The Estimation of Meta-Frontiers by Constrained Maximum Likelihood

Alexandre Repkine

The Estimation of Meta-Frontiers by Constrained Maximum Likelihood

Alexandre Repkine[1]

Abstract

Existing approaches to the meta-frontier estimation are largely based on the linear programming technique, which does not hinge on any statistical underpinnings. We suggest estimating meta-frontiers by constrained maximum likelihood subject to the constraints that specify the way in which the estimated meta-frontier overarches the individual group frontiers. We present a methodology that allows one to either estimate meta-frontiers using the conventional set of constraints that guarantees overarching at the observed combinations of production inputs, or to specify a *range* of inputs within which such overarching will hold. In either case the estimated meta-frontier coefficients allow for the statistical inference that is not straightforward in case of the linear programming estimation. We apply our methodology to the world's FAO agricultural data and find similar estimates of the meta-frontier parameters in case of the same set of constraints. On the contrary, the parameter estimates differ a lot between different sets of constraints.

[1] Department of Economics, Korea University, Seoul, Republic of Korea

1. Introduction

This paper is suggesting a methodology for estimating meta-frontiers within a framework of constrained maximum likelihood. Such estimation is a useful means of decomposing the level of productive inefficiency of a firm into two components: the one explained by the inefficient management practices, and the other one explained by the characteristics of the production environment in which firms operate. While the former type of productive inefficiency can be captured in the framework of technical efficiency estimation as in Aigner & al. (1977), the latter inefficiency type requires an estimation of the overarching envelope of a group of stochastic frontiers. The key assumption in this framework is that different groups of firms operate under statistically different stochastic frontiers that in turn are lying below the common envelope called meta-frontier.

The concept of meta-frontier has its roots in the concept of a meta-production function that is dating back to an early paper by Hayami & Ruttan (1971), where it is regarded as an envelope of the neoclassical production functions estimated for the individual groups of observations. Based on the concept of a stochastic frontier outlined in Aigner et al. (1977), O'Donnell & al. (2008) define *group frontiers* to be the boundaries of restricted technology sets where restrictions derive from the lack of economic infrastructure, institutions and general development of the economic environment. In contrast, the *meta-frontier* is defined as a boundary of the unrestricted technology set that the perfectly efficient firms in their respective groups would have attained, should the economic environment in which they operate have been the global best-practice one. Within the meta-frontier framework the comparison of firms' technical efficiency levels is making more sense since a clear distinction is being made between the inability to adopt the best operating practice *in one's group of*

firms (technical efficiency) and the lack of favorable operating environment *of one's group of firms* relative to the best environment available in the world (technological gap). We provide an overview of the meta-frontier framework in Section 2.

Fried & al. (2008) underscores the importance of making a distinction between differences in the private management practices, and differences in the operating environment. Estimating technical efficiency levels against a pooled sample technological frontier will attribute all inefficiency to the shortcomings of managerial practices, while in fact the inefficiency can be caused by a deficient technological environment. Thus, failing to recognize the fact that different groups of firms may be facing different operating environments will in general result in one's inability to correctly compare the estimated efficiency levels (e.g. by means of FRONTIER by Coelli (1992)) among firms belonging to two different groups. An interesting empirical study by Zeitsch & al. (1994) likens the comparison of technical efficiency levels for firms operating under different technological frontiers to the (in)famous comparison between apples and oranges. Another reason why proper attribution of differences in economic performance to managerial practices or operating environment is important is that such attribution allows one to design incentive-based economic policies depending on the relative importance of either type of the productive inefficiency. O'Donnell & al. (2008) present a general analytical framework in which meta-frontiers are defined and estimated using both parametric and non-parametric methods.

Battese & Rao (2002) offered an operational methodology for estimating a meta-frontier in the stochastic frontier framework based on the estimates of the individual stochastic group frontiers. However, their methodology provided no guarantee that the estimated meta-frontier will stay above the estimated group frontiers. Battese & al. (2004) have presented a practical

way to solve this problem by suggesting a non-parametric method of meta-frontier estimation. The framework they offer is a two-step approach. In the first step, one identifies the groups of firms (or other decision-making units) in the sample in order to estimate technical efficiency scores relative to these group frontiers. In the second step, based on the estimated parameters of the deterministic group frontiers, one looks for an envelope frontier that approaches the estimated group frontiers most closely without falling below any one of them at the observed combinations of input factors. The estimation procedure in the second step is based on the minimization of either the sum of the distances between a meta-frontier and the group frontiers, or the sum of those distances' squares. In the latter case the estimation can be done by constrained OLS. Statistical inference about the estimated meta-frontier parameters in the non-parametric cases is usually made on the basis of bootstrapping or simulation. We will further refer to the linear-programming based methodologies of meta-frontier estimations as LP methodologies.

In this study we attempt to extend the stochastic frontier framework for meta-frontier estimation in two ways. First, we suggest estimating meta-frontiers by constrained maximum likelihood (called CML from now on). Being a parametric one, the CML approach allows for a statistical inference on the estimated meta-frontier parameters. Such inference cannot be produced naturally in case of LP estimation, since the latter does not hinge on any statistical assumptions. The proposed CML methodology is essentially solving the same problem of minimizing the distance between the meta-frontier and the estimated group frontiers as the one handled by the LP methodology, at the same time providing for a possibility of statistical inference making.

Second, we extend the set of constraints that can be imposed on the estimated meta-frontier

so that it may overarch group frontiers either at the observed input combinations, or within an arbitrary range of those inputs, or both. Although the LP methodology results in a meta-frontier that never falls below any one of the group frontiers at any of the observed combinations of inputs, this meta-frontier *may* fall below one or several group frontiers at an input combination *within a cuboid in the space of the production inputs.* The overarching property may be violated e.g. in case of estimating the flexible functional forms such as translog that may not be monotonic in a connected set of input factors that includes the observed combinations of inputs (Henningsen and Henning, 2009). However, even the imposition of monotonicity requirement in a specific domain of input factors does not preclude the situation where one or several group frontiers overarch the estimated meta-frontier for some unobserved combinations of inputs belonging to the observed range. In case one is interested in ensuring the overarching property of the meta-frontier within a specific range of production inputs, requiring the overarching to only occur at the observed input combinations is not sufficient.

