# Stochastic Production Frontier and Technical Inefficiency : Evidence from Tunisian Manufacturing Firms*

Rafik Baccouche[†]& Mokhtar Kouki[‡] [§]

January 2002 (Preliminary version January 2000)

**Abstract**

In this paper, a stochastic frontier production model is estimated on panel data, and technical inefficiency indices are computed for Tunisian manufacturing firms over the 1983-1993 time period. The most commonly used one-sided distributions of the inefficiency error term are specified, namely the truncated normal, the half-normal and the exponential distributions. A generalized version of the half-normal, which does not embody the zero-mean restriction, is also explored. For each distribution, the likelihood function and the counterpart of Jondrow et al. (1982) estimator of technical efficiency are explicitly stated. Based on our data set, formal tests lead to a strong rejection of the zero-mean restriction embodied in the half normal distribution. Our main conclusion is that the degree of measured inefficiency is very sensitive to the postulated assumptions about the distribution of the one-sided error term. The estimated inefficiency

* We are acknowledged to Tahar Abdessalem, Jean Pierre Florens, Farouk Kriaa, Mohamed Salah Rejeb, Rached Bouaziz and Jean Charles Rochet for helpful comments and suggestions. We are grateful to the "Secrétariat d'Etat à la Recherche Scientifique en Tunisie" for the financial support.

† University of Tunis 3.

‡ LEGI - Polytechnics of Tunisia and University of Tunis

§ Correspondence to Mokhtar Kouki: LEGI - Polytechnics of Tunisia, BP743, 2048 La Marsa Tunis - Tunisia. email: mokhtar.kouki@fsjegj.rnu.tn

indices are, however, unaffected by the choice of the functional form for the production function.
Keywords: Stochastic frontier; Farrell's technical inefficiency; Unbalanced panel data; Composed disturbance error; One-sided distribution.

# 1 Introduction

The purpose of this paper is to estimate firm-specific levels of technical inefficiency using a panel data of Tunisian manufacturing firms. This panel data set covers the time period 1983-1993. Our methodology applies the stochastic frontier analysis which became popular in the recent literature.

The basic difference between stochastic frontier model and the standard econometric model is that the former adds a one-sided distributed random variable to the usual stochastic disturbances term. This supplementary random variable is intended to take into account the amount by which observed output is less than potential output. This amount is a measure of technical inefficiency in the sense of Farrell (1957).

Econometric estimation of firm-specific technical inefficiency raises two problems. The first problem relates to the appropriateness of the postulated distribution for the one-sided error term, particularly if maximum likelihood estimation method is to be used. Although an extensive literature had been devoted to this question, the fact remains that there is little guidance as to the appropriate specification of the one-sided distribution. A sensitivity analysis of the results to alternative distributional assumptions must then be conducted.

Once the parameters of the model have been estimated, the second problem is how one can extract the inefficiency component from the estimated composed error term. Jondrow et al. (1982) showed that firm-specific estimates of inefficiency can be obtained through the distribution of the inefficiency term conditional on the estimate of the whole composed error term. The Jondrow et al. estimator is easily seen to be inconsistent when used with a single cross-section. Consistency may, however, be achieved using a panel data of firms covering a sufficiently long time period.

In our empirical work we consider three alternative distributions for the one-sided error term, which are most commonly used by econometricians. These distributions are the truncated normal, the generalized half-normal and the exponential distributions. The term “generalized” is used here to designate the distribution of the absolute value of a non-zero mean normally distributed variable. We are not aware of any empirical study which has used this distribution of which the standard half-normal (zero mean) is a special case.

This paper is organized as follows. A brief discussion of some measurement methods of technical inefficiency in the context of the stochastic frontier model is presented in the following section. In section 3, explicit formulas for the log-likelihood functions and for the estimators of firm-specific inefficiency are given for the three distributions stated above, in the context of (unbalanced) panel data. Section 4 presents data description and our empirical results. The main conclusions are summarized in section 5.

# 2 Stochastic Frontier Model and Technical Inefficiency

The stochastic frontier model was introduced by Aigner, Lovell, and Schmidt (1977) and by Meeusen and Van der Broeck (1977) in order to avoid the shortcomings inherent to the deterministic frontier model[1]. It assumes a composed error term reflecting both the usual statistical random noise and technical inefficiency. Assuming a log-linear form, the stochastic frontier model may be defined as follows

$$y_{it} = \alpha + x_{it}\beta + \varepsilon_{it}, \tag{1}$$

where $y_{it}$ and $x_{it}$ denote, respectively, the logarithms of observed output and of a row vector of inputs for the $i$ th firm in the $t$ th time period; $\alpha$ and $\beta$ are the unknown parameters to be estimated and $\varepsilon_{it}$ is the stochastic error term which is assumed to behave in a manner consistent with the stochastic frontier concept, i.e.

$$\epsilon_{it} = v_{it} - u_i. \tag{2}$$

[1] The deterministic frontier model assumes only one-sided error term reflecting technical inefficiency, so that any statistical noise due to misspecification of the model or measurement error in the variables will be translated into inefficiency.

The disturbances $v_{it}$ consist of random shocks in the production process beyond the firms control and they are taken to be independent and identically distributed ($i.i.d.$) across observations as $N(0, \sigma_v^2)$. The random variables $u_i$, which are assumed to be non-negative, reflect the shortfall of actual output from the efficient frontier; they are also assumed to be $i.i.d.$ as well as independent of $v_{it}$ and of factor inputs. With this specification, the quantity $e^{\alpha + x_{it}\beta + v_{it}}$ specifies the efficiency stochastic production frontier as defined by Aigner et al. (1977), while the case $v_{it} = 0$ leads to the deterministic efficiency frontier model in Aigner and Chu (1968). In either case technical efficiency for each unit is measured by $TE_i = e^{-u_i}$, which, as $u_i \geq 0$, lies between zero and one.

The strength of the stochastic frontier is that it provides a method to separate random disturbances caused by inefficiency in a firm's behavior from other uncontrolled random shocks. However, given that $u_i$ is not observable, direct estimation of $TE_i$ is not possible even when $\alpha$ and $\beta$ are known.

