

Syracuse University

SURFACE

Economics - Faculty Scholarship

Maxwell School of Citizenship and Public
Affairs

2000

Multiple Comparisons with the Best, with Economic Applications

William C. Horrace

University of Arizona

Peter Schmidt

Michigan State University

Follow this and additional works at: <https://surface.syr.edu/ecn>



Part of the [Economics Commons](#)

Recommended Citation

Horrace, William C. and Schmidt, Peter, "Multiple Comparisons with the Best, with Economic Applications" (2000). *Economics - Faculty Scholarship*. 23.

<https://surface.syr.edu/ecn/23>

This Article is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE. It has been accepted for inclusion in Economics - Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

MULTIPLE COMPARISONS WITH THE BEST, WITH ECONOMIC APPLICATIONS

William C. Horrace, University of Arizona
 Peter Schmidt, Michigan State University

SUMMARY

In this paper we discuss a statistical method called *multiple comparisons with the best*, or MCB. Suppose that we have N populations, and population i has parameter value θ_i . Let $\hat{\theta}_{(N)} = \max_{i=1, \dots, N} \theta_i$, the parameter value for the 'best' population. Then MCB constructs joint confidence intervals for the differences $[\theta_{(N)} - \theta_1, \theta_{(N)} - \theta_2, \dots, \theta_{(N)} - \theta_N]$. It is *not* assumed that it is known which population is best, and part of the problem is to say whether any population is so identified, at the given confidence level. This paper is meant to introduce MCB to economists. We discuss possible uses of MCB in economics. The application that we treat in most detail is the construction of confidence intervals for inefficiency measures from stochastic frontier models with panel data. We also consider an application to the analysis of labour market wage gaps

1. INTRODUCTION

Empirical research often involves comparisons. For instance, one may wish to compare the effectiveness of several different drugs in the treatment of a disease, or crop yields for a variety of fertilizers, or productivity levels of different firms, or earnings of individuals exposed to a variety of types of training. Typically this would involve the calculation and comparison of some outcome measure or measures for the various populations of interest. Statistically speaking, calculation of an outcome measure is naturally viewed as the estimation of some parameter indexing outcomes, such as mean lifetime or mean earnings, and the comparison of these estimates would often involve statistical hypothesis testing and/or the construction of confidence intervals, so as to reflect the statistical uncertainty about the parameter values.

To be more precise, suppose that there are N populations, with parameter value θ_i and estimate $\hat{\theta}_i$ for population i ($i = 1, \dots, N$). Then, for given populations i and j , under standard assumptions one can generally test the hypothesis $\theta_i = \theta_j$, or construct a confidence interval for the difference $\theta_i - \theta_j$. For example, under the assumption of random samples from independent normal populations with equal variance, such a hypothesis test or confidence interval would involve the Student- t distribution.

In this paper, our interest will be on multiple comparisons rather than on individual ones. The simplest type of multiple comparison procedure is referred to as *multiple comparisons with a control* (or *MCC* for short), and constructs a joint confidence interval for the vector of differences $[\theta_N - \theta_1, \theta_N - \theta_2, \dots, \theta_N - \theta_{N-1}]$, where population N is the 'control' population, such as the current drug treatment. The motivation for such a multiple comparison procedure is what Hochberg and Tamhane (1987) refer to as the *multiplicity effect*, which is simply the fact that, if enough comparisons are made, some of them are more or less sure to be 'significant'. That is, with multiple individual comparisons, it is difficult or impossible to assess the 'significance' of an individual difference without knowing how many comparisons were made and how the particular individual difference was selected from all those considered. For example, if we read in the newspaper that left-handed individuals who eat broccoli for lunch on Tuesday have rates of colon cancer that are significantly different from the overall population, we would not be impressed by the 'significance' of this difference at usual confidence levels like the 5% level if we knew that 10,000 similar comparisons were made and only the most 'significant' differences were reported. From the perspective of hypothesis testing, recognition of the multiplicity effect argues for the use of a test of the joint hypothesis $\theta_1 = \theta_2 = \dots = \theta_N$ to control the size of the overall testing procedure. From the perspective of the construction of confidence intervals for differences, recognition of the multiplicity effect argues for the construction of a joint confidence interval for the vector of differences, which MCC provides.

This paper deals with an extension of MCC called *multiple comparisons with the best* (or *MCB*), which constructs a joint confidence interval for the vector of differences from the unknown 'best' population parameter. That is, suppose that the problem is parameterized in such a way that bigger θ is better than smaller θ , in the sense of a more favourable outcome, such as longer expected lifetime or greater expected earnings. Now define $\hat{\theta}_{(N)} = \max_{i=1, \dots, N} \theta_i$, so that (N) is the index of the best population and $\theta_{(N)}$ is the best value of θ in the N populations. Then MCB constructs a joint confidence interval for the vector of differences $[\theta_{(N)} - \theta_1, \theta_{(N)} - \theta_2, \dots, \theta_{(N)} - \theta_N]$. This is different from and more difficult than the MCC problem, because the relevant control population is assumed to be

the best population, and the identity of the best population is *not* assumed to be known. MCB also provides a subset of populations which (at the specified confidence level) contains the best population. Thus MCB tells us, with a specified level of confidence, which populations may be best, and it provides upper and lower bounds on the deviations of all N values of θ from the best value.

The purpose of this paper is to introduce MCB to economists and to illustrate its use in a few applications. One application is to the 'wage gap' literature, in which earnings of various groups are compared to the earnings of the highest-earning group. Another application is to the measurement of productive efficiency from stochastic frontier models with panel data. MCB has been applied to the stochastic frontier model in an earlier paper (Horrace and Schmidt, 1996), which was concerned with confidence statements for efficiency measures from a wider variety of stochastic frontier models, but the earlier paper did not attempt to develop or explain MCB methods in the way that the present paper does.

This paper is organized as follows. Section 2 gives a brief historical account of MCB and presents the main results for MCC and MCB under the 'standard' assumptions of balanced random sampling from independent, normal populations with equal variance. Section 3 discusses MCB in the more difficult and less well-explored case that the estimates $\hat{\theta}$ have a general covariance structure. Section 4 discusses issues in applying MCB in the regression model with panel data. Section 5 gives empirical examples of MCB. Finally, Section 6 draws some conclusions and suggests areas for additional research.

2. STANDARD MCB

2.1 Introduction

As in Section 1 above, we suppose that we are interested in a vector of parameters $\theta = [\theta_1, \theta_2, \dots, \theta_N]'$, where the interpretation is that each θ_i corresponds to some distinct population. Correspondingly we suppose that we have a vector of estimates $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N]'$. Generally the estimate $\hat{\theta}_i$ will be based on some number of observations, say T_i , on population i , so that we effectively have a panel data setting. However, at the present level of generality we will simply make assumptions about the distributional properties of $\hat{\theta}$, without explicit reference to the nature of the sample or method of estimation.

In this section we will describe *multiple comparisons with a control (MCC)* and *multiple comparisons with the best (MCB)*. We will focus on the simplest and most commonly used versions of these techniques, which we will refer to as standard MCC and MCB. We make the following assumptions.

Assumption 1: $\hat{\theta}$ is distributed as $N(\theta, \sigma^2 C)$, with C known.

Assumption 2: Either σ^2 is known or we have an estimate $\hat{\sigma}^2$, independent of $\hat{\theta}$, such that $\hat{\sigma}^2/\sigma^2$ is distributed as χ^2_ν/ν , with ν (degrees of freedom) known.

Assumption 3: $C = kI_N$, with k known.

Assumptions 1 and 2 will be maintained throughout the paper. MCC and MCB are generally viewed as exact (finite-sample) procedures, but we could consider an asymptotically valid version that replaced the normality assumption in Assumption 1 with asymptotic normality. From the panel data perspective, this would correspond to the case of $T \rightarrow \infty$ with N fixed. Similarly, in thinking about Assumption 2, in virtually all applications we can envisage, there will be enough degrees of freedom that the scale parameter σ^2 can effectively be taken as known.

Assumption 3 is the defining assumption for *standard* MCC and MCB. It will be maintained throughout this section, but will be relaxed in the next section and in some of our subsequent empirical work.

A simple model to which standard MCC and MCB are applicable, and which is the basis for the usual discussions of standard MCC and MCB, is the 'balanced one-way model' (e.g. Hsu, 1996, p. 43). Here we have N independent normal populations with equal variance but different means, and a random sample of size T from each population. Thus we have mutually independent observations y_{it} , $i = 1, \dots, N$, $t = 1, \dots, T$, where y_{it} is distributed as $N(\mu_i, \sigma^2)$. Then Assumptions 1, 2 and 3 hold, with $\theta_i = \mu_i$, $\theta_i = \bar{y}_i$, $k = 1/T$, $\hat{\sigma}^2$ equal to the usual pooled variance estimate, and $\nu = N(T-1)$. This is (perhaps) a model of broad empirical applicability, but it is not necessarily general enough for applications to the regression model, as we shall see.

2.2 Standard MCC

MCC is a type of multiple comparison procedure. There are several other types, including multiple comparisons with the best, which is the subject of this paper, and all pairwise comparisons, which we will not discuss. The literature on multiple comparisons first 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 good historical account. Shortly thereafter a related body of literature on 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 as described in

this section were primarily due to Dunnett (1955, 1964). A good textbook treatment can be found in Hsu (1996, Chapter 3).

MCC is concerned with the following vector of differences:

$$\delta = [\delta_1, \delta_2, \dots, \delta_{N-1}]' = [\theta_N - \theta_1, \theta_N - \theta_2, \dots, \theta_N - \theta_{N-1}]' \quad (1)$$

For any integer m , let I_m be the $m \times m$ identity matrix, and e_m be an $m \times 1$ vector of ones. We define the following matrices:

$$D = [-I_{N-1}, e_{N-1}], \quad B = DCD' \quad (2a)$$

$$R = (N-1) \times (N-1) \text{ matrix with typical element } R_{ij} = B_{ij}/(B_{ii}B_{jj})^{1/2} \quad (2b)$$

These matrices arise naturally in considering δ and its estimate $\hat{\delta}$. We have $\delta = D\theta$ and $\hat{\delta} = D\hat{\theta}$. Since $\hat{\theta}$ is distributed as $N(\theta, \sigma^2 C)$ by Assumption 1, it follows that $\hat{\delta}$ is distributed as $N(\delta, \sigma^2 B)$. That is, B is (up to proportionality) the variance matrix of $\hat{\delta}$; and R is the corresponding correlation matrix.

