Panel Stochastic Frontier Analysis with Dependent Error Terms

Rachida El Mehdi and Christian M. Hafner[®]

Receiving Time: May 21, 2021 Acceptance Time: October 21, 2021

ABSTRACT

In presence of panel data, technical efficiency is used to compare the performances of Decision-Making Units (DMUs). The novelty of this paper is the consideration of the dependence between the two error terms in the case of panel data and the introduction of time effect models in the Stochastic Frontier Analysis (SFA). Hence, our SFA model considers the balanced panel case, several models describing the evolution of the inefficiency over time and the dependence between the two error terms. The inefficiency and noise terms being dependent, a copula function which reflects the dependence between them is included in their joint density. The model is estimated by maximum likelihood and the Akaike Information Criterion (AIC) is used for model selection. Moreover, a likelihood ratio test is performed for the nested models. A bootstrap algorithm is proposed for statistical inference on the Technical Efficiency (TE) measures. Results for Moroccan policy of the production and sales of drinking water from 2001 to 2007 identify the most and least efficient provinces, and a generally positive trend of estimated TE measures.

Key words: *Bootstrap, Copulas, Efficiency, Panel data, Stochastic frontier analysis*
JEL Classifications : C13, C18, D24

1. INTRODUCTION

When panel data are available, it is recommended to use the structure of the data to estimate technical efficiencies in a Stochastic Frontier Analysis (SFA) because a panel contains more information than a single cross section. Furthermore, as noted in Schmidt and Sickles (1984) and Kumbhakar and Lovell (2000) some strong distributional assumptions used in the cross-sectional data case can be relaxed with the panel data and the technical efficiency can be estimated consistently when T, the number of time observations for each Decision-Making Unit (DMU), is large. Hence, repeated observations can be considered as a substitute for some strong distributional assumptions. They can also constitute a weakening of the independence assumption between the technical inefficiency term and the regressors.

An overview of the research on panel SFA models reveals that Jondrow et al. (1982) generalized the cross-sectional model to the panel data model and used the conditional

[®] Rachida El Mehdi, SmartICT Lab, National School of Applied Sciences, Mohammed First University, (email: r.elmehdi@ump.ac.ma), Tel: +212666146503
Christian M. Hafner, Louvain Institute of Data Analysis and Modelling in Economics and Statistics, and ISBA, Universit´e catholique de Louvain, (email: christian.hafner@uclouvain.be)

expectation of the inefficiency term u given the realized value of the error ϵ to estimate efficiency. Schmidt and Sickles (1984) and Kumbhakar and Lovell (2000) have proposed for the balanced panel case, models where they supposed that technical efficiency varies across producers but is either constant or varies through time for each producer. Battese et.al. (2000) adopted an unbalanced panel to investigate efficiency of labour in the Swedish banking industry. Kim and Lee (2006) assumed a time varying pattern of technical efficiency movements to analyze the productivity growth of several East Asian countries over a period of twenty years. However, all studies handled the panel SFA model with independence between noise and inefficiency terms. Recently, Smith (2008) handled the panel data model with dependent error components using a simulated example but without making inference on the estimated efficiency. Allowing for dependence is generally a desirable feature, because the efficiency of a given DMU at a given period of time might depend on whether or not the DMU was 'lucky', expressed by the random noise term. For example, if the DMU was unlucky in a particular period, it might attempt to compensate this by an increased efficiency, generating a dependence between noise and efficiency.

Furthermore, several studies have assessed the performance of water services such as Faria et al. (2005) and Tupper and Resende (2004), which compare the technical efficiency of Brazilian public and private companies in water supply; Sampaio et.al. (2005) which deals with the cost efficiency of the public water service in Portugal, and Vishwakarma and Kulshrestha (2010) which analyses the water supply utility of urban cities in India using the stochastic production frontier analysis. However, most of these studies use cross-sectional data. In the absence of such study in the water domain based on Moroccan data, the performance of the entities responsible of the water management in all regions will be measured by estimating the efficiencies in case of panel data and a nonparametric confidence interval will be proposed.

This research is an extension of El Mehdi and Hafner (2014b) to the panel data case. Our objective is to deal with the dependence of the error terms in the panel SFA approach using models for the time variation of efficiencies. We evaluate the efficiency and compare the DMU performances through an empirical data set on the water management in Morocco. Thus, in this work the production frontiers and panel data are considered to estimate technical efficiency when the two components of the error term are dependent. Efficiency being estimated, statistical inference is needed to draw reliable conclusions. Hence, this work presents also an associated procedure to build confidence intervals on the efficiencies in this considered case.

The remainder of the paper is organized as follows: Section 2 describes the model with the copula function, Section 3 presents the procedure of statistical inferences on the Technical Efficiency (TE) measure in the case of panel data with dependent error terms, and the last section presents results of an empirical analysis of the water area in all Moroccan regions with the numerical procedure estimation of technical efficiency in order to compare the $DMUs$. Finally we conclude by a summary of the results with some remarks and open issues.

2. EFFICIENCY MEASURES FOR PANEL DATA

The principle of the efficiency measure estimation for panel data is the same as that for cross-sectional data. We need however to make additional assumptions about the temporal pattern of inefficiency. There are also differences between the two procedures in terms of the simulated likelihood function definition.

When information about all is $DMUs$ available at T different time periods, it is preferable to use a stochastic frontier model which is adequate for panel data. In frontier analysis this model is more appropriate because even if it is not fundamentally different from the cross-sectional model, it has several advantages as it increases the degree of freedom to estimate parameters, provides consistent efficiency estimates when T is increasing and does not require that the inefficiencies are independent of the regressors.

2.1. The panel data production frontier model

The panel stochastic frontier model, when the inefficiencies are assumed to vary systematically with time, is specified as follows:

$$y_{it} = f\left(\underline{x}_{it}, \beta\right) + \epsilon_{it} = f\left(\underline{x}_{it}, \beta\right) + v_{it} - u_{it}\,,\ i = 1, 2, \dots, n\,;\ t = 1, 2, \dots, T \qquad (2.1)$$

where $y_{it} = log(Y_{it})$; Y_{it} : the observed output for DMU_i at the t^{th} time period (one output), so $Y_{it} \in \mathbb{R}_+$; $\underline{x}_{it} = log\left(\underline{X}_{it}\right)$; $\underline{X}_{it}$: a vector of length p that describes the observed inputs for observation i at time t, so $\underline{X}_{it} \in \mathbb{R}_+^p$ where p is the number of the inputs; β : a vector of unknown parameters to be estimated, $\beta \in \mathbb{R}^{l+(1\times T)}$ where l is the number of parameters excluding the time-varying intercepts. If intercepts are constant over time, then $\beta \in \mathbb{R}^{l+1}$. Moreover, ϵ_{it} is the error term for observation i at time t, $f\left(\underline{x}_{it}, \beta\right)$ the production frontier, n is the number of DMUs under study and T is the number of periods or the number of observations for each DMU.

The two components of the error term are motivated by the idea that deviations from the frontier might not be entirely under the control of the DMU and that the performance of a DMU is affected by these two components. Hence, the term ϵ_{it} is divided into two parts, the inefficiency term u_{it} which is constrained to be non-negative ($u_{it} \geq 0$) and the statistical noise term v_{it} which is usually a normal with zero as mean and σ_V as standard deviation ($v_{it} \sim N(0\,,\ \sigma_V^2)$).

