

Technical efficiency estimators in linear mixed effects stochastic production frontier models

Renato Salvatore

Abstract The influence of some external factors on the measure of technical efficiency of firms is a relevant question, because the analysis of their impact on the different levels of production can get two main features. Firstly, it allows to measure the efficiency rate of the production units that share the same level of the external factors, and, secondly, to recognize relevant factors that affect the efficiency in production itself. This study highlights the measure of firms technical efficiency, performed by the linear mixed effects model framework. Different levels of technical efficiency of firms can be evaluated by distribution-free quadratic estimators, as well as maximum likelihood estimators. The model parameters estimates in the application study are attained by the restricted maximum likelihood estimation method.

Key words: technical efficiency, stochastic frontier, linear mixed models, statistics for business and industry, multilevel data analysis

1 Introduction

In the context of the analysis of the stochastic frontier approach in the assessing of the economic production processes, several authors, such as Bauer and Hancock (1993), Simar et al. (1994), Kumbhakar and Lovell (2000), discuss models that account for exogenous variables as technical efficiency level characteristics, in order to analyze

¹

Renato Salvatore, DIMET Department, University of Cassino, Italy,
email: rsalvatore@unicas.it

This paper represents some of the results from the PRIN 2007 Italian National Research Project “*The use of auxiliary information in complex sampling survey designs*”, granted by the Italian Ministry of University and Education. The author thanks the Italian National Statistics Institute for supplying the data.

direct influences of external factors in the estimation of the frontier production function. Kumbhakar et al. (1991), Reifschneider and Stevenson (1991), Huang and Liu (1994), Battese and Coelli (1995), underline the relevance of non-neutral firms characteristics in estimating production frontiers, and proposed to estimate jointly technical efficiency and the exogenous variables effects at the same time. Caudill and Ford (1993), Caudill et al. (1995), studied influences of some exogenous variables on the production frontier function estimation through the heteroscedasticity of the one-sided efficiency term, as function of firm characteristics. All these approaches highlight the deep impact on the models efficiency estimates of these factors. Further, actually random effects models are extensively employed in the context of panel data stochastic frontier research. Greene (2005a, 2005b), extends the classic random effect model to panel data considering random firm-specific effects, in a so-called “true” random effects model, preserving the one-sided technical efficiency nuisance term. Some studies on categorized exogenous variables, as sources of efficiency in mixed models, are due to Gori et al., (2002), Grasseti et al., (2005), that use multilevel analysis in estimating production function in health economic data. Their works define the second-level (symmetric) error as an indicator of firm’s efficiency rate. Differences in estimated parameters of the production functions are explained by different levels of inefficiency, recognized in terms of within-group and between-group variability. Anyway, estimates of the average population production function are possible.

The model discussed in the present research has two main starting points, that are stated beforehand: a) the model is designed for including production technology as inputs in terms of continuous variables, and various sources of inefficiency in terms of categorical variables, that specify different clusters; b) the model can work without specific distributional assumptions related to technical inefficiency. Further, the model explains the presence of inefficiency as a portion of the model variance, based on the covariance matrices of the random effects and of the residual. If the variability due to specific efficiency parameters in the model is estimated to be negligible or zero, the model reduces to a standard linear mixed effects model with two sources of variation, related to the standard matrix of covariance parameters of random effects and of the measurement errors.

We introduce two distinctive types of parameters, connected with the problem of the technical efficiency estimation of firms: a “technical efficiency” matrix, in which the entries represent the departure from the population mean production function of the external factor-based technical inefficiency categories, and the between-subject technical inefficiency level-based heteroscedasticity rescaling parameter, that describe the sliding down of the production output of firms in the group considered, from the original condition of measurement error dispersion of units. Assuming that the measurement error component of the model is the same in every group, the portion model variability that exceed this limit is considered as due to the presence of inefficient firms or group of firms. From a statistical point of view, this starting point enable the researcher to estimate various forms of technical efficiency measures on units, as well as the frontier line. The last viewed as the population production function in absence of technical inefficiency of firms.

Empirical results of the application of the model on the Italian National Statistics Institute official wheat production survey data are discussed, compared with those from other two models in the literature, the Battese and Coelli (1992) model, and the multilevel model analysis of production, introduced by Gori and Grasseti (2005).