For the case of standard MCC, we also impose Assumption 3, so that $C = kI_N$ (with k known). Then simple algebra reveals that $B = k[I_{N-1} + e_{N-1}e_{N-1}']$. This is a matrix with diagonal elements equal to $2k$ and off-diagonal elements equal to k . It follows that the corresponding correlation matrix R has an *equicorrelated structure* with all correlations equal to $\frac{1}{2}$. That is, $R = E(\frac{1}{2})$, where $E(\frac{1}{2})$ is the $(N-1) \times (N-1)$ matrix defined as follows:

$$E(\frac{1}{2})_{ii} = 1; \quad E(\frac{1}{2})_{ij} = \frac{1}{2} \text{ for } i \neq j \quad (3)$$

That is, for the case of standard MCC, $B = (2k)R = (2k)E(\frac{1}{2})$

Under Assumptions 1, 2 and 3, $(\hat{\delta} - \delta)/(2k\hat{\sigma}^2)^{1/2}$ is distributed as multivariate Student- t , with dimension $N-1$, degrees of freedom ν , and correlation matrix $E(\frac{1}{2})$, which we will represent with the notation $t_{N-1, \nu, E(1/2)}$. (If σ^2 is assumed known, which is also possible under Assumption 2, we have a multivariate normal distribution of dimension $N-1$ and variance matrix $E(\frac{1}{2})$. This is subsumed in the previous discussion as the case corresponding to $\nu = \infty$.) Now, for a given confidence level $1 - \alpha$, we define the critical value d^* as the two-sided critical value for the maximal value (over its $N-1$ components) of the Student- t random variable. That is, if z is an $N-1$ dimensional random vector distributed as $t_{N-1, \nu, E(1/2)}$, we define d^* as the solution to

$$P(\max_{1 \leq j \leq N-1} |z_j| \leq d^*) = 1 - \alpha \quad (4)$$

Then we have the following set of joint confidence intervals:

$$1 - \alpha = P[\hat{\delta}_i - d^*(2k\hat{\sigma}^2)^{1/2} \leq \delta_i \leq \hat{\delta}_i + d^*(2k\hat{\sigma}^2)^{1/2} \quad \forall i = 1, \dots, N-1] \quad (5)$$

These are the MCC confidence intervals.

Obviously the intervals in equation (5) are two-sided. One-sided intervals are also possible but seem less likely to be of interest in econometric applications. They would be constructed in essentially the same way; of course, d^* would have to be a one-sided critical value.

The critical values d^* for standard MCC are widely tabulated (see e.g. Dunnett, 1964; Dunn and Massey, 1965; Hahn and Hendrickson, 1971; Bechhofer and Dunnett, 1988; or Hochberg and Tamhane, 1987). The MCC literature has typically considered rather small values of N , such as $N \leq 50$. Horrace (1998) gives values of d^* for larger values of N , with $\nu = \infty$ (which is not restrictive for large values of N). The values of d^* in Horrace (1998) are calculated by simulation, directly from the definition in equation (5). In the MCC literature, d^* is typically calculated by evaluating the probability in equation (5) numerically. This is feasible because, given the equicorrelated structure of R, the $N-1$ dimensional probability statement in equation (5) can be reduced to a double integral. Specifically, d^* is the solution in t to the following equation:

$$\int_0^\infty \int_{-\infty}^\infty \{\Phi^{N-1}[(z\rho^{1/2} + ts)(1-\rho)^{-1/2}] - \Phi^{N-1}[(z\rho^{1/2} - ts)(1-\rho)^{-1/2}]\} d\Phi(z) dQ_t(s) = 1 - \alpha \quad (6)$$

where $\rho = \frac{1}{2}$, Φ is the standard normal cdf, and Q_t is the cdf of the distribution of $(\chi^2_\nu/\nu)^{1/2}$. For economic applications, there is no compelling reason to prefer the numerical solution of equation (6) to a simulation, or vice versa; our simulations reproduced earlier tabulations where they overlapped. The earlier MCC literature emphasized deterministic (as opposed to stochastic) approximations for critical values because MCC was originally developed in a biomedical setting, and the nature of FDA regulation made it desirable for repeated analysis of the same data to reach exactly the same conclusions.

Naturally, the values of d^* increase with the number of comparisons $(N-1)$. Essentially, it takes a larger deviation to be judged significant if more comparisons are made. For example, for $\alpha = 0.05$ and $\nu = 1$, we have the usual value of 1.96 for $N = 2$, but 2.44 for $N = 5$, 2.81 for $N = 15$, 3.29 for $N = 100$ and 3.64 for $N = 500$.

2.3 Standard MCB

The point of MCB is to construct joint confidence intervals for the differences of each θ_i from the best, and also to construct a set of populations which could be the best. More precisely, we consider the populations in (unknown) rank order:

$$\theta_{(1)} \leq \theta_{(2)} \leq \dots \leq \theta_{(N)} \quad (7)$$

so that, as in Section 1, (N) is the index of the best population and $\theta_{(N)}$ is the standard of comparison for the θ_i . Then we seek upper and lower bounds L_i, U_i and a set S of indices of those populations in contention for the best, such that:

$$P\{(N) \in S \text{ and } L_i \leq \theta_{(N)} - \theta_i \leq U_i \forall i = 1, \dots, N\} \geq 1 - \alpha \quad (8)$$

In this section we consider the construction of the set S and the bounds L_i, U_i for the case of standard MCB, which corresponds to the imposition of Assumptions 1, 2 and 3 above. As with standard MCC, the primary example is the balanced one-way model discussed in Section 2.1.

MCB evolved in the early 1980s with the work of Jason Hsu. Hsu (1981) constructed parametric and non-parametric simultaneous one-sided confidence intervals for the balanced one-way model. Hsu (1984) constructed simultaneous two-sided MCB confidence intervals. Edwards and Hsu (1983) provided a general technique for adapting MCC intervals to MCB intervals. Hochberg and Tamhane (1987, Section 5.3) summarize the main results of these papers. A more detailed discussion is given in Hsu (1996, Chapter 4).

There is room for possible confusion over different kinds of confidence intervals found in the MCB literature. The intervals that we will discuss, as given in equation (8), are two-sided intervals for $\theta_{(N)} - \theta_i = \max_{j=1, \dots, N} \theta_j - \theta_i$. These are the intervals considered in the early MCB literature, and they seem most natural to us; they correspond directly to the parameters of interest in the applications we consider in this paper. An alternative considered in the more recent MCB literature is to consider confidence intervals for $\theta_i - \max_{j \neq i} \theta_j$. Hsu (1996, pp. 81–2) argues that these intervals are easier to interpret and that they simplify some of the derivations. Ultimately this is a matter of taste. However, Hsu also features confidence intervals for $\theta_i - \max_{j \neq i} \theta_j$ that are constrained to contain the point zero. These constrained intervals correspond to one-sided intervals for $\max_{j=1, \dots, N} \theta_j - \theta_i$, with an upper bound U_i but with $L_i \equiv 0$. We prefer two-sided intervals because, for example, we want to be able to place upper and lower bounds on the technical efficiency of inefficient firms.

To give the basic result of standard MCB, we define the following notation:

$$\theta_{(N)} - \theta_i = \max_{j=1, \dots, N} \theta_j - \theta_i$$

with d^* defined by equation (4):

$$S = \{i | \hat{\theta}_i \geq \max_{j=1, \dots, N} \hat{\theta}_j - h\} = \{i | \hat{\theta}_i \geq \hat{\theta}_j - h \forall j\} \quad (9b)$$

$$L_i = \max[0, \min_{j \in S} \hat{\theta}_j - \hat{\theta}_i - h] \quad (9c)$$

$$U_i = \max[0, \max_{j \notin S} \hat{\theta}_j - \hat{\theta}_i + h] \quad (9d) \dagger$$

Edwards and Hsu (1983) showed that, given these definitions, equation (8) holds; that is, $P\{(N) \in S \text{ and } L_i \leq \theta_{(N)} - \theta_i \leq U_i \forall i = 1, \dots, N\} \geq 1 - \alpha$. This is the fundamental result for standard MCB. It provides, at a confidence level of at least $1 - \alpha$, a set S of populations that contains the best population, and upper and lower bounds for each difference from the best.

We will not give a detailed proof, which can be found in Hsu (1996, pp. 105–6). However, the idea behind the proof is simple. We begin by considering the standard MCC problem in which the control population is the unknown best, indexed by (N) . Thus we have $1 - \alpha = P(E)$, where E is the event:

$$E = \{\hat{\theta}_{(N)} - \hat{\theta}_i - h \leq \theta_{(N)} - \theta_i \leq \hat{\theta}_{(N)} - \hat{\theta}_i + h \forall i \neq (N)\} \quad (10)$$

Now, it is easy to see that the event E implies the event $\{(N) \in S\}$, since E implies $\{0 \leq \theta_{(N)} - \theta_i \leq \theta_{(N)} - \theta_i + h \forall i\}$, which implies $\{\hat{\theta}_{(N)} \geq \hat{\theta}_i - h \forall i\} = \{(N) \in S\}$. Also, the condition $\forall i \neq (N)$ can be converted into simply $\forall i$, since the inequality clearly holds for $i = (N)$, since $h \geq 0$ so long as $\alpha \leq 0.5$. Thus E is equivalent to the event E^* :

$$E^* = \{(N) \in S \text{ and } \hat{\theta}_{(N)} - \hat{\theta}_i - h \leq \theta_{(N)} - \theta_i \leq \hat{\theta}_{(N)} - \hat{\theta}_i + h \forall i\} \quad (11)$$

In this event $\hat{\theta}_{(N)}$ is unknown since (N) is unknown. (In particular, $\hat{\theta}_{(N)}$ is the estimate of $\theta_{(N)}$, the value of θ for the best population; it is not necessarily equal to $\max_i \hat{\theta}_i$.) However, given E^* , we know that (N) is in the set S . Thus we can replace $\hat{\theta}_{(N)}$ on the left-hand side of the inequalities in equation (11) by $\min_{j \in S} \hat{\theta}_j$, and we can replace $\hat{\theta}_{(N)}$ on the right hand side of these inequalities by $\max_{j \in S} \hat{\theta}_j = \max_{j=1, \dots, N} \hat{\theta}_j$, to obtain an event that has probability at least as large

as $1 - \alpha$. All that remains then is a little logic to ensure non-negativity of the bounds, and to avoid setting a needlessly high bound for the population that corresponds to the maximal $\hat{\theta}_i$.

Recall that the MCC confidence interval for $\delta_i = \theta_N - \theta_i$ is $\hat{\delta}_i \pm h$, an interval of width $2h$. The MCB intervals may be narrower or wider, depending on whether L_i and/or $U_i = 0$. When both L_i and U_i are positive (the conditions for which will be discussed further in the next section), the MCB intervals are wider, with a width of $2h + \max_j \hat{\theta}_j - \min_{j \in S} \hat{\theta}_j$. That is, the uncertainty about which population is best adds to the width of the confidence intervals. However, if S consists of a single population, then the MCB intervals for the other populations are exactly the same as the MCC intervals with that population as the control. This occurs because in this case, at the specified confidence level, there is no uncertainty about the identity of the best population.

It should be noted that the MCB confidence intervals are conservative. Whereas the MCC statement (10) holds with a probability of exactly $1 - \alpha$, the MCB statement (8) holds with a probability of at least $1 - \alpha$, and the inequality occurs because of uncertainty about the identity of the best population. This uncertainty can be large when the θ s are estimated imprecisely and when $\theta_{(N)}$ is not much larger than one or more of the other θ s. In such cases the MCB statement would actually hold with much higher probability than $1 - \alpha$, and correspondingly the MCB intervals are (in principle) needlessly wide. There does not seem to be published evidence on the actual coverage probabilities of the MCB intervals. Ongoing work by Yangseon Kim and one of the authors provides Monte Carlo evidence on this question. For what seem to be empirically plausible parameter values, the MCB 90% probability statement often holds with a probability of over 99%. Thus one might hope to be able to improve on the MCB intervals, but how to do so is a matter for future research.