Furthermore, distributional assumptions will be imposed on both terms u_{it} and v_{it}. In particular, it is assumed that the components of the first are independently and identically positively distributed and the components of the second are independently and identically normally distributed. In addition, both terms are assumed continuous and independent of x_{it}. At first, it is supposed that the two terms are mutually independent and the model is estimated by the Maximum Likelihood (ML) method. At a second stage, they will be allowed to be dependent and the ML estimates of the first stage will be considered as initial values in the numerical optimization.

As proposed in the literature on panel SFA, the frontier model considers either the time-constant or the time-varying efficiency (see Schmidt and Sickles (1984) and Kumbhakar and Lovell (2000)). In our study we consider the frontier model with time-varying efficiency which is, in our opinion, more realistic and reflects the inefficiency variability over time. Nevertheless, we shall limit our analysis to the fixed intercept over time and to a comparison between some time-varying models in order to select one of them.

2.2. Time-varying efficiency

When T is large, the assumption of a time-constant inefficiency is typically not appealing, as one would expect that inefficient DMUs are forced to improve over time. So, a time varying inefficiency is needed and a random-effects model should be used. To define a random-effects model, one has developed an extension of the fixed-effects model to a more general model to get consistent estimators of u_i when T is large. Among these we refer to Jondrow et al. (1982), which derived panel generalizations of the conditional inefficiency predictors, Battese and Coelli (1988) where the term u_i has a more general truncated-normal distribution, and Battese et.al. (1989) which extend the model to allow unbalanced data.

The frontier model is called a random-effects model when it is described by

$$y_{it} = \beta_{0t} + \sum_{j=1}^{l} \beta_j \, x_{ijt} + v_{it} - u_{it} \; = \; \beta_{it} + \sum_{j=1}^{l} \beta_j \, x_{ijt} + v_{it} \qquad (2.2)$$

where $\beta_{it} = \beta_{0t} - u_{it}$ is the intercept for DMU$_i$ in the time period t and where β_{0t} is the intercept common to all DMUs in the time period t, see e.g. Kumbhakar (1990) and Kumbhakar and Lovel (2000). Of course, $n \times T$ parameters β_{it} should be estimated but Cornwell et.al. (1990) reduce this number to $3 \times n$. In the same way, Kumbhakar (1990) suggested a model in which the u_{it} are specified by the expression (2.5). He suggested estimating the model with the maximum likelihood method but does not provide an empirical application. Battese and Coelli (1992) suggested a time-varying model for unbalanced panel data with the exponential function of time for u_{it} in (2.3) bellow. They also proposed in their later work, Battese and Coelli (1995), a model where u_{it} follows a normal distribution truncated at zero. The ML estimation and the efficiency calculations of these cases have been included in the *FRONTIER* programs implemented by Coelli (1996).

Schmidt and Sickles (1984) suggested not to specify an implicit distribution for the inefficiency when the panel data are available and to estimate the fixed- effects model with the traditional panel data methods. In extension of this approach, Cornwell et.al. (1990) and Lee and Schmidt (1993) have developed an approach in which they introduce the variation of the effect of inefficiencies over time. Both of the latter approaches propose a variation of inefficiencies more flexible than proposed in (2.3) and (2.5).

In this research, the Kumbhakar (1990) and the Cornwell et.al. (1990) random-effects models will not be considered given the large number of parameters to be estimated, and so just one fixed intercept will be estimated. The Kumbhakar (1990) time effect expressed here by the formula (2.5), Battese and Coelli (1992) and a variety of time effects models are considered with a function $\eta(t) \geq 0$ that describes the evolution of inefficiency over time such that $u_{it} = \eta(t)u_i$. These models are denoted as

$$M_1 : u_{it} = [exp\{-\eta_1(t-T)\}]\, u_i \quad , \; u_i \sim N^+(0\,,\sigma_U^2)\,; \tag{2.3}$$

$$M_2 : u_{it} = [exp\{-\eta_1(t-T)\}]\, u_i \;; \tag{2.4}$$

$$M_3 : u_{it} = [1 + exp\{\eta_1 t + \eta_2 t^2\}]^{-1}\, u_i \;; \tag{2.5}$$

$$M_4\text{: } u_{it} = [1 + \eta_1(t-T)sin(t-T)]\, u_i \;; \tag{2.6}$$

$$M_5\text{: } u_{it} = [1 + \eta_1 sin(\eta_2 t)]\, u_i \;; \tag{2.7}$$

$$M_6\text{: } u_{it} = \left[1 + \eta_1 sin\big(\eta_2(t-T)\big)\right] u_i \;; \tag{2.8}$$

$$M_7\text{: } u_{it} = \left[1 + \eta_1(t-T)sin\big(\eta_2(t-T)\big)\right] u_i \;; \tag{2.9}$$

$$M_8\text{: } u_{it} = \left[\eta_0 + \eta_1 t + \tfrac{1}{2}\eta_2 t^2 + 2\textstyle\sum_{h=1}^{H}\big(a_h sin(ht) - b_h cos(ht)\big)\right] u_i; \tag{2.10}$$

where, for all models and except for M_1, $u_i \sim \mathrm{N}^+(\mu\,,\sigma_U^2)$ and μ is the mean of the original normal distribution. That indicates that the inefficiency term u_i is a normal truncated at zero with mean μ. Furthermore, M_1 and M_2 are the Battese and Coelli (1992, 1995) models and M_3 is the Kumbhakar (1990) model. The proposed $M_4 - M_7$ models include sinusoidal functions to allow for possible periodicity effects in the inefficiency. For example, M_7 models a time-varying amplitude of the sine function depending on the parameter η_1. The last considered model M_8 is the Fourier Flexible Form of Gallant (1984) which can closely approximate any smooth function $\eta(t)$ for sufficiently large H. In our study the Akaike Information Criterion (AIC) is used to select a model among M_1 to M_8. Moreover, a likelihood ratio test is performed for the nested models.

2.3. Model estimation

Considering the models described by (2.2) and (2.3)-(2.10), the methods used to estimate all models depend on the distributional assumptions. When v_{it} is i.i.d. normal, u_{it} is i.i.d. with positive support, v_{it} and u_{it} are mutually independent and independent from the regressors, the Maximum Likelihood Estimation (MLE) is feasible. Schmidt and Sickles (1984) conjectures that given suitable regularity conditions the ML estimates of (2.2) are consistent and asymptotically efficient as $n \to \infty$ regardless of T. In the particular, when $v_{it} \sim iid\ N(0\,,\ \sigma_V^2)$ and $u_{it} \sim iidN^+(0\,,\sigma_U^2)$, the MLE leads for $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{iT})'$, to the log-likelihood function, ignoring an additive constant,

$$l = \ln(L) = -\frac{n}{2} ln\, \sigma_*^2 - \frac{1}{2}\sum_{i=1}^{n} a_{*i} - \frac{n.T}{2}\ln \sigma_V^2 - \frac{n}{2} ln\, \sigma_U^2$$
$$+ \textstyle\sum_{i=1}^{n} ln\left[1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right], \tag{2.11}$$