Computation of $TE_i$ requires prior estimation of the non-negative error $u_i$, i.e. a method of decomposition of the entire error term $\varepsilon$ into its two individual components. Alternative solutions to this problem have been proposed in the recent literature, depending on whether we assume a particular distribution for $u_i$ or not[2]. However, such a decision is heavily conditioned by the nature of inefficiency we are interested in, i.e. relative or absolute. If inefficiency is to be defined relative to a completely efficient base firm, then no distributional assumption for $u_i$ is needed. In this case consistent residual-based estimation methods for firm-specific inefficiency may be used as suggested by Greene (1980) in the case of a single cross-section, and Cornwell et al. (1990) for panel data. If, however, absolute firm-specific inefficiency is to be computed, then prior assumption on the distribution of $u_i$ cannot be avoided. But, in such a case, only the population mean of technical inefficiency, i.e. $1 - E(TE_i)$, may be calculated using the estimated parameters of the maintained distribution. Identification of absolute firm-specific technical inefficiency is still impossible.

To by-pass this problem, Jondrow et al. (1982), and Kalirajan and Flinn (1983) propose to use the conditional expectation of $u_i$, given the entire error term $\varepsilon_i$, as an estimator of firm-specific technical inefficiency. This is very attractive since such an estimator is also the best predictor of $u_i$, i.e. the

[2] Greene (1994) offers an excellent survey of these methods. He also discusses the more general case where technical inefficiency is time-variant. See, also Cornwell et al. (1990).

minimum mean-squared-error predictor. However, it must be noted that, in the case of a single cross-section, this estimator is not consistent. This is because, while $\varepsilon_i$ contains only imperfect information about $u_i$, the variance of the conditional distribution of $u_i$ given $\varepsilon_i$ is independent of sample size. The variability of $v_i$ remains no matter how large $N$ is. This deficiency is resolved when panel data is used because the irrelevant variability contained in $\varepsilon_{it}$, i.e. $v_{it}$, is being averaged over the number of periods $T$. So, consistent estimates of firm-specific inefficiency may be obtained when $T$ goes to infinity. Extension of Jondrow et al. predictor for use with panel data can be found in Battese and Coelli (1988), Kumbhakar (1988), and Battese, Coelli and Colby (1989).

Much of the criticism addressed to the estimates of absolute technical inefficiency relate to the plausibility of the distributional assumption that must be made for the one-sided error term. Empirical studies, as in Greene (1980, 1994) and Stevenson (1980) among many others, revealed that different specified distributions do give different estimates of technical inefficiency. In some cases the estimates of the input coefficients were also affected. Without prior information on the economic processes generating the inefficiency, the choice of any particular distribution cannot be justified. In practice, such information is in general not available and different distributions must be tried in order to assess the sensitivity of the results to alternative distributional specifications.

Given that $u_i$ is a non-negative random variable, numerous density functions can be specified for it. Some distributions are however more attractive than others. The half-normal, exponential and truncated-normal distributions are by far the most used distributions in empirical studies. As pointed out by Stevenson (1980), the major drawback of the first two distributions is to restrict the density of the inefficiency to be most concentrated near zero. This implies that "the likelihood of inefficient behavior monotonically decreases for increasing levels of inefficiency" [3]. Stevenson proposed instead a more flexible distribution of the inefficiency given by the truncated normal density function which does not restrict the mode to occur at zero. This is a natural extension in so far as it enables the testing of the adequacy of the zero-mode restriction[4].

[3] See Stevenson (1980), p. 58.

[4] Note that LM tests for the adequacy of the half-normal and truncated normal distributions have been explicitly derived by Lee (1983) conditional on the assumption that the distribution of $u_i$ belongs to the Pearson family of truncated distributions.

# 3 Alternative Specifications and ML Estimation

Prior to the computation of firm-specific inefficiency the unknown parameters of the production function must be estimated. Note that, apart from the asymmetry of the distribution of $\varepsilon$, equations (1) and (2) fit well into the standard random effects model in panel data literature. The asymmetry characterizing $\varepsilon_{it}$ doesn't matter, and in any case it can be easily avoided by adjusting the constant term $\alpha$.

The choice of the method of estimation must be assessed according to some basic assumptions concerning the disturbances term. For example, assuming exogenous factor inputs, the Feasible GLS technique may be used to derive consistent estimates of the frontier. The main advantage of this technique is that no distributional specification is needed for the one-sided error term for consistent estimates of the parameters. However, assuming the distribution of $u_i$ to be known, some efficiency gain over GLS may be achieved through the ML method. If factor inputs were to be correlated with the disturbances then neither GLS nor ML estimator would be consistent. In such a case, an efficient IV estimator of the type discussed by Hausman and Taylor (1981) must be used[5]. Recall however, that, given the parameters estimates, prior knowledge of the inefficiency distribution is required when absolute firm-specific inefficiency is sought.

In this study, we choose to focus attention exclusively on the MLE method in estimating the production frontier parameters. We use three different specifications for the distribution of $u_i$. These are the truncated normal, the generalized half-normal and the exponential distributions. The first two distributions are general enough in that they do not restrict the mode to occur at zero, and they both encompass the standard half-normal as a special case.

Before going to the results, we give in the following the explicit formulas we used in our programming procedures. These formulas are stated for the general case of unbalanced panel data, and encompass both the production frontier and the cost frontier cases. Indeed, as we define $\varepsilon_{it} = v_{it} + \delta u_i$, where $\delta$ is a switching parameter taking value of $-1$ for a production frontier and value of 1 for a cost frontier, we can move from one specification to another by assigning an appropriate value to $\delta$ wherever it appears.

In deriving our results, the following two assumptions are made: $(i)$ the

[5] See, Schmidt and Sickles (1984), and Cornwell et al. (1990).

random variables $v_{it}$ are assumed to be *i.i.d.* across observations as $N(0, \sigma_v^2)$, as well as independent of the $u_i$ random variables, (*ii*) $v_{it}$ and $u_i$ are both assumed to be independently distributed of the factor inputs in the model[6].

