic selection procedure. An obvious line for further research is to examine the applicability of these techniques to stochastic frontier models.

4.3 Egyptian Tileries - Seale (1990)

We analyze data previously analyzed by Seale (1990). For a complete discussion of the data see Seale (1990). He observed 25 Egyptian small-scale floor tile manufacturers over 3 week periods for 66 weeks, for a total of 22 separate observation periods. The set contains some missing data points, so the number of separate observation periods varies across firms, making this an *unbalanced* panel. The data were collected by the Non-Farm Employment Project in 1982-1983. The firms were located in Fayoum and Kalyubiya, Egypt. Inputs to the production of cement floor tiles are labor (labor-hours) and machines (machine-hours). Output is in square meters of tile.

The model was estimated by OLS, fixed effects and CGLS. The third moment of the OLS residuals was positive, so MLE was not attempted. Estimation results are given in Table 10. It may be noted that σ^2_u and σ^2_v are of similar magnitude. For this reason, and because the number of firms is similar to the number of periods per firm (for most firms), this data set has characteristics that put it in the middle ground between the Erwidodo rice farm data set (N much larger than T, σ^2_v larger than σ^2_u) and the Kumbhakar utilities data set (T larger than N, σ^2_u larger than σ^2_v). We should expect confidence intervals wider than for the utilities but narrower than for the rice farms.

Table 11 gives the BC confidence intervals based on the CGLS estimates, for all firms. As a general statement, these confidence intervals are considerably wider than for the utility data. They are perhaps a little narrower than the confidence intervals for the rice farm data, but this is not entirely clear because the general level of efficiency is lower than it was for the rice farm data.

We next consider the MCB intervals. Because the panel is unbalanced, different $\hat{\alpha}_i$ are based on different numbers of observations, and we cannot expect the equicorrelated structure to hold. However, we can still proceed with MCB if the product structure of equation (1.12) holds. This structure held approximately, and so we calculated the MCB intervals, which are given in Table 12. As was the case for the BC results, the confidence intervals are generally narrower then those for the rice farm data but wider then those for the utility data.

MCC and per comparison intervals for the fixed effects estimation are contained in Tables 13 and 14, respectively. Once again, they are not very different from the MCB intervals.

5. CONCLUSIONS

In this chapter we have shown how to construct confidence intervals for efficiency estimates from stochastic frontier models. We have done so under a variety of assumptions that correspond to those made to calculate the efficiency measures themselves. For example, given distributional assumptions for statistical noise and inefficiency, the Jondrow-Lovell-Materov-Schmidt or Battese-Coelli estimates are typically used, and confidence intervals for these estimates are straightforward. With panel data but without distributional assumptions, efficiency estimates are commonly based on the fixed-effects intercepts, and confidence intervals follow from the statistical literature on multiple comparisons with the best.

In our analysis of three panel data sets, we found confidence intervals that were wider than we would have anticipated before this study began. The efficiency estimates

are more precise (and the confidence intervals are narrower) when T is large and when σ_u^2 is large relative to σ_v^2, and they are less precise when T is small and when σ_u^2 is small relative to σ_v^2. However, frankly, in all cases that we considered the efficiency estimates were rather imprecise. We suspect that, in many empirical analyses using stochastic frontier models, differences across firms in efficiency levels are statistically insignificant, and much of what has been carefully "explained" by empirical analysts may be nothing more than sampling error.

This is a fairly pessimistic conclusion, though it may turn out to be overly pessimistic when more empirical analysis is done. It is therefore important to stress that deterministic methods like DEA are not immune from this pessimism. Efficiency "measures" from DEA or other similar techniques are subject to the same sorts of uncertainty as are our estimates. The only difference is that we can clearly assess the uncertainty associated with our estimates while, at present, it is less clear how to assess the uncertainty associated with the DEA measures. In our opinion this should continue to be a high-priority item on the DEA research agenda.

CHAPTER THREE

ESTIMATION OF A CONDITIONAL VECTOR ERROR CORRECTION SUBMODEL FOR COUNTERFACTUAL POLICY ANALYSIS

1. INTRODUCTION

Within the last few years there have been several analyses performed that investigate the behavior of an economy under alternative sets of policy rules. These exercises, called counterfactual policy analyses, typically involve estimation of a VAR or VECM specification of an economy under a different set of monetary policy rules than the historical rules that generated the data. See for example McCallum (1988, 1990, 1993), Judd and Motley (1992, 1993) and Bordo, Choudri and Schwartz (1994). However, as Rasche (1995) points out "what is not acknowledged in these studies is that the technique which is applied to the VAR (VECM) to construct the counterfactual analysis severely limits the admissible structure of the economic model that could have generated the historical data".

Rasche identifies pitfalls associated with the practice of "VAR transplantation" where one a) estimates a reduced form VAR (VECM) in p variables, b) removes a subset of n equations representing the policy rule(s), c) supplements the remaining p-n equations with a new policy rule(s) and reanalyzes the system. He concludes that the usefulness of VAR transplantation is quite limited. Specifically, "the selected policy variable cannot enter contemporaneously into any of the equations of the economic structure except the policy rule". This is clearly entirely too restrictive.

In a subsequent note, Rasche proposes an alternative approach to the problem that overcomes the aforementioned "pitfall" of transplantation. He proposes structural (as opposed to reduced form) estimation of the remaining n-p equations using a subset limited information maximum likelihood (LIML) technique introduced by Rubin (1948) and further developed by Hood and Koopmans (1953) and Dhrymes (1970). This chapter is concerned with detailing this estimator which involves concentration of a multivariate Gaussian likelihood function. Under cointegration the resulting likelihood is the product of two generalized least-variance ratios, one of which is exactly the likelihood derived by Johansen (1988, 1991). The resulting estimator can be thought of as a generalization of Johansen's (1988, 1991) reduced form estimator to the structural case, where the analyst is only concerned with estimating a subset of the equations in the system.

This chapter is organized as follows. Section 2 derives the LIML estimator for counterfactual policy analysis under the cointegration hypothesis and then under no cointegration. Section 3 summarizes and concludes. Notation used in chapters one and two does not apply here, and vice versa.

2. ESTIMATION

2.1 Under Cointegration

Consider the following error-correction representation of a structural VAR model in p dimensions.

$$\Delta Y_t A + \Sigma_{i=1}^{q} \Delta Y_{t-i} \Gamma_i + Y_{t-1} \beta \alpha' + X_t \Theta = \epsilon_t \qquad t = 1, \ldots, T \tag{3.1}$$

where ϵ_t (t = 1, ..., T) are independent p-dimensional Gaussian variables with mean zero and variance matrix Σ, and X_t are s-dimensional instruments appearing only in the last n < p equations. The Y_t are cointegrated with β, the matrix of cointegrating vectors, and α', the matrix of error correction coefficients, being pxr matrices. The first q data points Y_0, ..., Y_{q-1} are considered fixed, while the parameters A, Γ_i, ..., Γ_q and Σ vary without restriction. The last n equations represent the historical policy rules, while the first m = p - n equations are the subset of interest. What is typically done in a transplantation analysis is to estimate the reduced form of equation (3.1) and discard the last n equations. We use limited information maximum likelihood techniques to estimate the first m equations in *structural form*.

Let $Y_{\bullet t} = [\Delta Y_{t-1}, \ldots, \Delta Y_{t-q}]$, $\Gamma_\bullet' = [\Gamma_i', \ldots, \Gamma_q']$, and for the moment let us suppress β by writing $Z_{t-1} = Y_{t-1}\beta$. Collecting t vertically, equation (3.1) becomes,

$$\Delta Y A + Y_\bullet \Gamma_\bullet + Z_{-1}\alpha' + X\Theta = \epsilon \tag{3.2}$$

We are interested in estimating the coefficients of the first m = p-n equations (a submodel). To this end we partition the coefficient matrices as follows,

$$A = [A_I \; A_{II}]; \qquad A_I \text{ (pxm)}, \; A_{II} \text{ (pxn)}$$

$$\Gamma_\bullet = [\Gamma_{\bullet 1} \; \Gamma_{\bullet 2}]; \qquad \Gamma_{\bullet 1} \text{ (mxq)}, \; \Gamma_{\bullet 2} \text{ (nxq)}$$

$\alpha' = [\alpha_1' \ \alpha_2']$ $\qquad$ α_1' (rxm), α_2' (rxn)

$\Theta = [0, C]$ $\qquad$ Θ (sxp), C (sxn)

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$ $\qquad$ Σ_{11} (mxm), Σ_{12} (mxn), Σ_{21} (nxm), Σ_{22} (nxn)

where C is unrestricted. The analyses that follow also require the following data matrices:

$W = [Y_\bullet \ X]; \qquad F = [Y_\bullet \ X \ Z_{-1}]; \qquad J = [Y_\bullet \ Z_{-1}];$

$N = [\Delta Y \ Y_\bullet \ X \ Z_{-1}];$

$R_{rs} = T^{-1}r's \qquad r, s = \Delta Y, Y_\bullet, Z_{-1}, X, F, J, N;$

$\lambda^i = R_{\Delta Y \Delta Y} - R_{\Delta Yi}(R_{ii})^{-1}R_{i\Delta Y} \qquad i = F, J.$