2 The model

2.1 Definition and properties

The linear mixed effects model considered in this paper works with cross-sectional data, aggregated by some effects such as interaction effects, nested effects, and specified in random intercept or random-coefficient models, as well as multilevel and hierarchical models (Demidenko, (2004)). This approach to the frontier production function analysis is “stochastic”, in the sense that any individual firm technical efficiency is estimated on the basis of the error components of the model. We define a production function in terms of random effect model, without considering one-sided errors, in the following way:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{v}_i, \quad (1)$$

where y_{ij} are the logarithm of output for firm j ($j=1, \dots, n_i$) in the group (subject) i ($i=1, \dots, N$), $\mathbf{x}_{ij}^T$ is a $1 \times p$ vector of inputs, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population parameters (fixed effects), $\mathbf{Z}_i$ is the design matrix of random effects. Further, $\mathbf{u}_i$, $\mathbf{v}_i$ are, respectively, a $q \times 1$ vector of random effects with zero mean and covariance matrix $\mathbf{G}$, and a $n_i \times 1$ vector of within-subject error terms with independent components. $\mathbf{u}_i$ and $\mathbf{v}_i$ are independent. In the case of normality distribution of $\mathbf{u}_i$ and $\mathbf{v}_i$, we have:

$$\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_N) \stackrel{iid}{\sim} N_q(\mathbf{0}, \mathbf{G}), \quad \mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_N) \stackrel{iid}{\sim} N_q(\mathbf{0}, \mathbf{R}_i), \quad i = 1, \dots, N$$

$\mathbf{G}$ and $\mathbf{R}_i$ are positive definite matrices, and, even though the observations across the groups are independent, one may specify the model (1) in one-equation structure. Letting $N_T = \sum_{i=1}^N n_i$, $\mathbf{P} = \mathbf{I}_N \otimes \mathbf{G}$, $\mathbf{S} = \oplus_{i=1}^N \mathbf{R}_i$, with symbols $\otimes$ and $\oplus$ that denote, respectively, the Kronecker product and the direct sum, if we set $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_N^T)$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_N^T)$, $\mathbf{Z} = \oplus_{i=1}^N \mathbf{Z}_i$, the model in (1) can be compactly specified as a model with a block-diagonal covariance structure (Jiang, (2007)):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{v}, \quad \text{with} \quad \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \sim \begin{pmatrix} \mathbf{0}_{qN} \\ \mathbf{0}_{N_T} \end{pmatrix}, \begin{bmatrix} \mathbf{P} & \mathbf{0}_{qN \times N_T} \\ \mathbf{0}_{N_T \times qN} & \mathbf{S} \end{bmatrix} \quad (2)$$

The model (2) has, in fact, the following block-diagonal covariance matrix:

$$\begin{aligned}\text{cov}(\mathbf{y}) = \mathbf{V} &= \mathbf{S} + \mathbf{Z}\mathbf{P}\mathbf{Z}^T = \mathbf{V}_1 \oplus \dots \oplus \mathbf{V}_N = \\ &= \text{diag}\left\{\mathbf{I}_{n_1} \otimes (\mathbf{R}_1 + \mathbf{Z}_1\mathbf{G}\mathbf{Z}_1^T), \dots, \mathbf{I}_{n_N} \otimes (\mathbf{R}_N + \mathbf{Z}_N\mathbf{G}\mathbf{Z}_N^T)\right\}\end{aligned}$$

The BLUE and BLUP estimates of, respectively, $\boldsymbol{\beta}$ and $\mathbf{u}$ are:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}, \quad \hat{\mathbf{u}}_i = \mathbf{G}\mathbf{Z}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) \\ \mathbf{y}_i^A &= \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{v}_i,\end{aligned}\tag{3}$$

with the superscript “A” that means “subject-Averaged”. Standing this definition of the problem, we as first step assume that $\mathbf{R}_i = \sigma_v^2 \mathbf{I}$, letting $\mathbf{S} = \oplus \sigma_v^2 \mathbf{I}_{n_i} = \sigma_v^2 \mathbf{I}_{N_T}$. The covariance matrix can be easily parametrized by a vector $\boldsymbol{\gamma}$, e.g. $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\gamma})$. Now consider the following $n_i \times 1$ vector that can accommodate the firms technical inefficiency-based variability through the data:

$$\mathbf{b}_i = \mathbf{W}_i \mathbf{u}_i + k_i \mathbf{v}_i,\tag{4}$$

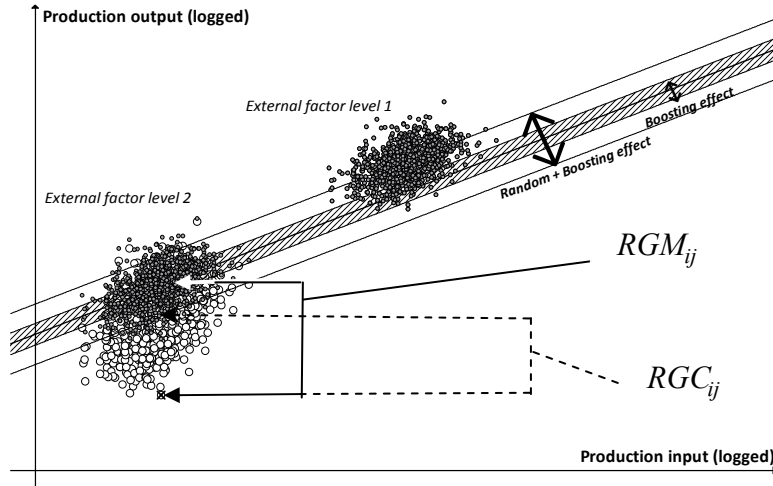
with w_{ij} as entries of the matrix $\mathbf{W}_i$ that are parameters of technical efficiency related to the random effects $\mathbf{u}_i$, and the k_i 's as parameters that account for heteroscedasticity among the subjects, due to different levels of within-subject (“internal”) technical inefficiency. The model (3) is now:

$$\mathbf{y}_i = \mathbf{y}_i^A + \mathbf{b}_i = \mathbf{y}_i^A + \mathbf{W}_i \mathbf{u}_i + k_i \mathbf{v}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{Z}_i^* \mathbf{u}_i + k_i^* \mathbf{v}_i,\tag{5}$$

with $\mathbf{Z}_i^* = \mathbf{Z}_i + \mathbf{W}_i = \mathbf{Z}_i + \mathbf{Z}_i \mathbf{E}$, and $k_i^* = 1 + k_i \geq 1$, where $\mathbf{E}$ and k_i^* can be obtained rescaling the within-subject variances. The $q \times q$ matrix $\mathbf{E}$, not necessarily symmetric, is a matrix of parameters, to be estimated, here called “technical efficiency” matrix. Parameters in $\mathbf{E}$ show the components of the prediction values induced by the $\mathbf{u}_i$'s, due to the presence of between-subject technical efficiency level, measured on the estimated population-averaged production function. If $\mathbf{E}$ is a diagonal matrix, its entries (parameters) define the vector of the “boosting effects” on the mean production, estimated by $\hat{\mathbf{E}} \hat{\mathbf{u}}_i$, made by the technical inefficiency levels on the random effects vector components. The boosting effects parameters, namely the values in $\text{diag}(e_{11}, e_{22})$, can be positive or negative, due to the circumstance that the within-group inefficiency-based variability inflate or play down the random effect variability (see a trivial example of this on Figure 1). When we search for an unstructured efficiency matrix $\mathbf{E}$, the off-diagonal elements of $\hat{\mathbf{E}}$ can explain the effect due to the interaction between the inflated (or deflated) random effects on intercept and slopes. The conditional residuals for the model (5), are the following:

$$\hat{\mathbf{v}}_i = (1 + \hat{k}_i)^{-1} (y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} + \hat{\mathbf{Z}}_i^* \hat{\mathbf{u}}_i) \quad (6)$$

Figure 1: This figure shows the increased variability of the group n.2, going from the original grey points (the only measurement error situation) to the white ones, due to the presence of the low technical efficiency-based firms. The figure highlights the “boosting effects” on the intercept, the Global Marginal (RGM) and the Global Conditional (RGC) model residuals.



2.2 Estimation, residuals, and technical efficiencies

Parameters in the model (5) can be estimated following the guidelines of the existing estimation methods in the literature. Distribution-free quadratic estimation methods, like the Method of Moments (MM), Minimum Variance Quadratic (MINQUE), Variance Least Squares (VLS) estimators can be employed, as well as likelihood function-based estimators like ML (maximum likelihood) or REML (restricted (residual) maximum likelihood)-types estimators. For example, considering the VLS estimation method, starting from the empirical covariance matrix $\hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T$, based on the fixed effects OLS estimates, we must minimize the following function:

$$V_{LS} = \sum_{i=1}^N \text{trace}(\hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T - k_i^* \sigma_v^2 \mathbf{I} - (\mathbf{Z}_i + \mathbf{Z}_i \mathbf{E}) \mathbf{G} (\mathbf{Z}_i + \mathbf{Z}_i \mathbf{E})^T)^2 = \min., \quad (7)$$

where N is the total number of the clusters. The VLS is a non-iterative method, and the unbiased estimates are easily obtained from the system of the derivatives of the function in (7). Setting $k_i^* = 1$, $\mathbf{E} = \mathbf{0}$ in the relation (7), we have the first estimates of the covariance parameters σ_v^2 and $\mathbf{G}$. After the evaluation of the gradient and the

hessian matrix of the VLS function respect to $(k_i^*, \mathbf{E})$, we have the desired estimates. We can employ also alternating least squares minimizing respect to k_i^* , and then solving the derivative of (7) respect to $\text{vec}(\mathbf{E})$. Finally we have the final estimates of σ_v^2 and $\mathbf{G}$ on the model (5), employing the estimates $(\hat{k}_i^*, \hat{\mathbf{E}})$. Alternatively, ML or REML estimates can be achieved, assuming normal distribution. For example, in the case of the REML, we must maximize the following residual version of the maximum likelihood:

$$l_{reml} = \frac{N_T - p}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{X}^T \mathbf{X}| - \frac{1}{2} \log |\mathbf{V}^*| - \frac{1}{2} \log |\mathbf{X}^T \mathbf{V}^* \mathbf{X}| - \frac{1}{2} \mathbf{y}^T \mathbf{P}^* \mathbf{y}, \quad (8)$$

with $\mathbf{P}^*$ the projection matrix $\mathbf{P}^* = \mathbf{V}^{*-1} - \mathbf{V}^{*-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{*-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{*-1}$ and $\mathbf{V}^* = \mathbf{S} + \mathbf{Z}^* \mathbf{P} \mathbf{Z}^{*T}$.

The estimators $(\hat{k}_i^*, \hat{\mathbf{E}})$ are consistent. It is easy to prove that the sufficient condition for consistency of $\hat{k}_i^*$ is that the total number of units at each cluster diverges. For the estimates of $\text{vec}(\hat{\mathbf{E}})$, consistency is defined by ML methods and by the quadratic-type estimators. Under normality assumption, it is possible to obtain the asymptotic covariance matrix for the model (5), letting $k_i^* = \hat{k}_i^*$, $\mathbf{E} = \mathbf{0}$, and then obtain an unique solution for $As(\mathbf{E})$.

One of the main features of the model in (5), is that individual technical efficiency can be estimated directly from the model residuals. We distinguish the standard mixed model residuals, here denoted "global" residuals, from "efficiency" residuals. While the global residuals predict the errors of the model, considering at the same time the measurement errors and the deviation from the production function due to the presence of inefficient firms, the efficiency residuals predict the part of the model error due to different inefficiency levels among the subjects. Remembering the relation (6), we have (see also Figure 1): A) the "global" marginal residuals (RGM), B) the "global" conditional residuals (RGC), C) the "efficiency" marginal residuals (REM) and, D) the "efficiency" conditional residuals (REC). The following are the formulas:

- A. $RGM_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} = \hat{\mathbf{z}}_{ij}^{*T} \hat{\mathbf{u}}_i + \hat{k}_i^* \hat{v}_{ij} = \mathbf{z}_{ij}^T \hat{\mathbf{u}}_i + \hat{v}_{ij} + REM_{ij}$, that predict the marginal global errors of the model;
- B. $RGC_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} - \hat{\mathbf{z}}_{ij}^{*T} \hat{\mathbf{u}}_i = \hat{k}_i^* \hat{v}_{ij} = \hat{v}_{ij} + REC_{ij}$, that predict the conditional global errors;
- C. $REM_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} - \mathbf{z}_{ij}^T \hat{\mathbf{u}}_i - \hat{v}_{ij} = \mathbf{z}_{ij}^T \mathbf{E} \hat{\mathbf{u}}_i + \hat{k}_i \hat{v}_{ij} = \mathbf{z}_{ij}^T \mathbf{E} \hat{\mathbf{u}}_i + REC_{ij}$, that predict the "efficiency error" from the population averaged production function;
- D. $REC_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} - \hat{\mathbf{z}}_{ij}^{*T} \hat{\mathbf{u}}_i - \hat{v}_{ij} = \hat{k}_i \hat{v}_{ij}$, that predict the error from the subject-specific production function.