**Proposition 1** Under the assumptions stated above, the log-likelihood functions of the stochastic production (or cost) frontier for the three considered distributions for $u_i$ are given by:

**Case 1**: The Truncated normal distribution (Battese and Coelli (1988))

$$\begin{aligned} \mathcal{L}_1 = & -\frac{1}{2}\sum_{i=1}\ln(1+\lambda T_i) - \ln(\sigma_u)\sum_{i=1} T_i - \frac{1}{2}\ln(2\pi)\sum_{i=1} T_i - \frac{N}{2}\gamma^2 \\ & -\frac{1}{2}\lambda\sum_{i=1}\sum_{t=1}\left(\frac{\epsilon_{it}}{\sigma_u}\right)^2 + \frac{1}{2}\sum_{i=1}(1+\lambda T_i)\left(\frac{\mu_{i1}}{\sigma_u}\right)^2 \\ & +\sum_{i=1}\ln\Phi\left(\sqrt{1+\lambda T_i}\,\frac{\mu_{i1}}{\sigma_u}\right) - N\ln(\Phi(\gamma)) \end{aligned} \tag{3}$$

where: $\lambda = \frac{\sigma_u^2}{\sigma_v^2}, \gamma = \frac{\mu}{\sigma_u}$, and $\mu_{i1} = \frac{\mu}{1+\lambda T_i} + \delta\frac{\lambda T_i}{1+\lambda T_i}\frac{\bar{\epsilon}_i}{\sigma_u}$. The functions $\phi$ and $\Phi$ are, respectively, the density function and distribution function of the standard normal.

**Case 2:** The generalized half-normal distribution

$$\begin{aligned} \mathcal{L}_2 = & -\frac{1}{2}\sum_{i=1}\ln(1+\lambda T_i) - \ln(\sigma_u)\sum_{i=1} T_i + \frac{1}{2}\ln(\lambda)\sum_{i=1} T_i - \frac{1}{2}\ln(2\pi)\sum_{i=1} T_i \\ & -\frac{N}{2}\gamma^2 - \frac{1}{2}\lambda\sum_{i=1}\sum_{t=1}\left(\frac{\epsilon_{it}}{\sigma_u}\right)^2 + \frac{1}{2}\sum_{i=1}(1+\lambda T_i)\left(\frac{\mu_{i1}}{\sigma_u}\right)^2 \\ & +\sum_{i=1}\ln\left(\Phi\left(\sqrt{1+\lambda T_i}\,\frac{\mu_{i1}}{\sigma_u}\right) + \psi\Phi\left(\sqrt{1+\lambda T_i}\,\frac{\mu_{i2}}{\sigma_u}\right)\right) \end{aligned} \tag{4}$$

where : $\lambda$, $\gamma$ and $\mu_{i1}$ as defined in the truncated normal case; $\mu_{i2} = -\frac{\mu}{1+\lambda T_i} + \delta\frac{\lambda T_i}{1+\lambda T_i}\frac{\bar{\epsilon}_i}{\sigma_u}$, with $\bar{\epsilon}_i = T_i^{-1}\sum_{t=1}^{T_i}\epsilon_{it}$ and $\psi = \exp\left(-2\delta\gamma_2\frac{\lambda T_i}{1+\lambda T_i}\frac{\bar{\epsilon}_i}{\sigma_u}\right)$.

[6] This may be justified using the Zellner, Kmenta and Drèze (1966) assumption that firms maximize expected profit, and that disturbances in the first order conditions are independent of $\varepsilon$.

**Case 3:** The exponential distribution.

$$
\begin{aligned}
\mathcal{L}_3 \; = \; & -\frac{1}{2}\sum_{i=1} \ln(T_i) - \sum_{i=1} T_i \ln(\sigma_v) + N\ln(\gamma_2) - \frac{1}{2}\sum_{i=1} (T_i - 1)\ln(2\pi) \\
& -\frac{1}{2}\sum_{i=1}\sum_{t=1}\left(\frac{\epsilon_{it}}{\sigma_v}\right)^2 + \frac{1}{2}\sum_{i=1} T_i \left(\frac{\mu_{i3}}{\sigma_v}\right)^2 + \sum_{i=1} \ln\Phi\left(\frac{\mu_{i3}}{\sigma_v}\sqrt{T_i}\right) \qquad (5)
\end{aligned}
$$

where: $\gamma_2 = \theta\sigma_v$ and $\mu_{i3} = -\frac{\theta\sigma_v^2}{T_i} + \delta\,\bar{\epsilon}_i$

It can be easily verified from either (3) or (4) that, taking $\mu = 0$ in $\mu_{i1}$ and $\mu_{i2}$ leads to the log-likelihood function corresponding to the standard half-normal distribution as defined by Aigner et al.. A standard likelihood ratio test can thus be conducted to test the validity of the half-normal specification.

In order to derive the counterpart of Jondrow et al. estimator of technical inefficiency, the conditional distributions of $u_i$, given sample values of the random vector $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT_i})'$, must be stated. For the truncated normal and the exponential distributions, it can be shown that the conditional distribution of $u_i$ given $\varepsilon_i$ is defined by the truncation (at zero) of the normal distribution with mean $\mu_{i1}$ and variance $(1+\lambda T_i)^{-1}\sigma_u^2$ for the truncated normal case, and with mean $\mu_{i3}$ and variance $\sigma_v^2/T_i$ for the exponential case. For the generalized half-normal case, it can be shown that the conditional distribution is given by

$$
f(u_i/\varepsilon_i) = \frac{1}{\sigma_i}\frac{\phi\left(\frac{u_i-\mu_{i1}}{\sigma_i}\right) + \phi\left(\frac{u_i-\mu_{i2}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{i1}}{\sigma_i}\right) + \Phi\left(\frac{\mu_{i2}}{\sigma_i}\right)} \qquad (6)
$$

where $\sigma_i = \frac{\sigma_u}{\sqrt{1+\lambda T_i}}$.

The following theorem generalize the Jondrow et al. result to the case of panel data using the alternative distributions mentioned above.

**Proposition 2** For the conditional distributions stated above, the conditional expectations of $TE_i = e^{-u_i}$, given $\varepsilon_i$, are given by

**Case 1:** The truncated normal distribution (Battese and Coelli (1988))

$$
E(TE_i/\varepsilon_i) = e^{\delta\,\mu_{i1}+\frac{\sigma_i^2}{2}}\frac{\Phi\left(\frac{\mu_{i1}}{\sigma_i} + \delta\,\sigma_i\right)}{\Phi\left(\frac{\mu_{i1}}{\sigma_i}\right)} \qquad (7)
$$

**Case 2:** The generalized half-normal distribution:

$$E(TE_i/\varepsilon_i) = e^{\delta\mu_{i1} + \frac{\sigma_i^2}{2}} \frac{\Phi\left(\frac{\mu_{i1}}{\sigma_i} + \delta\sigma_i\right) + \eta\,\Phi\left(\frac{\mu_{i2}}{\sigma_i} + \delta\sigma_i\right)}{\Phi\left(\frac{\mu_{i1}}{\sigma_i}\right) + \Phi\left(\frac{\mu_{i2}}{\sigma_i}\right)} \tag{8}$$

where $\eta = \exp(-2\delta\frac{\mu}{1+\lambda T_i})$.
**Case 3:** The exponential distribution.

$$E(TE_i/\varepsilon_i) = e^{\delta\mu_{i3} + \frac{\sigma_v^2}{2T_i}} \frac{\Phi\left(\frac{\sqrt{T_i}\mu_{i3}}{\sigma_v} + \delta\frac{\sigma_v}{\sqrt{T_i}}\right)}{\Phi\left(\frac{\sqrt{T_i}\mu_{i3}}{\sigma_v}\right)} \tag{9}$$

Note that, in equations (7)-(9), the quantities $\mu_{ij}$, $j = 1, 2, 3$, are defined as above with $\varepsilon_i$ replaced by its sample counterpart $e_i$, and with $\overline{\varepsilon}_i = \overline{e}_i$.