which leads to the technical efficiency (TE) estimate for all $i = 1, \dots, n$

$$TE_{it} = E(exp\{-u_{it}\}|\epsilon_i)\ ,$$
$$= \frac{1-\Phi\left(\eta(t)\sigma_* - \frac{\mu_{*i}}{\sigma_*}\right)}{1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)} exp\left\{-\eta(t)\mu_{*i} + \tfrac{1}{2}\eta^2(t)\sigma_*^2\right\}, \tag{2.12}$$

where L is the likelihood function, $\eta(t)$ is the time function, $\sigma_*^2 = \frac{\sigma_V^2\sigma_U^2}{\sigma_V^2+\sigma_U^2\sum_t \eta^2(t)}$, $\mu_{*i} = \frac{(\sum_t \eta(t)\epsilon_{it})\sigma_V^2}{\sigma_V^2+\sigma_U^2\sum_t \eta^2(t)}$, $a_{*i} = \frac{1}{\sigma_V^2}\left[\sum_t \epsilon_{it}^2 - \frac{\sigma_U^2(\sum_t \eta(t)\epsilon_{it})^2}{\sigma_V^2+\sigma_U^2\sum_t \eta^2(t)}\right]$ and Φ is the standard normal cdf.

In comparison with the cross-section data, calculation of the log-likelihood function for panel data is similar to El Mehdi and Hafner (2014b), but it changes at the level of computing the ϵ_i density which is $g(\epsilon_i)$ where $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{iT})'$. In the expression of $g(\epsilon_i)$, the joint density $f(u_i, v_i)$ of u_i and v_i is replaced by $f_1(u_i)\, f_2(v_i) = f_1(u_i) \prod_t f_2(\epsilon_{it} + \eta(t)u_i)$. Of course, this last expression is integrated by u_i to get $g(\epsilon_i)$.

When u_i and v_i are dependent, their joint density when panel data is available becomes

$$f_1(u_i)f_2(v_i)\, c_\theta\big(F_1(u_i), F_2(v_i)\big) =$$
$$f_1(u_i) \prod_t f_2(\epsilon_{it} + \eta(t)u_i) \prod_t c_\theta\big(F_1(u_i), F_2(\epsilon_{it} + \eta(t)u_i)\big) \qquad (2.13)$$

where c is a bivariate copula density which expresses the dependence between the two variables u_i and v_i, and $F_1(u_i)$ and $F_2(v_i)$ are two uniform variables which are the cdf of $f_1(u_i)$ and $f_2(v_i)$ respectively and called the margins. The independence case is a special case of this model when the copula is the product copula, for which $c(.,.) = 1$. But for general copula functions, the ML estimation will become more complicated.

Given that the v_{it} are supposed independent and identically distributed, the density of ϵ_i becomes

$$g(\epsilon_i) = \int_0^{+\infty} f(\epsilon_i, u_i)\, du_i = \int_0^{+\infty} f_1(u_i) \prod_t A_{it}\, du_i = E\left(\prod_t A_{it}\right), \qquad (2.14)$$

where $\mathrm{A_{it}} = \mathrm{f_2(\epsilon_{it} + \eta(t)u_i)\; c_\theta\big(F_1(u_i), F_2(\epsilon_{it} + \eta(t)u_i)\big)}$. See the Appendix A for more details. Therefore, assuming the independence across DMUs, the log-likelihood function can be written as

$$l(\vartheta) = \log L(\vartheta) = \log L\left(\sigma_U, \sigma_V, \theta, \beta_0, \beta, \underline{\eta}_k\right)$$
$$= \sum_{i=1}^n \log g_\vartheta(\epsilon_i) = \sum_{i=1}^n \log g_\vartheta\left(y_i - \left(\beta_0 + \sum_{j=1}^l \beta_j x_{ij}\right)\right), \qquad (2.15)$$

where $\beta_0 = (\beta_{01}, \dots, \beta_{0t}, \dots, \beta_{0T})'$, $\beta = \left(\beta_1, \dots, \beta_j, \dots, \beta_l\right)'$ are vectors with a length equal respectively to the time periods T and the number of inputs l, $\underline{\eta}_k$ is a vector of k parameters in the time-varying function and where $x_{ij} = \left(x_{ij1}, \dots, x_{ijt}, \dots, x_{ijT}\right)'$ and $y_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$. For simplicity, all intercepts β_{0t}, $t = 1, \dots, T$ are considered the same and denoted by β_0 in the empirical analysis.

Generally, the expression of the function $l(\vartheta)$ is complex in the dependence case and to obtain analytical derivatives becomes a tedious or even impossible task in several cases. So, the log-likelihood is optimized numerically using the *mle* function in the R software and using the simplex numerical method called the Nelder-Mead method.

Once the parameters $\left(\sigma_U, \sigma_V, \theta, \beta_0, \beta, \underline{\eta}_k\right)$ are estimated, the technical efficiency can be estimated using the expected value of $(exp\{-u_{it}\}|\epsilon_i)$ as

$$TE_{it} = E[exp\{-u_{it}\}|\epsilon_i] \ = \ E(exp\{-\eta(t)u_i\}\textstyle\prod_t A_{it}\ ) \ / \ E(\prod_t A_{it}\ ) \ . \qquad (2.16)$$

See the Appendix A for more details. Given again the complexity of the TE_{it} expression, the expectation will be estimated for a large number m of Monte Carlo draws by

$$\widehat{TE}_{it} \cong \left[\tfrac{1}{m}\textstyle\sum_{j=1}^{m}\left(exp\{-\eta(t)u_j\}\prod_t A_{ijt}\ \right)\right] / \left[\tfrac{1}{m}\sum_{j=1}^{m}\left(\prod_t A_{ijt}\ \right)\right]. \qquad (2.17)$$

The following section develops an algorithm based on the bootstrap to construct confidence intervals for the estimated technical efficiencies.

3. INFERENCE FOR THE TECHNICAL EFFICIENCY MEASURE

Since the technical efficiencies of each DMU at each time t are unknown and estimated by $\widehat{TE}_{it}$, an inference about them is required. To build the confidence interval at a level α, given that the true sampling distribution is not available, we see that a modified algorithm of the parametric bootstrap Algorithm#3 of Simar and Wilson (2010) adapted to the dependence case and to the panel framework is more appropriate. Hence, we developed a procedure to estimate the associated confidence bounds when v is normal and u is half-normal which can be generalized for any positive distribution of u such as the truncated-normal. The method is easy to apply but it is quite computationally intensive. The steps are the following:

1. Estimate $\vartheta = \left(\sigma_U, \sigma_V, \theta, \beta_0, \beta, \underline{\eta}_k\right)$ according to (2.15), using the observed $\left(\underline{x}_{it}\ , y_{it}\right)$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ and using a numerical optimization procedure to get $\hat{\vartheta} = \left(\hat{\sigma}_U, \hat{\sigma}_V, \hat{\theta}, \hat{\beta}_0, \hat{\beta}, \underline{\hat{\eta}}_k\right)$ and to compute the point estimates $\widehat{TE}$ as described before