We thus believe it is a reasonable extension to the methodology of meta-frontier estimation to let the latter overarch group frontiers not only at the observed combinations of inputs, but also within a particular cuboid containing those inputs. For example, in case one wanted to extrapolate the meta-frontier estimation results for a projected production plant in an industry, it would be natural to require that the predicted (and unobserved) best-practice output for this hypothetical plant given by an output level on the corresponding group frontier may not exceed the global best practice given by the estimated meta-frontier. We show that, in order for a meta-frontier to overarch the group frontiers everywhere within a specific domain of inputs, it is sufficient to require that it may only overarch the group frontiers in every vertex of a cuboid containing this domain.

We apply the CML methodology to the Food and Agriculture Organization (faostat.fao.org) database on 101 countries in order to compare the meta-frontier estimation results with the ones obtained by conventional linear programming methodology. We find that in case the meta-frontier is required to overarch group frontiers at the observed combinations of inputs, the two methodologies yield almost identical results. However, it is not clear at this stage whether this is an empirical coincidence or a theoretical regularity. In case the meta-frontier were required to overarch group frontiers within a cuboid containing all of the observed input combinations, both LP and CML resulted in a significant outward shift of the meta-frontier, proportionately driving down the estimated meta-technology ratios and meta-efficiency levels. If a researcher thinks it is enough for her purposes that the meta-frontier may overarch group frontiers only at the observed input combinations, the CML estimation easily accommodates this requirement as well (i.e. the "conventional" set of constraints may be used with the CML estimation as well).

This paper is organized as follows. In Section 2 we briefly present the meta-frontier estimation methodology. Section 3 presents an operational method of constrained maximum likelihood (CML) so that the resulting meta-frontier overarches the group frontiers either at the observed combinations of the production inputs, or within an arbitrary cuboid containing those combinations. Section 4 presents and discusses empirical specifications and results. Section 5 concludes.

2. The Meta-Frontier Production Function Framework

As mentioned in the Introduction, the key assumption behind the meta-frontier framework is that the production units in different groups operate under (statistically) different stochastic production frontiers that in turn fall short of an envelope called stochastic meta-frontier. Battese et al. (2004) define their stochastic frontier model for N_k units in group $k = 1..K$ as:

$$y_{it}^k = f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^k\right) e^{v_{it}^k - u_{it}^k} \tag{2.1}$$

where y_{it} is the output of production unit $i = 1..N$ belonging to group k at time period $t = 1..T$, assuming M production factors in the stochastic production function in each group, x_{jit}^k being the $j-th$ input of production unit i in group k at time period t. β^k is a vector of group k's frontier parameters. The composite error term $v_{it}^k - u_{it}^k$ consists of the symmetric normally distributed term $v_{it}^k \sim i.i.d.N\left(0, \sigma_{vk}^2\right)$ and the one-sided non-negative term u_{it}^k associated with technical inefficiency of production unit i, that is $u_{it}^k \sim \left|N\left(0, \sigma_{uk}^2\right)\right|$.

Stochastic frontier is then given by the value of output that would be obtained should there be no inefficiency in the production units' behavior whatsoever, i.e. if $u_{it}^k = 0, \forall i, t, k$. The benchmark level of output on the stochastic frontier would be then given by $f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^k\right) e^{v_{it}^k}$, and the level of technical efficiency for the production unit i in group k in time period t can be written as follows:

$$TE_{it}^k = \frac{f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^k\right) e^{v_{it}^k - u_{it}^k}}{f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^k\right) e^{v_{it}^k}} = e^{-u_{it}^k} \tag{2.2}$$

The stochastic frontiers in (2.1) for the separate groups $k = 1..K$ can be estimated by several software packages, e.g. FRONTIER 4.1 (Coelli, 1992), STATA, or LIMDEP. It is then necessary to test whether the K groups are characterized by statistically different production technologies by applying a likelihood ratio test that compares the value of the log-likelihood function estimated for the pooled sample to the sum of the values of the individual log-likelihooods from the estimation of all group stochastic frontiers (Moreira and Bravo-Ureta, 2010).

The meta-frontier model as defined by Battese et al. (2004) is a deterministic parametric function of the observed production inputs such that its value at every observed combination of inputs is greater than or equal to the predicted values of the deterministic group frontiers for all observations. Using the asterisk to refer to the meta-frontier, the latter then will be defined as follows:

$$
\begin{cases}
y_{it}^* = f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^*\right) \\
f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^*\right) \geq f\left(x_{1it}^k, x_{2it}^k, ..., x_{Mit}^k; \beta^k\right) \\
\forall i, k, t
\end{cases}
\tag{2.3}
$$

In case the production function in (2.1) is log-linear, (2.3) becomes:

$$
\begin{cases}
\ln\left(y_{it}^*\right) = \left(\ln x_{it}^k\right)' \beta^* \\
\left(\ln x_{it}^k\right)' \beta^* \geq \left(\ln x_{it}^k\right)' \beta^k \\
\forall i, t, k
\end{cases}
\tag{2.4}
$$

where $\left(\ln x_{it}^k\right)'$ refers to the row-vector of the logs of the observed production inputs for

production unit i in group k at time period t.

To estimate meta-frontiers in (2.4), Battese et al. (2004) offer either minimizing a sum of squared deviations of the distance between the meta-frontier and the value of the group stochastic frontiers at the observed input vectors, or minimizing the sum of the absolute values of such deviations. The authors report similar results for the two approaches. In the empirical part of this study we opt for the minimization of absolute deviations when estimating the meta-frontiers since the latter procedure is more easily implemented compared to its quadratic counterpart, and there has been little evidence so far why minimizing the sum of absolute or squared deviations might produce superior results.