# 4 Empirical Analysis

## 4.1 Data Description

The data used in our empirical work stem from the "Enquête Annuelle d'Entreprises" which has been annually conducted since 1983 by the "Institut National des Statistiques". Unfortunately, because the surveys conducted beyond 1993 were not available to us, our panel set is limited to 11 years and covers the period $1983 - 1993$. The total number of observations is 8191, corresponding to 1125 individual firms. These firms belong to six manufacturing sectors, namely Food, Construction Materials Ceramics and Glass, Mechanical and Electrical, Chemicals, Textiles, and Miscellaneous. Summary statistics on the main variables used in the regression are given in Table A1 in the Appendix. Table A2 gives both the distribution of sector samples over the 11 years and the distribution of sector units by the number of times they are observed over the entire period. We can see that 160 firms are staying in the sample for the entire period, while 27 firms are observed only once. The percentage of firms which are observed over at least 5 years is about 83.5.

Note that, the volume of output, $Y$, is defined as value added measured at factor costs and at constant prices with 1990 as a base year. The volume of fixed capital, $K$, is also measured at constant prices using the same base year. Labor, $L$, is defined as total employment and is measured by the number of employees. And finally, the capital vintage, $agek$, is measured as the average age of equipments. A detailed information on the procedure used to construct these variables may be found in Zribi (1995).

## 4.2 Frontier Specification

The production technology is assumed to be described by the transcendental logarithmic form, T-L, suggested by Christensen et al. (1973), which provides a second-order Taylor-series approximation to any twice-differentiable production function. The advantage of this flexible specification is twofold. Firstly, it allows the elasticities of output to vary across firms and across periods. Secondly, it allows to test, through exclusion restrictions on some parameters, the empirical plausibility of the restrictive Cobb-Douglas, C-D, specification. Formally the model we estimate can be written as

$$\begin{aligned} y_{it} = {} & \alpha + \gamma_1 t + \gamma_2 t^2 + (\beta_k + \gamma_k t) k_{it} + (\beta_l + \gamma_l t) l_{it} + (\beta_a + \gamma_a t) agek_{it} \\ & + 0.5\beta_{kk} k^2 + \beta_{kl}(k_{it} \times l_{it}) + \beta_{ka}(k_{it} \times agek_{it}) + 0.5\beta_{ll} l_{it}^2 \\ & + \beta_{la}(l_{it} \times agek_{it}) + 0.5\beta_{aa} agek_{it}^2 + \delta_{food} S_1 + \delta_{cmcg} S_2 + \delta_{chim} S_4 \\ & + \delta_{text} S_5 + \delta_{Misc} S_6 + v_{it} - u_i \end{aligned} \qquad (10)$$

where $y$, $k$ and $l$, denote, respectively, logarithms of $Y$, $K$ and $L$. The time variable $t$ is included as a regressor in order to catch neutral technical progress in production[7]. This implies that some of the shifts in the production frontier are allowed to occur independently of changes in inputs. The dummy variables $S_i$ are introduced in order to take into account some sectorial heterogeneity, and their effects must be interpreted in comparison with the omitted Mechanical and Electrical sector.

## 4.3 Empirical Results

In this section, the translog frontier model defined by (10) is estimated, along with the Cobb-Douglas frontier, by maximum likelihood method using alternative distributions for the one-sided error term[8]. These distributions are

[7] See Solow (1957).

[8] The use of the ML method supposes that the specific effects is random rather than fixed. This is the case here since the Hausman test statistic equals 199.42 which is largely greater than the $\chi^2_{14}$ critical value.

the truncated normal, the generalized half-normal ($\mu \neq 0$), the half-normal ($\mu = 0$), and the exponential. Since we are interested particularly in the measurement of technical inefficiency, the frontier parameters estimates are left out from the text and reported in Tables A4 and A5 in the Appendix[9]. A close look at these Tables reveals that there are no substantial changes in the parameters estimates over the alternative distributions for $u_i$. In particular, estimates obtained with the truncated normal and the generalized half-normal are the same[10]. Note also that pooled ordinary least squares estimates, shown in the last column, are quite similar to ML estimates, particularly in comparison with the exponential case. This similarity in the parameters estimates suggests that the shape and the location of the stochastic frontier are not sensitive to the distributional assumption of the inefficiency term.

Another interesting result that must be pointed out is that, in view of the asymptotic t-statistics, almost all the estimated parameters appear to be highly significant. As a consequence, a substantial reduction in the value of the likelihood function, given in the bottom row of Table 1, occurred when we moved from the T-L form to the C-D form. For example, for the truncated model, the value of the likelihood ratio test statistic, corresponding to the null hypothesis of a C-D form equals 408.2, which exceeds the $\chi^2_{10}$ critical value by a large amount. The C-D technology is thus rejected. The null hypothesis of absence of technological progress is also rejected without ambiguity, according to the likelihood ratio statistic. The same conclusion holds independently of the hypothesized distribution for $u_i$.

Table 1 below gives estimates of the main parameters characterizing the alternative distributions of the inefficiency one-sided error term. If these parameters happen to be statistically non significant, this error term can be altogether ignored. In this case the restricted ML estimation reduces to ordinary least squares.

Before coming to formal hypotheses testing, some insights regarding the importance of inefficiency may be pointed out. Indeed, recall that according to our parametrization, $\lambda$ is an indicator of the relative variability of the two sources of random errors, i.e. $\lambda = \sigma_u^2/\sigma_v^2$. Based on the translog frontier, a simple calculation reveals that, for the truncated normal case, the estimated

[9] The computer programs we used are available from the authors upon request. These programs were written using Gauss procedure.