2. For $i = 1, 2, \ldots, n$, draw $u_i^* \sim N^+(0\ , \hat{\sigma}_U^2)$ and $v_{it}^* \sim N(0\ , \hat{\sigma}_V^2)$, $t = 1, 2, \ldots, T$ such that u_i^* and v_{it}^* are dependent with dependence characterized by the Clayton copula. Then compute $y_{it}^* = \hat{\beta}_0 + \sum_{j=1}^{l}\hat{\beta}_j x_{ijt} + v_{it}^* - \hat{\eta}(t)u_i^*$.
   There are several procedures to generate the pair (u_i^*, v_{it}^*) according to the Clayton copula, we use the one described in Nelsen (1999), page 41. The four steps of this procedure are

   a. Draw $T + 1$ independent uniform random variables $w_{1i},\ h_{2i1}\ , \ldots, h_{2it}, \ldots, h_{2iT}$, such that $w_{1i} \sim U(0, 1)$ and $h_{2it} \sim U(0, 1)$ for $t = 1, 2, \ldots, T$.

   b. Set $w_{2it} = \left[w_{1i}^{-\hat{\theta}}\left(h_{2it}^{-\hat{\theta}/(1+\hat{\theta})} - 1\right) + 1\right]^{-1/\hat{\theta}}$ for all $t = 1, 2, \ldots, T$.

   c. Set $u_i^* = F_1^{-1}(w_{1i})$ and $v_{it}^* = F_2^{-1}(w_{2it})$ for all $t = 1, 2, \ldots, T$ and where F_1 and F_2 are the cdf of the $N^+(0\ , \hat{\sigma}_U^2)$ and $N(0\ , \hat{\sigma}_V^2)$ respectively.

d. Repeat steps a to c n times to generate $n \times T$ pairs (u_i^*, v_{it}^*).

3. Using the pseudo-data $\mathcal{P}_{b,n}^* = \{(\underline{x}_{it}, y_{it}^*)\}_{i=1}^n$, compute bootstrap estimates $\hat{\vartheta}_b^* = argMax_{\vartheta\in\Theta}\, l(\vartheta|\mathcal{P}_{b,n}^*)$ after replacing y_{it} by y_{it}^* in (2.15) and then compute the bootstrap estimates $\widehat{TE}_b^*$ using (A.4) after replacing ϵ by $\epsilon_b^* = y - \hat{\beta}_0^* - \hat{\beta}^*.x$, where $\underline{x}_{it}$ and y_{it} represent the observed data.

4. Repeat steps 2 and 3, B times to obtain estimates $\mathcal{B}^* = \{\hat{\vartheta}_b^*\}_{b=1}^B$. Therefore, use $\mathcal{B}^*$ to get $\xi^* = \{\widehat{TE}_b^*\}_{b=1}^B$. Each individual i is described by a sub-matrix of ξ^* denoted ξ_i^*, it has T rows and B columns.

   For each individual i at time period t (row t of the ξ_i^* matrix, denoted ξ_{it}^*, $i = 1, 2, \dots, n$, compute the $(\alpha/2)$ and the $(1 - \alpha/2)$ quantiles for ξ_{it}^* by considering its B components. The $100 \times (1 - \alpha)$ percentile bootstrap confidence interval of the statistic of interest TE is obtained by the probability $P\big((\xi_{it}^*)_{\alpha/2} < TE_{it} < (\xi_{it}^*)_{1-\alpha/2}\big) = 1 - \alpha$.

   Hence using the $100 \times \left(\frac{\alpha}{2}\right)$ and $100 \times \left(1 - \frac{\alpha}{2}\right)$ percentiles, we define the lower and the upper bounds of the confidence interval as $TE_{it} \in \big[(\xi_{it}^*)_{\alpha/2}, (\xi_{it}^*)_{1-\alpha/2}\big]$.

We note that the estimation procedure presented in Section 2.3 leads sometimes to a positive skewness of the composite error term which consequently leads to biased parameter estimates and to biased technical efficiencies estimates because all of these latter will be close to one. If this is the case, the procedure presented in this section allows us to overcome this problem.

To perform our procedure, a simulation example is proposed. The model describing data is supposed to be log-linear where there are one input and one output such that for all $i = 1, 2, \dots, n$ and for all $t = 1, 2, \dots, T$, we have $log(Y_{it}) = \beta_0 + \beta_1 log\big(10(1 + X_{it})\big)$ where $X_{it} \sim U(0, 1)$ and parameters will be set to $\beta_0 = log(10)$ and $\beta_1 = 1$. As for the noise term and the inefficiency term, they are supposed to be normal as usual for the first such that $v_{it} \sim N(0, \sigma_V^2)$ with $\sigma_V = 0.5$ and half-normal for the second such that $u_{it} = \eta(t)u_i$ and $u_i \sim N^+(0, \sigma_U^2)$ with $\sigma_U = 1$ and the two components are dependent using the Clayton copula with dependence parameter $\theta = 1$. The time varying function is supposed to be $\eta(t) = exp\{-\eta_1(t - T)\}$ with $\eta_1 = -0.1$. We suppose that $n = 50, T = 10$ and $B = 500$. To compute the true efficiencies, the number of simulations to approximate numerically the integral is set to $m = 10000$ which is large enough to have a good approximation of the expectation in equation (A.4) evaluated at the true values.

The bootstrap procedure shows that all estimated efficiencies are covered by their confidence intervals. The percentage for the true efficiencies is evaluated at 87%, 91.2% and 100% for a significance level of 10%, 5% and 1% respectively, so that the bootstrap coverage ratio is reasonably close to the nominal level given our moderate number of bootstrap replications.

4. EMPIRICAL ANALYSIS OF MOROCCAN DRINKING WATER SUPPLY

Moroccan data was subject of the frontier analysis as in El Mehdi and Hafner (2014a, 2014b). The data set analyzed in this study and chosen to illustrate our methodology contains information on the water production and its sales to the subscribers and to the municipal utilities called the self-governance in Morocco. The national public company called the National Office of the Drinking Water (ONEP according to its French abbreviation) ensures the largest part of the production, the pipe and the distribution of water in the entire national territory. It produces more than 80 percent of the country's drinking water. This sector depends mainly on the domestic consumption. So, to strengthen water resources and to rationalize its use, starting in the 1980s it led to certain administrative and technical actions such as an information campaign and the installation of individual water meters in households, in order for example to reduce wasting.

We are interested in this practical case in the efficiency of certain national participants in the management of this particular and vital good. To do so, the Farrel technical efficiency rate will be estimated using a panel data set in order to compare the performance of certain Moroccan provinces with respect to their produced quantities in the sector and to their water sales. We shall analyze, thus, the degree of efficiency of every producing entity of water in order to situate it among the others at the national level and we shall know how much should be its sales to attain efficiency. Efficiency being an estimate, we also provide confidence intervals.

The considered variables in this study are the ONEP's sales and the number of the subscribers as inputs and the water production as output. Both of sales and production are evaluated in thousand cubic meter (1000 m^3). Therefore, the model will be a simple model with two independent variables and the frontier function chosen to describe the production technology is the translog function. As for the copula function, the Archimedean Clayton copula is used because it is popular in empirical applications, it is flexible and easy to construct, and it nests the independence copula as a special case, see for example Bhat and Eluru (2009).