The efficiency frontier can be then defined by the two cases below:

- a. $\max_{i,j}(\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^T \hat{\mathbf{u}}_i + \mathbf{z}_{ij}^T \mathbf{E} \hat{\mathbf{u}}_i + REC_{ij}) = \max_{i,j}(\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^T \hat{\mathbf{u}}_i + REM_{ij})$, when we decide on significance of the random effects designed by the model (5);
- b. $\max_{i,j}(\mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + REM_{ij})$, otherwise

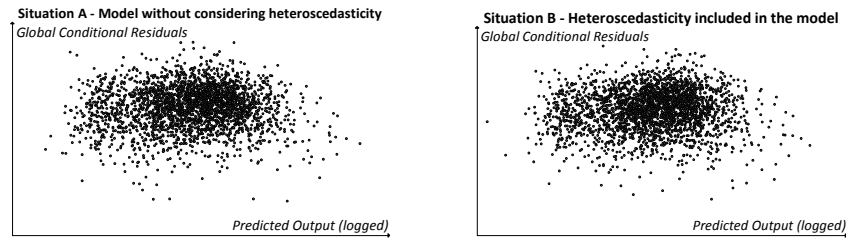
Note that the efficiency rates of firms, conditioned to the specific subject, is defined by the maximum value of the conditional “efficiency” residuals (REC) allocated in their own group. In case of b.), the frontier is specified as “REM frontier”. The last is obtained considering only the boosting effects (based on the entries of the efficiency matrix $\hat{\mathbf{E}}$) and the internal rates of efficiency $\hat{k}_i$.

There are several efficiency measures that can be estimated by the model (5). For all of them, we can utilize the formula: $efficiency = \exp(predict - \max(predict))$. We have a marginal efficiency, respect to the all units, and this also respect to the REM frontier. The model further enables to measure “relative” efficiencies, when the value of $\max(predict)$ is of the own cluster of membership for the examined firm. Finally, we can measure “global” efficiency levels, when we overlap the clusters and define the $\max(predict)$ as the maximum value of the departures from each group means. The last can be very useful in the comparative analysis of all the production units, e.g. without considering the component of the efficiency measure due to the position of the different subjects.

3 Empirical results

The application considers the official 2005 Farm Structure Survey data from the Italian National Statistics Institute, on wheat farmers in Italy. Data are considered on $n = 2603$ specialized farms in wheat production, and are classified in $N = 94$ different cross-classified levels of the external factors categorical variables, the administrative province and the altimetrical area of membership. Recent studies demonstrated that there are significative differences in the mean level of wheat production, after combining the different categories of these two territorial factors.

Figure 2: Plots of the RGC’s - global conditional residuals (externally Studentized).



The application compares the Cobb-Douglas wheat production function in the three models: a) the Battese and Coelli model (denoted with BC) under the 1992 specification on the aggregated data, b) a multilevel model under the Gori-Grassetti specification (denoted by GG), in which the second-level data are those from the interaction (crossed) effects due to the administrative province and the altimetrical area to which each farm (j) belongs, and c), the model presented in this paper (denoted by LME), with the same external factors and the same second-level effects as in b). We further employ the REML estimation method in order to get the model parameters highlighted by equation (5), assuming normality of the random effects and of the residual error for the GG and the LME models, on the basis of the l_{reml} as in (8).

Table 1: Parameters estimates in the three proposed stochastic production function models: **BC** (Battese and Coelli, 1992 specification), **GG** (multilevel mixed model, Gori and Grassetti, 2005), **LME** (the model discussed in this paper)

<i>Parameter</i>	<i>BC</i>	<i>GG</i>	<i>LME</i>
β_0 , with: ($\beta_{0,OLS} = 2.18082$)	3.30535	2.5381	2.7613
β_1 , with: ($\beta_{1,OLS} = 0.46483$)	0.40735	0.3312	0.3002
β_2 , with: ($\beta_{2,OLS} = -0.41365$)	-0.40796	-0.3798	-0.3822
σ_v^2 , with: ($\sigma_{OLS}^2 = 0.59659$)	1.2940	2.0621	0.001153
γ (σ_u^2 / σ_v^2)	0.8659	-	-
g_{11} (the variance of the rand. eff. on the intercept β_0)	-	0.2482	0.2117
g_{22} (the variance of the rand. eff. on the slope parameter β_1)	-	-	0.002928
θ (the dispersion parameter for the heteroscedasticity in the data)	-	-0.2835	-0.2935
e_{11} (the technical efficiency parameter in E on the intercept β_0)	-	-	-0.22504
e_{22} (the technical efficiency parameter in E on the slope parameter β_1)	-	-	0.181743
k_i^* (average)	-	-	383.7537
k_i^* (max)	-	-	4330.46
$-2 \times \text{residual log likelihood}$	-	5302.6	5013.7
Mean efficiency	0.51321	0.15982	0.1712
Mean REM efficiency	-	-	0.12660
Mean “global” efficiency	-	-	0.31512
Mean efficiency by WLS before BC	0.29415	-	-