While the focus of the current paper is on multiple comparisons, we note in passing that it is also possible to make *marginal* (or univariate) comparisons with the best. Kim and Schmidt (1999, unpublished manuscript) show how to construct a set S and upper and lower bounds L_i^m, U_i^m such that $P\{(N) \in S \text{ and } L_i^m \leq \theta_{(N)} - \theta_i \leq U_i^m\} \geq 1 - \alpha$, where this statement holds for a single, given value of i .

2.4 Possible Results from MCB

The easiest case to discuss is the one in which the set S of possibly best populations has a single element, say $S = \{i\}$. This occurs if $\hat{\theta}_i > \max_{j \neq i} \hat{\theta}_j + h$, with h defined in equation (9a); that is, if $\hat{\theta}_i$ is the largest value of $\hat{\theta}$ in the sample and it exceeds the second-largest value by at least h . In this case, at confidence level $1 - \alpha$, population i is clearly the best. As noted in the previous section, this implies that the MCB intervals for populations $j \neq i$ are exactly the same as the MCC intervals with population i as the control.

When $S = \{i\}$, it is easy to see that $L_i = U_i = 0$. In fact, $U_i = 0$ if and only if $S = \{i\}$, because the condition for $U_i = 0$ is $\max_{j \neq i} \hat{\theta}_j + h < \hat{\theta}_i$, which is the same as the condition for $S = \{i\}$. Similarly, $S = \{i\}$ is equivalent to $L_j > 0 \forall j \neq i$, as is easily seen from the definition of L_j in equation (9c). Thus the following three conditions are equivalent: (a) $S = \{i\}$. (b) $L_i = U_i = 0$. (c) $L_j > 0 \forall j \neq i$.

Things are a little more complicated when S contains multiple populations. In this case $U_i > 0$ for all i , since $U_i = 0$ for any i would imply that i was the only population in S . Furthermore, it is easy to see that, if $i \in S$, $L_i = 0$. Interestingly, however, the converse is not true. We can have $L_i = 0$ even for i that are not in S . The condition for $L_i = 0$ is that $\hat{\theta}_i \geq \min_{j \in S} \hat{\theta}_j - h$, which is weaker than the condition for $i \in S$, which is that $\hat{\theta}_i \geq \max_{j=1, \dots, N} \hat{\theta}_j - h$. That is, population i is in S if $\hat{\theta}_i$ is within h of the maximal value of $\hat{\theta}$, whereas $L_i = 0$ if $\hat{\theta}_i$ is within h of the minimal value of $\hat{\theta}$ for populations in S .

Thus, when S contains multiple populations, our sample is split into possibly three groups: (a) populations in S ; these have $0 = L_i < U_i$; (b) populations not in S , but close to S ; these also have $0 = L_i < U_i$; (c) populations not in S and not close to S ; they have $0 < L_i < U_i$.

3. MCB WITH GENERAL COVARIANCE STRUCTURE

3.1 Mechanics of MCB with General Covariance Structure

In this section we consider MCC and MCB when $\hat{\theta}$ has a general covariance structure. We will call this *general MCB*. Thus we maintain **Assumption 1**, that $\hat{\theta}$ is distributed as $N(\theta, \sigma^2 C)$, with C known; but we do not maintain **Assumption 3**, so we do not assert that C is proportional to I_N . As before we also maintain **Assumption 2** so that a suitable estimate $\hat{\sigma}^2$ is available.

General MCB seems to be much less well explored than standard MCB. For example, Edwards and Hsu (1983), Hochberg and Tamhane (1987) and Hsu (1996, Chapter 4) discuss the generalization of standard MCB in

terms of the unbalanced one-way model, which differs from the balanced one-way model discussed in Section 3.1 only in that the sample sizes in the different populations are allowed to be unequal. Unequal error variances across populations would fit the same pattern. This imposes some structure on the matrix C , whereas in this section we will assume only that C is known.

We consider general MCB in two steps. The first step is to perform N general MCC problems, with each of the N populations taking its turn as the control. The second step is to convert these MCC intervals into MCB intervals. All the difficulty relative to the case of standard MCB lies in the first step.

We begin with the problem of performing MCC using population j as the control; we will repeat this for $j = 1, \dots, N$. We define δ_j as the $(N-1) \times 1$ vector whose typical element is $\theta_j - \theta_i$; explicitly, $\delta_j = [\theta_j - \theta_1, \dots, \theta_j - \theta_{j-1}, \theta_j - \theta_{j+1}, \dots, \theta_j - \theta_N]^T$. We can write $\delta_j = D_j \theta$ where D_j has e_{N-1} as its j th column, and the remaining columns are the columns (in order) of $-I_{N-1}$. Suppose that $\hat{\theta}$ is the estimate of θ , so that $\hat{\delta}_j = D_j \hat{\theta}$ is the estimate of δ_j . Then, if the variance matrix of $\hat{\theta}$ is $\sigma^2 C$, the variance matrix of $\hat{\delta}_j$ is $\sigma^2 B_j$, where $B_j = D_j C D_j'$. Let R_j be the correlation matrix corresponding to B_j . In the special case of standard MCC, R_j did not depend on j , and it took the form of the equicorrelation matrix $E(\frac{1}{2})$. In the general case R_j has no specific form, though it is easily calculated. Finally, we define $\hat{\sigma}_{\delta_j} = \hat{\sigma}_{\delta_j} = [\hat{\sigma}^2(C_{jj} + C_{jj}' - 2C_{jj})]^{1/2}$, the standard deviation of the element of $\hat{\delta}$ that equals $\hat{\theta}_j - \hat{\theta}_i$.

The MCC confidence intervals with population j as a control will take the form:

$$1 - \alpha = \mathbb{P}[\hat{\theta}_j - \hat{\theta}_i - h_{ji} \leq \theta_j - \theta_i \leq \hat{\theta}_j - \hat{\theta}_i + h_{ji} \forall i \neq j] \quad (12)$$

where $h_{ji} = d_{ji}^* \hat{\sigma}_{\delta_j}$ and where d_{ji}^* is the critical value from the appropriate multivariate t distribution. Specifically, if z is an $(N-1)$ -dimensional random vector distributed as t_{N-1, R_j} , then d_{ji}^* is defined as the solution to:

$$\mathbb{P}(\max_{1 \leq i \leq N-1} |z_i| \leq d_{ji}^*) = 1 - \alpha \quad (13)$$

The solution d_{ji}^* generally depends on j (the control population), because the correlation matrix R_j generally depends on j . This will generally need to be calculated via a simulation, since the $(N-1)$ -dimensional integral implicit in equation (13) will simplify only in special cases. Tabulation is impossible except in special cases that dictate the form of R_j in terms of a small number of parameters. To calculate the d_{ji}^* by simulation, one possibility is to do N distinct simulations, one for each value of j . When N is large, this will be computer time-consuming both because the number of simulations is large and because each of the individual simulations is complicated (involving $(N-1)$ -dimensional probability statements). A computationally more efficient procedure would calculate all N values of the d_{ji}^* in one large simulation (that is, based on one set of pseudo-random draws). Specifically, suppose we make a large number of draws ($m = 1, \dots, M$) of the N -dimensional random variable a_m from the $t_{N, C}$ distribution. Now we can calculate the $(N-1)$ -dimensional draws $b_{j,m} = D_j a_m$, $j = 1, \dots, N$; whereas before D_j is the differencing matrix with respect to the j th entry. Thus $b_{j,m}$ is a draw from t_{N-1, R_j} where $B_j = D_j C D_j'$. We require draws from t_{N-1, R_j} where R_j is the correlation matrix corresponding to B_j . Since $R_j = S_j B_j S_j'$ where S_j is the diagonal matrix with i th diagonal element equal to $B_{jii}^{-1/2}$, the required draws are $S_j b_{j,m}$. This just corresponds to dividing the i th element of $b_{j,m}$ by $B_{jii}^{1/2}$. Then, for given j , sort as before to find the appropriate quantile. While most of the calculations must be done separately for each j , they are all based on the same set of random draws. This procedure is computationally efficient because most of the computer time is used in drawing the basic random variables, not in manipulating them after they are drawn.

Once we have calculated the allowances h_{ji} ($i, j = 1, \dots, N$) that are part of the general MCC procedures, it is easy to move on to MCB intervals. The logic is essentially as before, and indeed the attraction of the argument of Edwards and Hsu (1983) is that we can convert MCC intervals into MCB intervals independently of how we constructed the MCC intervals; all that is needed is that they are valid.

Define the following notation:

$$L_i^j = \hat{\theta}_j - \hat{\theta}_i - h_{ji}, \quad U_i^j = \hat{\theta}_j - \hat{\theta}_i + h_{ji} \quad (14)$$

With this notation the MCC confidence intervals (12) can be rewritten as $\mathbb{P}(L_i^j \leq \theta_j - \theta_i \leq U_i^j \forall i \neq j) = 1 - \alpha$. That is, L_i^j and U_i^j are the lower and upper MCC bounds for $\theta_j - \theta_i$, when MCC is performed with population j as the control. The last phrase, about which population is the control, matters for general MCC, although it did not matter for standard MCC. In the case of standard MCC, where all the h_{ji} are the same, we have $L_i^j = -U_i^j$ so that the lower bound for $\theta_j - \theta_i$ when j is the control is the negative of the upper bound for $\theta_j - \theta_i$ when i is the control. In the general case, however, $h_{ji} \neq h_{ji}$ and so $L_i^j \neq -U_i^j$.

We now define the further notation:

$$S = \{i \mid U_j^i \geq 0 \forall j \neq i\} = \{i \mid \hat{\theta}_i \geq \hat{\theta}_j - h_{ij} \forall j \neq i\} \quad (15a)$$

$$L_i = \max[0, \min_{j \in S} L_j^i] = \max[0, \min_{j \in S} (\hat{\theta}_j - h_{ji}) - \hat{\theta}_i] \quad (15b)$$

$$U_i = \max[0, \max_{j \neq i} U_j^i] = \max[0, \max_{j \neq i} (\hat{\theta}_j + h_{ji}) - \hat{\theta}_i] \quad (15c)$$

Then we have the MCB result: $P[(N) \in S \text{ and } L_i \leq \theta_{(N)} - \theta_i \leq U_i \forall i] \geq 1 - \alpha$. For the proof, see Edwards and Hsu (1983, p. 966). As in the standard case, the intuition is straightforward. The lower bound is taken by minimizing the MCC lower bounds over all control populations that might be best (are in S), while the upper bound is taken by maximizing the MCC upper bounds over the populations in S .

These results are similar in form to the corresponding results for standard MCB, as given in equation (9). They reduce to the results for standard MCB in the special case that the h_{ij} do not depend on i or j .