We use a subset LIML technique to transform equation (3.2) to allow concentration of the coefficients of the last n equations, while leaving the coefficients of the first m equations undisturbed and ensuring that the two subsystems are mutually

independent. This approach provides us with estimates of the structural parameters of the m equations of interest, while purging the subsystem of the stochastic effects of the n policy rules. This idea of subset LIML was first considered by Rubin (1948) and Koopmans and Hood (1953). For a complete explanation of this technique, see Dhrymes (1970). The resulting "concentrated" log-likelihood function per Dhrymes (equation 7.3.49, p335) is given by:

$$(3.3)\quad (-2/T)\ln L = c_0 + \ln |\lambda^F| + \ln |\Sigma_{11}| - \ln |A_1'\lambda^F A_1| + tr\Sigma_{11}^{-1} M'R_{NN}M$$

where $M' = [A_1' \; \Gamma_{\bullet 1}' \; 0' \; \alpha_1]$. To concentrate $\Gamma_{\bullet 1}$ and α_1, we examine the last term of equation (3.3). Specifically,

$$M'R_{NN}M = [A_1' \; \Gamma_{\bullet 1}' \; 0' \; \alpha_1]'(N'N/T)[A_1' \; \Gamma_{\bullet 1}' \; 0' \; \alpha_1]'$$

$$= [A_1' \; \Gamma_{\bullet 1}' \; \alpha_1]' \begin{bmatrix} R_{\Delta Y \Delta Y} & R_{\Delta YJ} \\ R_{J\Delta Y} & R_{JJ} \end{bmatrix} [A_1' \; \Gamma_{\bullet 1}' \; \alpha_1]'.$$

Let $\delta' = [\Gamma_{\bullet 1}' \; \alpha_1]$, then

$$(3.4)\quad M'R_{NN}M = A_1'R_{\Delta Y \Delta Y}A_1 + A_1'R_{\Delta YJ}\delta + \delta'R_{J\Delta Y}A_1 + \delta'R_{JJ}\delta.$$

Substituting equation (3.4) into (3.3) and taking the derivative w.r.t δ

$$\partial \ln L/\partial \delta = -(T/2)[2R_{J\Delta Y}A_I + 2R_{JJ}\delta]\Sigma_{11}^{-1} = 0$$

implying

(3.5) $$\delta = -(R_{JJ})^{-1}R_{J\Delta Y}A_I.$$

Substituting equation (3.5) into (3.4),

(3.6) $$M'R_{NN}M = A_I'[R_{\Delta Y\Delta Y} - R_{\Delta YJ}(R_{JJ})^{-1}R_{J\Delta Y}]A_I = A_I'\lambda^J A_I$$

Substituting equation (3.6) into (3.3) and taking the derivative w.r.t Σ_{11}

$$\partial \ln L/\partial \Sigma_{11} = -(T/2)[-\Sigma_{11}^{-1} + \Sigma_{11}^{-1} A_I'\lambda^J A_I \Sigma_{11}^{-1}] = 0,$$

implying

(3.7) $$\Sigma_{11} = A_I'\lambda^J A_I .$$

Substituting equations (3.6) and (3.7) into (3.3),

$$(-2/T)\ln L = c_1 + \ln | \lambda^F | + \ln | A_I'\lambda^J A_I | - \ln | A_I'\lambda^F A_I |$$

and suppressing c_1

$$(3.8) \qquad L^{-2/T}(A_I) = \frac{| A_I'\lambda^J A_I |}{| A_I'\lambda^F A_I |} | \lambda^F |$$

If β is known a priori, then equation (3.8) represents the final likelihood function. However, its form precludes a first derivative solution for A_I. Fortunately, the function is a generalized least variance ratio, solvable using the usual canonical techniques. We defer its solution until latter. It is important to notice that the likelihood function in equation (3.8) is only attainable because of the restrictions on Θ; without them A_I is not identified. Now, in general, β is unknown so we make it explicit in our definition of λ^F. First, let $F = [W\ Y_{-1}\beta]$, so,

$$R_{FF} = (1/T) \begin{bmatrix} W'W & W'Y_{-1}\beta \\ \beta'Y_{-1}'W & \beta'Y_{-1}'Y_{-1}\beta \end{bmatrix}$$

Let

$$R_{FF}^{-1} = \begin{bmatrix} R_{FF}^{11} & R_{FF}^{12} \\ R_{FF}^{21} & R_{FF}^{22} \end{bmatrix}$$

Then

$$FR_{FF}^{-1}F' = WR_{FF}^{11}W' + WR_{FF}^{12}\beta'Y_{-1}' + \beta Y_{-1}R_{FF}^{21}W' + \beta Y_{-1}R_{FF}^{22}Y_{-1}'\beta'$$

Defining the projection matrix:

$$P_i = i(i'i)^{-1}i'; \qquad i = W, J, F$$

and the sum of squared error matrix:

$$S_{ij}^{r} = (1/T)i'[I - P_r]j; \qquad r = W, J, F; \qquad i, j = \Delta Y, Y_{-1}, X$$

and using the rules of partitioned inverse,

$$\begin{aligned} FR_{FF}^{-1}F' &= (1/T)\{P_W + P_W Y_{-1}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'Y_{-1}'P_W \\ &\quad - Y_{-1}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'Y_{-1}'P_W - P_W Y_{-1}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'Y_{-1}' \\ &\quad + Y_{-1}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'Y_{-1}'\} \end{aligned}$$

$$(3.9) \qquad FR_{FF}^{-1}F' = (1/T)\{P_W + (I - P_W)Y_{-1}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'Y_{-1}'(I - P_W)\}$$

We can write $\lambda^F = \Delta Y'[I/T - FR_{FF}^{-1}F']\Delta Y$. Substituting equation (3.9) into this expression,

$$(3.10) \qquad \lambda^F(\beta) = S_{\Delta Y\Delta Y}^{W} - S_{\Delta YY-1}^{W}\beta[\beta'S_{Y-1Y-1}^{W}\beta]^{-1}\beta'S_{Y-1\Delta Y}^{W}$$

Using results for the determinant of a partitioned matrix, equation (3.10) yields:

$$(3.11)\quad |\lambda^F(\beta)| = |S^W_{\Delta Y \Delta Y}| \frac{|\beta'\{S^W_{Y-1Y-1} - S^W_{Y-1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY-1}\}\beta|}{|\beta' S^W_{Y-1Y-1}\beta|}$$

Substituting equation (3.11) into (3.8) and making β explicit in λ^J,

$$(3.12)\quad L^{-2/T}(A_I, \beta) = |S^W_{\Delta Y \Delta Y}| \frac{|A_I'\lambda^J(\beta)A_I|}{|A_I'\lambda^F(\beta)A_I|} \frac{|\beta'\{S^W_{Y-1Y-1} - S^W_{Y-1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY-1}\}\beta|}{|\beta' S^W_{Y-1Y-1}\beta|}$$

Equation (3.12) is the product of two generalized least variance ratios: one in A_I and β and another in β only. The ratio in β only is the ratio derived by Johansen (1991). So this problem can be thought of as a generalization of the Johansen likelihood. Indeed, if we exclude instruments, X, from the specification, $\lambda^F = \lambda^J$, A_I is not identified and the problem reduces to the Johansen result. This raises the question of how many instruments are required for identification. We defer the answer to this question until later. Notice that

$$\lambda^F = (1/T)[\Delta Y'\Delta Y - \Delta Y'F(F'F)F'\Delta Y]$$

$$\lambda^F = (1/T)\Delta Y'[I - F(F'F)F']\Delta Y$$

$$\lambda^F = S^F_{\Delta Y \Delta Y}$$

and by similar logic $\lambda^J = S^J_{\Delta Y \Delta Y}$.

From the rules of projection matrices

$$I - P_F = I - P_J - (I - P_J)X[X'(I - P_J)X]^{-1}X'(I - P_J)$$

Premultiplying by $\Delta Y'/T$ and postmultiplying by ΔY

$$\lambda^F = \lambda^J - S^J_{\Delta YX}[S^J_{XX}]^{-1}S^J_{X\Delta Y}$$

Let $$\phi = S^J_{\Delta YX}[S^J_{XX}]^{-1}S^J_{X\Delta Y}$$

then, making β explicit.

(3.13) $$\lambda^J(\beta) = \lambda^F(\beta) + \phi(\beta)$$

Substituting (3.13) into equation (3.12) and suppressing the $| S^W_{\Delta Y\Delta Y} |$ term,

(3.14) $$L^{-2/T}(A_I, \beta) =$$

$$\frac{| A_I'\{\lambda^F(\beta) + \phi(\beta)\}A_I |}{| A_I'\lambda^F(\beta)A_I |} \quad \frac{| \beta'\{S^W_{Y\text{-}1Y\text{-}1} - S^W_{Y\text{-}1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1} S^W_{\Delta YY\text{-}1}\}\beta |}{| \beta'S^W_{Y\text{-}1Y\text{-}1}\beta |}$$

$$= f(A_I, \beta)g(\beta)$$

We now have symmetry in the two variance ratios and would like to minimize this function, so the problem can be thought of as two eigenvalue problems. The double

eigenvalue problem of equation (3.14) can be transformed to a single block diagonal eigenvalue problem, which can generate a block diagonal solution. Define:

$$G = \begin{bmatrix} A_I & 0 \\ 0 & \beta \end{bmatrix}$$

$$D = \begin{bmatrix} \lambda^F(\beta) + \phi(\beta) & 0 \\ 0 & S^W_{Y\text{-}1Y\text{-}1} - S^W_{Y\text{-}1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY\text{-}1} \end{bmatrix}$$

$$B = \begin{bmatrix} \lambda^F(\beta) & 0 \\ 0 & S^W_{Y\text{-}1Y\text{-}1} \end{bmatrix}$$

then equation (3.14) can be written as:

$$(3.15)\quad L^{-2/T}(A_I, \beta) = |G'DG| \, |G'BG|^{-1}$$

We now discuss solutions to equations (3.14) and (3.15) using the LIML estimator and a 2-step estimator. LIML estimation of the parameters β and A_I involves minimization of equation (3.14). Equivalently, it involves minimization of equation (3.15) subject to the block diagonality of G. The 2-step estimator involves i) estimating β by minimizing $g(\beta)$ into equation (3.14), then ii) using this estimate to minimize $f(A_I, \hat{\beta})$ with respect to A_I. Under certain conditions these estimators are equivalent.

