

Documento de trabajo

E2004/31

A spreading method to improve efficiency prediction

centrA: Fundación
Centro de
Estudios
Andaluces

**Rafaela Dios-Palomares
Jose Miguel Martínez Paz**



centrA: Fundación
Centro de
Estudios
Andaluces

A spreading method to improve efficiency prediction.

Rafaela Dios-Palomares^{1,3}
Universidad de Córdoba

Jose Miguel Martínez Paz^{2,3}
Universidad de Murcia

RESUMEN

En los análisis de eficiencia mediante modelos de frontera estocástica, la variable de error compuesto incluye el componente de ineficiencia, lo cual hace que las predicciones individuales no puedan ser hechas directamente por medio del error estimado por estos modelos. Para resolver este problema, Jondrow et al (1982), y Battese y Coelli (1988) desarrollaron, de forma separada, dos procedimientos diferentes, basados en la esperanza de la distribución condicional. Aunque los dos estimadores son diferentes, ambos manifiestan el problema de presentar una concentración con respecto a la distribución de eficiencia teórica.

El estudio de las propiedades de ambos estimadores permiten concluir que el valor del parámetro gamma tiene una gran influencia en este efecto, produciendo un truncamiento de la distribución que puede ser mayor del 50%, de manera que los valores extremos de la eficiencia nunca serán pronosticados por los estimadores considerados.

En este trabajo se propone un procedimiento que extienda las eficiencias con el fin de minimizar el efecto de la concentración. El estudio de MonteCarlo desarrollado demuestra que los valores corregidos con este procedimiento tiene un comportamiento mejor que los estimadores originales.

Palabras clave: Eficiencia, Modelos frontera, Métodos Montecarlo

ABSTRACT

In efficiency analysis by means of a stochastic frontier production function, the composite error variable includes the inefficiency component. For this reason, individual prediction cannot be made directly from an estimation of the error in the model. In order to solve this problem, Jondrow et al (1982), and Battese and Coelli (1988) separately developed two different procedures, based on the expectation operator of the conditional distributions. Although the two predictors are different, each suffers from a shrinkage effect with respect to the distribution of theoretical efficiency.

Our study of the behaviour of these two predictors leads us to conclude that the value of the gamma parameter has a great influence on the above-mentioned effect, producing a truncation of the distribution that could be more than 50%, so that the extreme values of the efficiency can never be estimated by the predictors considered.

We also propose a method that spreads out the predicted efficiencies in order to minimise the shrinkage effect. The Monte Carlo results demonstrate that the corrected predictions have a better behaviour than the original predictors.

Key words: Efficiency, Frontier models, Monte Carlo methods.

JEL classification: C15

¹ Departamento Estadística. Escuela Técnica Superior de Ingenieros Agrónomos y Montes. Universidad de Córdoba. Avda. Menendez Pidal s/n, 14080 Córdoba (Spain), Tfn. 34 957218479, rdios@uco.es

² Departamento de Economía Aplicada. Facultad de Economía y Empresa. Universidad de Murcia. Campus de Espinardo - 30100 Murcia (Spain) Tel: 34 96836931, jmpaz@um.es

³ Grupo de Eficiencia y Productividad EFIUCO, Universidad de Córdoba, España
<http://www.uco.es/grupos/efiuco/>

Grupo de Eficiencia y Productividad de centra

1. Introduction

This paper presents the results of a study that aimed to evaluate an important problem that emerges in the framework of technical efficiency analysis that employs the stochastic frontier model. The Stochastic Frontier Production Function was proposed by Aigner, Lovell and Schmidt (1977) and independently by Meeusen and Van den Broeck (1977). The general model of an SFP function is as follows:

$$y_i = X_i * \mathbf{b} + e_i \quad (i=1,2,\dots, N) \quad \varepsilon_i = v_i - u_i$$

where y_i denotes the output for observation i , X_i is the vector of inputs for observation i , \mathbf{b} is a vector of parameters, N is the sample size, and the variable error e_i collects the differences between the observed values and the systematic part of the model.

The composed error assumes that the error variable e is generated as the difference between a stochastic variable v (not controllable, symmetric, and defined between $-\infty$ and ∞) and the inefficiency variable u , which will always be positive and asymmetric. The V variable is represented by a normal distribution with a mean of zero and variance s_v^2 , while the distributions usually employed to specify the u component are the half-normal and the truncated normal.

Once a specification for the u component has been assumed, the production frontier model is estimated by assigning the corresponding distribution to e as the difference between v and u . The residues are obtained as an immediate result of the estimation. ($e_i = y_i - X_i * \hat{\mathbf{b}}$), and they can be regarded as estimates of the error terms. However, the problem of separating these estimates into separate estimates of the constituent parts persisted until 1982.