[10] Recall that, contrary to the Cobb-Douglas case, the individual parameters in the translog model are not directly interpretable as elasticities.

variance of $u_i$ is about 44.7 percent of the estimated entire disturbance variance, $V(u_i) + \sigma_v^2$ which is equal to 0.12[11]. The figure is very different for the half-normal distribution ($\mu = 0$) where the estimated variance of $u_i$ equals 0.021, leading to 23.5 percent of the estimated total variance. This percentage dramatically reduces to 4.6 percent in the exponential case.

Formal tests on the significance of the random variable $u_i$ in the frontier model, i.e. $\lambda = \mu = 0$, may be conducted through the generalized likelihood ratio statistic. From the bottom row of Table 1, we can see that the value of the likelihood function is the same for both the truncated normal and the generalized half-normal cases. Thus, the same conclusion will apply to both.

For the translog case, the absolute value of the restricted maximum likelihood is equal to 2723.4. Hence, the negative of twice the logarithm of the likelihood ratio is equal to 2331.4, which is by far greater, at any significance level, than the $\chi_2^2$ critical value given in Table 1 of Kodde and Palm (1986)[12]. We conclude that the one-sided error term can not be ignored. Next we test, through the restriction $\mu = 0$, whether the truncated normal (or the generalized half-normal) is more reliable than the half-normal. From the corresponding values of the likelihood functions, it can be easily seen that the half-normal distribution is strongly rejected in favor of the alternative hypothesis. Note that all these conclusions were to be expected given the high values of the asymptotic t-statistics which are given in parenthesis in Table 1.

---

[11] It is worth noting here that the true variance of the truncated random variable $u_i$ is not equal to $\sigma_u^2$, but rather to:

$$V(u_i) = \left\{ 1 - \frac{\mu}{\sigma_u} \frac{\phi(\frac{\mu}{\sigma_u})}{\Phi(\frac{\mu}{\sigma_u})} - \left[ \frac{\phi(\frac{\mu}{\sigma_u})}{\Phi(\frac{\mu}{\sigma_u})} \right]^2 \right\} \sigma_u^2.$$

For the half normal case, $\mu = 0$, this variance reduces to $\left[1 - \frac{2}{\pi}\right] \sigma_u^2$, see Greene (1994), page 27.

[12] This Table must be used instead of the usual $\chi^2$ Table, because the polar value, $\lambda = 0$, is on the boundary of the parameter space, not in its interior.

| Table 1: Estimated parameters of the distribution of $u_i$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Translog Frontier | | | | Cobb-Douglas Frontier | | | |
| | TN | HN $\mu \neq 0$ | HN $\mu = 0$ | EXP | TN | HN $\mu \neq 0$ | HN $\mu = 0$ | EXP |
| $\sigma_u$ | 0.239 (119.5) | 0.237 (47.4) | 0.416 (34.7) | – | 0.260 (32.5) | 0.258 (23.4) | 0.458 (38.2) | – |
| $\lambda$ | 0.861 (66.2) | 0.845 (20.6) | 2.568 (15.6) | – | 0.987 (14.5) | 0.970 (88.2) | 3.001 (16.9) | – |
| $\frac{\mu}{\sigma_u}$ | 2.344 (167.4) | 2.201 (62.9) | – | – | 2.302 (9.7) | 2.158 (196.2) | – | – |
| $\sigma_v$ | – | – | – | 0.397 (132.3) | – | – | – | 0.417 (139) |
| $\theta\sigma_v$ | – | – | – | 4.532 (15.0) | – | – | – | 4.513 (14.9) |
| $\lvert\mathcal{L}\rvert$ | 1557.7 | 1557.3 | 1624.4 | 4292.4 | 1761.8 | 1760.7 | 1836.1 | 4705.4 |

Using the estimated parameter values for the frontier production function, firm-specific technical efficiency were predicted for each model, using formulas (7)-(9). The results are summarized in Table 2. The columns Q1 and Q3 refer, respectively, to the first and third quartile.

| Table 2: Summary statistics on technical efficiency | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Frontier | Min | Q1 | Mean | Median | Q3 | Max | S-D | Min. CV |
| Truncated Normal | | | | | | | | |
| C-D | 30.6 | 45.9 | 56.3 | 54.4 | 65.9 | 93.4 | 13.1 | 5.369 |
| T-L | 33.8 | 49.0 | 58.3 | 57.0 | 66.3 | 92.9 | 12.3 | 5.665 |
| Half Normal ($\mu \neq 0$) | | | | | | | | |
| C-D | 35.5 | 52.2 | 63.0 | 61.6 | 73.8 | 94.8 | 13.4 | 3.103 |
| T-L | 38.6 | 55.7 | 65.3 | 64.6 | 74.3 | 94.9 | 12.5 | 3.102 |
| Half Normal ($\mu = 0$) | | | | | | | | |
| C-D | 36.7 | 57.5 | 70.1 | 69.6 | 84.3 | 97.6 | 15.4 | 4.322 |
| T-L | 38.8 | 61.3 | 72.5 | 72.4 | 84.3 | 97.4 | 14.3 | 4.482 |
| Exponential | | | | | | | | |
| C-D | 66.3 | 89.4 | 91.6 | 93.1 | 95.3 | 97.9 | 5.1 | 5.966 |
| T-L | 66.2 | 90.2 | 92.0 | 93.4 | 95.5 | 97.9 | 4.9 | 6.247 |
| COLS[13] | | | | | | | | |
| C-D | 19.2 | 32.9 | 41.7 | 39.8 | 49.5 | 87.4 | 11.8 | – |
| T-L | 16.8 | 36.4 | 45.0 | 43.8 | 52.1 | 90.0 | 11.6 | – |

[13] COLS estimates are calculated for each year relative to a base firm and then averaged over the entire period. That is why the maximum value is not equal to 100.

Two salient facts from Table 2 are worth emphasizing. Firstly, the estimates of technical efficiency using the C-D and T-L functional forms are strongly similar. Indeed, column four indicates that, given the distribution of $u_i$, the difference in sample mean values never exceeds 2.4 percentage points. This finding seems to be somewhat troublesome since the C-D form has been strongly rejected in favor of the T-L form. However, it is worth noting that, in our case, this is hardly a surprise since, on the one hand, the C-D and T-L forms led approximately to the same parameters estimates of the distribution of $u_i$, and that, on the other, residuals are being averaged over periods when technical efficiency is computed.