Assuming the log-linear form of the production function and denoting by $\hat{\beta}^k$ the vector of the estimated parameters of the k-th group frontier, it has become customary to solve the following linear programming problem to estimate the parameters of the latter:

$$\begin{cases} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \left(\ln x_{it}^k \right)' \beta^* - \left(\ln x_{it}^k \right)' \hat{\beta}^k \right| \\ s.t. \\ \left(\ln x_{it}^k \right)' \beta^* \geq \left(\ln x_{it}^k \right)' \hat{\beta}^k, \forall i, t, k \end{cases} \tag{2.5}$$

In section 4 we are going to offer an alternative to (2.5) that is based on the maximization of a constrained likelihood function.

After the parameter vector β^* has been estimated from (2.5), the meta-technology ratio can be computed according to O'Donnell et al. (2008) as a ratio for each observation of the output on the corresponding group frontier to the output level on the meta-frontier:

9

$$MTR_{it}^{k} = \frac{f\left(x_{1it}^{k}, x_{2it}^{k}, ..., x_{Mit}^{k}; \hat{\beta}^{k}\right)}{f\left(x_{1it}^{k}, x_{2it}^{k}, ..., x_{Mit}^{k}; \hat{\beta}^{*}\right)} \tag{2.6}$$

The technical efficiency TE_{it}^{*} of each production unit with respect to the meta-frontier will then be equal to the product of the technical efficiency of a production unit with respect to its group frontier TE_{it}^{k} and the meta-technology ratio MTR_{it}^{k}:

$$TE_{it}^{*} = TE_{it}^{k} \times MTR_{it}^{k} \tag{2.7}$$

The increasing values of the meta-technology ratio in this context would mean a decrease in the distance between the group frontier and the meta-frontier.

3. Constraining the Maximum Likelihood

In this section we are deriving the set of constraints under which the estimated meta-frontier overarches the group frontiers within an arbitrarily specified cuboid containing the actually observed inputs. In the next subsection we derive a set of 2^S constraints for S terms in the log-linear specification of the production function that correspond to such a requirement and prove that applying constrained maximum likelihood estimation to the pooled sample will indeed result in a meta-frontier overarching the group frontiers within the observed range of inputs. We then discuss the operational method of meta-frontier estimation.

3.1 Deriving a set of constraints for the likelihood function maximization

Let $k^i_{\min}$ and $k^i_{\max}$ be the empirically observed logarithms of the minimum and maximum levels of S variables $k^i, i = 1..S$, respectively, that enter the log-linear specification of the production function. For example, variables $k^i, i = 1..S$ can be both the logarithms of e.g. capital and labor as well as their squares or interaction terms. This generalization is needed to accommodate flexible functional forms such as translog. Clearly, once the minimum and the maximum of the original input factors are established, the minima and maxima of the squared, as well as the interaction, terms are automatically established. In this way the ensuing discussion refers to any log-linear functional forms, not just the Cobb-Douglas one. Without loss of generality, we will further refer to k_i-s as "production factors" or "input factors".

Let y^m be the logarithm of the output level on the meta-frontier and $y^g, g = 1..G$ be the logarithm of the level of output on one of the G group frontiers. The set of all vertices in the set of the observed input factors $\aleph = \{k_{min}^1, k_{max}^1\} \times \{k_{min}^2, k_{max}^2\} \times ... \times \{k_{min}^S, k_{max}^S\}$ is a Cartesian product of S sets, each one of which consists of two elements: the minimum and the maximum observed level of the input factor $i = 1..S$. The number of elements in $\aleph$ is equal to 2^S. Let us show that, if a meta-frontier overarches each one of the G group frontiers at each point in $\aleph$, this meta-frontier will overarch each group frontier at any interior point of the observed input range $k^i \in [k_{min}^i, k_{max}^i], i = 1..S$.

Consider a log-linear meta-frontier $y^m = a_0 + \sum_{i=1}^{S} a_i k^i$ and G log-linear individual group frontiers $y^g = b_0^g + \sum_{j=1}^{S} b_j^g k^j$, $g = 1..G$. Since the meta-frontier by definition overarches each individual group frontier in each point of the observed input set, it overarches each group frontier in each vertex of $\aleph$, as well. Consider $2^S \times G$ conditions of the form:

$$y^m = a_0 + \sum_{i=1}^{S} a_i k^i \geq b_0^g + \sum_{i=1}^{S} b_i^g k^i \quad , \tag{3.1}$$

where $k^i \in \aleph$.

We claim the following:

Proposition 1

If (3.1) is satisfied, $y^m = a_0 + \sum_{i=1}^{S} a_i k^i$ overarches each group frontier *for all levels of all*

input factors within an arbitrary cuboid in the space of the production factors, that is:

$$y^m = a_0 + \sum_{i=1}^{S} a_i k^i \geq b_0^g + \sum_{i=1}^{S} b_i^g k^i \; , k^i \in [k_{min}^i, k_{max}^i], i = 1..S \; . \tag{3.2}$$

Proof

We prove our claim by contradiction. Let us assume that there exists a group frontier

$g^*, 1 \leq g^* \leq G$, and a point $(k_0^1, k_0^2, ..., k_0^S)$ within the empirically observed range of the input

factors $k_0^i \in (k_{min}^i, k_{max}^i) \forall i = 1..S$, such that (3.2) does not hold, namely:

$$y^m = a_0 + \sum_{i=1}^{S} a_i k_0^i < b_0^{g^*} + \sum_{i=1}^{S} b_i^{g^*} k_0^i \tag{3.3}$$

In other words, we are assuming that there exists an interior point in the empirically

observed set of input factors for which the corresponding output on the group frontier g^* is

strictly greater than the meta-frontier output at that interior point.