3.2 Results from MCB with General Covariance Structure

The condition for population i to be in the set S (so that population i is possibly best) is that $U_j^i \geq 0$ for all $j \neq i$. That is, in the MCC confidence intervals with population i as a control, the upper bounds for $\theta_i - \theta_j$ are all non-negative; the prospect of $\theta_i \geq \theta_j$ is not ruled out for any j .

As in the case of standard MCB, if $i \in S$, then $L_i = 0$. (This is easily seen from equation (15b), since if $i \in S$, $\min_{j \in S} (\hat{\theta}_j - h_{ji}) - \hat{\theta}_i \leq -h_{ii} < 0$.) Thus $L_i > 0$ implies that i is not in S . However, as in the case of standard MCB, we may have $L_i = 0$ for populations i that are not in S . Thus, in general, MCB once again splits the sample into potentially three groups: populations in S , with $0 = L_i < U_i$; and populations not in S , but with $0 < L_i < U_i$.

The biggest difference between general and standard MCB is in the characterization of the case in which there is a single population in the set S of potentially best populations. As in Section 3.4, consider the following three conditions: (a) $S = \{i\}$; (b) $U_i = 0$; (c) $L_j > 0 \forall j \neq i$. These three conditions are equivalent for standard MCB but not for general MCB. To further explore the relationships between these three conditions, observe that the condition for $S = \{i\}$ is that i is in S but no other j is in S . Thus for $S = \{i\}$ we have $\hat{\theta}_i \geq \hat{\theta}_j - h_{ij} \forall j \neq i$ so that i is in S ; and also,

$$\text{for every } j \neq i, \hat{\theta}_j < \hat{\theta}_k - h_{jk} \text{ for some } k \quad (16)$$

so that j is not in S . Since the set S cannot be empty, in fact equation (16) is necessary and sufficient for the event $S = \{i\}$. The population k that dominates population j may be different for different values of j . A sufficient condition for $S = \{i\}$ is that condition (19) holds with $k = i$; that is,

$$\hat{\theta}_i > \max_{j \neq i} (\hat{\theta}_j + h_{ji}) \quad (17)$$

However, from equation (15c), this is a necessary and sufficient condition for $U_i = 0$. Thus $U_i = 0$ implies $S = \{i\}$, but not conversely. This is not a sensible outcome, since if one population is uniquely identified as best (at the given confidence level) we logically should have $0 = L_i = U_i$. Thus the definition of U_i in equation (15c) should be modified to require $U_i = 0$ if $S = \{i\}$. This point appears to have been missed in some of the MCB literature, including Edwards and Hsu (1983), but is recognized by Hsu (1996, p. 112).

Similarly, the event $L_j > 0 \forall j \neq i$ implies the event $S = \{i\}$, but not conversely. To see the first implication, we simply observe that S cannot be empty, and $L_j > 0$ implies that j is not in S . To proceed further, we observe that $L_j > 0$ if and only if $\hat{\theta}_j < \min_{i \in S} (\hat{\theta}_i - h_{ji})$. Since $L_j > 0 \forall j \neq i$, implies that S contains only the single population i , a necessary and sufficient condition for $L_j > 0 \forall j \neq i$ is

$$\hat{\theta}_i > \hat{\theta}_j + h_{ij} \forall j \neq i \quad (18)$$

or, equivalently, $\hat{\theta}_i < \hat{\theta}_j - h_{ji} \forall j \neq i$. It is evident that this is not implied by equation (16), so that $S = \{i\}$ does not imply that $L_j > 0 \forall j \neq i$.

Finally, there is no apparent connection between the events $U_i = 0$ and $L_j > 0 \forall j \neq i$. To see this, compare equations (17) and (18), which are identical except for that h_{ji} appears in (17) where h_{ij} appears in (18). Thus these conditions would be the same if $h_{ij} = h_{ji}$ for all j , as is the case in standard MCB, but not in general.

3.3 The Equicorrelated Case and Other Special Cases

For this section we return to the simplified notation of Section 2.2. Thus $\sigma^2 C$ is the variance matrix of $\hat{\theta}$, $\sigma^2 B = \sigma^2 D C D'$ is the variance matrix of $\hat{\delta}$, and R is the correlation matrix corresponding to B .

The essential requirement for the applicability of standard MCB is that $R = E(\frac{1}{2})$, as defined in equation (3). That is, the differences of the $\hat{\theta}_s$ must be equicorrelated, with correlation equal to $\frac{1}{2}$. Our Assumption 3, that $C = k I_N$, is sufficient but not necessary for $R = E(\frac{1}{2})$, and thus for the applicability of standard MCB. A weaker condition than

$C = kI_N$, that still implies $R = E(\frac{1}{2})$, is that the $\hat{\theta}_i$ be equicorrelated, with arbitrary correlation. More formally, we have the following result.

Result 1: The following conditions are equivalent, and imply $R = E(\frac{1}{2})$. (a) $C = bI_N + ae_Ne_N'$, for some scalars a and b . (b) $B = wE(\frac{1}{2})$ for some scalar w . (c) $\text{var}(\hat{\theta}_i - \hat{\theta}_j) = w \forall i, j$, for some scalar w .

This result is well known in the MCB literature and proof is given by Hsu (1996, p. 187). The scalars w in (b) and (c) are the same, and equal $2b$, where b is the scalar in (a). Hsu (1996, p. 187) refers to this as the variance-balanced case.

The conditions in Result 1 are weaker than Assumption 3, but they are still sufficient but not necessary for $R = E(\frac{1}{2})$. It is well known, and easy to verify, that $R = E(\frac{1}{2})$ is equivalent to the following structure for B :

$$B_{ii} = b_i^2, B_{ij} = \frac{1}{2}b_i b_j \text{ for } i \neq j, \text{ for some scalars } b_1, \dots, b_N \quad (19)$$

A condition on C that is equivalent to equation (19), and which is therefore necessary and sufficient for $R = E(\frac{1}{2})$, does not seem to be given in the MCB literature. To give such a condition, recall the definition of D in equation (2a) above, and define $D^+ = D(DD')^{-1}$, the Moore-Penrose inverse of D (which exists in this form because $DD' = I_N + e_Ne_N'$ is non-singular). Now consider the following condition:

$$C = D^+B(D^+)' + ae_Ne_N' \quad (20) \quad \square$$

for some scalar a and some B of the form given in equation (19). Then we have the following result:

Result 2: The following are equivalent. (a) $R = E(\frac{1}{2})$. (b) B is of the form given in equation (19). (c) C is of the form given in equation (20).

To prove this result, we note first that $DD^+ = I_{N-1}$ and $De_N = 0$. Then (c) implies (b) because, if C is of the form in equation (20), $B = DCD'$ is of the form of equation (19). To prove that (b) implies (c), define $E = C - D^+B(D^+)'$, so $C = D^+B(D^+)' + E$. Then $B = DCD' = B + DED'$. This implies that E must lie in the null space of D , or $E = ae_Ne_N'$ for some scalar a .

Results 1 and 2 are of theoretical interest, and may be of use in practical applications in deciding whether standard MCB is applicable, or approximately applicable. However, they are of somewhat limited practical use because, given the matrix C (the variance matrix of $\hat{\theta}$), the matrix R is easily calculated. To see whether standard MCB is applicable or approximately applicable, it is an easy matter to calculate R and see whether $R = E(\frac{1}{2})$, either exactly or approximately.

Two additional special cases that are featured prominently in the MCB literature are *one-way structure* and *product structure* (see e.g. Hsu, 1996, Section 7.1). The model is said to have product structure if the matrix R (the correlation matrix of $\hat{\delta}$, the vector of differences of estimated θ_i) has the form: $R_{ij} = \lambda_i \lambda_j, i \neq j$, for some set of scalars $\lambda_1, \dots, \lambda_{N-1}$. The model is said to have one-way structure if $\text{var}(\hat{\theta}_i - \hat{\theta}_j) = a_i + a_j$ for $i, j = 1, \dots, N$ and for some set of positive scalars a_1, \dots, a_N . One-way structure is sufficient but not necessary for product structure. Specifically, one-way structure implies product structure with $\lambda_i = [a_N / (a_i + a_N)]^{1/2}$. The motivation for the terminology is that one-way structure arises when C , the variance matrix of $\hat{\theta}$, is diagonal with i th diagonal element a_i . This occurs in the unbalanced one-way model, in which we are estimating the means of a set of independent normal populations, but the populations have different variances or different numbers of observations (or both). It may also occur in the regression model with panel data, as we will discuss in the next section. In the presence of product structure, some significant simplifications are possible relative to the purely general case. Specifically, the integral defining the MCC critical values can be expressed as a double integral (instead of an $N-1$ -dimensional integral) (see e.g. Hsu, 1996, p. 63, equation (3.15) for the form of this integral).

3.4 Approximations

If we cannot appeal to any special structure of the variance matrix C , the general MCC and MCB procedures described in Sections 3.1 and 3.2 apply, but are computer time-intensive. An alternative is to replace the exact but complicated general MCC procedure with a conservative approximation. Such approximations provide an easily calculated upper bound on the significance point for the $F_{N-1, r, R}$ distribution. That is, they provide a value d_j^* such that $P(\max_{1 \leq i \leq N-1} |z_i| \leq d_j^*) \geq 1 - \alpha$, whereas the exact MCC procedure finds d_j^* such that this probability equals $1 - \alpha$. Hsu (1996, Section 7.2.1 and Appendix A) discusses a number of conservative approximations, based on the Bonferroni inequality, Scheeffé's inequality, Slepian's inequality, an inequality due to Sidák (1967), an inequality due to Hunter (1976) and Worsley (1982), and a so-called 'factor analytic method' due to Hsu (1992), for which the one-factor case corresponds to product structure as discussed in Section 3.3; in this case exact results are possible. Matejcek (1992, unpublished manuscript) discusses some of these inequalities as well as a method suggested by Tamhane (1977) based on Banerjee's inequality, a procedure using a moment-based approximation to the Behrens-

Fisher problem, also suggested by Tamhane (1977), and a new technique based on a heteroscedastic selection procedure. McCann and Edwards (1996) provide a brief survey and a new procedure based on a path length inequality due to Naiman (1986). The above-mentioned approximations are all numerical (deterministic) and, as stated earlier, there is no compelling reason to prefer numerical approximations over simulated (stochastic) ones in an economic setting. Some techniques that employ Monte Carlo Methods to approximate critical points are provided in Foutz (1981), Edwards and Berry (1987) and Naiman and Wynn (1992). The salient feature of these techniques is that when N is large they produce approximate critical values quickly, compared to the potentially time-consuming process of generating critical values directly from simulation of equation (13).

Many of these techniques are available using standard commercial software. For example, see Tobias (1996) for a discussion of MCB software in SAS.

4. MCB IN THE PANEL DATA REGRESSION MODEL

In this section we consider the panel data regression model:

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} \quad i = 1, \dots, N, t = 1, \dots, T \quad (21)$$

We suppose that the parameters of interest for the MCB analysis are the intercepts, which we express as the vector $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]'$. (In Section 5 we will give some empirical examples in which these are indeed the parameters of interest.) These can be ordered in the usual way as $\alpha_{(1)} \leq \alpha_{(2)} \leq \dots \leq \alpha_{(N)}$. Then we denote the differences of intercepts from the best as $\delta_i = \alpha_{(N)} - \alpha_i$, for which MCB will provide joint confidence intervals.