Hence, the data represent sales, the number of subscribers and the water production for a set of 50 provinces through 15 Moroccan regions for a duration of seven years from 2001 until 2007. Six provinces among a total of 56 were omitted because of lack or unavailability of data as indicated in the statistical yearbooks of the corresponding years published annually by the Statistics Direction and which have as source the ONEP entity in this area. These six provinces are Tan Tan, Inezgane-Ait-Melloul, Sidi Youssef Ben Ali and Al Ismailia in respectively Guelmim-Es-Semara, Souss-Massa-Daraâ, Marrakech-Tensift-AlHaouz and Meknès-Tafilalet regions and all provinces of Grand-Casablanca region which are Casablanca and Mohammedia.

Being complex, the optimization of the log-likelihood function is performed using numerical optimization in three steps:

Step 1. The model (2.1) is fitted using the pooled-OLS regression without the technical inefficiency term. Hence, the inefficiency is null ($u_{it} = 0$) and the time effect is zero.

Step 2. The OLS parameters of step 1, except the intercept β_0 which is biased, are used as initial values in this step to estimate numerically the model (2.1) using the maximum likelihood estimation (MLE) assuming independence between u and v. The β_0 parameter is adjusted by shifting it according to the Corrected Ordinary Least Squares procedure (COLS) used in the R frontier package, see e.g. Coelli (1995).

Step 3. In this last step, MLE numerical optimization is performed using the step 2 estimates as initial values. As for the initial value of the copula parameter θ, a grid of values for θ is given. Given that the models are the same (the difference is just the value of θ), the comparison of the log-likelihood function is done directly (without estimation) and hence the value which gives the highest log-likelihood value is chosen as initial value of θ in this step.

About the model, the full translog function with two exogenous variables is considered in this analysis. Inputs are the number of subscribers and the total of sales. So, we will have the following number of parameters: one for the intercept, two for the variables, two for their terms squared and one for their interaction. The others are σ_U, σ_V, θ and $\underline{\eta}_k$ and μ is added when the truncated-normal distribution is considered. The estimation of the model with the full translog function in the case of the independence, using the frontier package of the R software, has revealed that the number of subscribers variable, its squared term and the interaction term are not significant and consequently we do not reject that their coefficients are equal to zero. For this reason, only the sales and its square term will be included in the model. Hence, the total number of parameters included in the final considered model in the dependence case is at least seven parameters.

Furthermore, the full translog model is not used because the minimal *AIC* criterion which is $-2logLik + 2k$, where k is the number of parameters to be estimated in the model, is smaller for the restricted function evaluated at -78.0919 in comparison with the full function evaluated at -75.3816; and being nested, the likelihood ratio test rejects the full model in favor of the restricted one at the 5% level of significance. Estimates are used as initial values in the numerical optimization in the case where v_{it} is normal and u_{it} is half-normal or truncated-normal when dependence between them is considered.

According to the expressions of the time-varying function and to the inefficiency distribution, Panel Stochastic Frontier models described by (2.1) and (2.3) - (2.10) are estimated and the model M_7 which has the time-varying expression (2.9) is selected according the minimum *AIC* criterion with a log-likelihood value evaluated at 595.041 as pointed out in Table 4.1. In addition, M_4 and M_7 being nested, the latter one was not rejected according to the likelihood ratio test.

First of all, it is noted that for this chosen model the time effect is significant, μ is not equal to zero and the two terms of inefficiency u and noise v are dependent. The parameter θ is not close to zero and hence the Clayton copula does not approach the Product copula related to the independence of the two error terms. All other parameters are statistically significant in the model as pointed out in Table 4.2.

Model	*M*1	*M*2	*M*3	*M*4	*M*5	*M*6	*M*7	*M*8
logL	413.384	418.518	400.134	502.000	433.728	492.155	595.041	165.156
df	7	8	9	8	9	9	9	14
AIC	-812.768	-821.036	-782.268	-988.000	-849.456	-966.310	-1172.082	-302.312

Table 4.1 The AIC values of the estimated models

	Estimate	Std. Error	*t* value	$Pr(\|T\| > \|t\|)$
σ_U	3.2313	0.000400	8084.890	< 1e-16
σ_V	0.0539	0.000215	250.615	< 1e-16
$\beta 0$	3.7094	0.001028	3607.835	< 1e-16
$\beta 1$	1.6721	0.000161	10413.472	< 1e-16
$\beta 2$	-0.0858	0.000082	-1048.721	< 1e-16
$\eta 1$	0.0425	0.000668	63.676	< 1e-16
$\eta 2$	0.2276	0.000625	364.036	< 1e-16
θ	2.0389	0.000352	5789.226	< 1e-16
μ	0.0179	0.000607	29.452	< 1e-16
-2logL	-1190.083			

Table 4.2 The model estimation with correlated error terms

As for the efficiency scores, which are one of our major objectives, Table 4.3 presents the estimation results of the model (2.1) under assumptions (2.9) denoted M_7 and where v_{it} and u_{it} are correlated by the Clayton copula. It mainly shows that all provinces are technically inefficient. The most efficient are Rabat-Skhirate-Témara and El Jadida in respectively Rabat-Salé-Zemmour-Zaer and Doukkala-Abda regions with a mean score for the period greater than 0.65 and the most inefficient one is Ben Slimane province in Chaouia-Ouardigha region. Rabat-Skhirate-Témara is near the frontier and should on average increase its sales by just about 0.3% to be efficient. Moreover, Table 4.4 which summarizes the previous one shows also that only 20% of the provinces exceed the national efficiency mean for the entire period evaluated at 0.102 which is a very weak score. Even if the average is low, it has progressed but slowly from year to the next one which may indicate that the ONEP policy in the production and selling of water was not adequate during the seven studied years. Moreover, the efficiency standard deviation is large because it is estimated at 0.176. Hence, with the exception of one or two provinces, scores are very low and the dispersion is high which reflect the mediocre performance of the sector with respect to the relation between the ONEP's drinking water production and its sales.

It is clear also that the time effect on the efficiency scores is positive given that $\hat{\eta}_1$ and $\hat{\eta}_2$ are positive and the TE increases over the period under study as specified previously. Even if this effect is weak, it is statistically highly significant with a p-value less than 1e-16.