Consider the following system of inequalities:

$$
\begin{cases}
y^m = a_0 + \displaystyle\sum_{i=1}^{S} a_i k^i \geq b_0^{g^*} + \sum_{i=1}^{S} b_i^{g^*} k^i, \forall k^i \in \aleph \\[4mm]
y^m = a_0 + \displaystyle\sum_{i=1}^{S} a_i k_0^i < b_0^{g^*} + \sum_{i=1}^{S} b_i^{g^*} k_0^i
\end{cases}
\tag{3.4}
$$

The first part of (3.4) consists of 2^S inequalities and it is a subsystem of (3.1). It requires that the meta-frontier may not fall below the group frontier g^* in the vertices of the empirically observed input set (these vertices are the elements of $\aleph$.) The last inequality in (3.4) is reflecting our assumption that at least in one interior point in the observed set of inputs the meta-frontier does fall below the group frontier g^*. In what follows, we are going to demonstrate that (3.4) is inconsistent.

For convenience, set $b_i = b_i^{g^*}$. Multiplying the last inequality in (3.4) by -1 and summing it up with each one of the first 2^S inequalities in (3.4), we obtain the following:

$$
\begin{cases}
\displaystyle\sum_{i=1}^{S} (a_i - b_i)(k^i - k_0^i) > 0, \forall k^i \in \aleph \\[4mm]
a_0 + \displaystyle\sum_{i=1}^{S} a_i k_0^i < b_0 + \sum_{i=1}^{S} b_i k_0^i
\end{cases}
\tag{3.5}
$$

We now show that the first 2^S inequalities in (3.5) are inconsistent, which would imply that (3.3) does not hold for any interior point of the observed input factors, and meta-frontier indeed overarches the individual group frontiers everywhere in the observed input range.

Indeed, denoting $\Delta_i = a_i - b_i$, the first 2^S inequalities in (3.5) can be rewritten as follows:

$$\begin{cases} \Delta_1(k_{min}^1 - k_0^1) + \Delta_2(k_{min}^2 - k_0^1) + ... + \Delta_S(k_{min}^S - k_0^S) > 0 \\ \Delta_1(k_{min}^1 - k_0^1) + \Delta_2(k_{min}^2 - k_0^1) + ... + \Delta_S(k_{max}^S - k_0^S) > 0 \\ ... \\ \Delta_1(k_{max}^1 - k_0^1) + \Delta_2(k_{max}^2 - k_0^1) + ... + \Delta_S(k_{max}^S - k_0^S) > 0 \end{cases} \qquad (3.6)$$

We are claiming that the left hand side of at least one of the inequalities in (3.6) is strictly negative, thus rendering (3.6) inconsistent. Indeed, to that purpose, it suffices to find an inequality in (3.6) for which $sign(\Delta_i) = -sign(k_m^i - k_0^i)$, where $m = \{min, max\}$. Since each inequality in (3.6) corresponds to a point in the Cartesian product $\aleph = \{k_{min}^1, k_{max}^1\} \times \{k_{min}^2, k_{max}^2\} \times ... \times \{k_{min}^S, k_{max}^S\}$, (3.6) contains an inequality for which $sign(\Delta_i) = -sign(k_m^i - k_0^i) \forall i = 1..S, m = \{min, max\}$, so that the left hand side of this inequality is strictly negative.

In practical terms, one can find such an inequality by taking the following steps:

Step 1: Exclude all inequalities for which $sign(\Delta_1) = sign(k_m^1 - k_0^1), m = \{min, max\}$. That will leave one with 2^{S-1} inequalities.

Step 2: Exclude all inequalities for which $sign\Delta_2 = sign(k_m^2 - k_0^2), m = \{min, max\}$. That will leave one with 2^{S-2} inequalities.

…

Step S-1: Exclude all inequalities for which $sign(\Delta_{S-1}) = sign(k_m^2 - k_0^2), m = \{min, max\}$. At this stage one is left with two inequalities.

Step S: Exclude one of the two remaining inequalities for which

$$sign(\Delta_S) = sign(k_m^2 - k_0^2), m = \{\min, \max\}.$$

The left hand side of the single inequality one is left with after having performed step S is strictly negative, which is rendering (3.6) inconsistent. Hence, there does not exist a combination of input factor levels in the observed input range such that at this point some individual group frontier lies above the meta-frontier that satisfies conditions (3.1), which is completing our proof by contradiction.

End of Proof

It is important to notice that k_i-s in the specification (3.2) above may also be powers of the original production factors, as well as their interaction terms, which makes it easy for Proposition I to accommodate flexible functional specifications such as translog.

The cuboid of production inputs in (3.2) may be chosen arbitrarily so that it may contain a specific (not necessarily observed) range of production factors. For example, the cuboid's vertices can be chosen so that the meta-frontier will overarch group frontiers within one standard deviation from their means. Of course, the cuboid may be chosen to be the minimal multidimensional "box" that contains the observed levels of the production factors, too.

3.2 The operational method of estimating meta-frontiers using constrained maximum likelihood

The proof of Proposition 1 in subsection 3.1 is directly identifying the steps one has to follow in order to estimate the parameters of a meta-frontier based on the constrained likelihood maximization procedure. Namely:

Step 1. Identify the minimum and the maximum levels of all S input factors in the sample and form Cartesian product $\aleph = \{k_{min}^1, k_{max}^1\} \times \{k_{min}^2, k_{max}^2\} \times ... \times \{k_{min}^S, k_{max}^S\}$.