2.1.1 LIML Estimation

Ideally we would solve equation (3.15) for G using the usual multivariate techniques from the theory of partial canonical correlations and reduced rank regression (see Anderson (1951) and Tso (1981)). This is equivalent to solving for A_I and β simultaneously in equation (3.14). Since matrices D and B are functions of β, the usual techniques to produce a solution are not applicable; some numerical optimization method must be utilized to obtain a solution. Another complication is there are restrictions on G and, ultimately, on any estimate, $\hat{G}$. Specifically, the upper-right and lower-left blocks of the matrix G are zero. This, we shall see, causes the value of the likelihood in equation (3.15), $L^{-2/T}(A_I, \beta)$, to be larger than if G were allowed to vary without restrictions.

For now assume B and D are not fuctions of β. We can use Theorem 1 of Anderson (1951) to solve equation (3.15), yielding an estimate of the space spanned by G, $\hat{G} = [\hat{\theta}_1, \ldots, \hat{\theta}_r, \hat{\theta}_{p+1}, \ldots, \hat{\theta}_{p+m}]$, where $\hat{\Theta} = [\hat{\theta}_1, \ldots, \hat{\theta}_{2p}]$ are the eigenvectors of the equation

$$(3.16) \qquad | \sigma B - D | = 0.$$

normed by $\hat{\Theta}'B\hat{\Theta} = I$ and ordered $0 < \hat{\sigma}_1 < \ldots < \hat{\sigma}_r < \hat{\sigma}_{r+1} < \ldots < \hat{\sigma}_p < \hat{\sigma}_{p+1} \leq \ldots \leq \hat{\sigma}_{p+m} \leq \hat{\sigma}_{p+m+1} \leq \ldots \leq \hat{\sigma}_{2p}$. Upon substituting matrices for B and D, equation (3.16) becomes

$$\text{Det}\begin{bmatrix} (\sigma\text{-}1)\lambda^F - \phi & 0 \\ 0 & (\sigma\text{-}1)S^W_{Y\text{-}1Y\text{-}1} + S^W_{Y\text{-}1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY\text{-}1} \end{bmatrix} = 0$$

Multiplying by $\text{Det}\begin{bmatrix} I_p & 0 \\ 0 & -I_p \end{bmatrix} = (-1)^p \neq 0$

yields

$$\text{(3.17)} \quad \text{Det}\begin{bmatrix} (\sigma\text{-}1)\lambda^F - \phi & 0 \\ 0 & (1\text{-}\sigma)S^W_{Y\text{-}1Y\text{-}1} - S^W_{Y\text{-}1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY\text{-}1} \end{bmatrix} = 0$$

To minimize equation (3.15), subject to the restriction that $\hat{G}$ be block diagonal, we select the two sets of eigenvalues ranked $\hat{\sigma}_1 < \ldots < \hat{\sigma}_r$ and $\hat{\sigma}_{p+1} \leq \ldots \leq \hat{\sigma}_{p+m}$ and use their corresponding eigenvectors to form $\hat{G}$. Selecting the eigenvalues in this way is necessary to ensure satisfaction of the block diagonal restrictions on G, while ensuring a solution to equation (3.16) and minimization of the likelihood in equation (3.15). We now justify the selection of these specific eigenvalues.

For any σ_i, which is a solution to equation (3.17),

$$\text{(3.18)} \qquad D\hat{\theta}_i = \sigma_i B\hat{\theta}_i, \text{ or}$$

$$\text{(3.19)} \qquad (\sigma_i B - D)\hat{\theta}_i = 0.$$

Partition the eigenvectors so $\hat{\theta}_i' = [\hat{\theta}_{i,I}' \mid \hat{\theta}_{i,II}']$, and substituting for B and D, equation (3.19) becomes

$$(3.20)\quad \begin{bmatrix} [(\sigma_i-1)\lambda^F - \phi]\hat{\theta}_{i,I} \\ [(\sigma_i-1)S^W_{Y-1Y-1} + S^W_{Y-1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY-1}]\hat{\theta}_{i,II} \end{bmatrix} = 0$$

Premultiply by $\begin{bmatrix} I_p & 0 \\ 0 & -I_p \end{bmatrix}$

to get

$$\begin{bmatrix} [(\sigma_i-1)\lambda^F - \phi]\hat{\theta}_{i,I} \\ [(1-\sigma_i)S^W_{Y-1Y-1} - S^W_{Y-1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY-1}]\hat{\theta}_{i,II} \end{bmatrix} = 0.$$

This is solved when σ_i satisfies:

$$(3.21)\ (a)\qquad |\,(\sigma-1)\lambda^F - \phi\,| = 0 \quad \underline{\text{or}}$$

$$(b)\qquad |\,(1-\sigma)S^W_{Y-1Y-1} - S^W_{Y-1\Delta Y}[S^W_{\Delta Y\Delta Y}]^{-1}S^W_{\Delta YY-1}\,| = 0$$

Assume λ^F positive definite and ϕ positive semi-definite; then from equation (3.21a) it is true that $(\sigma_i-1) \geq 0$ (Dhrymes (1978), Proposition 62), implying $\sigma_i \geq 1$. There will be p of these eigenvalues. Also from equation (3.21b), it is clear that if

$S^{W}_{Y\text{-}1\Delta Y}[S^{W}_{\Delta Y\Delta Y}]^{-1}\,S^{W}_{\Delta YY\text{-}1}$ and $S^{W}_{Y\text{-}1Y\text{-}1}$ are positive definite, then $(1\text{-}\sigma_i) > 0$, implying $\sigma_i < 1$. There will be p of these eigenvalues. So, any single σ_i cannot simultaneously satisfy both parts of equation (3.21) (i.e. $\sigma_i \geq 1$ and $\sigma_i < 1$). So, the p eigenvalues that solve equation (3.21a) are distinct from those p eigenvalues that solve equation (3.21b). Additionally, it is important to point out that the p eigenvalues that solve equation (3.21a) will have p eigenvectors of the form $\hat{\theta}_i' = [\hat{\theta}_{i,I}' \mid 0']$, while the p eigenvalues that solve equation (3.21b) will have p eigenvectors of the form $\hat{\theta}_i' = [0' \mid \hat{\theta}_{i,II}']$. This is due to the block diagonal form of matricies B and D.

Now, we must select σ_i and the corresponding $\hat{\theta}_i$, to ensure that the likelihood of equation (3.15) is minimized while ensuring that the block diagonal restrictions on G are not violated. Theorem 1 of Anderson (1951) shows that to minimize equation (3.15) where the estimate of G is allowed to vary unrestricted, we select the m+r smallest eigenvalues and let their corresponding eigenvectors form our estimate of G. However, G is restricted block diagonal so some modification of this procedure is necessary. Consider four sets of eigenvalues. Consider $\hat{\sigma}_1 < \ldots < \hat{\sigma}_r$, the smallest eigenvalues, which are constrained between 0 and 1 by construction and which solve equation (3.21b). To maximize the likelihood we select these smallest eigenvalues and their eigenvectors which are of the form $\hat{\theta}_i' = [0' \mid \hat{\theta}_{i,II}']$, for all i = 1, ..., r. These r eigenvectors form the basis for the β estimate. Now, consider $\hat{\sigma}_{r+1} < \ldots < \hat{\sigma}_p$, the second smallest eigenvalues, which are also constrained between 0 and 1. Ideally, we would select these next largest eigenvalues and their eigenvectors to minimize equation (3.15), but these eigenvectors will not form a basis for an estimate of A_1, since they are of the form

$\hat{\theta}_i' = [0' \mid \hat{\theta}_{i,II}']$, for all i = r+1, ..., p. The form precludes an estimate of A_I, because of the position of the 0. Hence, none of these eigenvalues suffice. Consider $\hat{\sigma}_{p+1} \leq \ldots \leq \hat{\sigma}_{p+m}$, the next m largest eigenvalues. We select these eigenvalues and associated eigenvectors which are of the form $\hat{\theta}_i' = [\hat{\theta}_{i,I}' \mid 0']$, for all i = p+1, ..., p+m, forming a basis for A_I. The last set of eigenvalues ($\hat{\sigma}_{p+m+1} \leq \ldots \leq \hat{\sigma}_{2p}$) is irrelevant.

This procedure produces a likelihood function

$$L^{-2/T}(r, m, p) = \Pi_{i=1}^{r} \hat{\sigma}_i \quad \Pi_{j=p+1}^{p+m} \hat{\sigma}_j. \tag{3.22}$$

Equation (3.22) illustrates the effect of the block diagonal restriction. We can think of the ordering of the eigenvalues as four ranked sets. The set of lowest ordered eigenvalues ($\hat{\sigma}_1$ to $\hat{\sigma}_r$) has r members. The set of second lowest eigenvalues ($\hat{\sigma}_{r+1}$ to $\hat{\sigma}_p$) has p-r members. The set of third lowest eigenvalues ($\hat{\sigma}_{p+1}$ to $\hat{\sigma}_{p+m}$) has m members. The set of largest eigenvalues ($\hat{\sigma}_{p+m+1}$ to $\hat{\sigma}_{2p}$) has n members. Selecting restricted $\hat{G}$ is equivalent to selecting the set of smallest eigenvalues and the set of third smallest eigenvalues. Examining the likelihood of equation (3.22) we see that its value is the product of the eigenvalues of these selected sets. The first set of smallest eigenvalues is optimal in the sense that they minimize $L^{-2/T}(r, m, p)$, but the set of third smallest eigenvalues is a deviation from the usual cononical techniques and reflects the effect of the block diagonal restriction on $\hat{G}$.