Nr	DMUs Name	$\widehat{TE}i1$	$\widehat{TE}i2$	$\widehat{TE}i3$	$\widehat{TE}i4$	$\widehat{TE}i5$	$\widehat{TE}i6$	$\widehat{TE}i7$	$\overline{\widehat{TE}_\iota}$	$\overline{\widehat{bias}_\iota^*}$	$\overline{\hat{\sigma}_\iota^*}$	$\overline{\widehat{TE\_cor}_\iota}$	$\overline{Lower}$	$\overline{Upper}$
1	Oued Ed-Dahab	0.0143	0.0173	0.0211	0.0254	0.0294	0.0323	0.0333	0.0247	-0.0019	0.0118	0.0264	0.0084	0.0506
2	Boujdour	0.0057	0.0072	0.0092	0.0115	0.0137	0.0154	0.0160	0.0112	-0.0003	0.0039	0.0115	0.0051	0.0195
3	Laâyoune	0.0293	0.0344	0.0406	0.0473	0.0534	0.0578	0.0594	0.0460	-0.0054	0.0275	0.0517	0.0128	0.1105
4	Assa-Zag	0.0057	0.0072	0.0092	0.0114	0.0137	0.0153	0.0160	0.0112	0.0006	0.0040	0.0108	0.0057	0.0203
5	Es-Semara	0.0086	0.0107	0.0134	0.0164	0.0193	0.0215	0.0223	0.0160	0.0001	0.0071	0.0160	0.0067	0.0326
6	Guelmim	0.0297	0.0348	0.0411	0.0478	0.0540	0.0583	0.0599	0.0465	-0.0045	0.0253	0.0508	0.0143	0.1038
7	Tata	0.0096	0.0119	0.0148	0.0181	0.0212	0.0235	0.0244	0.0176	-0.0005	0.0073	0.0183	0.0072	0.0337
8	Agadir-Ida ou Tanane	0.1623	0.1762	0.1920	0.2076	0.2210	0.2302	0.2334	0.2032	-0.0280	0.2269	0.2032	0.0364	0.8463
9	Chtouka-Ait Baha	0.0097	0.0119	0.0148	0.0181	0.0212	0.0236	0.0244	0.0177	0.0001	0.0080	0.0176	0.0072	0.0361
10	Ouarzazate	0.0301	0.0352	0.0415	0.0483	0.0545	0.0589	0.0606	0.0470	-0.0021	0.0314	0.0488	0.0142	0.1251
11	Taroudannt	0.0271	0.0319	0.0378	0.0441	0.0500	0.0541	0.0557	0.0429	-0.0028	0.0269	0.0461	0.0127	0.1086
12	Tiznit	0.0179	0.0215	0.0259	0.0308	0.0354	0.0388	0.0400	0.0300	0.0001	0.0186	0.0299	0.0101	0.0762
13	Zagora	0.0145	0.0176	0.0214	0.0257	0.0298	0.0327	0.0338	0.0251	-0.0010	0.0120	0.0262	0.0092	0.0518
14	Kenitra	0.1200	0.1322	0.1460	0.1599	0.1721	0.1804	0.1833	0.1563	-0.0160	0.1420	0.1614	0.0355	0.5265
15	Sidi Kacem	0.0671	0.0759	0.0861	0.0967	0.1062	0.1128	0.1151	0.0943	-0.0087	0.0657	0.1007	0.0246	0.2559
16	Ben Slimane	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0025
17	Khouribga	0.0199	0.0238	0.0286	0.0339	0.0387	0.0423	0.0436	0.0330	-0.0045	0.0238	0.0375	0.0078	0.0920
18	Settat	0.0057	0.0072	0.0092	0.0115	0.0137	0.0154	0.0160	0.0113	-0.0012	0.0094	0.0125	0.0025	0.0355
19	Al Haouz	0.0111	0.0137	0.0169	0.0205	0.0239	0.0264	0.0274	0.0200	-0.0002	0.0091	0.0205	0.0078	0.0406
20	Chichaoua	0.0095	0.0118	0.0146	0.0179	0.0210	0.0233	0.0241	0.0175	-0.0002	0.0075	0.0176	0.0072	0.0343
21	El Kelaâ des Sraghna	0.0314	0.0368	0.0433	0.0502	0.0566	0.0611	0.0627	0.0489	-0.0034	0.0317	0.0525	0.0140	0.1264
22	Essaouira	0.0204	0.0243	0.0292	0.0346	0.0395	0.0431	0.0444	0.0336	-0.0019	0.0193	0.0361	0.0106	0.0791
23	Marrakech Ménara	0.3315	0.3485	0.3671	0.3849	0.3999	0.4098	0.4133	0.3793	-0.0585	0.3183	0.4022	0.0582	0.9839
24	Berkane - Taourirt	0.0820	0.0919	0.1033	0.1151	0.1255	0.1326	0.1352	0.1122	-0.0110	0.0949	0.1198	0.0262	0.3574
25	Figuig	0.0108	0.0133	0.0164	0.0199	0.0233	0.0258	0.0267	0.0195	-0.0021	0.0073	0.0216	0.0073	0.0336
26	Jerada	0.0135	0.0165	0.0202	0.0243	0.0281	0.0310	0.0320	0.0236	0.0003	0.0119	0.0234	0.0094	0.0516
27	Nador	0.0655	0.0741	0.0842	0.0947	0.1041	0.1106	0.1129	0.0923	-0.0095	0.0750	0.0989	0.0218	0.2844
28	Oujda	0.0551	0.0628	0.0720	0.0816	0.0901	0.0961	0.0983	0.0794	-0.0063	0.0708	0.0848	0.0184	0.2692
29	Khemisset	0.0355	0.0413	0.0483	0.0558	0.0626	0.0674	0.0692	0.0543	-0.0026	0.0409	0.0577	0.0148	0.1594
30	Rabat-Skhirate-Témara	0.9966	0.9968	0.9970	0.9971	0.9972	0.9973	0.9973	0.9970	-0.0419	0.3040	0.9970	0.1733	1.0000
31	El Jadida	0.6315	0.6448	0.6588	0.6720	0.6827	0.6898	0.6922	0.6674	-0.0966	0.3282	0.7533	0.1110	1.0000
32	Safi	0.0636	0.0721	0.0820	0.0924	0.1016	0.1080	0.1103	0.0900	-0.0077	0.0847	0.0931	0.0205	0.3187
33	Azilal	0.0215	0.0256	0.0307	0.0362	0.0413	0.0450	0.0463	0.0352	-0.0039	0.0170	0.0389	0.0113	0.0715
34	Beni Mellal	0.1289	0.1415	0.1558	0.1702	0.1826	0.1911	0.1942	0.1663	-0.0151	0.1671	0.1663	0.0362	0.6114
35	El Hajeb	0.0082	0.0103	0.0128	0.0158	0.0186	0.0207	0.0215	0.0154	-0.0008	0.0081	0.0163	0.0053	0.0341
36	Errachidia	0.0469	0.0539	0.0623	0.0710	0.0790	0.0845	0.0865	0.0692	-0.0069	0.0487	0.0762	0.0180	0.1906
37	Ifrane	0.0239	0.0283	0.0337	0.0396	0.0450	0.0489	0.0503	0.0385	-0.0026	0.0223	0.0415	0.0121	0.0909
38	Khénifra	0.0310	0.0362	0.0427	0.0496	0.0559	0.0604	0.0620	0.0482	-0.0045	0.0367	0.0535	0.0100	0.1385
39	Meknès El Menzeh	0.0452	0.0521	0.0602	0.0688	0.0765	0.0820	0.0840	0.0670	-0.0115	0.0495	0.0779	0.0000	0.1688
40	Boulmane	0.0109	0.0134	0.0165	0.0201	0.0234	0.0259	0.0268	0.0196	0.0001	0.0093	0.0191	0.0078	0.0412
41	Fès	0.3423	0.3594	0.3780	0.3958	0.4108	0.4207	0.4241	0.3902	-0.0587	0.3274	0.4020	0.0594	0.9844
42	Sefrou	0.0234	0.0277	0.0331	0.0389	0.0442	0.0481	0.0495	0.0378	-0.0014	0.0259	0.0392	0.0113	0.1032
43	Zouagha My Yacoub	0.0084	0.0104	0.0130	0.0160	0.0189	0.0210	0.0218	0.0156	-0.0005	0.0059	0.0161	0.0066	0.0281
44	Al Houceïma	0.0297	0.0349	0.0411	0.0479	0.0540	0.0584	0.0600	0.0466	-0.0056	0.0286	0.0523	0.0126	0.1144
45	Taounate	0.0304	0.0357	0.0420	0.0488	0.0551	0.0595	0.0611	0.0475	-0.0031	0.0252	0.0511	0.0156	0.1049
46	Taza	0.0279	0.0328	0.0388	0.0453	0.0512	0.0555	0.0570	0.0441	-0.0058	0.0245	0.0499	0.0123	0.0995
47	Chefchaouen	0.0199	0.0238	0.0286	0.0339	0.0388	0.043	0.0436	0.0330	-0.0021	0.0169	0.0349	0.0114	0.0713
48	Larache	0.0567	0.0646	0.0739	0.0836	0.0923	0.0984	0.1006	0.0815	-0.0112	0.0657	0.0909	0.0179	0.2481
49	Tanger	0.2548	0.2711	0.2891	0.3066	0.3215	0.3314	0.3349	0.3013	-0.0365	0.2979	0.3013	0.0501	0.9696
50	Tétouan	0.1522	0.1658	0.1811	0.1964	0.2096	0.2185	0.2217	0.1922	-0.0275	0.2205	0.1922	0.0354	0.8186