Throughout this section we assume that the errors v_{it} are i.i.d. $N(0, \sigma_v^2)$, and we treat the x_{it} as fixed (independent of v). Also, unless stated otherwise, we assume a balanced panel (T observations for each value of i). This model might be estimated in a number of different ways depending on what one is willing to assume about the α_i . We will consider the simplest and most generally relevant case, which is to treat the α_i as fixed parameters, in which case the model would typically be estimated by the within (fixed-effects) estimator.

We begin with a simple point. If β were known, we could write $(y_{it} - x_{it}'\beta) = \alpha_i + v_{it}$. This is the balanced one-way model to which standard MCB applies.

Since standard MCB would be applicable if β were known, it is reasonable to presume that standard MCB is a good approximation if β is estimated sufficiently precisely. More precisely, suppose that $\hat{\beta}$ is the within (fixed-effects) estimator, and we ignore its variability and apply standard MCB. As is well known, we can obtain the fixed effects estimates of $\hat{\beta}$ and the α_i by regressing y_{it} on x_{it} and a set of N dummy variables representing populations (values of i). Equivalently, we can obtain the within estimate $\hat{\beta}$ from the regression in deviations from individual means (i.e. by regressing $y_{it} - \bar{y}_i$ on $x_{it} - \bar{x}_i$), and then $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$. This leads to the following expression for $\hat{\alpha}_i$:

$$\hat{\alpha}_i = \alpha_i + \bar{v}_i - \bar{x}_i'(\hat{\beta} - \beta) \quad (22)$$

The variance of \bar{v}_i is of order T^{-1} , while the variance of $\hat{\beta} - \beta$ is of order $[N(T-1)]^{-1}$. As a result, we would expect standard MCB to be approximately valid when N is large, since the variability in $\hat{\beta} - \beta$ will be small relative to the variability in \bar{v}_i . (Similarly, in the case of an unbalanced panel, we would expect the one-way structure, and therefore the product structure, to hold approximately when N is large.) This is an important observation, because general MCB will be numerically very complicated when N is large.

MCB is not designed as an asymptotic procedure. Indeed, the problem of comparing N populations is hard to conceptualize unless N is fixed. However, since econometricians often think in terms of asymptotics, the following two comments may be helpful. First, as just noted, standard MCB may be approximately valid when N is large. Second, MCB assumes that the $\hat{\alpha}_i$ are normally distributed. This should be so if the errors v_{it} are normal or if T is large.

Whether or not N is large, general MCB can be applied. Let \bar{x} be the $NT \times K$ matrix of x 's expressed in deviations from individual means, so that its typical row is of the form $x_{it} - \bar{x}_i$. Let \bar{x} be the $N \times K$ matrix whose i th row is \bar{x}_i . Then the variance matrix of $\hat{\beta}$ is $\sigma_v^2(\bar{x}'\bar{x})^{-1}$, and correspondingly

$$V(\hat{\alpha}) = \sigma_v^2 C \text{ with } C = T^{-1}I_N + \bar{x}(\bar{x}'\bar{x})^{-1}\bar{x}'. \quad (23) \quad \square$$

Given the matrix C , we can calculate the matrices B and R , and we can see whether or not R approximately equals $E(\frac{1}{2})$. If it does, standard MCB is approximately applicable. If not, general MCB or an approximation will be required.

Interestingly, in our empirical work we often found that C approximately had an equicorrelated structure, so that R was approximately equal to $E(\frac{1}{2})$. To understand why this is not a surprising outcome, consider the following argument. For $B = DCD'$, the variance matrix of $\hat{\delta}$, we have

$$\begin{aligned}
B_{ij} &= C_{ij} - C_{iN} - C_{jN} + C_{NN} \\
&= (\bar{x}_i - \bar{x}_N)(\bar{x}'\bar{x})^{-1}(\bar{x}_j - \bar{x}_N) \\
&= [(\bar{x}_i - \mu_x) - (\bar{x}_N - \mu_x)](\bar{x}'\bar{x})^{-1}[(\bar{x}_j - \mu_x) - (\bar{x}_N - \mu_x)]' \quad \square
\end{aligned} \tag{24}$$

Given independent populations, we would expect terms like $(\bar{x}_i - \mu_x)(\bar{x}'\bar{x})^{-1}(\bar{x}_j - \mu_x)'$ to be small for $i \neq j$, since they essentially measure the covariance of the x s across populations i and j . However, the same type of terms with $i = j$ should not necessarily be small, since they essentially measure a variance. Each element of B_{ij} for $i \neq j$ contains only one 'variance' term, namely $(\bar{x}_N - \mu_x)(\bar{x}'\bar{x})^{-1}(\bar{x}_N - \mu_x)'$. However, B_{ii} contains two 'variance' terms, namely $(\bar{x}_N - \mu_x)(\bar{x}'\bar{x})^{-1}(\bar{x}_N - \mu_x)'$ and $(\bar{x}_i - \mu_x)(\bar{x}'\bar{x})^{-1}(\bar{x}_i - \mu_x)'$. Thus, heuristically, diagonal elements of B should be roughly twice as large as off-diagonal elements, and R should be roughly equal to $E(\frac{1}{2})$. We will discuss this point further in our empirical section.

5. EMPIRICAL APPLICATIONS

5.1 Stochastic Frontier Models

In this section we demonstrate the use of MCB to construct confidence intervals for measures of technical efficiency in stochastic frontier models with panel data. We use three previously analysed panel data sets. For these data sets, Horrace and Schmidt (1996) constructed confidence intervals for efficiency measures using a number of different procedures (i.e. under a number of different assumptions), including standard MCB. Here the emphasis will be on whether standard MCB was appropriate, and whether general MCB would make a difference. We find that general MCB is called for in one of the three cases, and that it then makes a moderate difference.

Our three data sets possess rather different characteristics, most notably in the relative numbers of firms and time periods. In the first data set, the number of time periods is much smaller than the number of firms. In the second, the number of time periods is approximately equal to the number of firms, and the number of time periods is different across firms. In the third data set the number of time periods is larger than the number of firms. These characteristics affect the results in predictable ways. For example, only when the number of time periods is large and the number of firms is small are we able to identify a unique best firm.

We begin with a brief discussion of the stochastic frontier model with panel data. Stochastic frontier models were originally due to Aigner *et al.* (1977) and Meeusen and van den Broeck (1977). These models were based on cross-sectional data and strong distributional assumptions. Models with alternative distributional assumptions have been presented by Stevenson (1980), Greene (1990) and many others. The use of panel data in frontier models was first suggested by Pitt and Lee (1981) and Schmidt and Sickles (1984). More recent treatments include Cornwell and Schmidt (1995) and Greene (1997). This discussion will follow Schmidt and Sickles (1984) and Horrace and Schmidt (1996).

Consider the following logarithmic specification (e.g. Cobb-Douglas or translog) of a production function:

$$y_{it} = \beta_0 + x'_{it}\beta + v_{it} - u_i \quad i = 1, \dots, N, t = 1, \dots, T \quad (25) \square$$

Here i indexes firms or productive units and t indexes time periods. y_{it} is the scalar dependent variable representing the logarithm of output for the i th firm in period t , β_0 is a scalar intercept, x_{it} is a $K \times 1$ vector of functions of inputs (e.g. in logarithms for the Cobb-Douglas specification), β is a $K \times 1$ vector of coefficients and v_{it} is an i.i.d. error term with zero mean and finite variance. The u_i satisfy $u_i \geq 0$, and $u_i > 0$ is an indication of technical inefficiency. Note that u_i is time-invariant. For a logarithmic specification such as this the technical efficiency of the i th firm is given by $TE_i = \exp(-u_i)$, so technical inefficiency is $1 - TE_i$.

Different treatments of this model follow from different assumptions. For example, Pitt and Lee (1981) discussed maximum likelihood estimation, under the assumptions that the x_{it} are fixed, the v_{it} are i.i.d. normal, the u_i are i.i.d. according to a specified distribution (e.g. half-normal), and the u_i and v_{it} are independent. Further discussion can be found in Schmidt and Sickles (1984) or Horrace and Schmidt (1996). In this section, we treat the u_i as fixed. If we let $\alpha_i = \beta_0 - u_i$, equation (25) becomes the standard fixed effects regression model given in equation (21) above. Then $\hat{\beta}$ and $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta}$ are obtained as described in Section 4 above.

Given the estimates $\hat{\alpha}_i$, one can then define $\hat{\beta}_0 = \max_{i=1, \dots, N} \hat{\alpha}_i$ and $\hat{u}_i = \beta_0 - \hat{\alpha}_i$. Loosely speaking, these estimates are consistent as N and T both approach infinity. We require $T \rightarrow \infty$ so that $\hat{\alpha}_i \rightarrow_p \alpha_i$; whereas in equation (22) the term $(\hat{\beta} - \beta) \rightarrow_p 0$ as either $N \rightarrow \infty$ or $T \rightarrow \infty$, the term \bar{v}_i is the average of T values of v_{it} and $T \rightarrow \infty$ is required for it to converge in probability to zero. Furthermore, we require $N \rightarrow \infty$ (and some restrictions on the distribution of the u_i) so that $\max_{i=1, \dots, N} \alpha_i \rightarrow_p \beta_0$. Since $\max_{i=1, \dots, N} \alpha_i = \beta_0 - \min_{i=1, \dots, N} u_i$, we basically need the density of u to be non-zero in a neighborhood of zero so that the minimum of the u_i converges to zero as the number of u_i increases. These arguments were given by Schmidt and Sickles (1984). A more rigorous treatment of the

asymptotics for this model is given by Park and Simar (1994), who show that, in addition to $N \rightarrow \infty$ and $T \rightarrow \infty$, we need to require $T^{-1/2} \ln N \rightarrow 0$ in order to ensure the consistency of $\hat{\beta}_0$. This latter requirement limits the rate at which N can grow relative to T , in order to ensure that the upward bias induced by the 'max' operation disappears asymptotically. If N increased too rapidly relative to T , the increased bias due to maximizing over more $\hat{\alpha}_i$ could dominate the convergence of each individual $\hat{\alpha}_i$ to the corresponding α_i .