The study of efficiency by means of the SFP model was significantly advanced by the contributions of Jondrow et al. (Henceforth, JLMS, 1982) and Battese and Coelli (BC, 1988) who produced expressions involving conditional distributions given e_i , to draw inferences about individual efficiencies. These estimators were a cornerstone of this methodology since they allow matching and benchmarking of results. However, they are widely used regardless of their possible shortcomings.

The JLMS formulation is as follows:

The conditional distribution of u_i , given e_i , is $N(\mu^*, \sigma^{*2})$ variable truncated at zero.

The distributions of v_i and u_i are:

$$v_i \approx N(0, s_v^2) \quad \text{and} \quad u_i \approx \left[N(0, s_u^2) \right], \text{ where}$$

$$s^2 = s_u^2 + s_v^2 \quad m_* = -\frac{s_u^2 e}{s^2} \quad s_*^2 = \frac{s_u^2 s_v^2}{s^2}$$

and “ \approx ” means “distributed as” and m_* and s_*^2 are the mean and the variance of the conditional distribution of u given e , respectively.

Hereafter, we drop the subscript (i) for the sake of simplicity.

Jondrow et al. defined I as $I = \frac{s_u}{s_v}$ and produced the following two point estimators:

$$E(u|e) = s_* \left[\frac{f\left(\frac{eI}{s}\right)}{1 - F\left(\frac{eI}{s}\right)} - \left(\frac{eI}{s}\right) \right]$$

$$M(u|e) = -e \left(\frac{s_u^2}{s^2} \right) \quad \text{if } e \leq 0$$

$$= 0 \quad \text{if } e > 0$$

where “ $|$ ” means “conditioned to” and $f(\cdot)$ and $F(\cdot)$ are the probability density and distribution functions, respectively, of a standard normal random variable.

The results of recent Monte Carlo experiments in which the JLMS performance was analysed can be found in Coelli (1995), Kumbhakar and Löthgren (1998) and Dios-Palomares et al. (2002). In these papers the bias and variance of the mean efficiency estimate were studied, considering the gamma parameter as a source of variation. The conclusion was that both the bias and the precision present their worst results when gamma assumes central values: i.e. near 0.5.

Battese and Coelli (1988) proposed the alternative point estimator for individual estimated efficiency:

$$E(\exp\{u\}|e) = \left[\frac{1 - F\left(\frac{s_* - m_*}{s_*}\right)}{1 - F\left(-\frac{m_*}{s_*}\right)} * \exp\left(-m_* + \frac{1}{2}s_*^2\right) \right]$$

With regard to the BC predictor, we have previously performed comparative Monte Carlo experiments (Dios-Palomares et al, 2003) and observed its unbiased behaviour which enables it to perform better than JLMS estimator.

Nevertheless, both predictors suffer from a severe problem of shrinkage, due to the nature of the procedure employed, which treats the expectation as the point estimate of the whole conditioned distribution (See Figure 1). Therefore, both the rank and the

variance of the empirical estimated efficiency will always be lower than those of the true efficiency.

We regard this as an important deformation in the predicted distribution of individual efficiency. For this reason we have performed a Monte Carlo experiment in order to achieve two objectives. First, we evaluated the shrinkage effect, after which we proposed a spreading method that corrects the above-mentioned effect. The following section explains the methodology. Section 3 reports the results and section 4 presents our main conclusions.

2.- Methodology

Design of the Monte Carlo study

The main parameters of the experiment are: s^2 (the variance of e) and γ (variance ratio). The random terms $v_i, i=1,\dots,N$, are drawn from a normal distribution $N(0, s_v^2)$ and the technical inefficiency terms u_i , from the half normal distribution. e_i is obtained by means of the expression $e_i = v_i - u_i (i=1,2,\dots,N)$.

No regression is involved and we assume that the usually unobservable e and its two components are known beforehand.

Ten values of s^2 ($s^2=0.1,0.2,\dots,1$) and nine variance ratios γ ($\gamma=0.1,0.2,\dots,0.9$) are considered in the research. In each combination, 10000 Monte Carlo reiterations are involved.

The subscript J denotes the JLMS method. The subscript B denotes the BC approach.

In order to compare the performance of the estimators under study, true efficiency distributions were attained by raising e to $-u$. Efficiency estimated values were also obtained by means of the JLMS and BC expressions, taking the corresponding ε values as if they had been known in advance.