Step 2. Form a set of constraints (3.1)

Step 3. Solve the following constrained maximum likelihood problem:

$$
\begin{cases}
Max \sum_{i=1}^{N} \ln L(a^0 + \sum_{j=1}^{S} a^j \ln k^j, \sigma^2) \\
s.t. \\
y^m = a_0 + \sum_{j=1}^{S} a_j k^j \geq b_0^g + \sum_{j=1}^{S} b_j^g k^j, \forall k^j \in \aleph, \forall g = 1..G
\end{cases}
\tag{3.7}
$$

where $L(\bullet)$ is the likelihood function corresponding to a specific observation, N is the number of observations, and σ^2 is the error's variance. The maximization in (3.7) is based on the assumption of an error term that follows normal distribution with a zero mean and finite variance.

The number of constraints in (3.7) is $2^S \times G$, but it can be reduced down to 2^S if we notice that, in fact, it is enough to require that our meta-frontier is above the highest group frontier at each vertex since that will automatically guarantee the meta-frontier output is above any other group frontier at these vertices. Of course, in each vertex the maximum output can correspond to different group frontiers. Thus, (3.7) is equivalent to

$$\begin{cases} Max \sum_{i=1}^{N} \ln L(a^0 + \sum_{j=1}^{S} a^j \ln k^j, \sigma^2) \\ s.t. \\ y^m = a_0 + \sum_{i=1}^{S} a_i k^i \geq \widetilde{y}(k^i) \\ \widetilde{y}(k^i) = Max\{b_0^1 + \sum_{i=1}^{S} b_i^1 k^i, b_0^2 + \sum_{i=1}^{S} b_i^2 k^i, ..., b_0^G + \sum_{i=1}^{S} b_i^G k^i\} \end{cases} \qquad (3.8)$$

In other words, it suffices to require that the meta-frontier overarches the segmented frontier $\widetilde{y}(k^i)$ in each vertex of the specified cuboid containing the original combinations of the input factors.

It is important to notice that, since the linear programming methodology results in the closest possible linear envelope of the observed outputs (given the constraints), any other linear meta-frontier will be overarching this envelope or at least coincide with it. The implication is that the estimated meta-technology ratios are going to be smaller, reflecting the larger distance between the meta-frontier and group frontiers when the meta-frontier is estimated by (3.8). As a result, the efficiency levels with respect to the meta-frontier are going to be smaller compared to the case when meta-frontiers are estimated by means of the linear programming methodology.

18

4. Empirical Framework and Results

Similarly to O'Donnell & al. (2008), we are employing the database on agricultural production compiled by the United Nations' Food and Agriculture Organization in order to empirically estimate the group and meta-frontiers using the two alternative methods: the LP method and the constrained maximum likelihood estimation proposed in this paper. In what follows, we discuss the construction of the dataset, describe the empirical methodology we used for estimating the group and meta-frontiers, and present our empirical results.

4.1 Dataset

We download our data from the Food and Agriculture Organization's website for dissemination of statistical data on the countries' agricultural performance, faostat.fao.org. The output variable is defined as net production of crops and livestock, measured in thousands of international dollars. The three input factors are labor, land, and machines. Labor is total economically active population in agriculture, thousand people. Land includes arable land, land under permanent crops and land under permanent pasture, thousand hectares. Machinery is understood in terms of the total number of wheeled and crawler tractors, excluding garden tractors.

There are 101 countries in our panel, covering the five years between 1986 and 1990, resulting in 505 observations. We divide these countries into the following four groups: Advanced Economies, African, Latin American, and Asian countries. The division into these four country groups was performed for the illustrative purposes as to the suggested estimation

methodology in this paper rather than in the framework of a serious attempt of examining productive efficiency of the world agriculture. Table 1 below specifies the exact grouping of the countries. Table 2 provides basic summary statistics of our dataset.

Table 1.

Groups of Countries for the Empirical Analysis.

Advanced Economies	Africa	Asia	Latin America
Australia, Austria	Algeria, Angola	Bangladesh, Cambodia	Antigua and Barbuda, Argentina
Belgium, Canada	Burkina Faso, Burundi	China, India	Bahamas, Barbados
Denmark, Finland	Cameroon, Chad	Indonesia, Iran	Belize, Bolivia
France, Germany	Cote d'Ivoire, Egypt	Iraq, Israel	Brazil, British Virgin Islands
Greece, Ireland	Ethiopia, Ghana	Laos, Mongolia	Chile, Colombia
Italy, Japan	Guinea, Kenya	Myanmar, Nepal	Costa Rica, Cuba
Korea Republic of, Mexico	Madagascar, Malawi	Pakistan, Papua New Guinea	Dominica, Dominican Republic
Netherlands, New Zealand	Mali, Morocco	Philippines, Saudi Arabia	Ecuador, El Salvador
Norway, Portugal	Mozambique, Niger	Sri Lanka, Syria	Grenada, Guadeloupe
Spain, Sweden	Nigeria, Rwanda	Thailand, Vietnam	Guatemala, Guyana
Switzerland, Turkey	Senegal, South Africa		Haiti, Honduras
United Kingdom, United States	Sudan, Tunisia		Jamaica, Martinique
	Uganda, Zimbabwe		Nicaragua, Panama
			Peru, Puerto Rico
			Trinidad and Tobago, Uruguay
			Venezuela

Table 2.

Summary Statistics.

	Advanced Economies		Africa		Asia		Latin America	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Output (mn I$)	14800	25500	1481	2118	11400	33300	2564	7736
Labor (thousand)	1770	2936	3235	4021	27202	88238	980	2644
Land (mn Ha)	47.62	116.21	21.86	25.38	37.25	95.06	15.07	43.37
Machines (thousand)	607.19	966.01	10.39	25.93	73.91	201.66	33.89	119.37
Livestock (mn heads)	110.00	173.00	36.90	54.70	82.80	228.00	65.10	205.00

Note: I$ is international dollars; we are employing 505 observations in all.