The assumption of positive semi-definiteness for the matrix ϕ will be made clear. Consider the rank of ϕ. When rank(ϕ) = 0, there are p eigenvalues equal to 1, and they

represent the p largest eigenvalues. This causes the selection of the set of third smallest eigenvalues (and their corresponding eigenvectors) to be ambiguous, so $\hat{G}$ is not uniquely defined. When $0 < \text{rank}(\phi) < n$, there are more than m eigenvalues equal to 1, and again $\hat{G}$ is not uniquely defined. When $\text{rank}(\phi) = n$, there are exactly m eigenvalues equal to 1 and $\hat{G}$ is unique. When $\text{rank}(\phi) > n$, there are less than m eigenvalues equal to 1, and $\hat{G}$ is unique.

Now relax the assumption that B and D are not functions of β. Consider the case where, $\text{rank}(\phi(\beta)) = n$. Not only does this produce a unique $\hat{G}$, but all the eigenvalues comprising the A_I portion of G will be unity and, hence, independent of β. The corresponding vectors of the A_I portion of $\hat{G}$ will, in general, not be independent of β. However, a two-stage estimate consisting of first estimating β in equation (3.21b), then using this estimate in equation (3.21a) to solve for an estimate of A_I is equivalent to LIML when $\text{rank}(\phi(\beta)) = n$. This can be shown by examining equation (3.20). To minimize equation (3.15) while observing the restrictions on G, we select the set of first smallest eigenvalues and the set of third smallest eigenvalues. As we have seen, the two sets of eigenvalues have no common elements. In this sense their selections are independent. This independence causes the problem to be separable, insofar as equation (3.20) can be solved by independently solving the two characteristic equations (3.21). The only complication is the fact that λ^F and ϕ are functions of β. So, finding the eigenvectors of equation (3.20) reduces to finding the eigenvectors of equation (3.21), namely $\hat{\theta}_{i,I}$, $i = p+1, \ldots, p+m$ and $\hat{\theta}_{i,II}$, $i = 1, \ldots, r$. Suppose we are interested in the estimate of β only and solve equation (3.21b) to get the set of smallest eigenvalues and $\hat{\theta}_{i,II}$, $i = 1, \ldots, r$. This selection is optimal in the sense that these eigenvalues ensure that

the likelihood is maximized. Now suppose we use this β estimate to solve equation (3.21a), and based on this estimate, rank(ϕ) = n. Then the set of third smallest eigenvalues will be optimal in the sense that these eigenvalues ensure that the likelihood is maximized, because they all equal their lower bound of 1. Thus, both sets of eigenvalues are optimal. The estimate of A_I consists of the same $\hat{\theta}_{i,I}$, i = p+1, ..., p+m, as before. Thus, when rank($\phi(\beta)$) = n, this two stage procedure produces a maximized likelihood function, which is equal to the likelihood of the LIML estimate.

The equality of likelihoods of this two-stage estimate and the LIML estimate also occurs when rank($\phi(\beta)$) = 0 and when rank($\phi(\beta)$) < n, but in these cases $\hat{G}$ is not uniquely defined in the sense that the selection of the likelihood-maximizing eigenvalues is ambiguous. When rank($\phi(\beta)$) > n, $\hat{G}$ is unique, but the selection of the estimate of A_I may not be optimal. Therefore, the case when rank($\phi(\beta)$) = n is of particular interest. Specifically, when rank($\phi(\beta)$) = n, the LIML estimator of equation (3.15) is equivalent to a two-stage estimator described in the following section.

2.1.2 A Two-stage Estimator

We investigate an estimator based on a likelihood function conditional on r in two steps when rank($\phi(\beta)$) = n, as follows. First minimize the variance ratio, g(β), using the usual Johansen (1991) technique. This yields $\hat{\beta}_r = [\hat{v}_1, \ldots, \hat{v}_r]$, where $\hat{V} = [\hat{v}_1, \ldots, \hat{v}_p]$ are the eigenvectors of the equation

$$| \rho S^{W}_{Y-1Y-1} - S^{W}_{Y-1\Delta Y} [S^{W}_{\Delta Y\Delta Y}]^{-1} S^{W}_{\Delta YY-1} | = 0,$$

normed by $\hat{V}'S^{W}_{Y-1Y-1}\hat{V} = I$ and ordered $1 > \hat{\rho}_1 > \ldots > \hat{\rho}_p > 0$. Then the likelihood is maximized by the usual eigenvalue product, and equation (3.14) becomes

$$(3.23) \qquad L^{-2/T}(A_I, r \mid \hat{\beta}_r) = \frac{\mid A_I'\{\lambda^F(\hat{\beta}_r) + \phi(\hat{\beta}_r)\}A_I \mid}{\mid A_I'\lambda^F(\hat{\beta}_r)A_I \mid} \quad \Pi_{i=1}^{r}(1 - \hat{\rho}_i)$$

$$= f(A_I, \hat{\beta}_r)g(\hat{\beta}_r)$$

This procedure is that of Johansen, and (under additional restrictions) produces consistent estimates of the eigenvalues ρ_i, the dimension of the cointegrating space, $sp(\beta)$, and the estimates of α and Σ. As proved in Johansen (1988), the restrictions necessary for consistency are: 1) Y_t is integrated of order 1, and 2) α and β are full column rank. That is, under these restrictions,

$$plim(\hat{\rho}_1, \ldots, \hat{\rho}_r) = \rho_1, \ldots, \rho_r \qquad \text{and} \qquad plim(sp(\hat{\beta}_r)) = sp(\beta).$$

Additionally, Hargreaves (1994) points out that in the probability limit all the r eigenvalues of the cointegrating space are distinct.

In a fashion analogous to Johansen (1991) we can find $\hat{A}_I = [\hat{\tau}_1^r, \ldots, \hat{\tau}_m^r]$, where $\hat{U} = [\hat{\tau}_1^r, \ldots, \hat{\tau}_p^r]$ are the eigenvectors of the equation

$$\mid \omega\lambda^F(\hat{\beta}_r) - \phi(\hat{\beta}_r) \mid = 0$$

normed by $\hat{U}'\lambda^F(\hat{\beta}_r)\hat{U} = I$ and ordered $\hat{\omega}_p^r \geq \ldots \geq \hat{\omega}_{m+1}^r > \hat{\omega}_m^r = \hat{\omega}_{m-1}^r =, \ldots, \hat{\omega}_1^r = 0$. Notice that the eigenvalues and eigenvectors are functions of r, the number of cointegrating vectors. This reflects the fact that the likelihood value is conditional on $\hat{\beta}_r$. Also the order statistic for $\hat{\omega}_j^r$ is the reverse of that for $\hat{\rho}_i$ ($\hat{\omega}_p^r$ is the largest eigenvalue while $\hat{\rho}_p$ is the smallest), because $\lambda^F(\hat{\beta}_r)$ and $\phi(\hat{\beta}_r)$ are summed in $f(A_1, \hat{\beta}_r)$, while the analogous matrices are differenced in $g(\beta)$. The maximized likelihood becomes

$$(3.24) \qquad L^{-2/T}(m, r) = \Pi_{j=1}^{m}(1 + \hat{\omega}_j^r)\,\Pi_{i=1}^{r}(1 - \hat{\rho}_i)$$

When $\text{rank}(\phi(\hat{\beta}_r)) = n$, the m smallest eigenvalues selected, $\hat{\omega}_m^r, \ldots, \hat{\omega}_1^r$, all equal zero for all values of r, because $\phi(\hat{\beta}_r)$ is not full-rank. This particular rank condition reduces the problem to that of Johansen insofar as the likelihood function of equation (3.24) has $\Pi_{j=1}^{m}(1 + \hat{\omega}_j^r) = 1$ and becomes:

$$L^{-2/T}(m, r) = \Pi_{i=1}^{r}(1 - \hat{\rho}_i)$$

Since this two-stage procedure is equivalent to the LIML estimator when $\text{rank}(\phi(\beta)) = n$, we would like to derive the asymptotic properties of the estimates when this condition holds. As previously stated, under certain conditions $\text{sp}(\phi(\hat{\beta}_r))$ is consistent. We claim that the space spanned by $\hat{A}_1$ is also consistent. In lieu of a proof, we present the following argument for why this may be. Johansen (1988) shows the T-consistency of $\text{sp}(\phi(\hat{\beta}_r))$ in the presence of short-run dynamics (Y.Γ.). It is reasonable to believe that if the instruments, X_t, are stationary, then the same result of T-

consistency of $sp(\phi(\hat{\beta}_r))$ will hold in this analysis. Let us take this fact as given. Since Johansen shows this in the presence of short-run dynamics it is reasonable to believe that an estimate of a short-run dynamic parameter, A_I, would also be consistent when it is based on a T-consistent estimate of an unknown long-run dynamic parameter, β. This claim is supported by the fact that the usual LIML estimate of a structural parameter, say A_I, is $T^{1/2}$-consistent (Dhrymes (1970), p335), while $sp(\phi(\hat{\beta}_r))$ (the estimate upon which A_I estimate is based) is T-consistent. This claim is further supported by the fact that in Johansen (1988) the T-consistent estimate of $sp(\phi(\hat{\beta}_r))$ yields consistent estimates of reduced form parameters in that particular model. Formal proof of this claim will be undertaken in future research.

