Title: Metafrontier Frameworks for the Study of Firm-Level Efficiencies and Technology Ratios

Authors: Associate Professor Christopher J. O'Donnell[1]
School of Economics
University of Queensland
Brisbane, QLD 4072
Australia
Email: c.odonnell@economics.uq.edu.au

Professor D.S. Prasada Rao
School of Economics
University of Queensland
Brisbane, QLD 4072
Australia

Dr. George E. Battese
School of Economics
University of New England
Armidale, NSW 2351
Australia

Running Title: Metafrontier Frameworks

Keywords: Stochastic Frontier Analysis, Data Envelopment Analysis, Technical Efficiency

---

[1] Corresponding author.

Metafrontier Frameworks for the Study of Firm-Level Efficiencies and Technology Ratios *

Christopher J. O'Donnell, D.S. Prasada Rao, George E. Battese

Abstract. This paper uses the concept of a metafrontier to compare the technical efficiencies of firms that may be classified into different groups. The paper presents the basic analytical framework necessary for the definition of a metafrontier, shows how a metafrontier can be estimated using non-parametric and parametric methods, and presents an empirical application using cross-country agricultural sector data. The paper also explores the issues of technological change, time-varying technical inefficiency, multiple outputs, different efficiency orientations, and firm heterogeneity.

* The authors thank the editor, associate editor and the referees for comments and suggestions that have contributed to improving the paper.

1. Introduction

Firms in different industries, regions and/or countries[2] face different production opportunities. Technically, they make choices from different sets of feasible input-output combinations. These so-called *technology sets* differ because of differences in available stocks of physical, human and financial capital (e.g., type of machinery, size and quality of the labour force, access to foreign exchange), economic infrastructure (e.g., number of ports, access to markets), resource endowments (e.g., quality of soils, climate, energy resources) and any other characteristics of the physical, social and economic environment in which production takes place. Such differences have led efficiency researchers to estimate separate production frontiers for different groups of firms. For example, separate frontiers have been estimated for universities in Canada (McMillan and Chan, 2004), Australia (Worthington and Lee, 2005) and the United Kingdom (Glass, McKillop and Hyndman, 1995), and for bank branches in South Africa (O'Donnell and van der Westhuizen, 2002) and Spain (Lovell and Pastor, 1997).

After using data on a group of firms to estimate a production frontier, it is common and straightforward to measure the relative performance of firms within the group (e.g., Canadian universities). However, there is often considerable interest in measuring the performance of firms across groups (e.g., comparing efficiency levels in Canadian universities with efficiency levels in Australian universities). Unfortunately, such comparisons are only meaningful in the limiting special case where frontiers for different groups of firms are identical. As a general rule, efficiency levels measured relative to one frontier (e.g., the Canadian frontier) cannot be compared with efficiency levels measured relative to another frontier (the Australian frontier).

One of the aims of this paper is to develop a formal theoretical framework for making efficiency comparisons across groups of firms. We do this by measuring efficiency

[2] In this paper, we use the term *firm* in a generic sense (i.e., to refer to all types of production entities and decision-making units) and we restrict our attention to cross-sectional variations in technology sets (i.e., variations across industries and countries). The concepts and methods we discuss are also relevant and valid when technology sets vary over time.

relative to a common *metafrontier*, defined as the boundary of an unrestricted technology set. We also define *group frontiers* to be the boundaries of restricted technology sets, where the restrictions derive from lack of economic infrastructure and/or other characteristics of the production environment, as discussed above. Importantly, the metafrontier envelops the group frontiers.[3] Thus, efficiencies measured relative to the metafrontier can be decomposed into two components: a component that measures the distance from an input-output point to the group frontier (the common measure of technical efficiency); and a component that measures the distance between the group frontier and the metafrontier (representing the restrictive nature of the production environment).

Estimates of technical efficiency are often used to design programs for performance improvement. These programs involve changes to the *management and structure of the firm*. Estimates of the gap between group frontiers and the metafrontier can also be used to design programs for performance improvement, but these programs involve changes to the *production environment*. Governments can change characteristics of the production environment by, for example, building roads and ports, deregulating financial markets, and relaxing labour laws. Firms in some industries (e.g., finance and manufacturing) can also change characteristics of their production environment through relocation (witness the number of U.S. clothing firms that have relocated their manufacturing operations to south-east Asia, and the number of telephone call centres that now operate out of India). Governments have reduced capacity to change some of the physical and cultural characteristics of the production environment (e.g., soil quality), and firms in some industries (e.g., agriculture) may have limited capacity to move their operations to more productive locations. In such cases, measures of the gaps between group frontiers and the metafrontier are informative, but may be of little use in designing performance-improvement programs.

Another aim of this paper is to show how metafrontiers and group frontiers can be estimated using data envelopment analysis (DEA) and stochastic frontier analysis

[3] Thus, the metafrontier is related to the concept of the metaproduction function defined by Hayami and Ruttan (1971, p. 82): “the metaproduction function can be regarded as the envelope of commonly conceived neoclassical production functions”.

(SFA) techniques. Battese and Rao (2002) present an SFA approach to the estimation of metafrontiers that is implicitly underpinned by two different data-generating mechanisms, one that explains deviations between observed outputs and (fixed) group frontiers, and another that explains deviations between observed outputs and the metafrontier (also fixed). The problem with this approach is that points on the estimated metafrontier may lie below points on the estimated group frontiers. Battese, Rao and O'Donnell (2004) solve the problem by specifying a single data-generating process that explains deviations between observed outputs and group frontiers, and by defining the metafrontier to be a function that envelops the deterministic components of the group frontiers. However, they only consider estimation of the metafrontier using one type of SFA methodology. This paper considers both DEA and (alternative) SFA approaches to estimating both metafrontiers and group frontiers, and for decomposing differences in performance into technical efficiency and technology gap effects. To highlight differences between the DEA and SFA approaches, we use country-level data drawn from the Food and Agriculture Organization (FAO) of the United Nations to make inter-regional comparisons of agricultural efficiency.

The final aim of the paper is to confirm that our basic analytical framework and estimation methods can be extended to deal with issues such as technological change, time-varying inefficiency effects, and multiple outputs. The ability to deal with these issues means that the metafrontier approach can be adapted for use in empirical contexts where relatively sophisticated DEA and SFA methods are already routinely applied.

The structure of the paper is as follows. In Section 2, we explain the microeconomic theory underpinning metafrontiers, and show how distances between observed data points and the metafrontier can be decomposed into what we refer to as *metatechnology ratios* and conventional radial measures of technical efficiency. In Section 3, we explain how group frontiers and metafrontiers can be estimated using DEA and SFA methods. SFA estimation is especially complicated by the theoretical requirement that the metafrontier envelops the group frontiers. In Section 4, we illustrate the differences between the DEA and SFA approaches by estimating agricultural production frontiers for countries in Africa, the Americas, Europe and Asia (i.e., four groups). In Section 5, we discuss extensions to the methodological framework to accommodate technological change, time-invariant inefficiency effects, multiple outputs, and alternative efficiency orientations. In Section 6, we summarise the paper and identify two opportunities for further research.

2. The Basic Analytical Framework

Efficiency measurement is deeply rooted in production theory and the concept of distance functions. In this section, we define the metafrontier and group frontiers in terms of output sets and output distance functions. We then show how output distance functions can be used to define technical efficiencies and metatechnology ratios.

2.1 The Metafrontier

Let **y** and **x** be nonnegative real output and input vectors of dimension $M\times 1$ and $N\times 1$, respectively. The *metatechnology set* contains all input-output combinations that are technologically feasible. Formally:

$$T = \{(\mathbf{x},\mathbf{y}) : \mathbf{x} \geq \mathbf{0}; \mathbf{y} \geq \mathbf{0};\ \mathbf{x} \text{ can produce } \mathbf{y}\}. \tag{1}$$

Associated with this metatechnology set are input and output sets. For example, the output set is defined for any input vector, **x**, as:

$$P(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in T\}. \tag{2}$$

We refer to the boundary of this output set as the output *metafrontier*. We assume the output set satisfies the standard regularity properties listed in Färe and Primont (1995).