The evaluation measures

In order to evaluate the shrinkage effect we calculated the following two measures:

* The Rank is the difference between the 97.5 and 2.5 percentiles of each empirical distribution:

$R_j = \text{Effj}_{97.5} - \text{Effj}_{2.5}$ $j = \text{TE, J, and BC}$

* The Coverage is the ratio between both the estimated index and the true efficiency ranks:

$$CJ = \frac{RJ}{RT} ; CB = \frac{RB}{RT}$$

The spreading method

We have based the spreading method on the establishment of a correspondence between the estimated and true efficiency distributions. We done this by standardising both distributions.

We shall refer to the true and estimated efficiencies and their corresponding means and standard deviations (SD) by the designations given them in Table 1.

Table 1: Variables and Statistics Designations

	Name	Mean	Std. Deviation
True efficiency	TE	TM	TSD
Estimated index	EI	EM	ESD

The standardised true efficiency is expressed by $\frac{TE - TM}{TSD}$

and the standardised estimated index by $\frac{EI - EM}{ESD}$

With the above-mentioned correspondence established, we can write the following expression:

$$\frac{TE - TM}{TSD} = \frac{EI - EM}{ESD}$$

Furthermore , we may assume that the best predictor of efficiency would be the one most similar to the true efficiency. We therefore replace the latter in the formulated equation with the corrected efficiency and then isolate it as a function of the remainder of the known parameters, which gives us the following expressions:

$$\frac{CE - TM}{TSD} = \frac{EI - EM}{ESD}$$

$$CE = \left(\frac{EI - EM}{ESD} \right) * TSD + TM$$

This last expression enables us to spread the estimated index in order to correct the shrinkage effect.

3 Results

In order to attain the first objective, we calculated and studied the empirical distributions of the three measures of efficiency that we are attempting to compare: the True Efficiency (ET), and the J and BC predictors.

In order to analyse the results, we have represented the three empirical distributions in the same graph. We also present three different figures (2, 3 and 4) corresponding to different values of the parameters which we think may influence the shrinkage effect: i.e. s^2 and γ .

With respect to the true efficiency, it is worth noting that parameters γ and s^2 both have an influence on its distribution, the real effect having been produced by the σ_u^2 parameter, which is the result of multiplying the above-mentioned parameters.

The graphs show that when σ_u^2 is low, the empirical density of the efficiency is found in the highest zone of the (0,1) interval, while no values occur in the zone next to zero. Nevertheless, from σ_u^2 about 0,1, the empirical density is spread over the entire (0,1) range. This implies that the mean efficiency is, in some way, related to the σ_u^2 parameter; the lower this is, the higher the variance of u.

With respect to the behaviour of the distributions of the predictors, the graphs show that the above-mentioned distributions have suffered a considerable degree of shrinkage with respect to the true efficiency distribution.

Looking at the histograms of both predictors, it is clear that there is no significant difference between their performances in terms of the shrinkage effect, although it should be noted once again that the J predictor has a bias. As far as the shrinkage effect is concerned, the figures show that γ is the more influential parameter. Figure 5 illustrates the ranks of the three distributions studied.

The figure shows that the performances of the two predictors are certainly similar. We can also see that the differences between the ranks of the true efficiency

and that of the predicted one depend on the γ value, these differences being higher, the lower the value of γ .

The coverage measure is also shown in Figure 6. Once again, it can be seen that the two predictors studied perform in a similar fashion. The evaluation of the shrinkage effect as a function of γ can also be observed. The graph shows that coverage may be only 35% when γ is low (0.1).

This result can be rather dramatic, given that it means that only 35% of the true distribution is predicted, in other words, that the shrinkage effect is 65%.

These results suggest that it might be of interest to devise a method of correcting this shrinkage effect.

Turning our attention to the second objective, we employed a spreading method to correct the estimated indices obtained in our Monte Carlo experiment, using the calculated empirical distribution, and Tables 2 and 3 show the results for two particular experimental points. These two tables correspond to the experimental points $\sigma^2 = 0.1$; $\gamma=0.1$ and $\sigma^2 = 1$; $\gamma=0.9$, respectively.

Table 2: Empirical distribution statistics for $s^2 = 0.1$ and $g=0.1$

$\sigma^2 = 0,1$ $\gamma = 0,1$	EJ	EB	CEJ	CEB	ET
Mean	0,924	0,925	0,927	0,927	0,927
Std dev	0,011	0,010	0,053	0,053	0,053
Min	0,872	0,875	0,665	0,664	0,703
Max	0,950	0,951	1,061	1,060	1,000

Table 3: Empirical distribution statistics for $s^2 = 0.9$ and $g=1$

$s^2 = 0,9$ $g = 1$	EJ	EB	CEJ	CEB	ET
Mean	0,523	0,538	0,539	0,539	0,539
Std dev	0,211	0,212	0,245	0,245	0,245
Min	0,035	0,037	0,006	0,000	0,036
Max	0,896	0,900	0,972	0,957	0,999

The figures in these tables show the main statistics for five different distributions that correspond to the two estimated efficiencies, the two corrected efficiencies and the

true efficiency. As we have pointed out, the less the value of γ , the greater the shrinking effect..