4.2 Estimation Methodology

We start by estimating the local group stochastic production frontiers by employing the translog functional form of (2.1) for each country group:

$$\ln y_{it}^k = \beta_0^k + \beta_N^k \ln N_{it} + \beta_L^k \ln L_{it} + \beta_M^k \ln M_{it} + \beta_{NN}^k \left(\ln N_{it}\right)^2 + \beta_{LL}^k \left(\ln L_{it}\right)^2 + \beta_{MM}^k \left(\ln M_{it}\right)^2 + $$
$$ + \beta_{NL} \ln N_{it} \ln L_{it} + \beta_{NM} \ln N_{it} \ln M_{it} + \beta_{LM} \ln L_{it} \ln M_{it} + v_{it}^k - u_{it}^k \tag{4.1}$$

where subscripts i and t refer to countries and time periods, respectively, while $k = 1..4$ is indexing the four country groups. The three production factors in (4.1) are labor N, land L, and machines M, measured as described in section 4.1. The $\beta_\bullet^k$ are unknown parameters to be estimated for each country group.

Random variables v_{it}^k are symmetric shocks representing the purely random part of each observation's inefficiency, while the random variables u_{it}^k are reflecting the inefficient managerial behavior. We assume a half-normal distribution for u_{it}^k. Technical efficiency levels are computed as $TE_{it} = e^{-\hat{u}_{it}}$, where $\hat{u}_{it}$ are the estimated inefficiency components of the composite error term (Battese and Coelli, 1988).

Estimating a Meta-Frontier by Linear Programming

We first estimate the meta-frontier, as defined in (2.4), by solving a linear program (2.5). In other words, we are minimizing the sum of absolute deviations of the meta-frontier from the group frontiers subject to the requirement that the former may overarch the latter at the observed combinations of the production inputs:

$$\begin{cases} \underset{\beta^*}{Min} \sum_{i=1}^{9} \sum_{t=1}^{5} \left| \left(\ln x_{it}^k \right)' \beta^* - \left(\ln x_{it}^k \right)' \hat{\beta}^k \right| \\ s.t. \\ \left(\ln x_{it}^k \right)' \beta^* \geq \left(\ln x_{it}^k \right)' \hat{\beta}^k, \forall i, t, k \end{cases}$$

$$(4.2)$$

where x_{1t}^k, x_{2t}^k, and x_{3t}^k are the actually observed levels of labor, land, and machines, respectively, in year t for country group k. β^* is the vector of the meta-frontier parameters. As mentioned in the introduction, (4.2) is currently one of the more popular methods of meta-frontier estimation (see e.g. O'Donnell et al, 2008).

Estimating a Meta-Frontier by Constrained Maximum Likelihood

In this study we are offering a method of meta-frontier estimation by means of constrained maximum likelihood (CML). As mentioned in the Introduction, constrained maximization of the likelihood function corresponding to the estimated meta-frontier is a parametric methodology allowing for the statistical inference regarding the estimated parameters, which does not come naturally in case of the linear programming estimation.

We are maximizing the same likelihood function as in Battese and Rao (2002) *subject to a set of constraints* that guarantees the meta-frontier overarches the observed group frontiers in a particular set of production factor combinations. As is well noticed in Battese et al. (2004), the unconstrained maximization will implicitly employ the assumption that the data generating mechanism for the group stochastic frontiers is different from that for the meta-frontier resulting in the latter failing to overarch the group frontiers at all of the observed output levels. By introducing the constraints that guarantee the meta-frontier's domination over the group frontiers either at the observed combinations of inputs or in an arbitrarily defined connected set thereof, we guarantee the single data generating mechanism for both group frontiers and the meta-frontier (that is, the one defined by (2.1) and (2.3)) .

Formally, we are solving the following maximization problem in order to estimate the meta-frontier parameters:

$$
\begin{cases}
\underset{\beta^*}{Max}\ln L = \sum_{s=1}^{505}\left[\ln\phi\left\{\frac{y_s - \mu_s}{\sigma_\varepsilon}\right\} - \ln\sigma_\varepsilon\right] \\
\mu_s = E(y_s) = x_s'\beta^* \\
y_s = \ln Y_{it}^k \\
s.t \\
x_s'\beta^* \geq x_s'\hat{\beta}^k
\end{cases}
\qquad (4.3)
$$

where $s = 1..505$ is indexing all of the observations in our sample, y_s is the observed level

of output, x_s' is a vector of production factors for a particular country in a particular year,

and $\hat{\beta}^k$ is the vector of the estimated stochastic frontier parameters in country group k.

$\phi(\bullet)$ is a standard normal probability distribution function.

The constraints in (4.3) can stipulate that the estimated meta-frontier be overarching the

country group frontiers either at the actually observed input combinations (in which case (4.3)

becomes a constrained maximum likelihood counterpart to (2.5)), or within any cuboid

containing the observed inputs.

The solution to (4.3) is found by running the CML program in GAUSS on the data. The

complete program text (except for the constraints in order to save the space) can be seen in

the Appendix.

4.3 Estimation Results

We use empirical specification (4.3) for the constrained maximum likelihood estimation of the meta-frontiers, and specification (4.2) for the meta-frontier estimation based on the linear programming approach. Our purpose in this subsection is not to investigate the world's agricultural productivity, but to illustrate the differences between the two methodologies using real-world data.

We used a range of software packages in order to obtain our empirical results. STATA's *frontier* command was employed to obtain technical efficiency scores relative to the group frontiers. The already mentioned GAUSS procedure was written in order to estimate a meta-frontier by means of the constrained maximum likelihood methodology offered in this paper (see Appendix). The Risk Solver Platform for MS Excel helped us estimate meta-frontiers by linear programming in (4.2).

The generalized likelihood test for the null hypothesis that the estimated four group frontiers are statistically indistinguishable from each other confirmed that all four stochastic frontiers are, indeed, representing four distinct technological environments. In what follows, we will present two sets of estimates corresponding to the different type of associated constraints for both LP and the CML procedures. Table 3 allows for a comparison of the meta-technology ratios and meta-efficiencies obtained using the two methodologies in case the meta-frontier is required to overarch group frontiers only at the observed combinations of input factors. Table 4 presents the same type of comparison based on the requirement for the meta-frontier to overarch individual group frontiers *everywhere in the cuboid containing the observed range of inputs*. The estimates of technical efficiency levels relative to the group

frontiers are by construction the same for both types of estimation and constraints and are presented in Tables 3 and 4 for convenience.