Since the main focus of this paper is to measure efficiency, it is convenient to represent the technology using the *output metadistance function*, defined as:

$$D(\mathbf{x},\mathbf{y}) = \inf_{\theta}\{\theta > 0 : (\mathbf{y}/\theta) \in P(\mathbf{x})\}. \tag{3}$$

This function gives the maximum amount by which a firm can radially expand its output vector, given an input vector. The distance function inherits its regularity properties from the regularity properties of the output set. An observation (**x**, **y**) can be considered technically efficient with respect to the metafrontier if and only if $D(\mathbf{x},\mathbf{y}) = 1$.

2.2 Group Frontiers

It is also possible to conceptualise the existence of sub-technologies that represent the production possibilities of groups of firms. We consider the case where the universe of firms can be divided into *K* (>1) groups, and we suppose that resource, regulatory or other environmental constraints may prevent firms in certain groups from choosing from the full range of technologically feasible input-output combinations in the metatechnology set, *T*. Rather, the input-output combinations available to firms in the *k*-th group are contained in the group-specific technology set:

$$T^k = \left\{(\mathbf{x},\mathbf{y}) : \mathbf{x} \geq \mathbf{0}; \mathbf{y} \geq \mathbf{0};\ \mathbf{x} \text{ can be used by firms in group } k \text{ to produce } \mathbf{y}\right\}. \quad (4)$$

The *K* group-specific technologies can also be represented by the following group-specific output sets and output distance functions:

$$P^k(\mathbf{x}) = \left\{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in T^k\right\},\ k = 1, 2,\ldots, K; \text{ and} \quad (5)$$

$$D^k(\mathbf{x},\mathbf{y}) = \inf_\theta \left\{\theta > 0 : (\mathbf{y}/\theta) \in P^k(\mathbf{x})\right\},\ k = 1, 2, \ldots, K. \quad (6)$$

We refer to the boundaries of the group-specific output sets as *group frontiers*. If the output sets, $P^k(\mathbf{x})$, *k = 1, 2, …, K,* satisfy standard regularity properties then the distance functions, $D^k(\mathbf{x},\mathbf{y})$, *k = 1, 2, …, K,* also satisfy standard regularity properties. Irrespective of the properties of these sets and functions, it is clear that

R.1 If $(\mathbf{x},\mathbf{y}) \in T^k$ for any k then $(\mathbf{x},\mathbf{y}) \in T$;

R.2 If $(\mathbf{x},\mathbf{y}) \in T$ then $(\mathbf{x},\mathbf{y}) \in T^k$ for some k;

R.3 $T = \{T^1 \cup T^2 \cup ... \cup T^K\}$; and

R.4 $D^k(\mathbf{x}, \mathbf{y}) \geq D(\mathbf{x}, \mathbf{y})$ for all $k = 1, 2, \ldots, K$.

These properties follow from the fact that the group-specific output sets, $P^k(\mathbf{x})$, $k = 1, 2, \ldots, K$, are subsets of the unrestricted output set, $P(\mathbf{x})$. This is illustrated in Figure 1[4] where we depict the production possibilities available to single-input, single-output firms from three different groups. The group-*k* frontier is labelled *k*-*k*' and is assumed to be convex (k = 1, 2, 3). If the three groups are exhaustive (i.e., if K = 3) then the group-specific frontiers envelop all the input-output combinations that could be produced by *any single firm*, implying the metafrontier is the nonconvex piecewise frontier, 1-*B*-3'. However, if the three groups are not exhaustive, then other input-output combinations may be feasible and the metafrontier could conceivably be the convex frontier, *M*-*M*'. Thus,

R.5 Convex $P(\mathbf{x})$ does not necessarily imply convex group output sets, $P^k(\mathbf{x})$, $k = 1, 2, \ldots, K$; and vice versa.

[FIGURE 1 NEAR HERE]

2.3 Technical Efficiencies and Metatechnology Ratios

Recall from Section 2.1 that an observation ($\mathbf{x}$, $\mathbf{y}$) is technically efficient with respect to the metafrontier if and only if $D(\mathbf{x}, \mathbf{y}) = 1$. More generally, an output-orientated measure of the *technical efficiency* of an observed pair ($\mathbf{x}$, $\mathbf{y}$) with respect to the metatechnology is:

$$TE(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}, \mathbf{y}). \qquad (7)$$

[4] We thank Gunnar Breustedt from the Department of Agricultural Economics at the University of Kiel for his comments made on the metafrontier, as proposed in Battese, Rao and O'Donnell (2004). These comments, which were made in a personal communication to George Battese, have influenced our discussion of Figure 1.

For example, $D(\mathbf{x},\mathbf{y})$ = 0.6 indicates that the output vector, **y**, is 60 per cent of the maximum output that could be produced by a firm using the input vector, **x**.

We can also measure technical efficiency with respect to the group-*k* frontier. Specifically, an output-orientated measure of technical efficiency with respect to the technology of group *k* is:

$$TE^k(\mathbf{x},\mathbf{y}) = D^k(\mathbf{x},\mathbf{y})\,. \qquad (8)$$

For example, if $D^k(\mathbf{x},\mathbf{y})$ = 0.8 then output is 80 per cent of the maximum output that could be produced by a firm using the input vector, **x**, and group-*k* technology.

Result R.4 states that the group-*k* output distance function, $D^k(\mathbf{x},\mathbf{y})$, can take a value no less than the output metadistance function, $D(\mathbf{x},\mathbf{y})$. This is another way of saying that the metafrontier envelops the group-*k* frontier. Whenever a strict inequality is observed between the group-*k* distance function and the metadistance function, we can obtain a measure of how close the group-*k* frontier is to the metafrontier. Specifically, the output-orientated *metatechnology ratio*[5] for group-*k* firms is defined as:

$$MTR^k(\mathbf{x},\mathbf{y}) = \frac{D(\mathbf{x},\mathbf{y})}{D^k(\mathbf{x},\mathbf{y})} = \frac{TE(\mathbf{x},\mathbf{y})}{TE^k(\mathbf{x},\mathbf{y})}. \qquad (9)$$

Using the numerical examples above, where the technical efficiency of (**x**, **y**) with respect to the metatechnology was 0.6 and the technical efficiency with respect to the group-*k* frontier was 0.8, the metatechnology ratio is found to be 0.75 (= 0.6/0.8). This means that, given the input vector, the maximum output that could be produced by a firm from group *k* is 75 per cent of the output that is feasible using the metatechnology.

[5] Battese, Rao and O'Donnell (2004, p. 94) refer to this measure as the "technology gap ratio". However, *increases* in the (technology gap) ratio imply *decreases* in the gap between the group frontier and the metafrontier. To avoid confusion, we use the term "metatechnology ratio" here.

Equation (9) provides for the following convenient decomposition of the technical efficiency of a particular input-output combination:

$$TE(\mathbf{x},\mathbf{y}) = TE^k(\mathbf{x},\mathbf{y}) \times MTR^k(\mathbf{x},\mathbf{y}). \tag{10}$$

This shows that technical efficiency measured with reference to the metafrontier (representing the existing state of knowledge) can be decomposed into the product of technical efficiency measured with reference to the group-*k* frontier (representing the existing state of knowledge and the physical, social and economic environment that characterises group *k*) and the metatechnology ratio for group-*k* (which measures how close the group-*k* frontier is to the metafrontier). Policies and programs for efficiency improvement are often targeted at either firms (e.g., education and training programs) or characteristics of the firm's operating environment (e.g., construction of railways and port facilities). The decomposition given by equation (10) is useful because it allows policymakers to better assess the potential payoffs from these different types of programs.

Finally, the decision to assume a convex or nonconvex metafrontier has implications for measures of efficiency and the metatechnology ratio. For example, consider a firm from group 2 that produces at the input-output combination labelled *A* in Figure 1. If the metafrontier is the nonconvex frontier, 1-*B*-3', then our measures of technical efficiency and the metatechnology ratio are

$$TE(A) = \frac{0C}{0E} \approx 0.67\,, \tag{11}$$

$$TE^2(A) = \frac{0C}{0D} \approx 0.74, \text{ and} \tag{12}$$

$$MTR^2(A) = \frac{TE(A)}{TE^2(A)} = \frac{0C/0E}{0C/0D} = \frac{0D}{0E} \approx 0.90. \tag{13}$$

However, if the metafrontier is the convex function labelled *M*-*M*', our measures are:

$$TE(A) = \frac{0C}{0F} \approx 0.60, \quad (14)$$

$$TE^2(A) = \frac{0C}{0D} \approx 0.74, \text{ and} \quad (15)$$

$$MTR^2(A) = \frac{TE(A)}{TE^2(A)} = \frac{0C/0F}{0C/0D} = \frac{0D}{0F} \approx 0.81. \quad (16)$$

Thus, the convexity property leads to lower metatechnology ratios and lower measures of efficiency with respect to the metafrontier.