The above discussion regards zero as the absolute minimum value of the u_i , and hence β_0 as the absolute maximal value of the α_i , over any possible sample (essentially, as $N \rightarrow \infty$). This can be distinguished from the minimal value of the u_i and maximal value of the α_i in a given sample of size N , and this distinction is relevant when N is small, and therefore treated as fixed for purposes of asymptotics. For fixed N , let the ordering of the α_i be as before: $\alpha_{(1)} \leq \alpha_{(2)} \leq \dots \leq \alpha_{(N)}$, so that $\alpha_{(N)} = \max_{i=1, \dots, N} \alpha_i$, with $\alpha_{(N)} \leq \beta_0$. Similarly, let the ordering of the u_i in reverse order be $u_{(N)} \leq u_{(N-1)} \leq \dots \leq u_{(1)}$, so that $u_{(N)} = \min_{i=1, \dots, N} u_i$, with $u_{(N)} \geq 0$. Then clearly $\alpha_{(i)} = \beta_0 - u_{(i)}$ or equivalently $u_{(i)} = \beta_0 - \alpha_{(i)}$. This corresponds to comparing the α_i to the absolute standard β_0 . The alternative is to compare the α_i to the within-sample standard $\alpha_{(N)}$. This leads to the definition: $u_i^* = \alpha_{(N)} - \alpha_i = u_i - u_{(N)}$, so that $0 \leq u_i^* \leq u_i$. (Note that u_i^* corresponds to δ_i of Section 4.)

The relevance of this distinction to the present discussion is straightforward. MCB naturally takes the number of populations (N) as fixed, and generates a confidence interval for the vector of differences relative to the best population. That is, MCB will construct joint confidence intervals for the u_i^* , not for the u_i . Similarly, the joint confidence intervals for the u_i^* can be easily converted into joint confidence intervals for the values of $TE_i^* = \exp(-u_i^*)$, which we distinguish from $TE_i = \exp(-u_i)$. The intervals for TE_i^* are a monotonic transformation of the intervals for $\alpha_{(N)} - \alpha_i$. It should be noted that this transformation will cause lower bounds to become upper bounds, and vice versa.

Koop *et al.* (1997) have provided a Bayesian analysis of the stochastic frontier model that has some strong parallels to MCB. They consider four different models, depending on different assumptions about the effects. Their Standard Individual Effects (SIE) model uses an uninformative prior for α and β and is therefore very similar in spirit to the usual classical fixed effects model. They describe how to calculate the marginal posterior distribution of u_i^* (u_i^{REL} , in their notation) and TE_i^* (r_i^{REL}), so that they make marginal rather than multiple comparisons with the best, but presumably multiple comparisons could also be made. An interesting question that we have not yet pursued is how different such a Bayesian analysis based on an uninformative prior would be from MCB, in typical applications. Incidentally, both Koop *et al.* (1997) and Horrace and Schmidt (1996) note that construction of confidence intervals for u_i rather than u_i^* fundamentally requires an assumed distribution for u_i . The reader is referred to those papers for more discussion of inference on u_i given such a distributional assumption.

5.2 Empirical Examples of MCB in Stochastic Frontier Models

Indonesian rice farms

We analyse data previously analyzed by Erwidodo (1990, unpublished manuscript), Lee (1991, unpublished manuscript), Lee and Schmidt (1993) and Horrace and Schmidt (1996). For a complete discussion of the data see Erwidodo (1990, unpublished manuscript). One hundred and seventy-one rice farms in Indonesia were observed for six growing seasons, so $N = 171$ and $T = 6$, in our previous notation. 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.

The model is a Cobb-Douglas (loglinear) production function, with some additional dummy variables. Output is measured as kilograms of rice produced. Inputs included in the specification are seed (kg), urea (kg), trisodium phosphate (TSP) (kg), labour (labour-hours) and land (hectares). The specification also included the following dummy variables. DP equals 1 if pesticides were used and 0 otherwise. $DV1$ equals 1 if high yield varieties of rice were planted. $DV2$ equals 1 if mixed varieties were planted; the omitted category represents that traditional varieties were planted. DSS equals 1 for the wet season. There were several other time invariant dummy variables which had to be excluded from the analysis to preclude multicollinearity.

The estimated coefficients are not presented here, because they are not relevant to this discussion; they can be found in Horrace and Schmidt (1996, p. 268). The estimate of a , while relevant, consists of 171 elements and is too large to be presented in its entirety. Some salient facts concerning $\hat{\alpha}$ are:

$$\begin{aligned}
 \text{Largest } \hat{\alpha}_i &: \max_{j=1, \dots, 171} (\hat{\alpha}_j) = \hat{\alpha}_{164} = 5.556 \\
 \text{Second largest } \hat{\alpha}_i &: \max_{j \neq 164} (\hat{\alpha}_j) = \hat{\alpha}_{118} = 5.486 \\
 \text{Median } \hat{\alpha}_i &: \hat{\alpha}_{15} = 4.966 \\
 \text{Smallest } \hat{\alpha}_i &: \min_{j=1, \dots, 171} (\hat{\alpha}_j) = \hat{\alpha}_{45} = 4.550 \\
 \text{Second smallest } \hat{\alpha}_i &: \min_{j \neq 45} (\hat{\alpha}_j) = \hat{\alpha}_{117} = 4.586, j = 1, \dots, 171, j \neq 45 \\
 & s^2 = 0.1076, \text{ degrees of freedom} = 846
 \end{aligned}$$

Some of the results from standard MCB are presented in Table I. They are based on the 90% confidence level ($\alpha = 0.10$), and a simulated critical value for $t_{170, 0.05, E(1/2)}$ of 3.18. This implies an MCB 'allowance' of $h = 0.6022$. The set S of possibly efficient firms contains 92 of the 171 farms. The value of $\min_{i \in S} \hat{\alpha}_i$ was 4.955. Because all the remaining 79 farms had values of $\hat{\alpha}_i$ that were within 0.6022 of 4.955, all the farms had upper bounds for TE_i equal to one (i.e. lower bounds for $\alpha_{(N)} - \alpha_i$ equal to zero). The lower bounds for TE_i ranged from 0.5874 for $i = 145$ to 0.2001 for $i = 45$. It would take up too much space to present results for all 171 farms, so Table I presents the results for only eight farms: the three with the largest $\hat{\alpha}_i$; the farm with the median value of $\hat{\alpha}_i$ ($i = 15$); the farm with the smallest value of $\hat{\alpha}_i$ for $i \in S$ ($i = 119$); and the three farms with the smallest $\hat{\alpha}_i$. For each of these eight farms, we give the value of $\hat{\alpha}_i$, the standard error of $\hat{\alpha}_i$, and the MCB lower and upper bounds for the confidence intervals for TE_i .

These confidence intervals are very wide and indicate that the point estimates of efficiencies contain too much statistical noise to be taken very seriously. Horrace and Schmidt (1996) performed some sensitivity analyses to try to understand the reason for such wide confidence intervals. The intervals are shortened, but not by a large amount, if we do MCC with farm $i = 164$ as a control (i.e. treat the identity of the most efficient firm as known); if we reduce the number of farms for which we construct confidence intervals; or if we change the confidence level from 0.90 to 0.75. The main reason why the confidence intervals are wide is that the $\hat{\alpha}_i$ are not estimated very precisely, and this is due to a small value of T and a relatively large value of σ_v^2 .¹ Our conclusion is that analysis of technical efficiency for these data is difficult, and would require stronger assumptions than the fixed effects model makes.

Table I. Ninety per cent MCB confidence intervals for technical efficiency, Indonesian rice farms

Firm	$\hat{\alpha}_i$	Std error	L bound	U bound
164	5.556	0.260	0.5874	1.0000
118	5.486	0.253	0.5105	1.0000
163	5.484	0.258	0.5094	1.0000
15	4.966	0.260	0.3034	1.0000
119	4.955	0.254	0.3002	1.0000
143	4.598	0.254	0.2101	1.0000
117	4.586	0.245	0.2075	1.0000
45	4.550	0.243	0.2001	1.0000

We now turn to the question of whether standard MCB is appropriate for this analysis. As discussed in Section 4, we might hope that standard MCB will be appropriate, because N is fairly large. A direct check on the applicability of standard MCB is to calculate the correlation matrix R of the vector of differences $\hat{\delta}_i$, and see how close it is to $E(\frac{1}{T})$. When we calculated R with farm 164 as the control, we obtained an average correlation of 0.4967, with a standard deviation of 0.0139. The correlations of elements of $\hat{\delta}$ were close enough to $\frac{1}{T}$ to convince us that standard MCB was approximately valid. This is a good thing, because with $N = 171$, general MCB would have been extremely time-consuming.

Texas utilities

In this section we reanalyse data previously analyzed by Kumbhakar (1996) and Horrace and Schmidt (1996). Kumbhakar estimated a cost function, whereas we will estimate the production function. The data set consists of observations on $N = 10$ major privately owned Texas electric utilities observed annually over $T = 18$ years from 1966 to 1983, and includes information on annual labour, capital and fuel (inputs) for electrical power generation (output). 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 is a Cobb-Douglas production function. The within-estimates of the regression coefficients of this model are given in Horrace and Schmidt (1996, p. 272). We note that the estimate of σ_v^2 is 0.0029, which is much smaller than in the case of the Erwidodo data. Table II gives the intercepts $\hat{\alpha}_i$, their standard errors, and the results of standard MCB at the 90% confidence level. The critical value of 2.42 implies an MCB allowance of $h = 0.0434$. Because $\hat{\alpha}_5$ exceeds all the other $\hat{\alpha}_i$ by at least 0.0434, $S = \{5\}$; that is, firm 5 is identified as the best firm. Because there is a unique best firm, the upper and lower bounds for the efficiency levels of the other firms are not

¹ It may be noted that the standard errors of the $\hat{\alpha}_i$ give a good idea of the precision of estimation of the α_i , but not of the precision of estimation of differences of the α_i , which is what is relevant for MCB. Because the $\hat{\alpha}_i$ are strongly positive correlated, the standard error of the difference ($\hat{\alpha}_i - \hat{\alpha}_j$) is actually typically smaller than the standard error of $\hat{\alpha}_i$ or $\hat{\alpha}_j$. For example, for $i = 164$ and $j = 118$, Table I reports standard errors of 0.260 and 0.253, while the difference has a standard error of 0.192. Nevertheless, the standard error of the difference is large relative to the difference ($5.556 - 5.486 = 0.07$), so that there is considerable uncertainty about the level of inefficiency even if it were known that firm 164 was most efficient. See Horrace and Schmidt (1996) for more detail.

equal to one. The bounds for the efficiency levels of the firms are much tighter than they were for the Erwidodo data. For example, firm 8 has confidence interval for $TE_{\hat{\alpha}_i}$ of [0.7847, 0.8562], which is arguably tight enough to be informative. The bounds are tighter in this case than in the previous analysis primarily because differences of the α_i are estimated more precisely, and this is so because T is larger and σ_i^2 is smaller.²

Table II. Ninety per cent standard MCB confidence intervals for technical efficiency, Texas utilities

Firm	$\hat{\alpha}_i$	Std error	L bound	U bound
5	-4.9952	0.269	1.0000	1.0000
3	-5.0826	0.249	0.8772	0.9571
10	-5.1451	0.249	0.8241	0.8991
1	-5.1760	0.262	0.7990	0.8717
8	-5.1940	0.264	0.7847	0.8562
9	-5.2105	0.270	0.7719	0.8421
2	-5.2176	0.264	0.7664	0.8362
7	-5.2362	0.252	0.7523	0.8208
6	-5.2366	0.263	0.7520	0.8205
4	-5.2669	0.268	0.7296	0.7960

Of course, while a larger value of T improves precision of estimation, it also makes the assumption of time-invariant efficiency less plausible. Various models that allow time-varying efficiency exist (e.g. Kumbhakar, 1990; Battese and Coelli, 1992; Lee and Schmidt, 1993) and MCB would in principle be applicable; for each time period, we simply need to be able to calculate the joint distribution of the N estimated intercepts. As a rough test of the need for such a model, we did a Chow test of the hypothesis that the slopes and intercepts are all constant (against the alternative that one or more differs between the first nine observations and the last nine). We found $F_{13,154} = 12.4$, which is significant at any reasonable level, and indicates the need to consider seriously that something has changed over time. Further pursuit of these considerations is empirically relevant but not within the scope of this paper.