Table 3.

Constrained Maximum Likelihood and Linear Programming Estimation of Meta-Technology Ratios and Meta-Efficiencies: Overarching at the Observed Input Combinations

	Constrained Maximum Likelihood				Linear Programming			
	Mean	**Standard Deviation**	**Min**	**Max**	**Mean**	**Standard Deviation**	**Min**	**Max**
	Group Efficiency				**Group Efficiency**			
Advanced Economies	95.92%	0.30%	95.22%	96.44%	95.92%	0.30%	95.22%	96.44%
Africa	78.52%	6.26%	62.48%	88.92%	78.52%	6.26%	62.48%	88.92%
Asia	63.18%	11.94%	23.85%	82.36%	63.18%	11.94%	23.85%	82.36%
Latin America	65.28%	16.40%	20.27%	89.02%	65.28%	16.40%	20.27%	89.02%
	Meta-Technology Ratio				**Meta-Technology Ratio**			
Advanced Economies	27.32%	16.94%	2.43%	86.09%	27.32%	16.94%	2.43%	86.08%
Africa	24.42%	11.90%	8.04%	57.76%	24.41%	11.90%	8.04%	57.75%
Asia	48.18%	13.63%	27.17%	86.49%	48.18%	13.63%	27.17%	86.49
Latin America	42.01%	14.16%	17.23%	97.58%	42.01%	14.16%	17.23%	97.56%
	Meta-Efficiency				**Meta-Efficiency**			
Advanced Economies	36.47%	16.17%	7.22%	73.9%	36.46%	16.17%	7.22%	73.87%
Africa	26.13%	8.40%	10.22%	48.37%	26.13%	8.40%	10.22%	48.37%
Asia	31.36%	9.86%	9.48%	64.76%	31.35%	9.86%	9.48%	64.75%
Latin America	22.36%	8.46%	4.74%	38.70%	22.35%	8.46%	4.74%	38.70%

For both estimation techniques the meta-technology ratios for the four country groups are virtually identical to each other, reflecting the fact that the coefficients of the two meta-frontiers are coinciding up to the fourth digit after the decimal point, except for the intercept where the fourth digits are different. As a result, meta-efficiency levels are coinciding as well for the two estimation procedures. The advanced economies are estimated to be located closer to the meta-frontier relative to the other country groups, even if the meta-technology ratio is

the highest for the Asian countries. The number of constraints for both estimation procedures is 505, which is equal to the total number of observations in our sample.

Table 4 below compares the results of the CML and LP estimations in case the meta-frontier is required to overarch group frontiers everywhere within the observed range of production inputs. As shown in section 3.1, this requirement is fulfilled whenever the meta-frontier overarches the segmented frontier (defined as the maximum of all group frontiers at each point) in the vertices of the observed range of inputs defined as elements of a Cartesian product of the two-element sets containing the minimum and the maximum levels of each term entering the translog specification (4.1). The number of constraints in case of the latter requirement is equal to 2^S, where S is the number of input factors. In our case, the number of constraints is equal to $2^9 = 512$.

Table 4.

Constrained Maximum Likelihood and Linear Programming Estimation of Meta-Technology Ratios and Meta-Efficiencies:
Overarching *Within the Range of* the Observed Inputs.

	Constrained Maximum Likelihood				Linear Programming			
	Mean	Standard Deviation	Min	Max	Mean	Standard Deviation	Min	Max
	Group Efficiency				**Group Efficiency**			
Advanced Economies	95.92%	0.30%	95.22%	96.44%	95.92%	0.30%	95.22%	96.44%
Africa	78.52%	6.26%	62.48%	88.92%	78.52%	6.26%	62.48%	88.92%
Asia	63.18%	11.94%	23.85%	82.36%	63.18%	11.94%	23.85%	82.36%
Latin America	65.28%	16.40%	20.27%	89.02%	65.28%	16.40%	20.27%	89.02%
	Meta-Technology Ratio				**Meta-Technology Ratio**			
Advanced Economies	4.77%	2.75%	0.69%	16.28%	8.96%	10.64%	0.11%	65.27%
Africa	4.27%	1.30%	1.87%	9.03%	6.23%	6.01%	0.44%	53.31%
Asia	9.07%	2.69%	4.53%	19.83%	11.30%	7.41%	1.22%	47.88%
Latin America	7.51%	1.04%	4.61%	10.89%	11.54%	10.69%	0.64%	94.41%
	Meta-Efficiency				**Meta-Efficiency**			
Advanced Economies	7.56%	3.16%	2.76%	15.63%	14.94%	11.66%	0.40%	55.63%
Africa	3.97%	0.98%	2.24%	6.20%	6.45%	3.20%	0.73%	14.44%
Asia	5.46%	1.94%	2.32%	9.66%	8.44%	6.75%	0.75%	33.34%
Latin America	4.54%	1.19%	1.30%	6.45%	5.37%	5.47%	0.23%	26.33%

Both meta-technology ratios and, correspondingly, the meta-efficiency scores are lower compared to their counterparts in Table 3, reflecting changes in the set of constraints underlying the estimation in both LP and CML case. Since the modified set of constraints requires the meta-frontier to overarch group frontiers on a larger set of input factor combinations (not only in *a subset of the observable range*!), both LP and CML meta-frontiers are pushed "higher", resulting in lower meta-technology ratios. However, distinctly from the case presented in Table 3, the meta-frontier parameters are now estimated to be

different, which shows up in the different estimates of efficiency levels for the two procedures.