3. Estimation

Empirical measurement of technical efficiencies and metatechnology ratios requires an empirical description of the technology. Once data on the inputs and outputs for random samples of firms from the different groups are available, we can estimate the metafrontier and the group frontiers using DEA or SFA. In this section, we consider the problem of estimating the metafrontier when panel data are available on single-output firms. The case of multiple-output firms is discussed in Section 5.

3.1 Data Envelopment Analysis

Two types of DEA models are found in the mainstream efficiency literature – an input-orientated model and an output-orientated model. In the input-orientated case, DEA defines the frontier by holding output levels constant and seeking the maximum proportional reduction in input usage that is compatible with the technology set. In the output-orientated case, the DEA method holds inputs constant and seeks the maximum possible proportional increase in outputs. The two measures give the same technical efficiency scores if the technology exhibits constant returns-to-scale (CRS), but different scores when the technology exhibits variable returns-to-scale (VRS).

It is possible to construct a convex *group-k frontier* by applying the DEA method to all the observed inputs and outputs of firms in that group. If group k consists of data on L_k firms and there are T periods, the VRS output-orientated DEA problem is:

$$
\begin{aligned}
\max_{\phi_{it},\boldsymbol{\lambda}_{it}} \quad & \phi_{it} \\
\text{s.t.} \quad & \phi_{it} y_{it} - \mathbf{y}'\boldsymbol{\lambda}_{it} \le 0 , \\
& \mathbf{X}\boldsymbol{\lambda}_{it} - \mathbf{x}_{it} \le 0 , \\
& \mathbf{j}'\boldsymbol{\lambda}_{it} = 1 \\
\text{and} \quad & \boldsymbol{\lambda}_{it} \ge 0 . \qquad (17)
\end{aligned}
$$

where

y_{it} is the output quantity for the i-th firm in the t-th period;
$\mathbf{x}_{it}$ is the $N\times 1$ vector of input quantities for the i-th firm in the t-th period;
$\mathbf{y}$ is the $L_kT\times 1$ vector of output quantities for all L_k firms in all T periods;
$\mathbf{X}$ is the $N\times L_kT$ matrix of input quantities for all L_k firms in all T periods;
$\mathbf{j}$ is an $L_kT\times 1$ vector of ones;
$\boldsymbol{\lambda}_{it}$ is an $L_kT\times 1$ vector of weights; and
ϕ_{it} is a scalar.

The value of ϕ_{it} that solves the linear program (LP) defined by equation (17) can be shown to be no less than one and provides information on the technical efficiency of the i-th firm in the t-th period. Specifically, $\phi_{it} - 1$ is the proportional increase in outputs that could be achieved when the input quantities of the i-th firm in the t-th period are held constant. Thus, $1/\phi_{it}$ is (an estimate of) the output-orientated technical efficiency measure given by equation (8). The value of $\boldsymbol{\lambda}_{it}$ that solves the LP defined by equation (17) provides information on the *peers* of the i-th firm in the t-th period. These peers are efficient points that define the facet of the frontier onto which the inputs and outputs of the i-th firm in the t-th period are projected. Solving the LP defined by equation (17) separately for every firm in the group in every time period identifies all the facets on the group-k frontier. In practice, the tedious work of solving a different LP for every firm in every period is usually undertaken using purpose-built software packages such as DEAP (see Coelli, 1996b).

A convex *metafrontier* can be identified by applying the DEA method to the inputs and outputs of all $L = \sum_k L_k$ firms in all T periods. The structure of the metafrontier LP is identical to that of equation (17) except that $\mathbf{X}$ is of dimension $N \times LT$, and $\mathbf{y}$ and $\boldsymbol{\lambda}_{it}$ are $LT \times 1$. Solving this metafrontier LP separately for every firm in the sample in every period yields estimates of firm efficiencies with respect to the metafrontier, and also identifies all the facets on the metafrontier. The value of ϕ_{it} that solves the group-k problem can be shown to be no greater than the value of ϕ_{it} that solves the metafrontier problem.[6] Thus, firms will be no more technically efficient when they are assessed against the metafrontier than against the group frontiers, and the metafrontier will never lie below any of the group frontiers.

Finally, having estimated the technical efficiencies of firms with respect to the metafrontier and group frontiers, it is straightforward to estimate the metatechnology ratio at observed input and output levels using equation (9).

3.2 Stochastic Frontier Analysis

SFA involves parameterising the frontier and estimating it using econometric techniques. A stochastic *group-k frontier* model is

$$y_{it} = f(x_{1it}, x_{2it}, \ldots, x_{Nit}; \beta^k) e^{V_{it}^k - U_{it}^k} \tag{18}$$

where x_{nit} is the n-th input quantity of the i-th firm in the t-th period; β^k is an unknown parameter vector associated with the k-th group; the V_{it}^ks represent statistical noise and are assumed to be independently and identically distributed as $N(0, \sigma_{vk}^2)$-random variables; and the U_{it}^ks represent inefficiency and are defined by

[6] The dual to the group-k maximisation LP is a minimisation problem with $L_k T$+1 constraints, while the dual to the metafrontier maximisation LP is a minimisation problem with LT+1 constraints. The constraints in the group-k dual problem are a subset of the constraints in the metafrontier problem, so the minimised value of the former is no greater than the minimised value of the latter.

the truncation (at zero) of $N(\mu_{it}^k, \sigma_k^2)$-distributions, where the μ_{it}^ks are defined by some appropriate inefficiency model (e.g., the model of Battese and Coelli, 1995)[7]. If the exponent of the frontier production function is linear in the parameter vector, β^k, then the model can be written

$$y_{it} = f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta^k) e^{V_{it}^k - U_{it}^k} \equiv e^{\mathbf{x}_{it}' \beta^k + V_{it}^k - U_{it}^k} \quad (19)$$

where $\mathbf{x}_{it}$ is now a vector of (transformations of) inputs for the *i*-th firm in the *t*-th period. Data on the inputs and outputs of firms in the *k*-th group can be used to obtain either least-squares or maximum-likelihood (ML) estimates of the unknown parameters of this frontier. ML estimates are programmed to be calculated in the software package FRONTIER (Coelli, 1996a). Following estimation, the technical efficiency of the *i*-th firm in the *t*-th period with respect to the group-*k* frontier can be obtained using the result:

$$TE_{it}^k = \frac{y_{it}}{e^{\mathbf{x}_{it}' \beta^k + V_{it}^k}} = e^{-U_{it}^k}. \quad (20)$$

A predictor for this technical efficiency measure, as proposed in Battese and Coelli (1988), is also programmed to be calculated in FRONTIER.

A deterministic *metafrontier* production function is

$$y_{it}^* = f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta) \equiv e^{\mathbf{x}_{it}' \beta} \quad (21)$$

where y_{it}^* is the metafrontier output and β is a vector of metafrontier parameters satisfying the constraints

$$\mathbf{x}_{it}' \beta \geq \mathbf{x}_{it}' \beta^k \text{ for all } k = 1, 2, \ldots, K. \quad (22)$$

7 An alternative model that accounts for heteroscedasticity is proposed by Wang (2002).

Two features of the model given by equations (19) to (22) are noteworthy. First, the constraints given by (22) imply that the metafrontier function cannot fall below any of the group frontiers. Second, the model is underpinned by only one data-generating process. This contrasts with the stochastic metafrontier model of Battese and Rao (2002) that assumes a different data-generating mechanism for the metafrontier than for the different group frontiers. Their stochastic metafrontier can be conveniently estimated using FRONTIER (using the inputs and outputs of all firms in all groups in all time periods), but there is no guarantee that the estimated metafrontier will envelop the estimated group frontiers.