The second salient fact that emerges from Table 2 is that technical efficiency estimates appear to be highly sensitive to the maintained hypothesis on the distribution of the one-sided error term. First, the exponential case leads to the highest technical efficiency levels with a sample average rate of 91.2%, the half-normal ranked second with a rate of 72.5% and the generalized half-normal and truncated normal followed with rates of 65.3% and 58.3%, respectively. However, we have some reasons to believe that results obtained from the last two distributions are more reliable. Indeed, the half-normal imposes $\mu = 0$, which seems to be, in the light of our statistical tests, a very restrictive assumption. The exponential estimates seem meaningless: the minimum efficiency rate is about 66% and, as indicated by the first quartile Q1, 75% of the sample firms have a rate of technical efficiency greater than 89.4%.

The truncated normal and the generalized half-normal distributions give plausible estimates of technical efficiency, probably because their densities are not restricted to be more concentrated near zero. Which one of them is more reliable remains an open question.

Note that, the 'COLS' estimates stem from the deterministic frontier. They are obtained in compliance with Greene's (1980) suggestion, i.e. using least squares slope estimates on the pooled data and shifting the constant term up until no residual is positive and one is zero. Hence, the finding that 'COLS' method leads to important reduction in technical efficiency is not a surprise since the entire deviation from the frontier is attributed to inefficiency.

The last column of Table 2 reports the minimum value over the sample firms of the coefficient of variation, CV, associated to the estimates of technical inefficiency[14]. It shows that all our estimates of efficiency are statistically quite accurate.

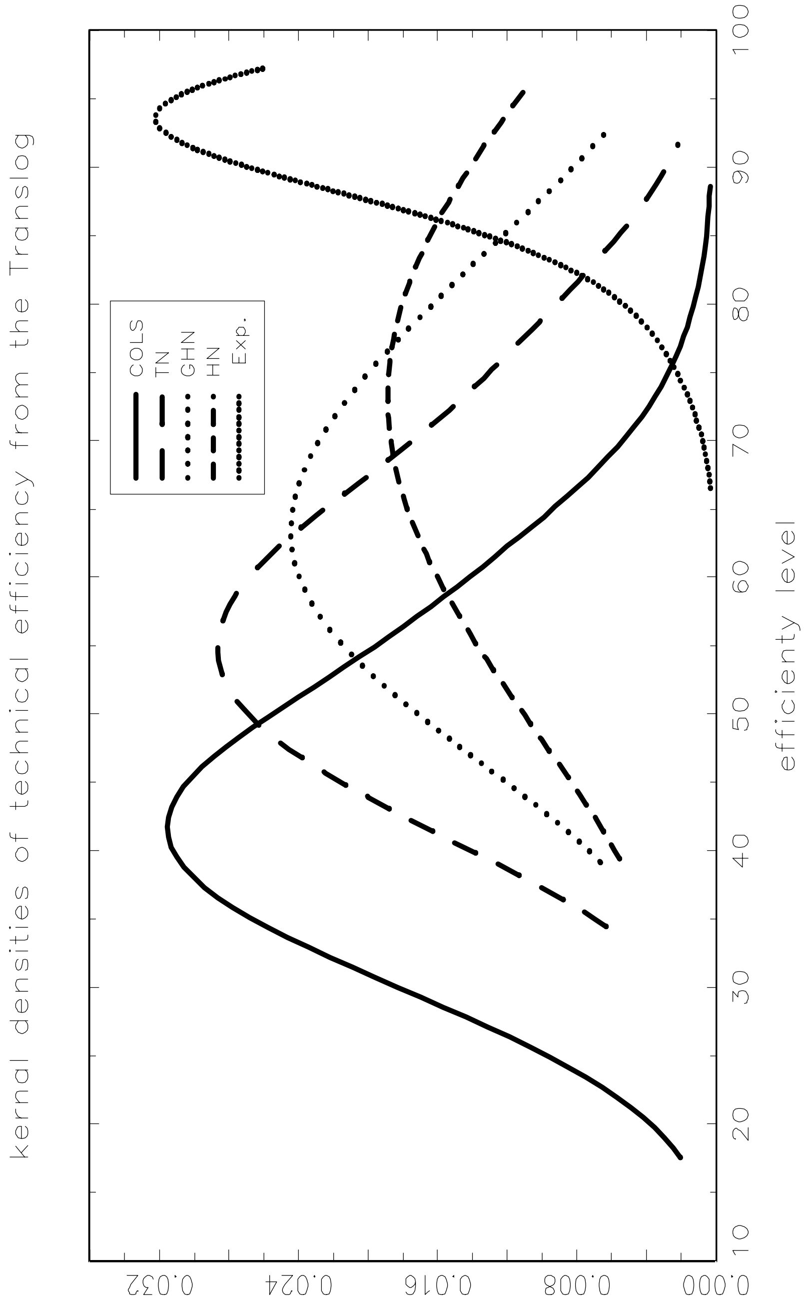
The figure above illustrates the estimated kernal densities of the estimated firm-technical efficiencies. It clearly reinforces our previous finding that different assumptions about the distribution of $u_i$ leads to quite different measures of firm-technical efficiency. In particular, the expected upward bias in the exponential case and the downward bias inherent to the deterministic case, i.e. COLS estimates, are apparent. The estimated densities for the three other cases deviate substantially from each other, with as expected that corresponding to the half-normal assumption is to the right of those of the truncated normal and the generalized half-normal assumptions.

---

[14] CV is the coefficient of variation of the technical efficiency. Note that the variance of technical efficiency is easily calculated for a truncated normal $X$ of mean $\mu$ and variance $\sigma^2$, using the following formula:

$$E(e^{aX}) = \exp(a\mu + .5a^2\ \sigma^2)\frac{\Phi(\frac{\mu}{\sigma}+a\sigma)}{\Phi(\frac{\mu}{\sigma})}.$$

This formula also applies, with a minor change, for the generalized half normal case.



kernal densities of technical efficiency from the Translog
COLS
TN
GHN
HN
Exp.
efficienty level
10
20
30
40
50
60
70
80
90
100
0.000
0.008
0.016
0.024
0.032

# 5 Concluding remarks

This paper provides alternative measures of technical inefficiency depending on what type of distributional assumptions for the one sided error term are considered. These assumptions include the most commonly used one-sided distributions, namely the half-normal, the exponential and the truncated normal distributions. A generalized version of the half-normal, which does not embody the zero-mean restriction, is also suggested.

Comparisons of the estimated technical inefficiency indices induced by all these distributions were attempted using a panel data on Tunisian manufacturing firms over the 1983-1993 time period.