The meta-technology ratios and meta-efficiency scores are estimated to be higher in case of an LP estimation compared to the CML one. In other words, the LP estimation procedure in this case has produced a lower meta-frontier compared to the CML one. One reason why the CML meta-frontier was estimated to be above the LP one is that the meta-frontier that is geometrically the closest one to the group frontiers need not be the likeliest one in the statistical sense.

Depending on the type of the constraints imposed on the meta-frontier and the type of the estimation procedure, the rankings of country groups in terms of their mean meta-technology ratios and meta-efficiency scores appear to be only slightly different. When the meta-frontier is required to overarch group frontiers only at the observed input combinations (Table 3), the Asian countries appear to be the closest to the global meta-frontier, while the African countries are located the furthest from it on average. In case of the meta-frontier overarching the group frontiers on a cuboid containing all possible input combinations *within the observed range*, Table 4, the Asian countries are still the closest to the meta-frontier in case of the CML estimation procedure, while the LP estimates put the Latin American countries marginally closer. The African countries keep on being located the furthest away from the meta-frontier. Advanced countries are estimated to be boasting the highest meta-efficiency levels relative to the other country groups on average in case of all four estimations.

5. Conclusion

The purpose of this study is to offer an alternative parametric methodology of estimating the meta-frontiers. Our estimation procedure is based on the constrained maximum likelihood (CML) methodology and a more flexible set of constraints on the meta-frontier in terms of the latter's ability to overarch the individual group frontiers. We applied our procedure to the dataset on the global agricultural production and found that the application of constrained maximum likelihood methodology results in the same or "higher" meta-frontier compared to the one obtained by means of the linear programming methodology.

Both the individual group frontiers and the meta-frontiers are estimated by maximum likelihood, which makes the process of estimating the meta-technology ratios and meta-efficiency scores inherently more consistent. Since the CML meta-frontier overarches the country group frontiers at the observed input combinations, the data generating mechanism in case of the proposed CML estimation is consistent with the data generating mechanism for the group frontiers.

The CML estimation approach allows for the statistical inference on the estimated parameters of the meta-frontier production function that cannot be provided naturally in case of the linear programming estimation, although bootstrapping and simulation may help alleviate the difficulty.

In our view, it is an interesting question whether there exist conditions under which the CML procedure produces the same meta-frontier (and, hence, the same meta-technology ratios and meta-efficiency scores) as does the LP procedure provided, of course, the same set

of constraints on the meta-frontier. However, the exploration of this issue is beyond the scope of this study.

References

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6, 21-37.

Battese, G.E., Coelli, T., 1988. Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. Journal of Econometrics, 38, 387-399.

Battese, G.E., Rao, D.S.P., 2002. Technology potential, efficiency and a stochastic metafrontier function. International Journal of Business and Economics, 1(2), 1-7.

Battese, G.E., Rao, D.S.P., O'Donnell, C.J., 2004. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. Journal of Productivity Analysis, 21(1), 91-103.

Coelli, T.J., 1992. A computer program for frontier production function estimation: FRONTIER version 2.0. Economics Letters, 39(1), 29-32.

Fried, H. O., Lovell, C.A.K., Schmidt, Sh. S., 2008. Efficiency and Productivity. In: Fried, Harold. O., Lovell, C.A.Knox, Schmidt, Shelton S., (Eds.). The Measurement of Productive Efficiency and Productivity Growth. New York: Oxford University Press, 7-8.

Hayami, Y., Ruttan, V.W., 1971. Agricultural development: an international perspective. Johns Hopkins University Press, Baltimore.

Henningsen, A., and Henning, Ch.H.C.A., 2009. Imposing regional monotonicity on translog production frontiers with a simple three-step procedure. Journal of Productivity Analysis, 32(3), 217-229.

Moreira, V.H., Bravo-Ureta, B.E., 2010. Technical efficiency and metatechnology ratios for dairy farms in three southern cone countries: a stochastic meta-frontier model. Journal of Productivity Analysis, 33(1), 33-35.

O'Donnell, C.J., Rao, D.S.P., Battese, G.E., 2008. Meta-frontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical Economics, 34, 231-255.

Zeitsch, J., Lawrence, D., Salerian, J., 1994. Comparing Like with Like in Productivity Studies: Apples, Oranges and Electricity. Economic Record, 70:209 (June), 162-70.

Appendix: The GAUSS Program for Constrained Maximum Likelihood

//Include CML library
library cml;
#include cml.ext;
cmlset;

//Input data from Excel spreadsheet
raw=xlsreadm("forgauss.xls","a1:k500",1,0);

//Define likelihood function: x contains parameters, z is data matrix
proc lpr(x,z);
   local s,m;
   s=x[11];
   m=z[.,2:11]*x[1:10,.];
   retp(ln(pdfn((z[.,1]-m)/sqrt(s)))-ln(sqrt(s)));
endp;

//Set initial values for parameters x
x0={12.18, -0.42, -0.11, 0.47, 0.04, 0.006, 0.016, 0.06, -0.05, -0.01, 0.12};

//Constrain the meta frontier to overhaul the local frontiers at specific combinations of inputs
_cml_C={1 7.466228 10.56328 11.32233 55.74455 111.583 128.1952 78.86789 84.53512 119.601 0,

<elements of the matrix of constraints C>

1 7.928046 9.469237 9.72675 62.85391 89.66646 94.60967 75.07255 77.11412 92.1049 0,
1 7.96137 9.473474 9.763076 63.38341 89.7467 95.31765 75.42183 77.72746 92.49024 0};

_cml_D={17.46638,

<elements of the constraints' right hand sides D>

16.65534,
16.67542};

//Constrain the standard deviation to be positive
_cml_Bounds={-50 50, -50 50, -50 50, -50 50, -50 50, -50 50, -50 50, -50 50, -50 50, -50 50, 0.001 50};

//Perform constrained likelihood maximization
{x,f,g,cov,ret}=CMLPrt(CML(raw,0,&lpr,x0));