An estimated metafrontier function that envelops the estimated group frontiers can be obtained by solving the optimisation problem:

$$\min_{\beta} \quad \sum_{i=1}^{L}\sum_{t=1}^{T}\left[\ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta) - \ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \hat{\beta}^k)\right]$$

$$\text{s.t.} \quad \ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta) \geq \ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \hat{\beta}^k), \text{ for all } i \text{ and } t; \qquad (23)$$

where $\hat{\beta}^k$ is the estimated coefficient vector associated with the group-*k* stochastic frontier. Since these estimated coefficient vectors are fixed for the above problem, an equivalent form of the LP defined by equation (23) is

$$\min_{\beta} \quad \sum_{i=1}^{L}\sum_{t=1}^{T}\ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta)$$

$$\text{s.t.} \quad \ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \beta) \geq \ln f(x_{1it}, x_{2it}, ..., x_{Nit}; \hat{\beta}^k), \text{ for all } i \text{ and } t. \qquad (24)$$

Furthermore, if the function *f(.)* is log-linear in the parameters (as assumed in the empirical application in this paper), the LP problem becomes:

$$\min_{\beta} \quad \bar{\mathbf{x}}'\beta$$

$$\text{s.t.} \quad \mathbf{x}_{it}'\beta \geq \mathbf{x}_{it}'\hat{\beta}^k \text{ for all } i \text{ and } t, \qquad (25)$$

where $\bar{\mathbf{x}}$ is the arithmetic average of the $\mathbf{x}_{it}$-vectors over all firms in all periods.

This optimisation problem, and a similar problem involving minimisation of a sum of squared deviations, is discussed in more detail in Battese, Rao and O'Donnell (2004). Standard errors for the estimators for the metafrontier parameters can be obtained using simulation or bootstrapping methods.

After solving the LP problem defined by equation (25), estimates of metatechnology ratios and technical efficiencies with respect to the metafrontier can be obtained using the following decomposition of equation (19):

$$y_{it} = e^{-U_{it}^k} \times \frac{e^{\mathbf{x}_{it}'\beta^k}}{e^{\mathbf{x}_{it}'\beta}} \times e^{\mathbf{x}_{it}'\beta + V_{it}^k}. \tag{26}$$

The first term on the right-hand side is the technical efficiency of the *i*-th firm in the *t*-th period with respect to the group-*k* frontier, defined by equation (20). The second term on the right-hand side is the metatechnology ratio for the *i*-th firm in the *t*-th period (in the *k*-th group):

$$MTR_{it}^k = \frac{e^{\mathbf{x}_{it}'\beta^k}}{e^{\mathbf{x}_{it}'\beta}}. \tag{27}$$

Estimating the metatechnology ratio is simply a matter of substituting estimates of β and β^k into equation (27). The constraints in the LP problem defined by equation (25) guarantee that metatechnology ratios estimated in this manner will lie in the unit interval. Finally, equations (9), (20), (26) and (27) together imply that the technical efficiency of the *i*-th firm in the *t*-th period with respect to the metafrontier is

$$TE_{it} = \frac{y_{it}}{e^{\mathbf{x}_{it}'\beta + V_{it}^k}}. \tag{28}$$

Thus, technical efficiency relative to the metafrontier is defined in an analogous way to equation (20) – it is the ratio of the observed output relative to the frontier output, adjusted for the corresponding random error. In practice, it is convenient to predict technical efficiency with respect to the metafrontier using the decomposition:

$$\hat{TE}_{it} = \hat{TE}_{it}^{k} \times \hat{MTR}_{it}^{k} \qquad (29)$$

where $\hat{TE}_{it}^{k}$ and $\hat{MTR}_{it}^{k}$ are the predictors discussed in connection with equations (20) and (27).

4. Empirical Application

In this section, we use country-level data drawn from the Food and Agriculture Organization (FAO) of the United Nations to make inter-regional comparisons of agricultural efficiency. Our main aim is not to provide a detailed analysis of global agricultural productivity, but rather to illustrate the concepts and methods discussed in Sections 2 and 3 above. Because our application is illustrative, we make several simplifying assumptions concerning, for example, technological change and the time-varying nature of technical inefficiency. Methods for relaxing some of these assumptions are discussed in Section 5.

4.1 Data

We use data exclusively drawn from the FAOSTAT system used for dissemination of statistics compiled at the Statistics Division of the FAO in Rome. The Statistics Division maintains a website where data on agriculture are made available to potential users. The site can be reached through the web address: www.fao.org.

The data set contains observations on 97 countries covering the five-year period from 1986 to 1990. These countries are evenly distributed across the globe and account for roughly 99 per cent of the world's population and agricultural output. To maintain the terminology of previous sections, the 97 countries in the data set can be thought of as *firms*, and these countries/firms are classified into four *groups*: Africa (27 countries); The Americas (21 countries); Asia (26 countries); and Europe (23 countries). These regional groupings are those defined by the FAO for its purposes of dissemination of information and data. We are not seeking to justify this grouping of the countries for a detailed analysis of agricultural productivity or efficiency. A complete list of the countries in each group is provided in Table 1. Australia, New Zealand and Papua New Guinea are included in Asia. The former USSR is included in Europe.

The data set comprises observations on one output variable (an aggregate of 185 agricultural commodities) and five input variables[8] (land, machinery, labour, fertiliser and livestock). A full description of the variables in our empirical application is provided in Appendix A.

4.2 DEA Results

DEA estimates of the group frontiers and the metafrontier were obtained using DEAP 2.1 (see Coelli, 1996b). All results were obtained using the VRS output-orientated DEA model given by equation (17). In the remainder of this paper, the acronyms DEA-K and DEA-MF are used to refer to DEA estimates of technical efficiencies relative to the group-*k* frontiers and the metafrontier, respectively. The acronym DEA-MTR is used to refer to DEA estimates of the metatechnology ratio.

Technical efficiencies and metatechnology ratios were estimated for each of the 97 countries in each of five years. Table 2 reports descriptive statistics for these estimates for selected countries, for the four groups, and for all countries combined. For example, the first row reports that the technical efficiency of Argentina with respect to the group frontier was estimated to vary between 0.904 and 1.000 during the five-year period, with an average value of 0.958 and a standard deviation of 0.035.

The DEA-K results reveal that the UK, for example, was about 97% efficient when measured relative to the European frontier. This means that average agricultural output in the UK is approximately 97% of the output that is possible using the same input levels and the production technology available in Europe. The DEA-MTR results reveal that the average metatechnology ratio for the UK is about 0.94. This means that the maximum output that could be produced using the inputs of the UK and the technology of Europe is about 94% of the maximum output that could be produced using the same inputs and the technology represented by the metafrontier.

[8] We restrict the number of outputs and inputs used in order to maintain degrees of freedom – for details, see Coelli *et al* (2005).

Observe from Table 2 that the technical efficiency of South Africa is quite high when measured with respect to the African frontier (0.964) but low when measured against the metafrontier (0.610). This difference implies a low metatechnology ratio. Indeed, the average DEA-MTR estimate for South Africa is 0.633, indicating that the maximum output that is feasible using the African technology (and the input levels used by South Africa) is only about 63% of the output that could be achieved using the technology represented by the metafrontier.

The results reported in Table 2 for different groups of countries suggest that, on average, African countries produce agricultural outputs under conditions that are more restrictive than in other regions of the world – the average metatechnology ratio (0.886) suggests that African countries could, at best, produce only 89% of the agricultural output that could be produced using the (unrestricted) metatechnology. The average metatechnology ratio for Asia (0.925) suggests that Asian countries could produce about 93% of the output that could be produced using the metatechnology (and the same inputs).

Finally, it is worth noting that many of the DEA-MF maximum values reported in Table 2 are equal to 1. For these countries, there must have been at least one year when they used an input-output combination that placed them at the point of tangency between their group frontiers and the metafrontier.[9] However, in the case of South Africa, for which we report a maximum DEA-K estimate of 1 and a maximum DEA-MTR estimate of 0.655, there must have been at least one year when it used an input-output combination that placed it on the African frontier, but well below the metafrontier.