2.2 Under No Cointegration

If there are no cointegrating relationships, then the problem reduces to subset LIML, and the likelihood function is that of equation (3.3). If the same restrictions on the Θ hold then the likelihood can be concentrated to the generalized least variance ratio of equation (3.8) but with β suppressed. Suppressing the $|\lambda^F|$ term, this ratio becomes

$$(3.25) \qquad L^{-2/T}(A_I) = \frac{|A_I'\{\lambda^F + \phi\}A_I|}{|A_I'\lambda^F A_I|}$$

which can be solved as described above. The generalized least variance ratio of equation (3.25) was first considered by Rubin (1948) and Koopmans and Hood (1953), but to our

knowledge has never been solved using canonical correlations. If we have enough instruments then, the maximized value of the likelihood is

$$(3.26) \qquad L^{-2/T}(m) = \Pi_{j=1}^{m}(1 + \hat{\omega}_j).$$

Notice that the $\hat{\omega}_j$ are no longer a function of r. In the case where $m = 1$, this problem reduces to single equation LIML which is identical to two-stage least squares (2SLS) under exact identification. So, an equally valid approach to this estimation when $m > 1$ would be to estimate each of m single equations separately using 2SLS. Then a meaningful question would be whether there any efficiency gains to the methods detailed herein over the single equation approach. We plan to explore this question in future research.

3. CONCLUSIONS

The structural estimates derived in the preceding section for β and A_1 can be used to derive estimates for the other structural parameters of the submodel of interest by substituting them back through the derivation. The resulting estimates are unique in that they permit counterfactual policy analysis without the structural limitations presented by Rasche (1995). However, this has come at a cost. Since the estimates are based on limited information (i.e. we have concentrated out the parameters of the last n equations), they are inefficient, but this is the price that is paid to ensure the mutual independence of the two subsystems.

We present a subset LIML estimator which is amenable to a numerical optimization technique such as Newton-Raphson. An obvious starting value for β in such an approach would be the estimate produced from the Johansen analysis of the reduced form of equation (3.1). If rank($\phi(\beta)$) equals n, the number of policy equations, then the LIML solution is equal to the two-stage estimator of section 2.1.2. If rank($\phi(\beta)$) > n, the two estimators are, in general, not equal.

The two-stage estimate involves performance of a modified Johansen (1988) technique involving the addition of stationary instruments in the first stage. This produces a consistent estimate of the matrix of cointegrating vectors, β. In the second stage A_1 is estimated using similar canonical methods. Another way to perform this analysis would be to perform single equation LIML on each of the m equations in the subset, given a first stage estimate of β. However we suspect that this would be less efficient than the subset LIML estimate derived here.

The preceding analysis represents an initial attempt to solve the problems inherent in counterfactual policy analysis. It does not profess to be comprehensive, but provides the fodder for additional research. While there are many unanswered questions, a few merit brief consideration here. First, equation (3.22) suggests likelihood ratio testing of the number of cointegrating vectors that would be similar in structure to that of Johansen. We would expect that such tests in this context to yield results consistent with the reduced form tests of Johansen. If these tests are not consistent, what does this imply about the tests and for the overall modelling approach? This, of course, remains to be seen. Second, will the estimator of β produced in the numerical analysis of the LIML

estimator be similar in magnitude and dimension to that of Johansen's reduced form estimate? If they are not, what can be inferred? Clearly there is much left to be done.

on te

mode

with

natur

best

diffe

quar

we

prec

rela

wer

fro

and

mo

to

wit

esti

inte

CONCLUSIONS TO THE DISSERTATION

The primary goal of this dissertation has been to develop a method for inference on technical efficiency estimates in the fixed effect formulation of the stochastic frontier model. We find that under certain conditions the methodology of multiple comparisons with the best is a useful tool in performing this inference. Additionally, the simultaneous nature of the resulting intervals provides us with the added benefit of inference on the best firm in the sample; a benefit not found in existing interval construction techniques.

Secondly, our comprehensive empirical analysis attempted to disentangle the different sources of uncertainty in efficiency estimates and to provide us with a quantification of this uncertainty. We found confidence intervals that were wider than we would have anticipated before this study began. The efficiency estimates are more precise (and the confidence intervals are narrower) when T is large and when σ_u^2 is large relative to σ_v^2. However, frankly, in all cases that we considered the efficiency estimates were rather imprecise. We suspect that, in many empirical analyses using stochastic frontier models, differences across firms in efficiency levels are statistically insignificant, and much of what has been carefully "explained" by empirical analysts may be nothing more than sampling error. This is a fairly pessimistic conclusion, though it may turn out to be overly pessimistic when more empirical analysis is done.

Lastly, we provide an initial attempt at solving the counterfactual policy "pitfall" with a subset LIML estimator. What comes out of the analysis are a seemingly useful estimator and several questions about the characteristics of this estimator. It is our intention to address these questions in subsequent research.

APPENDIX

APPENDIX

TABLE 1

Rice Farms - Cross Sectional Estimation Results

Variable	Period 1 COLS	Period 1 MLE	Period 6 COLS	Period 6 MLE
Constant*	5.8483	5.9540	5.3327	5.5525
Seed	0.0572	0.0583	0.0983	0.0975
Urea	0.1036	0.1028	0.2073	0.2105
TSP	0.0033	0.0034	0.0674	0.0693
Labor	0.1970	0.1970	0.2122	0.2007
Land	0.6372	0.6374	0.5194	0.5252
DP	0.0143	0.0138	-0.1480	-0.1453
DV1	-0.0857	-0.0861	0.0639	0.0568
DV2	0.0853	0.0853	0.1192	0.1175
DR1	0.2192	0.2173	-0.2642	-0.2695
DR2	-0.0325	-0.0330	-0.3744	-0.3712
DR3	0.1388	0.1385	-0.3815	-0.3712
DR4	0.0814	0.0817	-0.0613	-0.0591
DR5	0.1810	0.1829	-0.2864	-0.2668
$\hat{\sigma}^2_u$	0.0185	0.0174	0.0253	0.0462
$\hat{\sigma}^2_v$	0.0632	0.0579	0.0705	0.0564
$E(u_i)$	0.1084	0.1053	0.1268	0.1714
$\hat{\sigma}_*$	0.1195	0.1157	0.1364	0.1593

* The constant term reported for COLS is before correction by $E(u_i)$.

APPENDIX

TABLE 2A

Rice Farms - Confidence Intervals Based on JLMS Method, COLS Estimates - Period 1

Firm	Efficiency	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	0.9421	0.8242	0.9981	0.8482	0.9962	0.8843	0.9902
118	0.9390	0.8173	0.9980	0.8417	0.9959	0.8786	0.9895
108	0.9380	0.8151	0.9979	0.8396	0.9958	0.8768	0.9892
92	0.9029	0.7514	0.9956	0.7777	0.9912	0.8196	0.9782
87	0.9027	0.7511	0.9955	0.7774	0.9911	0.8193	0.9781
61	0.9026	0.7511	0.9955	0.7774	0.9911	0.8193	0.9781
79	0.8479	0.6811	0.9873	0.7067	0.9760	0.7484	0.9479
16	0.8441	0.6769	0.9863	0.7024	0.9744	0.7440	0.9450
143	0.8066	0.6398	0.9727	0.6642	0.9531	0.7042	0.9129

TABLE 2B

Rice Farms - Confidence Intervals Based on JLMS Method, MLE Estimates - Period 1

Firm	Efficiency	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	0.9452	0.8322	0.9982	0.8553	0.9964	0.8901	0.9908
118	0.9421	0.8253	0.9981	0.8489	0.9961	0.8845	0.9901
108	0.9411	0.8232	0.9980	0.8469	0.9960	0.8827	0.9899
19	0.9054	0.7578	0.9957	0.7835	0.9914	0.8243	0.9788
61	0.9053	0.7576	0.9957	0.7832	0.9914	0.8241	0.9787
87	0.9052	0.7575	0.9957	0.7832	0.9914	0.8240	0.9787
79	0.8481	0.6849	0.9866	0.7099	0.9750	0.7506	0.9465
16	0.8437	0.6803	0.9855	0.7051	0.9730	0.7457	0.9431
143	0.8040	0.6415	0.9694	0.6652	0.9487	0.7041	0.9078

APPENDIX

TABLE 3

Rice Farms - Panel Data Estimation Results

Variable	OLS	Fixed Effects	CGLS	MLE
Constant	5.0811		5.0639	5.2073
Seed	0.1358	0.1208	0.1327	0.1332
Urea	0.1196	0.0918	0.1133	0.1127
TSP	0.0718	0.0892	0.0761	0.0769
Labor	0.2167	0.2431	0.2230	0.2194
Land	0.4819	0.4521	0.4771	0.4810
DP	0.0077	0.0338	0.0140	0.0093
DV1	0.1755	0.1788	0.1772	0.1767
DV2	0.1356	0.1754	0.1444	0.1410
DSS	0.0489	0.0533	0.0492	0.0492
DR1	-0.0500		-0.0511	-0.0594
DR2	-0.0393		-0.0441	-0.0480
DR3	-0.0623		-0.0723	-0.0799
DR4	0.0248		0.0119	0.0150
DR5	0.0818		0.0751	0.0826
$\hat{\sigma}^2_u$			0.0214	0.0215
$\hat{\sigma}^2_v$		0.1076	0.1076	0.1070
$E(u_i)$			0.1166	0.1170
σ_u			0.0987	0.0987

APPENDIX

TABLE 4A

Rice Farms - Confidence Intervals Based on BC Method, CGLS Estimates - Panel Data

Firm	TE	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	0.9648	0.8848	0.9990	0.9025	0.9979	0.9280	0.9946
118	0.9642	0.8830	0.9989	0.9009	0.9979	0.9267	0.9945
5	0.9608	0.8744	0.9988	0.8930	0.9976	0.9204	0.9938
51	0.9002	0.7628	0.9944	0.7857	0.9890	0.8224	0.9738
38	0.9002	0.7628	0.9944	0.7857	0.9890	0.8224	0.9738
88	0.8999	0.7624	0.9943	0.7852	0.9889	0.8219	0.9737
142	0.7660	0.6287	0.9215	0.6485	0.8951	0.6810	0.8537
145	0.7615	0.6249	0.9167	0.6447	0.8901	0.6769	0.8488
143	0.7422	0.6089	0.8951	0.6281	0.8683	0.6595	0.8274