We now ask whether standard MCB is in fact appropriate for this data set. In fact, it is not, because the correlation matrix R is not close to $E(\frac{1}{T})$. For δ defined with firm 5 as the control, the average correlation between elements of $\hat{\delta}$ equalled 0.8599, which is not at all close to $\frac{1}{T}$, and ranged from 0.7374 to 0.9520, with a standard deviation of 0.0429. Similar results hold for δ defined with other firms as the control. As a result, standard MCB should not have been expected to be a good approximation to general MCB.

While it is not the purpose of this paper to be a corrigendum for Horrace and Schmidt (1996), it may be worth discussing the argument they used (erroneously) to justify standard MCB, so that others do not follow their example. They presented the following facts about $V(\hat{\alpha})$:

$$\begin{aligned} \text{Mean of } \bar{x}_i \text{Var}(\hat{\beta})\bar{x}'_i &= 0.06765 \\ \text{Standard deviation of } \bar{x}_i \text{Var}(\hat{\beta})\bar{x}'_i &= 0.00289 \\ \text{Maximum of } \bar{x}_i \text{Var}(\hat{\beta})\bar{x}'_i &= 0.07255 \\ \text{Minimum of } \bar{x}_i \text{Var}(\hat{\beta})\bar{x}'_i &= 0.06166 \end{aligned}$$

Based on these facts, they argued that the equicorrelated structure for $V(\hat{\alpha})$ approximately held, and that standard MCB was applicable. The approximate equality of the elements of $\bar{x}V(\hat{\beta})\bar{x}'$ is indeed relevant, but the relevant comparison is of the differences of these elements to $2(\sigma_i^2/T)E(\frac{1}{T})$. The mean of the elements of $\bar{x}V(\hat{\beta})\bar{x}'$ is irrelevant, because the variance matrix of $\hat{\delta} = D\hat{\alpha}$ is $(\sigma_i^2/T)DD' + D\bar{x}V(\hat{\beta})\bar{x}'D' = 2(\sigma_i^2/T)E(\frac{1}{T}) + D\bar{x}V(\hat{\beta})\bar{x}'D'$, and the differencing matrix D removes the mean from $\bar{x}V(\hat{\beta})\bar{x}'$ and leaves only the differences. For standard MCB to apply these differences need to be small relative to $2(\sigma_i^2/T)E(\frac{1}{T})$. Although the variation in the elements of $\bar{x}V(\hat{\beta})\bar{x}'$ was not large relative to the mean value of $\bar{x}_i \text{Var}(\hat{\beta})\bar{x}'_i$, it was quite large relative to

²We might note that the α_i themselves are not estimated more precisely for this data set than for the previous one. The standard errors of the $\hat{\alpha}_i$ are very similar in Tables I and II. However, the standard error of the difference $\hat{\alpha}_3 - \hat{\alpha}_5$ is only 0.0653, compared to 0.192 for the comparable difference for the rice farm data, and this comparison does not depend much on choice of firms. In the present data set more of the variability of $\hat{\alpha}$ is due to error in the estimation of β , and this largely differences away.

Table III. Ninety per cent general MCB confidence intervals for technical efficiency, Texas utilities

Firm	$\hat{\alpha}_i$	Std error	L. bound	U bound
5	-4.9952	0.269	0.9448	1.0000
3	-5.0826	0.249	0.7969	1.0000
10	-5.1451	0.249	0.7653	0.9845
1	-5.1760	0.262	0.7694	0.9844
8	-5.1940	0.264	0.7595	0.9731
9	-5.2105	0.270	0.7551	0.9720
2	-5.2176	0.264	0.7347	0.9425
7	-5.2362	0.252	0.7166	0.9203
6	-5.2366	0.262	0.7146	0.9169
4	-5.2669	0.268	0.7068	0.9102

$2(\sigma_i^2/T)E(\frac{1}{2})$, whose diagonal elements equal only 0.00032. As a result, the correlation matrix R is not very nearly equal to $E(\frac{1}{2})$.

As the preceding discussion makes clear, the moral of the story is simply to compute R and compare it to $E(\frac{1}{2})$ before proceeding with standard MCB. There is no obviously preferred metric for this comparison.

Table III reports the results of general MCB applied to the Kumbhakar data. The major difference from the standard MCB results is that the set S of possibly efficient firms now contains two firms ($S = \{5, 3\}$) rather than just one ($S = \{5\}$). For the firms not in S , this results in only minor changes in the lower bounds for technical efficiency, but it leads to significant increases in the upper bounds. For example, the confidence interval for TE_x was [0.7847, 0.8562] from standard MCB, but is now [0.7595, 0.9731] from general MCB. The increase in width is fundamentally due to the enlargement of the set of possibly efficient firms, and is not a general consequence of the difference between standard and general MCB. General MCB could yield a larger or smaller set S , or narrower or wider intervals, than standard MCB. What is clear is that, for this data set, standard versus general MCB matters. Given the lack of similarity of R to $E(\frac{1}{2})$, we trust the general MCB results, not those from standard MCB.

Given that standard MCB was not applicable, we could have considered the applicability of other procedures that are simpler (less computer time-consuming) than general MCB. For example, a referee pointed out correctly that we could have checked how close the correlation matrix R came to having the product structure discussed in Section 3.3 above. This comparison can be made using standard software, such as PROC FACTOR in SAS. Product structure would result in considerable simplifications relative to general MCB.

Egyptian tileries

In this section we analyse data previously analyzed by Seale (1990) and Horrace and Schmidt (1996). The data were collected by the Non-Farm Employment Project in 1982±3. The firms were located in Fayoum and Kalyubiya, Egypt. $N = 25$ Egyptian small-scale floor tile manufacturers were observed over 3-week periods for 66 weeks, for a total of 22 separate observation periods. The data contain some missing observations, so the number of separate observation periods varies across firms, making this an *unbalanced* panel. The total number of observations is 484, so T_i is on average about 19. For a complete discussion of the data see Seale (1990).

Table IV. Ninety per cent MCB confidence intervals for technical efficiency product structure, Egyptian tileries

Firm	$\hat{\alpha}_i$	Std error	L. bound	U bound
14	0.9881	0.324	0.7569	1.0000
24	0.9825	0.284	0.7499	1.0000
25	0.9770	0.297	0.7432	1.0000
3	0.9691	0.322	0.7345	1.0000
19	0.9687	0.341	0.7371	1.0000
22	0.9685	0.324	0.6765	1.0000
18	0.9399	0.283	0.7187	1.0000
5	0.9061	0.297	0.6763	1.0000
21	0.9043	0.327	0.6911	1.0000
16	0.9036	0.261	0.6601	1.0000
2	0.8834	0.320	0.6742	1.0000
23	0.8811	0.337	0.6777	1.0000
4	0.8769	0.306	0.6699	1.0000
17	0.8497	0.283	0.6567	1.0000
15	0.8121	0.330	0.6324	1.0000
11	0.6240	0.292	0.5257	1.0000
13	0.6071	0.319	0.5169	1.0000
12	0.5796	0.292	0.5029	1.0000
6	0.5494	0.292	0.4879	1.0000
10	0.5190	0.307	0.4530	1.0000
9	0.4604	0.280	0.4417	0.9428
1	0.4475	0.280	0.4276	0.9493
20	0.4256	0.313	0.4228	0.9186
7	0.3987	0.303	0.4197	0.8769
8	0.2818	0.277	0.3709	0.7855

Inputs to the production of cement floor tiles are labour (labour-hours) and machines (machine-hours). Output is in square metres of tile. The model is a Cobb-Douglas production function. The within-estimates of the regression coefficients are given in Horrace and Schmidt (1996, p. 276). The estimate of σ_v^2 equals 0.1147, which is comparable to the value for the Erwidodo data.

Horrace and Schmidt (1996) performed MCB under the assumption of the product structure discussed in Section 4.3. However, their results are incorrect due to an error in the calculation of the appropriate critical values. The corrected results for MCB assuming one-way structure and using the 90% confidence level are given in Table IV. All but five of the firms are in the set S, so we are not very sure about the identity of the best firm. The confidence intervals for the individual technical efficiencies are also fairly wide; they are much wider than for the Kumbhakar data, though not as wide as for the Erwidodo data.

The values of $\hat{\alpha}_i \sqrt{\hat{\sigma}_v^2 / T_i}$ range from 0.0587 to 0.1106, with a mean of 0.0857 and a standard deviation of 0.00941. This degree of variation does not appear small compared to the average value of $2\sigma_v^2 / T_i$, which is approximately 0.012 (based on an average value of T_i of approximately 19), and so we might suspect that general MCB would be called for. As explained in the previous section, we could have checked how close the correlation matrix, R , was to possessing a product structure. This would have involved finding a matrix, say \hat{R} , which possessed a product structure or a one-way structure and minimized some Euclidean norm criterion such as $\|R - \hat{R}\|$ or

Table V. Ninety per cent general MCB confidence intervals for technical efficiency, Egyptian tileries

Firm	$\hat{\alpha}_i$	Std error	L bound	U bound
14	0.9881	0.324	0.7357	1.0000
24	0.9825	0.284	0.7171	1.0000
25	0.9770	0.297	0.7221	1.0000
3	0.9691	0.322	0.7201	1.0000
19	0.9687	0.341	0.6949	1.0000
22	0.9685	0.324	0.7191	1.0000
18	0.9399	0.283	0.6678	1.0000
5	0.9061	0.297	0.6399	1.0000
21	0.9043	0.327	0.6736	1.0000
16	0.9036	0.261	0.6032	1.0000
2	0.8834	0.320	0.6588	1.0000
23	0.8811	0.337	0.6502	1.0000
4	0.8769	0.306	0.6462	1.0000
17	0.8497	0.283	0.6163	1.0000
15	0.8121	0.330	0.6126	1.0000
11	0.6240	0.292	0.5136	1.0000
13	0.6071	0.319	0.5065	1.0000
12	0.5796	0.292	0.4838	1.0000
6	0.5494	0.292	0.4770	1.0000
10	0.5190	0.307	0.4485	1.0000
9	0.4604	0.280	0.4263	0.9814
1	0.4475	0.280	0.4104	0.9875
20	0.4256	0.313	0.4183	0.9814
7	0.3987	0.303	0.4162	0.8869
8	0.2818	0.277	0.3587	0.8170

$\|R(\hat{R})^{-1} - I_{N-1}\|$. Given \hat{R} , we could have compared it to R , to determine if general MCB was indeed called for.