4.3 SFA Results

SFA estimates of the frontier models defined by equations (19) and (21) were obtained by assuming a translog functional form. The region-*k* frontier is defined by

[9] More precisely, since the metafrontier and group frontiers are formed as the intersection of several hyperplanes, there was at least one year when each of these countries operated at a point where the hyperplane of their group frontiers touched a hyperplane of the metafrontier. The results for the four groups of countries suggest that there was at least one country in each of the groups that operated at such a point.

$$\ln y_{it} = \beta_0^k + \sum_{j=1}^{5} \beta_j^k \ln x_{jit} + 0.5 \sum_{j=1}^{5} \sum_{m=1}^{5} \beta_{jm}^k (\ln x_{jit})(\ln x_{mit}) + V_{it}^k - U_{it}^k \tag{30}$$

where $\beta_{jm}^k = \beta_{mj}^k$ for all *j* and *m*. For simplicity, we assumed $U_{it}^k = U_i^k$ where the U_i^ks are half-normal random variables. Zero observations in the data set were handled using the approach suggested by Battese (1997).[10] SFA estimates of the parameters of the group frontiers were obtained using FRONTIER 4.1c (see Coelli, 1996a), while the parameters of the metafrontier were estimated using the SHAZAM code that is presented in Appendix B.

SFA estimates of technical efficiencies and metatechnology ratios are summarised in Table 3. In this table, we use the acronyms SFA-K and SFA-MF to refer to technical efficiencies relative to the estimated group-*k* frontiers and the estimated metafrontier, respectively. We also use the acronym SFA-POOL to refer to technical efficiencies obtained using the pooled model of Battese and Rao (2002), discussed in Section 3.2. Estimating this pooled model allows us to formally test for differences between the group frontiers. Specifically, the generalised likelihood-ratio test statistic for the null hypothesis that the group frontiers are identical is LR = 187.6. This has a *p*-value of 0.000 (using a chi-square distribution with 69 degrees of freedom), so we reject the null hypothesis that the group frontiers are the same. This implies that the use of a metafrontier framework is appropriate.

The SFA efficiency estimates reported in Table 3 are generally lower than the DEA estimates reported in Table 2. For the ten selected countries listed in the tables, the SFA estimates for the group frontiers are lower than the DEA estimates for eight of the ten countries. Thus, DEA seems to fit a tighter frontier than does SFA, as it does in most empirical applications.[11] Countries in the Americas are still estimated to be the most technically efficient when assessed against both the group frontier and the

[10] Two of the 130 observations in Asia (Group 3) had zero values for the fertiliser input, so for these observations the logarithm of the fertiliser variable is undefined. The Battese (1997) solution is to specify a different production frontier for these observations. Thus, for observations with zero values for the fertiliser input, we use a dummy variable to both remove any terms involving $\ln x_{4i}$ and allow for a different intercept.

[11] The DEA results on efficiency are probably misleading for an agricultural application in which random errors are highly significant, even though the technical inefficiency effects seem to be dominant.

metafrontier. However, the metatechnology ratio for Africa is now estimated to be higher than that of any other group (i.e., Africa is now found to be the least restrictive production environment).

There are some important differences between the SFA and DEA estimates of technical efficiencies and metatechnology ratios for selected countries. For example, Indonesia is now estimated to be relatively inefficient with respect to the Asian frontier (its average efficiency score is 0.563 using SFA compared with 0.997 using DEA), and the Indonesian metatechnology ratio is now quite low (0.830 compared with 1.000). These results suggest that the SFA estimate of the Asian frontier is considerably higher than the DEA estimate in the region of the input-output combinations used by Indonesia.[12]

Several of the estimated metatechnology ratios summarised in Table 3 have maximum values of 1. This means that several countries could have used their inputs to place themselves at points of tangency between the group frontiers and the metafrontier. Countries in this position (from among the ten selected countries) were Argentina, Australia, China, the Netherlands and the USA.

Finally, it is worth noting that there are some significant differences between the SFA-MF efficiency estimates and the SFA-POOL efficiency estimates, the latter having being obtained using the stochastic metafrontier model of Battese and Rao (2002). In many cases (e.g., Australia, China and the Netherlands), the SFA-MF estimates (i.e., 0.921, 0.933 and 0.932) are more plausible than the SFA-POOL estimates (0.347, 0.624 and 0.351).

5. Extensions to the Basic Framework

Sections 3 and 4 describe and illustrate basic non-parametric and parametric methods for estimating single-output group frontiers and the metafrontier using panel data. The recent efficiency literature explores several extensions to conventional non-parametric

[12] By a similar line of reasoning, the two estimated frontiers are much closer to each other in the region of the input-output combinations used by, for example, Australia.

and parametric frontier models and estimators, and, in this section, we consider some of these extensions in the context of the metafrontier.

5.1 Technological Change

The group frontiers and the metafrontier depicted in Figure 1 are boundaries of restricted and unrestricted production possibilities sets. Technological change generally causes one or more of these sets to expand (i.e., leads to an increase in the number of feasible input-output combinations), implying an outward movement in one or more group frontiers and/or the metafrontier. To allow for this possibility, we must modify our metafrontier estimation strategy. In the case of DEA, we would first solve the LP problem of equation (17) using only observations from period 1 (this yields an estimate of the period-1 frontier), then solve the problem again using observations from periods 1 and 2 (this yields a period-2 frontier that envelops the period-1 frontier), and so on until we eventually solve the problem using all the observations in the data set (this yields a period-T frontier that envelops the frontiers from all earlier periods). In the case of SFA, we would simply include a time trend in the deterministic component of the model – details can be found in Coelli *et al.* (2005).

5.2 Time-Invariant Inefficiency Effects

The SFA model defined by equation (18) and the DEA problem given by equation (17) both allow for variations in technical inefficiency effects over time. If we assume $U_{it}^{k} = U_{i}^{k}$ then the SFA model yields time-invariant estimates of technical efficiencies with respect to the group frontiers (this simplifying assumption underpinned the SFA results in Section 4). However, estimates of metatechnology ratios and technical efficiencies with respect to the metafrontier are still time-varying.

Unfortunately, the DEA problem defined by equation (17) cannot be modified to yield meaningful time-invariant technical efficiency estimates. When time-invariant DEA estimates are required, it is a common practice to simply take averages of the time-varying estimates.

5.3 Multiple-Outputs

The multiple-output version of the DEA problem defined by (17) is:

$$
\begin{aligned}
&\max_{\phi_{it},\boldsymbol{\lambda}_{it}} && \phi_{it} \\
&\text{s.t.} && \phi_{it}\mathbf{y}_{it} - \mathbf{Y}\boldsymbol{\lambda}_{it} \le 0, \\
& && \mathbf{X}\boldsymbol{\lambda}_{it} - \mathbf{x}_{it} \le 0, \\
& && \mathbf{j}'\boldsymbol{\lambda}_{it} = 1 \\
&\text{and} && \boldsymbol{\lambda}_{it} \ge 0.
\end{aligned} \tag{31}
$$

where $\mathbf{y}_{it}$ is an M×1 vector of output quantities for the i-th firm in the t-th period; $\mathbf{Y}$ is the $M \times L_k T$ matrix of output quantities for all L_k firms in all T periods; and all other scalars, vectors and matrices are exactly as they are defined in Section 3.1. Importantly, $1/\phi_{it}$ is still an output-orientated technical efficiency measure with respect to the group frontier or the metafrontier (depending on whether the LP is solved using the group data or all the data in the sample), and the metatechnology ratio can still be estimated using equation (9).

The multiple-output analogue of the single-output stochastic production frontier is a stochastic distance function. For example, a stochastic translog version of the group-k output distance function is:

$$
\begin{aligned}
-\ln y_{Mit} = {} & \beta_{k0} + \sum_{j=1}^{N} \beta_{kj} \ln x_{jit} + 0.5 \sum_{j=1}^{N} \sum_{s=1}^{N} \beta_{kjs} (\ln x_{jit})(\ln x_{sit}) \\
& + \sum_{m=1}^{M-1} \alpha_{kj} \ln\left(y_{mit}/y_{Mit}\right) + 0.5 \sum_{m=1}^{M-1} \sum_{n=1}^{M-1} \alpha_{kmn} \ln\left(y_{mit}/y_{Mit}\right) \ln\left(y_{nit}/y_{Mit}\right) \\
& \sum_{j=1}^{N} \sum_{n=1}^{M-1} \gamma_{kmn} \ln x_{jit} \ln(y_{nit}/y_{Mit}) + V_{it}^{k} + U_{it}^{k}
\end{aligned} \tag{32}
$$

where $U_{it}^{k} = -\ln D^{k}(x_{1it}, x_{2it}, \ldots, x_{Nit}, y_{1it}, y_{2it}, \ldots, y_{Mit})$ captures inefficiency. This model can be estimated using standard SFA methods, although there is debate in the literature concerning endogeneity bias – for details, see Atkinson, Färe and Primont

(1998) and Coelli (2000). Following estimation of the group-k distance functions, an estimate of the metafrontier can be obtained by solving a generalisation of the optimisation problem defined by equation (23).