Our main findings can be summarized as follows. Firstly, absence of technical inefficiency is wrongly rejected. This implies that the one-sided error term must be taken into account explicitly when econometric estimation of frontier function is in order. Secondly, estimates of technical efficiency seem to be insensitive to the degree of flexibility of the frontier production function. Indeed, the Cobb-Douglas technology gives almost the same estimates of efficiency as the translog functional form. Thirdly, different assumptions on the distribution of technical inefficiency imply quite different estimates of efficiency.

The estimated inefficiencies suggest that the restricted models, i.e. the exponential and, to a lesser degree, the half-normal, produce a very optimistic impression, with an average rate of efficiency of 92% and 72.5%, respectively. It is worth noting, however, that formal test led to a strong rejection of the zero-mean half-normal model.

Based on the truncated normal or the generalized half normal models, we find that Tunisian firms had been inefficient over the period 1983-1993; the average rate of inefficiency being approximately equal to 40%. Fifty percent of the sampled firms have a rate of efficiency between 58.3% and 66.3% for the truncated normal case; the counterpart is 65.3 and 74.3% for the generalized half normal case.

# Appendix

**Table A1:** Mean values

| Year | $Y$ | $K$ | $L$ | $agek$ |
|---|---|---|---|---|
| 1983 | 653.4 | 4383.7 | 118.1 | 7.2 |
| 1984 | 610.2 | 4241.1 | 105.9 | 7.1 |
| 1985 | 657.1 | 3904.4 | 105.6 | 7.4 |
| 1986 | 688.0 | 4203.8 | 99.5 | 7.8 |
| 1987 | 715.0 | 4405.8 | 101.5 | 8.0 |
| 1988 | 784.9 | 4668.4 | 107.0 | 8.1 |
| 1989 | 821.8 | 4564.9 | 108.3 | 7.9 |
| 1990 | 812.0 | 3912.7 | 105.8 | 7.9 |
| 1991 | 850.6 | 4310.0 | 108.4 | 7.9 |
| 1992 | 822.2 | 3705.9 | 101.9 | 8.4 |
| 1993 | 764.8 | 3548.1 | 93.3 | 8.7 |

**Table A2:** Sector samples by the number of time firms are observed over the period

| $T_i$ | $S1$ | $S2$ | $S3$ | $S4$ | $S5$ | $S6$ | **Total** |
|---|---|---|---|---|---|---|---|
| 1 | 5 | 1 | 17 | 1 | | 3 | 27 |
| 2 | 8 | 6 | 21 | 4 | 3 | 7 | 49 |
| 3 | 9 | 9 | 7 | 3 | 3 | 5 | 36 |
| 4 | 17 | 13 | 24 | 9 | 3 | 8 | 74 |
| 5 | 19 | 15 | 39 | 7 | 13 | 8 | 101 |
| 6 | 24 | 18 | 44 | 15 | 10 | 22 | 133 |
| 7 | 29 | 16 | 49 | 9 | 13 | 26 | 142 |
| 8 | 29 | 18 | 42 | 9 | 17 | 29 | 144 |
| 9 | 28 | 29 | 34 | 13 | 12 | 21 | 137 |
| 10 | 26 | 17 | 46 | 5 | 13 | 15 | 122 |
| 11 | 21 | 24 | 54 | 13 | 21 | 27 | 160 |
| Total | 215 | 166 | 377 | 88 | 108 | 171 | 1125 |

$S1$ = Food, $S2$ = ConstructionMaterials Ceramics and Glass, $S3$ = Mechanical and Electrical, $S4$ = Chemicals, $S5$ = Textiles, $S6$ = Miscellaneous

**Table A3:** Parameters estimates of the translog frontier

| $Var.$ | TN | HN ($\mu \neq 0$) | HN ($\mu = 0$) | EXP | OLS |
|---|---|---|---|---|---|
| $cte$ | 8.407 (652.218) | 8.367 (28.151) | 7.973 (24.750) | 8.003 (37.556) | 8.008 (35.153) |
| $k$ | −0.468 (25.200) | −0.468 (8.866) | −0.425 (6.890) | −0.537 (14.070) | −0.535 (12.897) |
| $l$ | 1.779 (33.600) | 1.778 (29.036) | 1.734 (24.734) | 1.857 (40.654) | 1.849 (37.465) |
| $agek$ | 0.045 (4.422) | 0.045 (4.217)) | 0.043 (2.504) | 0.098 (8.829) | 0.084 (7.344) |
| $t$ | −0.031 (3.009) | −0.031 (2.952°) | −0.031 (2.598) | −0.051 (3.747) | −0.043 (3.346) |
| $0.5k^2$ | 0.099 (29.459) | 0.098 (17.907) | 0.094 (13.797) | 0.105 (25.826) | 0.105 (23.670) |
| $kl$ | −0.117 (18.745) | −0.117 (16.733) | −0.112 | −0.114 (21.766) | −0.115 (20.291) |
| $k \times agek$ | −0.009 (9.256) | −0.009 (8.915) | −0.009 (5.680) | −0.014 (13.204) | −0.012 (11.688) |
| $k \times t$ | 0.010 (10.380) | 0.010 (10.251) | 0.010 (9.387) | 0.011 (9.192) | 0.011 (9.356) |
| $0.5l^2$ | 0.115 (10.078) | 0.115 (9.414) | 0.112 (8.100) | 0.091 (10.350) | 0.097 (10.211) |
| $l \times agek$ | 0.017 (10.145) | 0.017 (9.957) | 0.016 (8.381) | 0.017 (11.556) | 0.017 (10.944) |
| $l \times t$ | −0.012 (8.547) | −0.012 (8.406) | −0.012 (8.545) | −0.012 (6.952) | −0.012 (7.383) |
| $0.5inc^2$ | −0.001 (1.726) | −0.001 (1.749) | −0.001 (1.748) | −0.000 (1.409) | −0.000 (1.184) |
| $agek \times t$ | −0.002 (4.213) | −0.002 (4.186) | −0.002 (3.783) | −0.002 (5.275) | −0.002 (4.955) |
| $0.5t^2$ | −0.002 (2.556) | −0.002 (2.605) | −0.002 (2.608) | 0.001 (0.735) | −0.001 (0.594) |
| $s1$ | 0.116 (9.045) | 0.115 (6.374) | 0.139 (6.886)) | 0.128 (12.634) | 0.124 (10.846) |
| $s2$ | 0.254 (21.116) | 0.254 (13.235) | 0.259 (11.130) | 0.264 (22.904) | 0.260 (20.047) |
| $s4$ | 0.082 (5.390) | 0.081 (4.143) | 0.055 (2.108) | 0.102 (7.279) | 0.100 (6.341) |
| $s5$ | −0.207 (18.089) | −0.210 (11.620) | −0.188 (7.531) | −0.203 (15.722) | −0.211 (14.638) |
| $s6$ | −0.002 (0.205) | −0.002 0.117 | −0.031 (1.804) | 0.021 (1.989) | 0.015 (1.205) |
| $\mathcal{L}$ | 1557.7 | 1557.3 | 1624.4 | 4292.4 | |