The simple functional form considers two explanatory variables (covariates) in the LME model as inputs in wheat production: i) the tillable area (ta), and ii) the total

working days (twd), while the output is represented by the wheat production per hectare ($wprod$):

$$\ln(wprod)_{ij} = (\beta_0 + u_{0i} + e_{11}u_{0i}) + (\beta_1 + u_{1i} + e_{22}u_{1i}) \ln(ta_{ij}) + \beta_2 \ln(twd_{ij}) + v_{ij} \quad (7)$$

Two random effects are considered, on intercept (u_{0i}), and on the slope of the tillable area (u_{1i}). The covariance matrix of the random effects is of the form $\mathbf{G} = \text{diag}(g_{11}, g_{22})$. Further, the efficiency matrix is considered in the form $\mathbf{E} = \text{diag}(e_{11}, e_{22})$, with the only boosting effects on intercept and on the first slope parameter. First-level within-group heteroscedasticity error was corrected with the exponential-type variance function $\text{Var}(v_{ij}) = \sigma_v^2 |ta_{ij}|^{2\theta}$ (see Figure 2). Battese and Coelli model and Gori and Grassetto model specifications are referable to their own works in (Battese and Coelli, (1992)), and (Gori and Grassetto, (2005)). Maximum likelihood estimates for the Battese-Coelli model are reached after 8 iterations and a grid search for γ as in the formulas of COLS estimators by Battese and Coelli (1995), using the Frontier 4.1 Software. Estimates for models of Gori-Grassetto and the LME model are reached after 5 and 37 Newton-Raphson iterations, respectively, on the residual log-likelihood. The parameters are showed in the Table 1. All the models considered have the fixed effects parameters for which $H_0 : \beta_i = 0, i = 0, 1, 2$ is rejected. Considering the BC model, also the value of γ is significative ($H_0 : \gamma = 0$ is rejected), which specifies that the wheat farmers are not fully technically efficient. In terms of the application of the models GG and LME, the covariance parameters associated with the random effects (the intercept for the GG model, and the intercept and the first slope for the LME) are significative. Further, for the last models, random effects parameters are significative ($\alpha < 0.1$) for 49 areas on the 94 in total (the crossed values resulting from the interactions of administrative provinces and altimetrical areas) in the GG model (the random effects vector here is a scalar), while for the LME model we have only 17 areas significative ($\alpha < 0.1$) on the total 188 random effects considered (94×2 , because we have for the LME model a bivariate vector of random effects). Now considering the above estimates, they leads to some appraisals on the characteristics of the models.

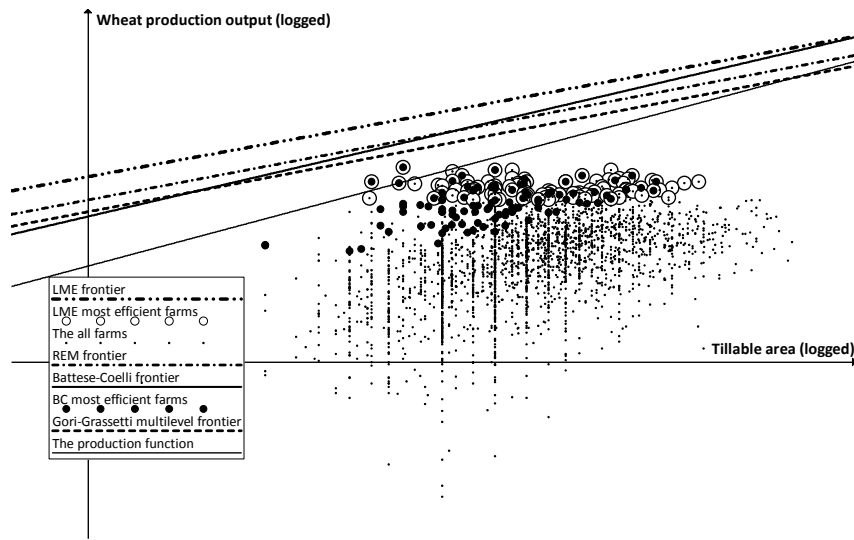
First we note that the estimation method coming from the COLS estimates in the BC model, in order to estimate the composite error vector $(\mathbf{v} - \boldsymbol{\tau})$, with $\boldsymbol{\tau}$ the one-sided technical inefficiency vector error component of half-normal type, has final estimates of β_0 and of the variance of v_j increased respect to the standard OLS estimates.