5.4 Alternative Orientations

The metafrontier approach discussed in this paper can accommodate alternative efficiency orientations. The input-orientated analogue of the VRS multiple-output DEA problem defined by equation (31) is given by:

$$\begin{aligned} \min_{\phi_{it},\boldsymbol{\lambda}_{it}} \quad & \phi_{it} \\ \text{s.t.} \quad & \mathbf{Y}\boldsymbol{\lambda}_{it} - \mathbf{y}_{it} \geq 0, \\ & \phi_{it}\mathbf{x}_{it} - \mathbf{X}\boldsymbol{\lambda}_{it} \geq 0, \\ & \mathbf{j}'\boldsymbol{\lambda}_{it} = 1 \\ \text{and} \quad & \boldsymbol{\lambda}_{it} \geq 0. \end{aligned} \tag{33}$$

Recall from Section 3.1 that the input-orientated and output-orientated problems yield identical estimates of technical efficiency under the assumption of constant returns to scale. To impose the CRS property on the estimated DEA frontier we simply remove the constraint that the $\boldsymbol{\lambda}_{it}$ s sum to one.

An input-orientated analogue of the group-k output distance function in equation (32) is also available for use in a metafrontier framework. Both types of distance functions are special cases of the directional distance function discussed by Chambers, Chung and Färe (1996).[13] Directional distance functions are particularly useful for measuring the efficiency of firms that produce both good and bad outputs. For example, Färe *et al.* (2005) use a quadratic directional distance function to measure the technical efficiencies of US electric utilities (producers of electricity and sulphur dioxide).

[13] The directional distance function is a variant of the shortage function of Luenberger (1992, 1995).

5.5 Identifying Groups

In most practical settings, firms can be grouped *a priori* on the basis of geographical, economic and/or political boundaries. In the absence of such boundaries, multivariate statistical techniques (e.g., cluster analysis) are available for determining numbers of groups and group membership.

When the analysis is conducted in an SFA framework, it is also possible to conduct statistical tests concerning the number of groups. For example, when the SFA model defined by equation (19) is employed, we can test whether groups *j* and *l* should be amalgamated into a single group by testing the joint null hypothesis $H_0 : \beta_m^j = \beta_m^l$ for all $m = 0,1,...,K'$. The LR test reported in Section 4.2 is a more general test of this type – it is a test of the null hypothesis that *all* (four) groups of firms (countries) can be amalgamated into a single group.

More generally, statistical methods for determining and/or confirming the number of groups (and group membership) range from estimating and testing the coefficients of dummy-variable models, to estimating and testing the parameters of mixtures models (e.g., Orea and Kumbhakar, 2004; O'Donnell and Griffiths, 2006).

6. Conclusion

This paper develops the concept of the metafrontier for the purpose of studying differences in efficiency across groups of firms. The metafrontier is defined as the boundary of an unrestricted technology set. Groups of firms that operate in resource-poor or highly-regulated production environments may only have access to a restricted technology set. We refer to the boundaries of these restricted sets as group frontiers.

In this paper, the closeness of group frontiers to the metafrontier are measured as metatechnology ratios for the different groups. These ratios are of considerable interest to managers and government policy-makers, not least because they measure the potential improvement in performance resulting from changes in the production environment. Governments can change the production environment by investing in physical, financial and human capital (e.g., by building ports and power stations, creating credit markets, or by investing in education and training); managers may have the capacity to change the production environment by moving their operations from one location to another.

The paper shows how group frontiers and the metafrontier can be estimated using DEA and SFA methodologies. Both approaches are popular in the efficiency literature. An empirical example using cross-country agricultural data provides evidence that agricultural producers in Europe operate in relatively restrictive production environments (using the SFA results).

Our empirical example illustrates the use of the DEA and SFA methodologies in a single-output context under various restrictive assumptions concerning the nature of the technology and firm inefficiency. However, the concepts and methods developed in the paper readily extend to multiple-output firms and models that are based on different assumptions concerning inefficiency and technological change.

In conclusion, we mention two opportunities for improving of our SFA estimator. First, observe that the linear program used to construct the SFA metafrontier (i.e., equation 23) gives equal weight to all firms, irrespective of how close they may be to their respective group frontiers. This is in contrast to the linear program used to construct the DEA metafrontier, which can be shown to give fully-efficient firms a weight of one and inefficient firms a weight of zero. This general idea can be implemented in a stochastic frontier context by multiplying the term in the square brackets in equation (23) by $T\hat{E}_{it}^{k}$. Second, econometric theory suggests that more efficient estimators of the parameters of both the group frontiers and the metafrontier can be obtained by estimating the group frontiers in a seemingly unrelated regression (SUR) framework. The size of the efficiency gains is an empirical question that seems worthy of investigation.

References

Atkinson SE, Färe R, Primont D (1998) Stochastic estimation of firm inefficiency using distance functions. Working Paper, Department of Economics, University of Georgia, Athens.

Battese GE (1997), A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. Journal of Agricultural Economics 48: 250-252.

Battese GE, Coelli TJ (1988), Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. Journal of Econometrics 38: 387-399.

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20: 325-332.

Battese GE, Rao DSP (2002) Technology potential, efficiency and a stochastic metafrontier function. International Journal of Business and Economics 1(2): 1-7.

Battese GE, Rao DSP, O'Donnell CJ (2004) A metafrontier production function for estimation of technical efficiencies and technology potentials for firms operating under different technologies. Journal of Productivity Analysis 21: 91-103.

Chambers RG, Chung Y and Färe R (1996) Benefit and distance functions. Journal of Economic Theory 70: 407-419.

Coelli TJ (1996a) A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimation. CEPA Working Papers, No. 7/96, Department of Econometrics, University of New England, Armidale.

Coelli TJ (1996b) A guide to DEAP version 2.1: A data envelopment analysis (computer) program. CEPA Working Papers No. 8/96, Department of Econometrics, University of New England, Armidale.

Coelli TJ (2000) On the econometric estimation of the distance function representation of a production technology. Discussion Paper 2000/42, Center for Operations Research and Econometrics, Universite Catholique de Louvain, Louvain-la-Neuve.

Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) Introduction to Efficiency and Productivity Analysis, 2nd Ed., Springer, Heidelberg.

Färe R, Primont D (1995) Multi-output Production and Duality: Theory and Applications. Kluwer Academic Publishers, Boston.

Färe R, Grosskopf S, Noh D-W, Weber W (2005) Characteristics of a polluting technology: theory and practice. Journal of Econometrics 126(2): 469-492.

Glass JC, McKillop DG, Hyndman N (1995) Efficiency in the provision of university teaching and research: An empirical analysis of UK universities. Journal of Applied Econometrics 10(1): 61-72.

Hayami Y, Ruttan VW (1970) Agricultural productivity differences among countries. American Economic Review 40: 895-911.

Hayami Y, Ruttan VW (1971) *Agricultural development: An international perspective.* Johns Hopkins University Press, Baltimore.

Luenberger DG (1992) Benefit functions and duality. Journal of Mathematical Economics 21(5): 461-481.

Luenberger D (1995) Microeconomic Theory. McGraw-Hill, Boston.

Lovell CAK, Pastor JT (1997) Target setting: An application to a bank branch network. European Journal of Operational Research 98: 290-299.

McMillan ML, Chan WH (2004) University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. University of Alberta Economics Working Paper Series, No. 2005-04.

O'Donnell CJ,Griffiths WE (2006) Estimating state-contingent production frontiers. American Journal of Agricultural Economics 88(1): 249-266.

O'Donnell CJ, van der Westhuizen G (2002) Regional comparisons of banking performance in South Africa. South African Journal of Economics 70(3): 485-518.

Orea L, Kumbhakar SC (2004) Efficiency measurement using a latent class stochastic frontier model, Empirical Economics 29: 169-183.

Rao DSP (1993) Inter-country comparisons of agricultural output and productivity. Social and Economic Development Paper, No. 112, Food and Agriculture Organization of the United Nations, Rome.

Wang H-J (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. Journal of Productivity Analysis 18: 241-253.