It is quite clear that the spreading method produces corrected efficiencies with distributions very close to the true efficiency as we expected. As we can see, both the mean and the standard deviation of the corrected indices have changed. In particular, we note that the standard deviations have increased and that their values are similar to the SD of the true efficiency.

Figure 7 presents a comparative representation of the five distributions. In order to illustrate the spreading method, we applied it to the data set used by Coelli in his Frontier 4.1 User's Manual (1996). We also calculated the J index by running the the same data set through the Limdep program (2000).

The spreading method requires the same information about the statistics of true efficiency and the corresponding estimated index. The statistics of the estimated index have been calculated directly from their estimated values as these appeared in the output file of the relevant software.

From the same output file we also took the estimated parameters σ^2 and γ . We then used the empirical values of the statistics that we obtained as a result of our Monte Carlo experiment, given these particular values for σ^2 and γ . In the case of the J index we had to calculate γ from the values of σ^2 and σ_u^2 .

Table 4 shows the values of the statistics used for the Coelli data set.

Table 4: Statistics and parameters

	Expectation	S. Deviation	σ^2	σ_u^2	γ
J et al	0.7324	0.1129	0.22	0.1789	0.8133
B y C	0.7406	0.1284	0.22		0.8
Empirical True Efficiency	0.735	0.1652	0.20		0.8

For the J index we extracted the values for the expectation, the SD, and the estimated parameters σ^2 and σ_u^2 from the output file of the Limdep software and calculated:

$$\gamma = (\sigma_u^2 / \sigma^2) = 0.8133$$

In the case of the BC index the values for the expectation, the SD and the estimated parameters ($\sigma^2=0.2$ and $\gamma=0.8$) were obtained directly from the Frontier 4.1 output file.

Then, given $\gamma=0.8$ and $\sigma^2=0.2$, our Monte Carlo results show that the expectation and SD of the empirical true efficiency distribution are 0.735 and 0.1652 respectively (see Appendix).

Table 5 collects the main statistics of the two estimated indices and their corresponding corrected indices. As we can see, they have been spread in the direction of the true efficiency distribution.

Table 5: Statistics of the original and corrected indices

	EJ	CEJ	EB	CEB
Mean	0.732	0.735	0.7406	0.7352
Std dev	0.129	0.1652	0.1284	0.1653
Min	0.345	0.2419	0.3513	0.2343
Max	0.935	0.9941	0.9374	0.9884
Count	60	60	60	60

This effect can also be seen in the Figures 8 and 9, where we have paired each index with its corrected equivalent.

4. Conclusions

In this study we have collected the results of research that was carried out with the aim of evaluating and correcting the shrinkage effect that appears when the Jondrow et al. and Battese and Coelli predictors are employed in a stochastic frontier framework.

We regard this as an important deformation of the predicted distribution of individual efficiency. For this reason we performed a Monte Carlo experiment in order to achieve two objectives. First we evaluated the shrinkage effect after which we suggested a spreading method to correct this effect.

We conclude that a marked shrinkage effect occurs in the estimated empirical distributions when the analysed predictors of individual efficiency are employed, and that both the BC and the JLMS predictors perform similarly with respect to the above-mentioned effect.

We also conclude that the shrinkage effect is primarily dependent on the γ parameter as follows: the lower the γ , the lower the coverage value.

With regard to the coverage measure, we point out that a 60% truncation in the data may occur when $\gamma = 0.1$.

Applying the proposed spreading method to the Coelli data set clearly corrects the shrinkage effect.

5. References

Aigner, D., Lovell, C., Schmidt, P., 1977, Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 21-37.

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

Coelli, T.J., 1996. A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, CEPA Working Papers, No. 7/96, ISBN 1 86389 4950, Department of Econometrics, University of New England, Armidale, pp. 33

Dios-Palomares, R., Ramos, A. and Roldán, J.A., 2002. A Monte Carlo Study On The Technical Efficiency Estimation In The Stochastic Frontier Model. *Quesito* v. 26, nº 3: 443-459.