Where $\overline{\widehat{TE}_\iota}$, $\overline{\widehat{bias}_\iota^*}$, $\overline{\hat{\sigma}_\iota^*}$, $\overline{\widehat{TE\_cor}_\iota}$, $\overline{Lower}$ and $\overline{Upper}$ are the means according T of their corresponding expressions TE_cor : Bias corrected efficiency

Table 4.3 Technical Efficiency scores for the 2001-2007 period and their confidence intervals

Region Name	[0, 0.2[	[0.2, 0.4[	[0.4, 0.6[	[0.6, 0.8[	[0.8, 1[
Oued Ed-Dahab - Lagouira	1				
Laâyoune-Boujdour-S. EL Hamra	2				
Guelmim - Es-Semara	4				
Souss - Massa - daraâ	5	1			
Gharb - Chrarda - Béni Hssen	2				
Chaouia - Ouardigha	3				
Marrakech - Tensift - Al Haouz	4	1			
Oriental	5				
Rabat-Salé-Zemmour-Zaer	1				1
Doukala-Abda	1			1	
Tadla - Azilal	2				
Meknès - Tafilalet	5				
Fès - Boulemane	3	1			
Taza - Al Hoceïma - Taounate	3				
Tanger - Tétouan	3	1			

Table 4.4 Provinces number in each region according to the mean of TE estimates
(The table is the same according to the median of TE estimates)

Inference on the technical efficiency measure is made using a parametric percentile bootstrap procedure to estimate robust confidence intervals of the statistic of interest. Bootstrap samples are obtained according to the step 2 of the procedure in Section 3 and Table 4.3 presents technical efficiency estimates for each province and for the 2001-2007 period, their means, their corrected bias and the mean of the estimated lower and upper confidence interval bounds following the rest of the steps of the same bootstrap procedure. So, for each province, the mean of $\widehat{TE}_i$ is bounded by the mean of $(\xi_{it}^*)_{\alpha/2}$ and the mean of $(\xi_{it}^*)_{1-\alpha/2}$ for the seven years. Inference performed with $B = 500$ bootstrap replications shows that TE estimates for all provinces are in their corresponding confidence intervals at a 5% significant level but with a relatively large range for that which have great TE scores as depicted in Table 4.3. However, the global average width of the intervals is 0.22 with fifteen provinces (30% of all) having a width bigger than this average.

Technical efficiencies being estimates and in order to know if the bias correction is needed, their bias were estimated using the bootstrapped efficiencies as defined in Daraio and Simar (2007, p.55) but using the median instead of the mean given that the TE distribution is skewed (mean estimated at 0.1024 is greater than median estimated at 0.0423). Then, for each individual i at time period t the bias is expressed by $\widehat{bias}^*\left(\widehat{TE}_{it}\right) = median\left(\widehat{TE}_b^*\right)_{it} - \widehat{TE}_{it}$ and the bias corrected estimate of TE is defined as $\widehat{TE\_cor}_{it} = \widehat{TE}_{it} - \widehat{bias}^*\left(\widehat{TE}_{it}\right)$. Generally, the correction is needed if $\left|\widehat{bias}^*\left(\widehat{TE}_{it}\right)\right| / \hat{\sigma}_{it}^* > 0.25$ as pointed out in Efron (1982), where $\widehat{bias}^*\left(\widehat{TE}_{it}\right)$ is the estimated bias of the bootstrap estimates and $\hat{\sigma}_{it}^*$ is their standard error. Indeed, the bias was important for forty-four DMUs among fifty and the ratio reached on average its maximum with 1.03 points for the Assa-Zag province. The correction has reduced the average of the scores of the period for six DMUs and has increased this average for the thirty-eight others. So, the efficiency scores were overestimated for the first ones and underestimated for the last ones. Furthermore, the means of the corrected TE are well in their confidence intervals.

5. CONCLUSION

This paper presented and proposed, at first, the panel stochastic frontier analysis when the error terms are dependent and secondly an associated confidence interval procedure on efficiency with an empirical study on the drinking water in Morocco. The analysis was performed using some previous time effect functions in the literature and using some proposed sinusoidal time effect functions. Given the appeal of the numerical optimization due to the complexity of the log-likelihood functions in the presence of dependence and given the use of the parametric bootstrap in the construction of the confidence intervals, the computation was highly intense.

The study has demonstrated that the consideration of the dependence between the error components was strongly recommended given that the copula does not approach the product copula. It has revealed that the proposed time effect model which expresses that the amplitude of the time function decreases and that the time effect disappears over time is appropriate for our data according to the AIC criterion. The results revealed also a positive and significant time effect on the technical efficiency scores. It showed that the bias was important for several entities and that the efficiency scores were underestimated for thirty-eight among the fifty provinces. It showed also a weak efficiency score and hence a weak bias corrected for almost all provinces, a very low national average of efficiency and therefore a weak performance of the ONEP's drinking water. This might be due to several factors such as public management of the sector, the lack of cooperation projects to supply rural communities, the waste due to damaged pipes and the free distribution of drinking water through public fountains either in some rural districts or in informal areas of large cities. Unfortunately, the effect of factors on efficiency can not be measured in the absence of data in this regard. If data will be available, a study of the effect of the environmental variables will be done with the aim to identify the main determinants of the inefficiency in this domain in Morocco.

Will the total privatization of the sector yield to an improved efficiency? Some experiences as in Portugal indicate the poor performance of the private management in comparison with the public one as shown in De Witte and Marques (2008). Otherwise, it should be noted that business creation and economic development of rural municipalities enable rural citizens to have a considerable income and to fund their partial or total need of water and pipes. In addition, as an open issue, it will be interesting to perform a frontier analysis on the drinking water quality management in the country and to compare the provinces performance in this regard.