Worthington AC, Lee BL (2005) Efficiency, technology and productivity change in Australian universities, 1998-2003. University of Wollongong School of Accounting and Finance Working Paper Series, No. 05/01.

Appendix A – Variable Descriptions for Empirical Application

Output (y_{it}): This variable is an aggregate of 185 agricultural commodity outputs, net of feed and seed. The output aggregate we use corresponds to the sum of the crop and livestock aggregates reported in Table 4 of Rao (1993). Those aggregates were constructed using international average prices (expressed in US dollars), derived using the Geary-Khamis method.

Land (x_{1it}): This variable is millions of hectares of arable land, land under permanent crops and land under permanent pasture. Arable land includes land under temporary crops, temporary meadows, land under market or kitchen gardens, and land temporarily fallowed (for less than five years). Land under permanent crops is the land cultivated with crops that need not be replanted after each harvest. This category includes land under flowering shrubs, fruit trees, nut trees and vines but excludes land under trees grown for wood or timber.

Machinery (x_{2it}): This input is measured as the total number of wheeled and crawler tractors used in agriculture, excluding garden tractors..

Labour (x_{3it}): This variable measures the economically-active population in agriculture, defined as the number of persons engaged in or seeking employment in the operation of a family farm or business, whether as employers, own-account workers, salaried employees or unpaid workers.

Fertiliser (x_{4it}): Following other studies (e.g. Hayami and Ruttan, 1970) on inter-country comparisons of agricultural productivity, we measure this input using thousands of tonnes of nitrogen (N), potassium (P_2O_2) and phosphate (K_2O) contained in the fertilisers that were applied.

Livestock (x_{5it}): This input is the number of buffaloes, cattle, pigs, sheep and goats, measured in sheep equivalents. The following conversion factors were used: 8.0 for buffalo and cattle; and 1.0 for sheep, goats and pigs.[14]

[14] These conversion figures are similar to those used in Hayami and Ruttan (1970).

Appendix B – SHAZAM Code

```
* The file parm.txt contains estimated parameters of group frontiers (by column)
* The file sfa#.txt contains n# data observations for group #
* Sections 1 and 3 are problem-specific.

* 1. SET NUMBERS OF PARAMETERS ETC.
gen1 nparms = 22
gen1 ngroups = 4
gen1 n1 = 135
gen1 n2 = 105
gen1 n3 = 130
gen1 n4 = 115

* 2. READ THE ESTIMATED PARAMETERS OF THE GROUP FRONTIERS
smpl 1 nparms
read (parm.txt) parm / rows = nparms cols = ngroups
do # = 1,ngroups
  dim b# nparms
  copy parm b# / fcols=#;# tcols = 1;1
endo

* 3. READ THE DATA AND CONSTRUCT DATA MATRICES AND VECTORS
gen1 j2 = n1+1
gen1 j3 = n1+n2+1
gen1 j4 = n1+n2+n3+1
gen1 k2 = n1+n2
gen1 k3 = n1+n2+n3
gen1 n = n1+n2+n3+n4
smpl 1 n
genr one = 1
genr dummy = 0
read (sfa1.txt) group t ly lx1-lx5 lx11-lx15 lx22-lx25 lx33-lx35 lx44-lx45 lx55
smpl j2 k2
read (sfa2.txt) group t ly lx1-lx5 lx11-lx15 lx22-lx25 lx33-lx35 lx44-lx45 lx55
smpl j3 k3
read (sfa3.txt) group t ly dummy lx1-lx5 lx11-lx15 lx22-lx25 lx33-lx35 lx44-lx45 lx55
smpl j4 n
read (sfa4.txt) group t ly lx1-lx5 lx11-lx15 lx22-lx25 lx33-lx35 lx44-lx45 lx55
smpl 1 n
matrix x =
one|dummy|lx1|lx2|lx3|lx4|lx5|lx11|lx12|lx13|lx14|lx15|lx22|lx23|lx24|lx25|lx33| &
     lx34|lx35|lx44|lx45|lx55
dim x1 n1 nparms x2 n2 nparms x3 n3 nparms x4 n4 nparms
copy x x1 / frows=1;n1 trows=1;n1
copy x x2 / frows=j2;k2 trows=1;n2
copy x x3 / frows=j3;k3 trows=1;n3
copy x x4 / frows=j4;n trows=1;n4
do # = 1,ngroups
  matrix yhat# = x#*b#
endo
matrix b = -(yhat1'|yhat2'|yhat3'|yhat4')'

* 4. OBTAIN AND PRINT PARAMETERS OF THE METAFRONTIER
stat x / means = xbar
matrix c = ((-xbar')|xbar')'
matrix A = (-x)|x
?lp c A b / iter = 5000 primal = bstar
dim beta1 nparms beta2 nparms
gen1 p1 = nparms+1
gen1 p2 = nparms*2
copy bstar beta1 / frows=1;nparms trows=1;nparms
copy bstar beta2 / frows=p1;p2 trows=1;nparms
matrix beta = beta1-beta2
print beta

* 5. OBTAIN AND PRINT METATECHNOLOGY RATIOS
do # = 1,ngroups
  matrix xbeta# = x#*beta
  matrix mtr# = exp(yhat#)/exp(xbeta#)
  stat mtr#
  print mtr#
endo
stop
```