Table V gives the results for general MCB (i.e. not assuming the one-way structure). They are not very different from the results in Table IV, so in this case general MCB does not make much difference, despite the fact that the one-way structure did not appear to hold very well.

5.3 Estimation of Labour Market Wage Gaps

To illustrate another potential application of MCB in economics, consider the following example given in more detail in Horrace (1997, unpublished manuscript). The labour market wage discrimination literature is concerned with explaining differences in wages across gender, race or ethnicity. The idea is to decompose an individual's wage into a human capital component and a discriminatory component using some form of regression analysis. The discriminatory component is called the 'wage gap' or 'wage differential'. Empirical studies are often concerned with estimating these wage gaps for different employment classifications and ranking them to determine which employment classifications are the most or least discriminatory. For example, Fields and Wol (1995) estimate wage gaps across 284 industry classifications such as public utilities, retail trade and agriculture. They then rank these industry wage gaps to show which ones possess the largest wage gaps and are, hence, the most discriminatory. However, they do so without performing any inference on the ranking.

Horrace (1997, unpublished manuscript) performs an analysis similar to the Fields and Wol study, but includes MCB inference on the wage gaps. More specifically, let θ_j represent the male-female wage gap in industry $j = 1, \dots, J$. Order these in the usual way: $\theta_{(1)} \leq \theta_{(2)} \leq \dots \leq \theta_{(N)}$, so that $\theta_{(1)}$ represents the smallest male-female wage gap. Then MCB confidence intervals can be constructed for all $u_j \equiv \theta_j - \theta_{(1)}$, a measure of relative non-discrimination. The analysis provides not only the MCB confidence intervals themselves, but the set S of industries that are 'least non-discriminatory' at a prespecified confidence level. The latter result has obvious policy implications for monitoring discriminatory industries for equal opportunity compliance, if the wage gaps are indeed interpretable as reflecting discrimination.

The basis of the empirical exercise is a sample from the March 1988 *Current Population Survey*. There are $N = 14$ industries selected, and a total of 53,669 individuals. The number of individuals per industry differs across industries, so this is essentially an unbalanced panel. A regression equation is estimated for which the dependent variable is log wage, while the explanatory variables are a large number of demographic variables (including variables representing educational and employment backgrounds), plus 13 dummy variables for industries and 14 interactions between the industry dummies and a dummy variable for gender. We view this basically as a data-descriptive model; Fields and Wol used a similar specification. The wage gap θ_j is defined as the coefficient of the interaction between the dummy for gender and the dummy for industry j . Because of the logarithmic specification, the measure of relative non-discrimination is actually e^{-u_j} .

Table VI gives the MCB results, for the 95% confidence level. These are the results from general MCB, since the covariance structure of the $\hat{\theta}_j$ did not support standard MCB. The set S of potentially best industries contained six members: Personal services; Transportation, communication and public utilities; Professional and related services; Business and repair service; Mining; and Agriculture, forestry and fisheries. While the point estimates of relative non-discrimination ($e^{-\alpha}$) ranged from 0.838 to one, the MCB intervals for all industries had an upper bound of one. That is, at the 95% confidence level, the possible range of the value of the relative non-discrimination parameter may be one for every industry. Even with an extremely large sample

Table VI. Ninety-five per cent MCB confidence intervals for wage gaps

Industry	$e^{-\alpha}$	L bound	U bound
Personal Services	1.0000	0.8876	1.0000
Entertainment	0.9886	0.8622	1.0000
Transportation, Communication & Public Utilities	0.9732	0.8740	1.0000
Professional & Related Services	0.9646	0.8708	1.0000
Business & Repair Service	0.9621	0.8625	1.0000
Public Administration	0.9111	0.8177	1.0000
Retail Trade	0.9073	0.8184	1.0000
Wholesale Trade	0.9043	0.8072	1.0000
Durable Good Manufacturing	0.8975	0.8085	1.0000
Mining	0.8867	0.7411	1.0000
Finance, Insurance & Real Estate	0.8818	0.7927	1.0000
Agriculture, Forestry & Fisheries	0.8734	0.7390	1.0000
Construction	0.8601	0.7599	1.0000
Non-Durable Good Manufacturing	0.8380	0.7541	1.0000

size there is much uncertainty in the ranking of the wage gaps, and differences in the measured discrimination levels across different industries may be nothing more than statistical noise.

6. CONCLUSIONS

This paper has provided an introduction to multiple comparisons with the best (MCB) for an audience of economists and econometricians, and has given some examples of its use. Our examples deal with the frontier production function (or efficiency measurement) problem, where MCB is very naturally applicable, and with the estimation of labour market wage gaps. However, we believe that MCB has other uses in economics. One potential application that comes to mind is the cross-country comparison of growth rates. Arie Kapteyn reports (personal communication) using MCB to construct confidence intervals for differences in intercepts in welfare expenditure functions, with municipalities in the Netherlands as the cross-sectional unit of observation. We expect that numerous similar applications may exist.

It is interesting to speculate about the possible connections between construction of confidence intervals and the point estimation problem. In the frontier production function literature, the point estimate of $\alpha_{(N)} - \alpha_i$ is $(\max_j \hat{\alpha}_j - \hat{\alpha}_i)$. It is clear that this estimate is biased upward, since $\max_j \hat{\alpha}_j$ is biased upward as an estimate of $\alpha_{(N)} = \max_j \alpha_j$. How to correct this bias is a challenging question. An intriguing fact is that the MCB confidence interval for $\alpha_{(N)} - \alpha_i$ is *not* centred on the above point estimate, and in that sense implicitly recognizes the bias inherent in the max operation. Edwards and Hsu (1983) suggest the midpoint of the MCB confidence interval as a point estimate, and argue but do not prove that it should be less biased than the usual point estimate. This idea bears further investigation.

The MCB literature also includes so-called non-parametric intervals, based on Wilcoxon statistics. These avoid the normality assumption, and may be useful when we do not wish to assume the normality of the errors and when T is small enough that asymptotic normality is not relevant for $\hat{\alpha}$. Some empirical experience with these procedures would also be useful.

ACKNOWLEDGEMENTS

The second author gratefully acknowledges the financial support of the National Science Foundation.

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.
- 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.

- 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.
- Bechhofer, R. E. and C. W. Dunnett (1988), *Percentage Points of Multivariate Distributions*, Volume 11 of *Selected Tables in Mathematical Statistics*, American Mathematical Society, Providence.
- Cornwell, C. and P. Schmidt (1995), 'Production frontiers and efficiency measurement', in L. Matyas and P. Sevestre (eds), *Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd edition, Kluwer Academic Publishers, Boston, MA.
- Desu, M. M. (1970), 'A selection problem', *Annals of Mathematical Statistics*, **41**, 1596-1603.
- Dunn, O. J. and F. J. Massey Jr (1965), 'Estimation of multiple contrasts using t-distribution', *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.
- Edwards, D. G. and J. J. Berry (1987), 'The efficiency of simulation-based multiple comparisons', *Biometrics*, **43**, 913-928.
- 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.
- Fabian, V. (1962), 'On multiple decision methods for ranking population means', *Annals of Mathematical Statistics*, **33**, 248-254.
- Fields, J. and E. N. Wolpin (1995), 'Interindustry wage differentials and the gender wage gap', *Industrial and Labor Relations Review*, **49**, 105-120.
- Foutz, R. V. (1981), 'Simultaneous tests for finite families of hypotheses', *Communications in Statistics: Theory and Methods*, **11**, 1839-1855.
- Greene, W. H. (1990), 'A gamma-distributed stochastic frontier model', *Journal of Econometrics*, **46**, 141-164.
- Greene, W. H. (1997), 'Frontier production functions', in H. Pesaran and P. Schmidt (eds), *Handbook of Applied Econometrics, Vol. II-Microeconomics*, Basil Blackwell, London.
- 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 decision (selection and ranking) rules', *Technometrics*, **7**, 225-245.
- 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.
- Harter, H. L. (1980), 'History of multiple comparisons', in P. R. Krishnaiah (ed.), *Handbook of Statistics*, Vol. I, North-Holland, Amsterdam, 617-622.
- Hochberg, Y. and A. C. Tamhane (1987), *Multiple Comparison Procedures*, Wiley, New York.
- Horrace, W. C. (1998), 'Tables of percentages of the k -variate normal distribution for large values of k ', *Communications in Statistics: Simulation and Computation*, **27**, 823-831.
- Horrace, W. C. and P. Schmidt (1996), 'Confidence statements for efficiency estimates from stochastic frontier models', *Journal of Productivity Analysis*, **7**, 257-282.
- 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.
- Hsu, J. C. (1992), 'The factor analytic approach to simultaneous inference in the general linear model', *Journal of Graphical and Computational Statistics*, **1**, 151-168.
- Hsu, J. C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman and Hall, London.
- Hunter, D. (1976), 'An upper bound for the probability of a union', *Journal of Applied Probability*, **13**, 597-603.
- Koop, G., J. Osiewalski and M. F. Steel (1997), 'Bayesian efficiency analysis through individual effects: hospital cost frontiers', *Journal of Econometrics*, **76**, 77-106.
- Kumbhakar, S. C. (1990), 'Production frontiers, panel data and time-varying technical inefficiency', *Journal of Econometrics*, **46**, 201-212.
- Kumbhakar, S. C. (1996), 'Estimation of cost efficiency with heteroscedasticity: an application to electric utilities', *Journal of the Royal Statistical Society, Series D*, **45**, 319-335.
- Lee, Y. H. and P. Schmidt (1993), 'A production frontier model with flexible temporal variation in technical efficiency', in H. O. Fried, C. A. K. Lovell and S. S. Schmidt (eds), *The Measurement of Productive Efficiency*, Oxford University Press, New York, pp. 237-255.

- McCann, M. and D. Edwards (1996), 'A path length inequality for the multivariate-t distribution, with applications to multiple comparisons', *Journal of the American Statistical Association*, **91**, 211-216.
- Meeusen, W. and J. Van den Broeck (1977), 'Efficient estimation from Cobb-Douglas production functions with composed error', *International Economic Review*, **18**, 435-444.
- Naiman, D. Q. (1986), 'Conservative confidence bands in curvilinear regression', *The Annals of Statistics*, **14**, 896-906.
- Naiman, D. Q. and H. P. Wynn (1992), 'Inclusion-exclusion Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics', *The Annals of Statistics*, **20**, 43-76.
- 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.
- Schmidt, P. and R. C. Sickles (1984), 'Production frontiers and panel data', *Journal of Business and Economic Statistics*, **2**, 367-374.
- 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.
- Sidak, Z. (1967), 'Rectangular confidence regions for the mean of multivariate normal distributions', *Journal of the American Statistical Association*, **62**, 626-633.
- 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 in Statistics, Series A*, **6**, 15-32.
- Tobias, R. (1996), 'Multiple comparisons with the best', *SAS Sample Library*, SAS Institute, Inc., Cary, NC.
- Worsley, K. J. (1982), 'An improved Bonferroni inequality and applications', *Biometrika*, **69**, 297-302.