TABLE 4B

Rice Farms - Confidence Intervals Based on BC Method, MLE Estimates - Panel Data

Firm	TE	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	0.9643	0.8837	0.9990	0.9013	0.9979	0.9270	0.9945
118	0.9639	0.8825	0.9989	0.9005	0.9978	0.9263	0.9944
5	0.9609	0.8746	0.9988	0.8931	0.9976	0.9206	0.9938
88	0.9010	0.7638	0.9945	0.7867	0.9892	0.8234	0.9742
51	0.9010	0.7638	0.9945	0.7867	0.9892	0.8234	0.9742
102	0.9001	0.7626	0.9944	0.7855	0.9890	0.8222	0.9738
142	0.7602	0.6238	0.9153	0.6435	0.8887	0.6757	0.8473
145	0.7564	0.6207	0.9111	0.6403	0.8845	0.6723	0.8432
143	0.7370	0.6045	0.8891	0.6236	0.8624	0.6549	0.8217

APPENDIX

TABLE 5A

Rice Farms - MCB Confidence Intervals - Panel Data

Firm	$\hat{\alpha}_i$	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	5.5561	0.5613	1.0000	0.5874	1.0000	0.6421	1.0000
118	5.4860	0.4878	1.0000	0.5105	1.0000	0.5580	1.0000
163	5.4838	0.4868	1.0000	0.5094	1.0000	0.5568	1.0000
166	4.9667	0.2902	1.0000	0.3037	1.0000	0.3320	1.0000
15	4.9656	0.2899	1.0000	0.3034	1.0000	0.3316	1.0000
40	4.9646	0.2896	1.0000	0.3031	1.0000	0.3313	1.0000
143	4.5982	0.2008	1.0000	0.2101	1.0000	0.2297	1.0000
117	4.5859	0.1983	1.0000	0.2075	1.0000	0.2269	1.0000
45	4.5496	0.1913	1.0000	0.2001	1.0000	0.2188	1.0000

TABLE 5B

Rice Farms - MCC Confidence Intervals - Farm 164 as Control - Panel Data

Firm	$\hat{\alpha}_i$	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	5.5561						
118	5.4860	0.4878	1.0000	0.5105	1.0000	0.5580	1.0000
163	5.4838	0.4868	1.0000	0.5094	1.0000	0.5568	1.0000
166	4.9667	0.2902	1.0000	0.3037	1.0000	0.3320	0.9266
15	4.9656	0.2899	1.0000	0.3034	1.0000	0.3316	0.9256
40	4.9646	0.2896	1.0000	0.3031	1.0000	0.3313	0.9247
143	4.5982	0.2008	0.7332	0.2101	0.7007	0.2297	0.6410
117	4.5859	0.1983	0.7243	0.2075	0.6921	0.2269	0.6332
45	4.5496	0.1913	0.6985	0.2001	0.6675	0.2188	0.6106

APPENDIX

TABLE 6A

Rice Farms - Subset MCB Confidence Intervals, N = 9 - Panel Data

Firm	$\hat{\alpha}_i$	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164	5.5561	0.6494	1.0000	0.6835	1.0000	0.7400	1.0000
118	5.4860	0.5644	1.0000	0.5940	1.0000	0.6432	1.0000
163	5.4838	0.5632	1.0000	0.5927	1.0000	0.6418	1.0000
166	4.9667	0.3358	0.9849	0.3534	0.9358	0.3827	0.8642
15	4.9656	0.3354	0.9837	0.3530	0.9347	0.3822	0.8632
40	4.9646	0.3351	0.9828	0.3527	0.9338	0.3819	0.8624
143	4.5982	0.2323	0.6813	0.2445	0.6473	0.2647	0.5978
117	4.5859	0.2296	0.6730	0.2415	0.6394	0.2615	0.5905
45	4.5496	0.2213	0.6490	0.2329	0.6166	0.2522	0.5695

TABLE 6B

Rice Farms - Per Comparison Confidence Intervals with α_{164} - Panel Data

Firm	Std Err	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
164							
118	0.1912	0.6409	1.0000	0.6807	1.0000	0.7483	1.0000
163	0.1898	0.6413	1.0000	0.6808	1.0000	0.7479	1.0000
166	0.1914	0.3812	1.0000	0.4049	0.7599	0.4451	0.6912
15	0.1940	0.3788	0.8102	0.4027	0.7622	0.4432	0.6924
40	0.1971	0.3761	0.8144	0.4002	0.7654	0.4412	0.6943
143	0.1915	0.2636	0.5584	0.2800	0.5257	0.3079	0.4782
117	0.1960	0.2581	0.5565	0.2745	0.5232	0.3025	0.4748
45	0.1941	0.2499	0.5347	0.2656	0.5030	0.2924	0.4569

APPENDIX

TABLE 7

Texas Utilities - Panel Data Estimation Results

Variable	OLS	Fixed Effects	CGLS	MLE
Constant	-3.7615		-5.0535	-5.0532
Labor	0.0881	-0.1291	-0.09660	-0.0775
Capital	0.3462	0.6275	0.5882	0.5856
Fuel	0.6406	0.5652	0.5807	0.5838
σ^2_u			0.0079	0.0266
σ^2_v		0.0029	0.0029	0.0029
$E(u_i)$			0.0709	0.1301
σ_u			0.0126	0.0126

APPENDIX

TABLE 8A

Texas Utilities - Confidence Intervals Based on BC Method, CGLS Estimates - Panel Data

Firm	TE	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
5	0.9982	0.9938	1.0000	0.9949	0.9999	0.9964	0.9998
3	0.9960	0.9866	0.9999	0.9887	0.9999	0.9919	0.9994
10	0.9649	0.9413	0.9885	0.9451	0.9848	0.9510	0.9788
1	0.9325	0.9097	0.9557	0.9133	0.9519	0.9190	0.9460
8	0.9167	0.8943	0.9396	0.8979	0.9358	0.9035	0.9300
9	0.8997	0.8777	0.9221	0.8812	0.9184	0.8867	0.9127
2	0.8973	0.8754	0.9197	0.8788	0.9160	0.8843	0.9103
6	0.8835	0.8619	0.9055	0.8653	0.9019	0.8707	0.8963
7	0.8788	0.8573	0.9006	0.8607	0.8971	0.8660	0.8915
4	0.8555	0.8346	0.8768	0.8379	0.8734	0.8431	0.8679

TABLE 8B

Texas Utilities - Confidence Intervals Based on BC Method, MLE Estimates - Panel Data

Firm	TE	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
5	0.9880	0.9685	0.9994	0.9721	0.9989	0.9776	0.9973
3	0.9793	0.9566	0.9979	0.9604	0.9962	0.9663	0.9923
10	0.9095	0.8872	0.9322	0.8907	0.9285	0.8963	0.9227
1	0.8649	0.8437	0.8865	0.8471	0.8830	0.8524	0.8775
8	0.8472	0.8264	0.8683	0.8297	0.8649	0.8349	0.8595
2	0.8322	0.8118	0.8530	0.8151	0.8496	0.8202	0.8443
9	0.8269	0.8066	0.8475	0.8098	0.8442	0.8149	0.8389
6	0.8214	0.8013	0.8419	0.8045	0.8386	0.8095	0.8334
7	0.8181	0.7981	0.8385	0.8012	0.8352	0.8062	0.8300
4	0.7873	0.7680	0.8069	0.7710	0.8037	0.7759	0.7987

APPENDIX

TABLE 9A

Texas Utilities - MCB & MCC Confidence Intervals - Panel Data

Firm	$\hat{\alpha}_i$	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
5	-4.9952	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	-5.0826	0.8730	0.9617	0.8772	0.9571	0.8837	0.9500
10	-5.1451	0.8201	0.9035	0.8241	0.8991	0.8302	0.8925
1	-5.1760	0.7951	0.8759	0.7990	0.8717	0.8049	0.8653
8	-5.1940	0.7809	0.8603	0.7847	0.8562	0.7905	0.8499
9	-5.2105	0.7681	0.8462	0.7719	0.8421	0.7776	0.8359
2	-5.2176	0.7627	0.8402	0.7664	0.8362	0.7721	0.8300
7	-5.2362	0.7487	0.8248	0.7523	0.8208	0.7579	0.8148
6	-5.2366	0.7484	0.8245	0.7520	0.8205	0.7576	0.8144
4	-5.2669	0.7261	0.7999	0.7296	0.7960	0.7350	0.7901

TABLE 9B

Texas Utilities - Marginal (Per Comparison) Intervals - Panel Data

Firm	Std Err	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
5							
3	0.0653	0.8063	1.0000	0.8230	1.0000	0.8501	0.9877
10	0.0549	0.7729	0.9586	0.7864	0.9421	0.8081	0.9169
1	0.0380	0.7747	0.8991	0.7840	0.8884	0.7989	0.8718
8	0.0356	0.7644	0.8789	0.7730	0.8691	0.7868	0.8539
9	0.0316	0.7593	0.8561	0.7666	0.8479	0.7783	0.8351
2	0.0401	0.7400	0.8660	0.7494	0.8551	0.7645	0.8383
7	0.0431	0.7222	0.8551	0.7320	0.8435	0.7478	0.8257
6	0.0442	0.7203	0.8566	0.7304	0.8447	0.7466	0.8464
4	0.0352	0.7113	0.8164	0.7293	0.8074	0.7319	0.7935

APPENDIX

TABLE 10

Tileries - Panel Data Estimation Results

Variable	OLS	Fixed Effects	CGLS
Constant	-0.2335		0.4732
Labor	1.1779	1.0075	1.0508
Machines	0.0347	0.0417	0.0448
σ^2_u			0.0842
σ^2_v		0.1147	0.1147
$E(u_i)$			0.2315