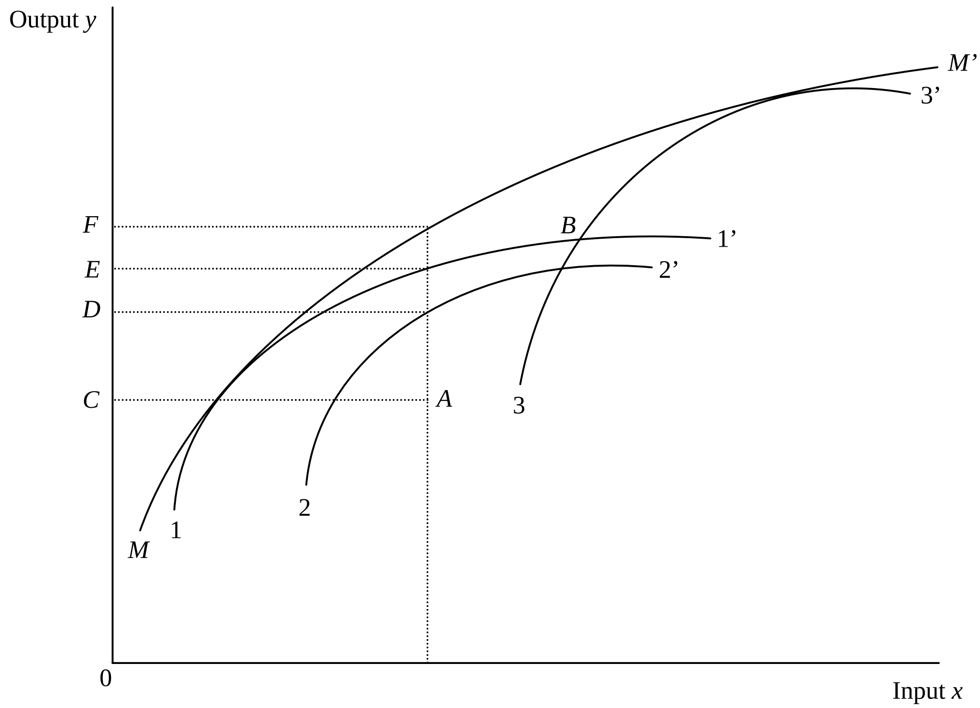

Figure 1: Technical Efficiencies and Metatechnology Ratios

Table 1. Countries and Regions for the Empirical Application

Country Code	Region Code [a]	Country	Country Code	Region Code [a]	Country
1	1	ALGERIA	51	3	SRI LANKA
2	1	ANGOLA	52	3	CHINA
3	1	BURUNDI	53	3	INDIA
4	1	CAMEROON	54	3	INDONESIA
5	1	CHAD	55	3	IRAN
6	1	EGYPT	56	3	IRAQ
7	1	ETHIOPIA PDR	57	3	ISRAEL
8	1	GHANA	58	3	JAPAN
9	1	GUINEA	59	3	CAMBODIA
10	1	COTE DIVOIRE	60	3	KOREA REP
11	1	KENYA	61	3	LAOS
12	1	MADAGASCAR	62	3	MALAYSIA
13	1	MALAWI	63	3	MONGOLIA
14	1	MALI	64	3	NEPAL
15	1	MOROCCO	65	3	PAKISTAN
16	1	MOZAMBIQUE	66	3	PHILIPPINES
17	1	NIGER	67	3	SAUDI ARABIA
18	1	NIGERIA	68	3	SYRIA
19	1	RWANDA	69	3	THAILAND
20	1	SENEGAL	70	3	TURKEY
21	1	SOUTH AFRICA	71	3	VIET NAM
22	1	SUDAN	72	4	AUSTRIA
23	1	TANZANIA	73	4	BEL-LUX
24	1	TUNISIA	74	4	BULGARIA
25	1	UGANDA	75	4	CZECHOSLOVAK
26	1	BURKINA FASO	76	4	DENMARK
27	1	ZIMBABWE	77	4	FINLAND
28	2	CANADA	78	4	FRANCE
29	2	COSTA RICA	79	4	GERMANY
30	2	CUBA	80	4	GREECE
31	2	DOMINICAN RP	81	4	HUNGARY
32	2	EL SALVADOR	82	4	IRELAND
33	2	GUATEMALA	83	4	ITALY
34	2	HAITI	84	4	NETHERLANDS
35	2	HONDURAS	85	4	NORWAY
36	2	MEXICO	86	4	POLAND
37	2	NICARAGUA	87	4	PORTUGAL
38	2	USA	88	4	ROMANIA
39	2	ARGENTINA	89	4	SPAIN
40	2	BOLIVIA	90	4	SWEDEN
41	2	BRAZIL	91	4	SWITZERLAND
42	2	CHILE	92	4	UK
43	2	COLOMBIA	93	4	YUGOSLAV SFR
44	2	ECUADOR	94	3	AUSTRALIA
45	2	PARAGUAY	95	3	NEW ZEALAND
46	2	PERU	96	3	PAPUA N GUIN
47	2	URUGUAY	97	4	USSR
48	2	VENEZUELA			
49	3	BANGLADESH			
50	3	MYANMAR			

[a] Region codes are: 1 = Africa; 2 = The Americas; 3 = Asia; 4 = Europe

Table 2. DEA Estimates of Technical Efficiencies and Metatechnology Ratios

	County or Group	Mean	St. Deviation	Minimum	Maximum
Technical Efficiency With Respect to the Group Frontiers (DEA-K)	Argentina	0.958	0.035	0.904	1.000
	Australia	0.987	0.020	0.955	1.000
	Brazil	0.962	0.043	0.888	1.000
	China	0.983	0.017	0.961	1.000
	India	0.996	0.009	0.979	1.000
	Indonesia	0.997	0.006	0.987	1.000
	Netherlands	0.993	0.010	0.977	1.000
	South Africa	0.964	0.040	0.899	1.000
	UK	0.967	0.015	0.948	0.990
	USA	0.970	0.038	0.908	1.000
	1 – Africa	0.788	0.217	0.247	1.000
	2 – The Americas	0.946	0.074	0.739	1.000
	3 – Asia	0.900	0.138	0.527	1.000
	4 – Europe	0.887	0.135	0.555	1.000
	All Countries	0.876	0.164	0.247	1.000
Metatechnology Ratio (DEA-MTR)	Argentina	0.998	0.002	0.996	1.000
	Australia	1.000	0.000	0.999	1.000
	Brazil	0.993	0.008	0.982	1.000
	China	0.998	0.002	0.995	1.000
	India	1.000	0.000	1.000	1.000
	Indonesia	1.000	0.000	1.000	1.000
	Netherlands	1.000	0.001	0.998	1.000
	South Africa	0.633	0.016	0.615	0.655
	UK	0.940	0.001	0.938	0.941
	USA	1.000	0.000	1.000	1.000
	1 – Africa	0.886	0.143	0.436	1.000
	2 – The Americas	0.907	0.147	0.498	1.000
	3 – Asia	0.925	0.100	0.633	1.000
	4 – Europe	0.892	0.143	0.425	1.000
	All Countries	0.903	0.134	0.425	1.000
Technical Efficiency With Respect to the Metafrontier (DEA-MF)	Argentina	0.957	0.035	0.904	1.000
	Australia	0.987	0.020	0.954	1.000
	Brazil	0.955	0.050	0.872	1.000
	China	0.981	0.018	0.960	1.000
	India	0.996	0.009	0.979	1.000
	Indonesia	0.997	0.006	0.987	1.000
	Netherlands	0.993	0.011	0.975	1.000
	South Africa	0.610	0.032	0.565	0.655
	UK	0.909	0.014	0.891	0.929
	USA	0.970	0.038	0.908	1.000
	1 – Africa	0.703	0.237	0.220	1.000
	2 – The Americas	0.861	0.171	0.474	1.000
	3 – Asia	0.839	0.185	0.435	1.000
	4 – Europe	0.793	0.186	0.425	1.000
	All Countries	0.795	0.207	0.220	1.000

Table 3. SFA Estimates of Technical Efficiencies and Metatechnology Ratios

	Country	Mean	St. Deviation	Minimum	Maximum
Technical Efficiency With Respect to the Group Frontiers (SFA-K)	Argentina	0.959	(a)	(a)	(a)
	Australia	0.950			
	Brazil	0.895			
	China	0.936			
	India	0.944			
	Indonesia	0.563			
	Netherlands	0.972			
	South Africa	0.935			
	UK	0.968			
	USA	0.917			
	1 – Africa	0.505	0.249	0.190	0.972
	2 – The Americas	0.824	0.137	0.519	0.981
	3 – Asia	0.719	0.195	0.362	0.981
	4 – Europe	0.823	0.151	0.514	0.982
	All Countries	0.707	0.233	0.190	0.982
Metatechnology Ratio (SFA-MTR)	Argentina	0.982	0.011	0.969	1.000
	Australia	0.969	0.034	0.924	1.000
	Brazil	0.799	0.020	0.772	0.823
	China	0.997	0.004	0.991	1.000
	India	0.740	0.041	0.696	0.788
	Indonesia	0.830	0.039	0.777	0.869
	Netherlands	0.959	0.030	0.918	1.000
	South Africa	0.607	0.006	0.601	0.615
	UK	0.559	0.007	0.552	0.568
	USA	0.987	0.007	0.982	1.000
	1 – Africa	0.752	0.206	0.308	1.000
	2 – The Americas	0.751	0.161	0.435	1.000
	3 – Asia	0.738	0.197	0.328	1.000
	4 – Europe	0.664	0.210	0.250	1.000
	All Countries	0.727	0.198	0.250	1.000
Technical Efficiency With Respect to the Metafrontier (SFA-MF)	Argentina	0.942	0.011	0.929	0.959
	Australia	0.921	0.032	0.878	0.950
	Brazil	0.715	0.018	0.691	0.736
	China	0.933	0.004	0.928	0.936
	India	0.698	0.038	0.657	0.744
	Indonesia	0.468	0.022	0.438	0.490
	Netherlands	0.932	0.029	0.892	0.972
	South Africa	0.568	0.005	0.562	0.575
	UK	0.541	0.007	0.535	0.550
	USA	0.906	0.006	0.901	0.917
	1 – Africa	0.362	0.185	0.122	0.972
	2 – The Americas	0.615	0.157	0.381	0.959
	3 – Asia	0.537	0.212	0.119	0.950
	4 – Europe	0.541	0.194	0.213	0.975
	All Countries	0.506	0.211	0.119	0.975
Technical Efficiency With Respect to the Pooled Frontier (SFA-POOL)	Argentina	0.952	(a)	(a)	(a)
	Australia	0.347			
	Brazil	0.648			
	China	0.624			
	India	0.400			
	Indonesia	0.716			
	Netherlands	0.351			
	South Africa	0.537			
	UK	0.583			
	USA	0.960			
	1 – Africa	0.418	0.191	0.169	0.975
	2 – The Americas	0.567	0.172	0.332	0.960
	3 – Asia	0.520	0.213	0.195	0.983
	4 – Europe	0.577	0.202	0.262	0.971
	All Countries	0.515	0.206	0.169	0.983

(a) There is little point reporting these blocks of numbers – the inefficiency effects in equation (30) are time-invariant, so the standard deviations are zero, and the minimums and maximums are equal to the means.