APPENDIX

1. Density of the error ϵ_i

Density of ϵ_i in the panel SFA when the dependence between the two error terms is considered and when $u_i \sim N^{+}(0\,,\sigma_U^2)$ becomes

$$g(\epsilon_i) = \int_0^{+\infty} f(\epsilon_i, u_i)\, du_i \tag{A.1}$$

$$= \int_0^{+\infty} f(\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{iT}, u_i)\, du_i$$

$$= \int_0^{+\infty} f_1(u_i) \prod\nolimits_t f_2(\epsilon_{it} + \eta(t)u_i)$$

$$.\ c_\theta\big(F_1(u_i), F_2(\epsilon_{i1} + \eta(1)u_i),\ \dots, F_2(\epsilon_{iT} + \eta(T)u_i)\big)\, du_i$$

$$= \int_0^{+\infty} f_1(u_i) \prod\nolimits_t f_2(\epsilon_{it} + \eta(t)u_i)\ .\prod\nolimits_t c_\theta\big(F_1(u_i), F_2(\epsilon_{it} + \eta(t)u_i)\big)\, du_i =$$

$$\int_0^{+\infty} f_1(u_i)\, .\prod\nolimits_t \big[f_2(\epsilon_{it} + \eta(t)u_i)\, .\, c_\theta\big(F_1(u_i), F_2(\epsilon_{it} + \eta(t)u_i)\big)\big]\ du_i$$

$$= \int_0^{+\infty} f_1(u_i) \prod\nolimits_t A_{it}\ \ du_i = E\left(\prod\nolimits_t A_{it}\ \right) \tag{A.2}$$

where $\mathrm{A_{it}} = \mathrm{f_2}(\epsilon_{\mathrm{it}} + \eta(\mathrm{t})u_{\mathrm{i}})\ \mathrm{c}_\theta\big(\mathrm{F_1}(u_{\mathrm{i}}), \mathrm{F_2}(\epsilon_{\mathrm{it}} + \eta(\mathrm{t})u_{\mathrm{i}})\big)$.

2. Technical efficiency for each DMU at time *t*

The associated technical efficiency of $\mathbf{DMU_{it}}$ is expressed as

$$TE_{it} = E[exp\{-u_{it}\}|\epsilon_i] \tag{A.3}$$

$$= E[exp\{-\eta(t)u_i\}|(\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{iT})]$$

$$= \int_0^{+\infty} exp\{-\eta(t)u_i\}\, f_1(u_i|(\epsilon_{i1}, \dots, \epsilon_{it}, \dots, \epsilon_{iT}))\, du_i$$

$$= \int_0^{+\infty} exp\{-\eta(t)u_i\}\ \frac{f(\,u_i\,,\ \epsilon_i)}{g(\epsilon_i)}\ du_i$$

$$= \frac{1}{g(\epsilon_i)} \int_0^{+\infty} exp\{-\eta(t)u_i\}\, f_1(u_i) \prod_t A_{it}\ \ du_i$$

$$= \frac{E(exp\{-\eta(t)u_i\} \prod_t A_{it}\ )}{E(\prod_t A_{it}\ )} \tag{A.4}$$

REFERENCES

Battese, G.E. and T.J. Coelli (1988). Prediction of firm level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38(3), 387-399.

Battese, G.E. and T.J. Coelli (1992). Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India. *The Journal of Productivity Analysis*, 3(1), 153-169.

Battese, G.E. and T.J. Coelli (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics*, 20, 325-332.

Battese, G.E., A. Heshmati, and L. Hjalmarsson (2000). Efficiency of labour use in Swedish banking industry: a stochastic frontier approach. *Empirical Economics*, 25(4), 623-640.

Battese, G.E., T.J. Coelli, and T.C. Colby (1989). Estimation of Frontier Production Functions and the Efficiencies of Indian Farms Using Panel Data from ICRISAT's Village Level Studies. *Journal of Quantitative Economics*, 5(2), 327-348.

Bhat, C. and N. Eluru (2009). A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling. *Transportation Research*, Part B, 43(7), 749-765.

Coelli, T.J. (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis*, 6(3), 247-268.

Coelli, T.J. (1996). A Guide to FRONTIER, Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation. Centre for efficiency and Productivity Analysis, CEPA Working Paper 96/07, Department of Econometrics, University of New England.

Cornwell, C., P. Schmidt, and R.C. Sickles (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46(1-2), 185-200.

Daraio, C. and L. Simar (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Springer.

De Witte, K. and R.C. Marques (2008). Designing incentives in local public utilities, an international comparison of the drinking water sector. *Social Science Research Network* SSRN 1084807.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathematics*, #38. Philadelphia: SIAM.

El Mehdi, R. and C.M. Hafner (2014a). Local government efficiency: The case of Moroccan municipalities. *African Development Review*, 26(1), 88-101.

El Mehdi, R. and C.M. Hafner (2014b). Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation (MATCOM)*, 102(C), 104-116.

Faria, R.C., G.S. Souza, and T.B. Moreira (2005). Public versus private water utilities: Empirical evidence for brazilian companies. *Economics Bulletin*, 8(2), 1-7.

Gallant, A. Ronald (1984). The fourier flexible form. *American Journal of Agricultural*

Economics, 66(2), 204-208.

Jondrow, J., C. A. Knox Lovell, I. S. Materov, and P. Schmidt (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2-3), 233-238.

Kim, S. and Y.H. Lee (2006). The productivity debate of East Asia revisited: a stochastic frontier approach. *Applied Economics, Taylor and Francis Journals*, 38(14), 1697-1706.

Kumbhakar, S.C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(1-2), 201- 211.

Kumbhakar, S.C. and C.A. Knox Lovell (2000). *Stochastic Frontier Analysis*. First edition. Cambridge University Press, United Kingdom.

Lee, Y.H. and P. Schmidt (1993). A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency. *The Measurement of Productive Efficiency: Techniques and Applications*. Edited by H. Fried, C.A.K. Lovell and S. Schmidt, Oxford University Press, pp. 237-255.

Nelsen, R. B. (1999). *An Introduction to Copulas*. First edition. Springer, New York.

Sampaio, A., C. Barros, and J. Ramajo (2005). Technical Inefficiency in Municipal Water Distribution Service: A Case Study for Portugal. *Anales de Economia Aplicada*, XIX Reuniốn Anual. Edições Asepelt (Associação de Economia Aplicada) Espanha Badajoz.

Schmidt, P. and R.C. Sickles (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, 2(4), 367-374.

Simar, L. and P.W. Wilson (2010). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, 29(1), 62-98.

Smith, M.D. (2008). Stochastic frontier models with dependent error components. *The Econometrics Journal*, 11(1), 172-192.

Tupper, H.C. and M. Resende (2004). Efficiency and regulatory issues in the Brazilian water and sewage sector: an empirical study. *Utilities Policy*, 12(1), 29-40.

Vishwakarma, A. and M. Kulshrestha (2010). Stochastic Production Frontier Analysis of Water Supply Utility of Urban Cities in the State of Madhya Pradesh, India. *International Journal of Environmental Sciences*, 1(3), 357-367.