While the other fixed effects, β_1, β_2 are similar to the OLS estimates. After the grid search and the maximization of the likelihood after the COLS specification, the value of γ is very high, denoting that the major part of the global error variance σ^2 is due to the variance of the one-sided technical inefficiency error. The mean efficiency is of about 0.51.

The multilevel model, denoted by GG, has some distinctive features: the estimation of the random effects on the intercept has its main consequence on estimate of the fixed effect β_1 (the slope parameter on the logged tillable area), while the estimate of β_0

and of β_2 are similar with the corresponding OLS estimates. A consistent number of significative random effects on intercept for each cluster leads to an appreciable variation, especially in the value of β_1 (this is a typical behavior in multilevel and, generally, in mixed models). Further, the major part of the model variance is allocated in the within-group variability. This estimate is about three times the estimated model variance of the OLS approach. We must consider the correction factor for the inclusion of the heteroscedasticity in the model, that over-estimates this parameter respect to the OLS estimate for the original data. The dispersion parameter, very similar to the corresponding parameter for the LME model, denotes an appreciable heteroscedasticity through the first-level units (the farms).

Figure 3: The most efficient firms under the Battese-Coelli, multilevel (Gori and Grasseti), and LME model, with a comparison of frontier functions.



The LME model has fixed effects parameters estimates that differ substantially from the other model parameters estimates, in the following aspects: a) β_0 is higher than that of the OLS and the GG model; b) the estimate for β_1 is smaller respect to the OLS and GG estimates; c) following the assumptions of the LME model, the residual variance, estimated rescaling all the within-group variances, is very small. This value corresponds to the estimation of the variability of the measurement error, supposed equal among groups, and specified as $\sigma_v^2 = \min_i \sigma_{vi}^2$. The covariance parameter g_{22} is very small, denoting small variations on the tillable area slope fixed parameter β_1 . The technical efficiency parameters e_{11}, e_{22} are of opposite sign: the estimated negative value of e_{11} denote that the estimated random effect of the intercept, u_{0i} , decreases in terms of the 22.5% of the random effect, due to the presence of groups with high within-subject inefficiency levels mainly where there is a high-mean efficiency. The

positive value of e_{22} denote that the random effects on the slope β_1 get the 18.2% of this amount, and so increase its original value because of the presence of the rather different levels of “internal” (within-subject) inefficiency among all the groups of farms. The “internal” inefficiency parameters k_i^* ’s have quite different values. The mean value of the k_i^* ’s show that the groups have an average value of within-subject variance about 384 times the variability of the less varying cluster (with $k_i^* = 1$). This means that in every group of farms considered, the almost whole residual error is explained by the model as “technical inefficiency” error component. Now take a look on the technical efficiency measures delivered by the three models. For the Battese-Coelli model, we have a mean efficiency of 0.51. This measure come from the OLS original estimates, but considering the BC model with the WLS estimates based on the same heteroscedasticity correction factor included in the GG and the LME models, we get a value of about 0.29. For the Gori-Grasseti multilevel model we have a mean efficiency of about 0.16, this value based on the all farms technical efficiency level respect to the best subject-averaged production function. Considering the LME model: a) the mean efficiency is of 0.17, increased respect to the GG model because of the specification of the boosting effects on parameters; b) the REM efficiency is of 0.13, without considering the significance of the random effects, and c) the “global” efficiency is of 0.32. Overlapping all the clusters centres, this technical efficiency measure compares at the same time all the farms, excluding the technical inefficiency component due to the own cluster membership. Note that this mean value is of the same order of the BC mean efficiency in the WLS regression setup. In particular this means that, respect to the standard stochastic frontier models, the basic hypothesis of the LME model, that refers to a joint minimal measurement error variability throughout the data, can leads to similar values of farms efficiency, without considering neither COLS estimates or one-sided error composite specification. Finally, Figure 3 shows the frontier lines estimated, together with the most 100 efficient farms for the three models. We have the in the highest positions the LME and the REM frontier lines, while the GG frontier is equivalent to the BC frontier. BC and OLS mean production function have a quite different slope parameter on tillable area. White points specify the most efficient farms in the LME and the GG models (the most 100 efficient farms are substantially the same for this two models), while in the dark bigger points there are the BC most efficient. Note that the application of the BC model on the aggregate data exclude a lot of farms at right in this plot.