**Table A4:** Parameter estimates of the Cobb-Douglas frontier

| $Var.$ | TN | HN ($\mu \neq 0$) | HN ($\mu = 0$) | EXP | OLS |
|---|---|---|---|---|---|
| $cte$ | 5.038 (63.941) | 4.994 (452.851) | 4.781 (94.042) | 4.308 (125.658) | 4.277 (114.086) |
| $k$ | 0.378 (63.725) | 0.378 (77.256) | 0.380 (77.235)) | 0.388 (114.177) | 0.387 (100.718) |
| $l$ | 0.718 (83.231) | 0.718 (67.445) | 0.719 (101.520) | 0.741 (152.019) | 0.738 (135.091) |
| $agek$ | $-0.029$ (18.002) | $-0.029$ (2.911) | $-0.029$ (18.710) | $-0.039$ (31.104) | $-.038$ (28.748) |
| $t$ | 0.033 (29.187) | 0.033 (6.755) | 0.033 (28.979) | 0.036 (24.661) | 0.035 (26.477) |
| $s1$ | 0.087 (4.299) | 0.086 (7.794) | 0.097 (5.380) | 0.093 (9.104) | 0.090 (7.629) |
| $s2$ | 0.267 (10.653) | 0.267 (24.179) | 0.285 (13.161) | 0.276 (23.734) | 0.270 (20.255) |
| $s4$ | 0.062 (1.740) | 0.062 (5.624) | 0.007 (0.264) | 0.079 (5.560) | 0.079 (4.882) |
| $s5$ | $-0.234$ (8.018) | $-0.237$ (21.425) | $-0.223$ (9.791) | $-0.238$ (18.355) | $-0.246$ (16.784) |
| $s6$ | $-0.017$ (1.720) | $-0.019$ (1.568) | 0.014 (0.712) | 0.008 (0.857) | 0.001 (0.053) |
| $\mathcal{L}$ | 1761.8 | 1760.7 | 1836.1 | 4705.4 | |

# References


[1] Aigner, D.J., Lovell, C.A.K. and Schmidt,P. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", Journal of Econometrics, 6, 21- 37.

[2] Aigner, D.J. and S. Chu (1968), "On estimating the Industry Production Function", American Economic Review, 55, PP. 826-839.

[3] Battese, G.E. and Coelli, T.J. (1988), "Prediction of Firm-Level Technical Efficiencies With a Generalised Frontier Production Function and Panel Data", Journal of Econometrics, 38, 387-399.

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

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

[6] Battese, G.E., Coelli, T.J. and Colby, T.C. (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, 327-348.

[7] Bauer, P.W. (1990), "Recent Developments in the Econometric Estimation of Frontiers", Journal of Econometrics, 46, 39-56.

[8] Christensen, L.R., D.W. Jorgensen and L.J. Lau(1973),"Transcendental Logarithmic Production Frontiers," Review of Economics and Statistics, 55, pp. 28-45.

[9] Cornwell C., P. Schmidt, and R.C. Sickles(1990), Production Frontiers With Cross-Sectional and Time-Series Variations in Efficiency Levels, Journal of Econometrics 46.

[10] Farrell, M. (1957), "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society A, general 120, pt. 3, pp. 253-281.

[11] Forsund, F.R., Lovell, C.A.K. and Schmidt, P. (1980), "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement", Journal of Econometrics, 13, 5-25.

[12] Greene, W.H. (1994), "Frontier Production Functions", in Fried, H.O., Lovell, C.A.K. and Schmidt, S.S.(Eds), The Handbook of Applied Econometrics, North Holland, vol. II.

[13] Hausman , J., and W. Taylor(1981),"Panel Data and Unobservable Individual Effects", Econometrica, 49, pp. 1377-1398.

[14] Jondrow, J.,. Lovell, C.A.K Materov, I.S. and Schmidt, P. (1982), "On estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model", Journal of Econometrics, 19, 233-238.

[15] Kalirajan, K., and J. Flinn(1983)"The Measurement of Farm Specific Technical Efficiency", Pakistan Journal of Applied Economics, 2, pp. 167-180.

[16] Kumbhakar S.C.(1990), Production Frontiers, Panel Data, and Time Varying technical inefficiency, Journal of Econometrics 46.

[17] Kodde David A. and Franz C. Palm (1986),"Wald Criteria for Jointly Testing Equality and Inequality Restrictions," Econometrica, 54, 1243-1248.

[18] Lee, L.,(1993), "A Test for Distributional Assumptions for a Stochastic Frontier Function", Journal of Econometrics, 22, pp. 245-267.

[19] Meeusen, W. and van den Broeck, J. (1977), "Efficiency Estimation from Cobb- Douglas Production Functions With Composed Error", International Economic Review, 18, 435-444.

[20] Schmidt, P. (1986), "Frontier Production Functions", Econometric Reviews, 4, 289- 328.

[21] Schmidt, P. and Lovell, C.A.K. (1980), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers", Journal of Econometrics 13, 83-100.

[22] Schmidt, P. and R.C. Sickles(1984), Production Frontiers and Panel Data, Journal of Business and Economic Statistics 2, 367-374.

[23] Solow, R. (1957),“Technical Change and the Aggregate Production Function”, Reviw of Economics and Statistics, 39, 312-320.

[24] Stevenson, R.E. (1980), ”Likelihood Functions for Generalised Stochastic Frontier Estimation”, Journal of Econometrics, 13, 57- 66.

[25] Zelner A., J. Kmenta and J. Drèze(1966),“Specification and Estimation of Cobb-Douglas Production Models,” Econometrica, 34, pp. 784-795.

[26] Zribi Y. (1995), “La dynamique du capital fixe productif des entreprises tunisiennes”, mimeo IEQ, N. 02-95.