APPENDIX

TABLE 11

Tileries - Confidence Intervals Based on BC Method, CGLS Estimates - Panel Data

Firm	TE	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
24	0.9499	0.8601	0.9981	0.8771	0.9961	0.9034	0.9903
14	0.9433	0.8482	0.9976	0.8656	0.9952	0.8928	0.9881
25	0.9425	0.8457	0.9976	0.8635	0.9952	0.8911	0.9880
19	0.9365	0.8359	0.9971	0.8538	0.9942	0.8850	0.9858
3	0.9341	0.83085	0.9969	0.8491	0.9939	0.8779	0.9851
22	0.9323	0.8280	0.9968	0.8463	0.9936	0.8753	0.9843
18	0.9296	0.8266	0.9963	0.8444	0.9928	0.8765	0.9826
16	0.9151	0.7885	0.9958	0.8103	0.9916	0.8449	0.9797
5	0.9025	0.7791	0.9935	0.7993	0.9875	0.8315	0.9713
21	0.9017	0.7889	0.9919	0.8070	0.9848	0.8359	0.9670
2	0.8992	0.7841	0.9917	0.8024	0.9834	0.8319	0.9660
23	0.8866	0.7742	0.9871	0.7917	0.9769	0.8198	0.9544
4	0.8854	0.7688	0.9878	0.7870	0.9779	0.8162	0.9555
17	0.8813	0.7686	0.9850	0.7860	0.9738	0.8140	0.9498
15	0.8405	0.7296	0.9587	0.7462	0.9404	0.7731	0.9098
11	0.7228	0.6285	0.8271	0.6425	0.8091	0.6652	0.7815
13	0.6988	0.6077	0.7997	0.6212	0.7823	0.6431	0.7556
12	0.6872	0.5976	0.7864	0.6109	0.7693	0.6325	0.7431
6	0.6748	0.5868	0.7723	0.5999	0.7554	0.6211	0.7297
10	0.6620	0.5534	0.7855	0.5692	0.7637	0.5950	0.7307
9	0.6437	0.5541	0.7436	0.5674	0.7262	0.5888	0.6998
1	0.6283	0.5314	0.7378	0.5456	0.7186	0.5686	0.6894
20	0.5973	0.5100	0.6952	0.5229	0.6781	0.5437	0.6521
7	0.5818	0.5059	0.6657	0.5172	0.6512	0.5354	0.6290
8	0.5399	0.4664	0.6216	0.4773	0.6074	0.4949	0.5858

APPENDIX

TABLE 12

Tileries - MCB Confidence Intervals - Panel Data

Firm	$\hat{\alpha}_i$	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
14	0.9881	0.7373	1.0000	0.7566	1.0000	0.7892	1.0000
24	0.9825	0.7291	1.0000	0.7484	1.0000	0.7803	1.0000
25	0.9770	0.7223	1.0000	0.7417	1.0000	0.7737	1.0000
3	0.9691	0.7136	1.0000	0.7330	1.0000	0.7651	1.0000
19	0.9687	0.7164	1.0000	0.7356	1.0000	0.7673	1.0000
22	0.9685	0.7132	1.0000	0.7326	1.0000	0.7647	1.0000
18	0.9399	0.6987	1.0000	0.7172	1.0000	0.7478	1.0000
5	0.9061	0.6557	1.0000	0.6748	1.0000	0.7064	1.0000
21	0.9043	0.6716	1.0000	0.6896	1.0000	0.7194	1.0000
16	0.9036	0.6386	1.0000	0.6585	1.0000	0.6916	1.0000
2	0.8834	0.6550	1.0000	0.6728	1.0000	0.7023	1.0000
23	0.8811	0.6588	1.0000	0.6763	1.0000	0.7051	1.0000
4	0.8769	0.6508	1.0000	0.6684	1.0000	0.6977	1.0000
17	0.8497	0.6384	1.0000	0.6553	1.0000	0.6833	1.0000
15	0.8121	0.6149	1.0000	0.6311	1.0000	0.6581	1.0000
11	0.6240	0.5113	1.0000	0.5246	1.0000	0.5468	1.0000
13	0.6071	0.5027	1.0000	0.5158	1.0000	0.5376	1.0000
12	0.5796	0.4891	1.0000	0.5018	1.0000	0.5230	1.0000
6	0.5494	0.4745	1.0000	0.4870	1.0000	0.5074	0.9772
10	0.5190	0.4387	1.0000	0.45207	1.0000	0.4741	0.9843
9	0.4604	0.4291	0.9673	0.4408	0.9417	0.4601	0.9022
1	0.4475	0.4146	0.9757	0.4266	0.9482	0.4466	0.9058
20	0.4256	0.4104	0.9433	0.4219	0.9176	0.4409	0.8779
7	0.3987	0.4081	0.8990	0.4188	0.8760	0.4365	0.8406
8	0.2818	0.3604	0.8057	0.3701	0.7848	0.3861	0.7522

APPENDIX

TABLE 14

Tileries - Per Comparison Confidence Intervals - Panel Data

Firm	Std Err	95% Lbnd	95% Ubnd	90% Lbnd	90% Ubnd	75% Lbnd	75% Ubnd
14							
24	0.1045	0.7958	1.0000	0.8248	1.0000	0.8725	1.0000
25	0.1137	0.7940	1.0000	0.8225	1.0000	0.8694	1.0000
3	0.1120	0.7950	1.0000	0.8223	1.0000	0.8672	1.0000
19	0.1100	0.7907	1.0000	0.8185	1.0000	0.8643	1.0000
22	0.1075	0.7943	1.0000	0.8216	1.0000	0.8665	1.0000
18	0.1225	0.7496	1.0000	0.7791	1.0000	0.8278	1.0000
5	0.1277	0.7173	1.0000	0.7468	1.0000	0.7955	1.0000
21	0.1060	0.7472	1.0000	0.7725	1.0000	0.8141	1.0000
16	0.1485	0.6869	1.0000	0.7198	1.0000	0.7747	1.0000
2	0.1078	0.7291	1.0000	0.7543	1.0000	0.7956	1.0000
23	0.1054	0.7309	1.0000	0.7556	1.0000	0.7960	1.0000
4	0.1148	0.7145	1.0000	0.7408	1.0000	0.7841	1.0000
17	0.1198	0.6886	1.0000	0.7151	1.0000	0.7587	0.9994
15	0.1047	0.6830	1.0000	0.7059	0.9962	0.7435	0.9459
11	0.1090	0.5612	0.8603	0.5808	0.8313	0.6130	0.7876
13	0.1034	0.5578	0.8367	0.5763	0.8098	0.6066	0.7694
12	0.1118	0.5339	0.8275	0.5530	0.7989	0.5845	0.7558
6	0.1090	0.5208	0.7984	0.5390	0.7715	0.5689	0.7310
10	0.1223	0.4922	0.7950	0.5116	0.7650	0.5435	0.7200
9	0.1194	0.4669	0.7456	0.4848	0.7181	0.5143	0.6768
1	0.1258	0.4552	0.7453	0.4736	0.7163	0.5040	0.6731
20	0.1115	0.4579	0.7089	0.4728	0.6844	0.5012	0.6477
7	0.1056	0.4510	0.7822	0.4662	0.6599	0.4912	0.6263
8	0.1173	0.3921	0.6210	0.4069	0.5985	0.4312	0.5648

APPENDIX

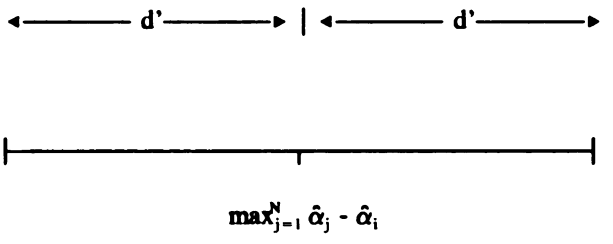
MCC INTERVAL

$\alpha_N = \alpha_{[N]}$

$i \neq N$

No uncertainty of population ranking

([N] known)



MCB INTERVAL

$i \neq [N]$

Uncertainty of population ranking

([N] unknown)

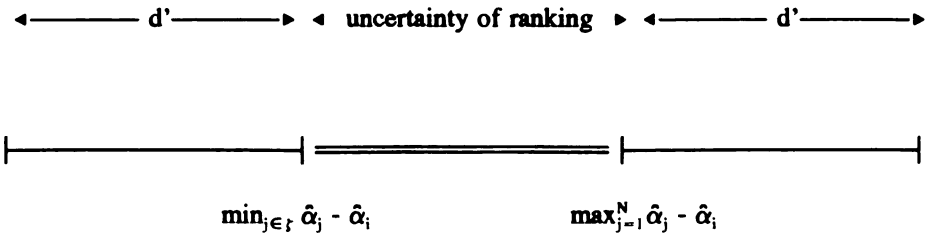


FIGURE 1. MCC and MCC Intervals

LIST OF REFERENCES

LIST OF REFERENCES

AIGNER, D.J., C.A.K. LOVELL and P. SCHMIDT (1977). Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics,* 6, 21-37.

ANDERSON, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics,* 22, 327-351.

BANKER, R. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, 1265-1273.

BANKER, R. (1995). Hypothesis testing using data envelopment analysis. *Journal of Productivity Analysis*, forthcoming.

BATTESE, G.E. and T.J. COELLI (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics,* 38, 387-399.

BATTESE, G.E. and T.J. COELLI (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3, 153-170.

BECKER, W.A. (1961). Comparing entries in random sample tests. *Poultry Science,* 40, 1507-14.

BECHHOFER, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics,* 25, 16-39.