4 Concluding remarks

The results indicate that the decomposition proposed of the model error can leads to a series of technical efficiency measures comparable with those form the standard stochastic frontier models for the aggregate data. The definition of “efficiency factors”, may have a more importance in the specification of the frontier production function than that the inputs. From a statistical point of view, the model discussed can give more information than other classic models, especially when we consider the technical efficiency matrix estimates and the internal technical efficiency parameters. Further the

technical efficiency estimates are consistent. From the standpoint of an economic analysis, the model permits to study production functions that can be very different among the groups, not only in terms of their own positions, but also in their general shapes. This fact, if it is recognized a number of highly significative random effects, highlights quite different production functions through the data. On the other hand, from an applicative point of view, the LME model can work also with technical inefficiency effects that are categorical in a nominal scale of measure, while the standard models like the Battese-Coelli (1995) or Huang-Liu (1994) can not. This is very useful in territorial economic analyses. Finally, we can estimate marginal and conditional efficiencies, respect to the aggregate data and to the specific group of firms, and, not least, the possibility of estimating the effects on the individual technical efficiency rates and on the frontier production function, caused by the different levels of internal technical inefficiency.

References

1. Battese, G.E., Coelli, T.J.: Frontier production functions, technical efficiency and panel data with application to paddy farmers in India, *Journal of Productivity Analysis*, 3, 153-169, (1992)
2. Battese, G.E., Coelli, T.J.: A model for technical inefficiency effects in a stochastic frontier production function for panel data, *Empirical Economics*, 20, 325-332, (1995)
3. Bauer, P.W., Hancock, D. : The efficiency of the Federal Reserve in providing check processing services, *Journal of Banking & Finance*, 17, 287-311, (1993)
4. Caudill, S.B., Ford, J.M.: Biases in frontier estimation due to heteroscedasticity, *Economic Letters*, 41, 17-20, (1993)
5. Caudill, S.B., Ford, J.M., Gropper, D.M.: Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity, *Journal of Business and Economic Statistics*, 13, 105-111, (1995)
6. Demidenko, E.: *Mixed Models: Theory and Applications*, Wiley, New York, (2004)
7. Greene, W.: Fixed and random effects in stochastic frontier models, *Journal of Productivity Analysis*, 23, 7-32, (2005a)
8. Greene, W.: Reconsidering heterogeneity in panel data estimators of the stochastic frontier model, *Journal of Econometrics*, 126, 269-303, (2005b)
9. Gori, E., Grassetti, L., Rossi, C.: Linear mixed models in efficiency analysis: evidence from validation procedures, *Italian Journal of Applied Statistics*, 14, 247-275, (2002)
10. Grassetti, L., Gori, E., Minotti, S.C.: Multilevel flexible specification of the production function in health economics, *Journal of Management Mathematics*, 16, 383-398, (2005)
11. Huang, C.J., Liu, J.T.: Estimation of a non-neutral stochastic production function, *Journal of Productivity Analysis*, 5, 171-180, (1994)
12. Jiang, J.: *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York, (2007)
13. Kumbhakar, S.C., Ghosh, S., McGuckin, J.T.: A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms, *Journal of Business and Economic Statistics*, 9, 279-286, (1991)
14. Kumbhakar, S.C., Lovell, C.A.K.: *Stochastic Frontier Analysis*, Cambridge University Press, New York, (2000)
15. Reifschneider, D., Stevenson, R.: Systematic departures from the frontier: a framework for the analysis of firm inefficiency, *International Economic Review*, 32, 715-723, (1991)
16. Simar, L., Lovell, C.A.K., van den Eeckaut, P.: Stochastic frontiers incorporating exogenous influences on efficiency, Discussion Paper n. 9403, Institut de Statistique, Université Catholique de Louvain, Louvain-la Neuve, Belgium, (1994)