BORDO, M.D., E.U. CHOUDRI and A.J. SCHWARTZ (1994). Could stable money have averted the great contraction? (mimeo).

CORNWELL, C. and P. SCHMIDT (1995). Production frontiers and efficiency measurement. In *Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd edition. (Eds. L. Matyas and P. Sevestre). Boston: Kluwer Academic Publishers.

CORNWELL, C., P. SCHMIDT and R.C. SICKLES (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46, 185-200.

DALAL, S.R. and V. SRINIVASAN. (1977). Determining sample size for pretesting comparative effectiveness of advertising copies. *Management Science*, 23, 1284-94.

DESU, M.M. (1970). A selection problem. *Annals of Mathematical Statistics*, 41, 1596-1603.

DHRYMES, P.J. (1970). Econometrics, Statistical Foundations and Applications. New York: Springer-Verlag.

DHRYMES, P.J. (1978). Mathematics for Econometrics. New York: Springer-Verlag.

DUNN, O.J. and F.J. MASSEY, Jr. (1965). Estimation of multiple contrasts using t-distributions. *Journal of the American Statistical Association*, 60, 573-583.

DUNNETT, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50, 1096-1121.

DUNNETT, C.W. (1964). New tables for multiple comparisons with a control. *Biometrics*, 20, 482-491.

DUNNETT, C.W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, 75, 796-800.

EDWARDS, D.G. and J.C. HSU (1983). Multiple comparisons with the best treatment. *Journal of the American Statistical Association*, 78, 965-971. Corrigenda (1984), *Journal of the American Statistical Association*, 79, 965.

ERWIDODO (1990). Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in West Java, Indonesia. Unpublished dissertation. Department of Agricultural Economics. Michigan State University, East Lansing.

FABIAN, V. (1962). On multiple decision methods for ranking population means. *Annals of Mathematical Statistics*, 33, 248-254.

GREENE, W.H. (1995). Frontier Production Functions. *Handbook of Applied Econometrics, Vol. II - Microeconomics*. (Eds. H. Pesaran and P. Schmidt). London: Basil Blackwell, forthcoming.

GREENE, W.H. (1990). A Gamma-distributed Stochastic Frontier Model. *Journal of Econometrics*, 46, 141-164.

GUPTA, S.S. (1956). On a decision rule for a problem of ranking means. *Institute of Statistics Mimeo Series,* No. 150, University of North Carolina.

GUPTA, S.S. (1965). On some multiple decisions (selection and ranking) rules. *Technometrics,* 7, 225-245.

GUPTA, S.S. and J.C. HSU (1977). Subset selection procedures with special reference to the analysis of two-way layout: Application to motor-vehicle fatality data. *Proceedings of the 1977 Winter Simulation Conference.* (Eds. H.J. Highland, R.G. Sargent and J.W. Schmidt) 81-85.

GUPTA, S.S., S. PANCHAPAKESAN and J.K. SOHN (1985). *Multiple Decision Procedures.* New York: John Wiley and Sons.

HAHN, G.J. and R.W. HENDRICKSON (1971). A table of percentage points of the distribution of the largest absolute value of k Student t variables and its applications. *Biometrika,* 58, 323-332.

HAMILTON, J.D. (1994). Time Series Analysis. Princeton University Press. Princeton, NJ.

HARGREAVES C. (Ed) (1994). A reveiw of methods of estimating cointegrating relationships. In *Nonstationary Time Series Analysis and Cointegration.* Oxford: Oxford University Press.

HARTER, H.L. (1980). History of multiple comparisons. *Handbook of Statistics, Vol I.* (Ed. P.R. Krishnaiah). Amsterdam: North Holland. 617-622.

HAUSMAN, J.A. (1978). Specification tests in Econometrics. *Journal of Econometrics,* 46, 1251-1271.

HOCHBERG, Y. and A.C. TAMHANE (1987). *Multiple Comparison Procedures.* New York: John Wiley and Sons.

HOOD, W.C. and T.C. KOOPMANS (Eds) (1953). Estimation of Linear Relationships. In *Studies in Econometric Methods.* Cowles Foundation for Research in Economics. Monograph No. 14. New York: Wiley.

HSU, J.C.. (1981). Simultaneous confidence intervals for all distances from the best. *Annals of Statistics,* 9, 1026-1034.

HSU, J.C. (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics,* 12, 1145-1150.

JOHANSEN, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control,* 12, 231-254.

JOHANSEN, S. (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59, 1551 - 1580.

JONDROW, J., C.A.K. LOVELL, I.S. MATEROV and P. SCHMIDT (1982). On the estimation of technical efficiency in the stochastic production function model. *Journal of Econometrics,* 19, 233-238.

JUDD, J.P. and B. MOTLEY (1992). Controlling inflation with an interest rate instrument. *Federal Reserve Bank of San Francisco Economic Review,* 3, 3-22.

JUDD, J.P. and B. MOTLEY (1993). Using a nominal GDP rule to guide discretionary monetary policy. *Federal Reserve Bank of San Francisco Economic Review,* 3, 3-11.

KNEIP, A. and L. SIMAR (1995). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis*, forthcoming.

KOROSTELEV, A., L. SIMAR and A. TSYBAKOV (1992). Efficient estimation of monotone boundaries. Working paper 9209, Institut de Statistique, Université Catholique de Louvain.

KOROSTELEV, A., L. SIMAR and A. TSYBAKOV (1995). On estimation of monotone and convex boundaries. *Publications de l'ISUP*, 39, 3-18.

KUMBHAKAR, S.C. (1990). Production frontiers, panel data and time-varying technical inefficiency. *Journal of Econometrics*, 46, 201-212.

KUMBHAKAR, S.C. (1994). Estimation of cost efficiency with heteroscedasticity: an application to electric utilities in Texas, 1966 - 1985. Unpublished manuscript. Department of Economics. University of Texas at Austin.

LEE, Y.H. (1991). Panel data models with multiplicative individual and time effects: applications to compensation and frontier production functions. Unpublished dissertation. Department of Economics. Michigan State University, East Lansing.

LEE, Y.H. and P. SCHMIDT (1993). A production frontier model with flexible temporal variation in technical efficiency. In *The Measurement of Productive Efficiency*. (Eds. H.O. Fried, C.A.K. Lovell and S.S. Schmidt). New York: Oxford University Press.

LOVELL, C.A.K (1993). Production frontiers and productive efficiency. In *The Measurement of Productive Efficiency*. (Eds. H.O. Fried, C.A.K. Lovell and S.S. Schmidt). New York: Oxford University Press.

LOVELL, C.A.K. and P. SCHMIDT (1988). A comparison of alternative approaches to the measurement of productive efficiency. In *Applications of Modern Production Theory*. (Eds. A. Dogramaci and R. Färe). Boston: Kluwer Academic Publishers.

MATEJCIK, F.J. and B.L. NELSON (1995). One stage multiple comparisons with the best for unequal variances. *The Proceedings of the Business and Economic Section of the American Statistical Association at the Joint Statistic Meetings, Toronto, 1994.* Forthcoming.

MATEJCIK, F.J. (1992). Heteroscedastic multiple comparison procedures for computer simulation. Unpublished dissertation. Department of Industrial and Systems Engineering. The Ohio State University, Columbus.

McCALLUM, B.T. (1988). Robustness properties of a rule for monetary policy. *Carnegie-Rochester Conference Series on Public Policy,* 29, 173-204.

McCALLUM, B.T. (1990). Targets, instruments and indicators of monetary policy. In *Monetary Policy for a Changing Financial Environment.* (Eds. W.S. Haraf and P. Cagan). Washington, D.C.: AEI Press.

McCALLUM, B.T. (1993). Specification and analysis of a monetary policy rule for Japan. *Bank of Japan Monetary and Economic Studies,* 11, 1-45.

McDONALD, G.C. (1977). Non-parametric selection procedures applied to state traffic fatality rates. *Proceedings of the 1977 Winter Simulation Conference.* (Eds. H.J. Highland, R.G. Sargent and J.W. Schmidt) 93-100.

MEEUSEN, W. and J. VAN DEN BROECK (1977). Efficient estimation from Cobb-Douglas production functions with composed error. *International Economic Review,* 18, 435-444.

MURPHY, K.M. and R.H. TOPEL (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business and Economic Statistics.*

OLSON, J.A., P. SCHMIDT and D.M. WALDMAN (1980). A Monte Carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics,* 13, 67-82.

PARK, B.U. and L. SIMAR (1994). Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89, 929-936.

PITT, M.M. and L.F. LEE (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics,* 9, 43-64.

RASCHE, R.H. (1995). Pitfalls in counterfactual analyses of policy rules. *Open Economies Review*, 6, 199-202.

RUBIN, H. (1948). Systems of Linear Stochastic Equations. Unpublished PhD. dissertation. University of Chicago, Chicago.

SEALE, J.L. (1990). Estimating stochastic frontier systems with unbalanced panel data: the case of floor tile manufactories in Egypt. *Journal of Applied Econometrics*, 5, 59-79.

SCHMIDT, P. (1985). Frontier production functions. *Econometric Reviews*, 4, 289-328.

SCHMIDT, P. and R.C. SICKLES (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics*, 2, 367-374.

SIMAR, L. and P.W. WILSON (1995). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Working paper 9503, Institut de Statistique, Université Catholique de Louvain.

STEVENSON, R.E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13, 57-66.

TAMHANE, A.C. (1977). Multiple comparisons in model I: one-way ANOVA with unequal variances. *Communications and Statistics, Series A*, 6, 15-32.

TSO, M.K.-S. (1981). Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society, Series B*, 42, 183-189.

WALDMAN, D (1982). A stationary point for the stochastic frontier likelihood. *Journal of Econometrics*, 18, 275-279.