Dios-Palomares, R., and Ramos, A., 2003. A Monte Carlo comparative analysis on the technical Efficiency Predictors. Working Papers Series no. 12/03. Department of Statistics, University of Cordova, Spain.

Jondrow, J., C. A. K. 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, 233-238.

Kumbhakar, S., and Löthgren M., 1998. A Monte Carlo Analysis of Technical Inefficiency Predictors. Working Paper Series in Economics and Finance, N° 229. Stockholm School of Economics.

Limdep v. 8 and Nlogit v. 3, 2000. Econometric Software, Inc. NY.USA

Meeusen, W., Broeck, J. van den, 1977, Efficiency Estimation from Cobb-Douglas Production Functions With Composed Error, *International Economic Review*, 18, 435-444.

Appendix .- Empirical Statistics of the Theoretical Efficiency

$\frac{?}{s^2}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1,0	Statistics
0.1	0.927	0.899	0.878	0.860	0.844	0.833	0.823	0.819	0.804	0.793	Mean
	0.053	0.070	0.084	0.099	0.108	0.114	0.117	0.127	0.132	0.138	Std dev
0.2	0.898	0.861	0.834	0.812	0.792	0.777	0.761	0.748	0.736	0.727	Mean
	0.073	0.092	0.113	0.127	0.139	0.149	0.156	0.161	0.167	0.173	Std dev
0.3	0.878	0.837	0.800	0.775	0.747	0.738	0.708	0.709	0.694	0.676	Mean
	0.085	0.113	0.138	0.145	0.162	0.169	0.177	0.180	0.190	0.193	Std dev
0.4	0.858	0.810	0.782	0.749	0.728	0.708	0.690	0.668	0.658	0.641	Mean
	0.098	0.129	0.143	0.159	0.170	0.182	0.191	0.193	0.211	0.215	Std dev
0.5	0.847	0.795	0.755	0.721	0.713	0.680	0.659	0.654	0.633	0.620	Mean
	0.107	0.136	0.162	0.169	0.182	0.195	0.202	0.202	0.215	0.223	Std dev
0.6	0.835	0.779	0.737	0.698	0.674	0.657	0.635	0.617	0.606	0.591	Mean
	0.116	0.149	0.171	0.187	0.191	0.206	0.210	0.216	0.222	0.229	Std dev
0.7	0.818	0.760	0.727	0.693	0.666	0.631	0.622	0.606	0.594	0.572	Mean
	0.121	0.154	0.178	0.187	0.198	0.212	0.220	0.226	0.229	0.235	Std dev
0.8	0.816	0.735	0.700	0.668	0.634	0.622	0.592	0.588	0.571	0.553	Mean
	0.127	0.165	0.183	0.196	0.207	0.217	0.232	0.232	0.238	0.238	Std dev
0.9	0.802	0.730	0.693	0.643	0.622	0.612	0.589	0.570	0.558	0.539	Mean
	0.133	0.167	0.188	0.204	0.220	0.224	0.228	0.231	0.241	0.245	Std dev

Figure 1: The shrinkage effect

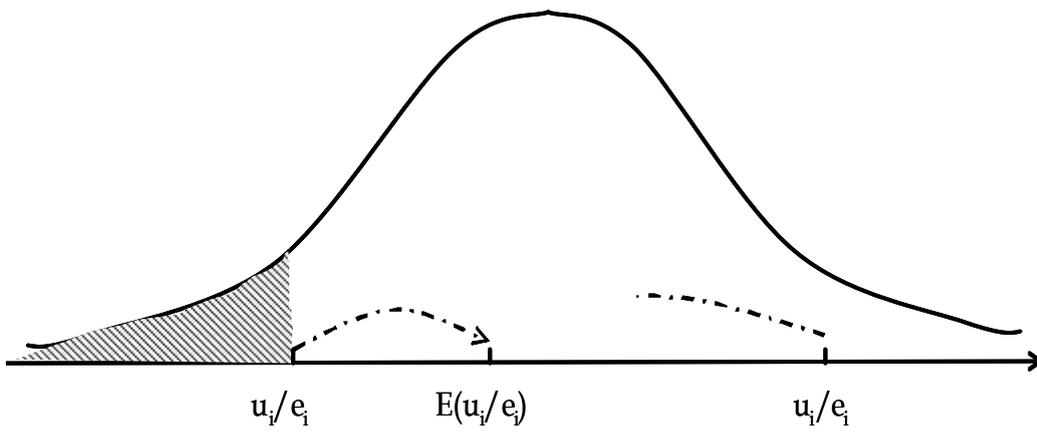


Figure 2.- Histogram of efficiency $s^2=0.1 ; g=0.1 ; s_u^2=0.01$

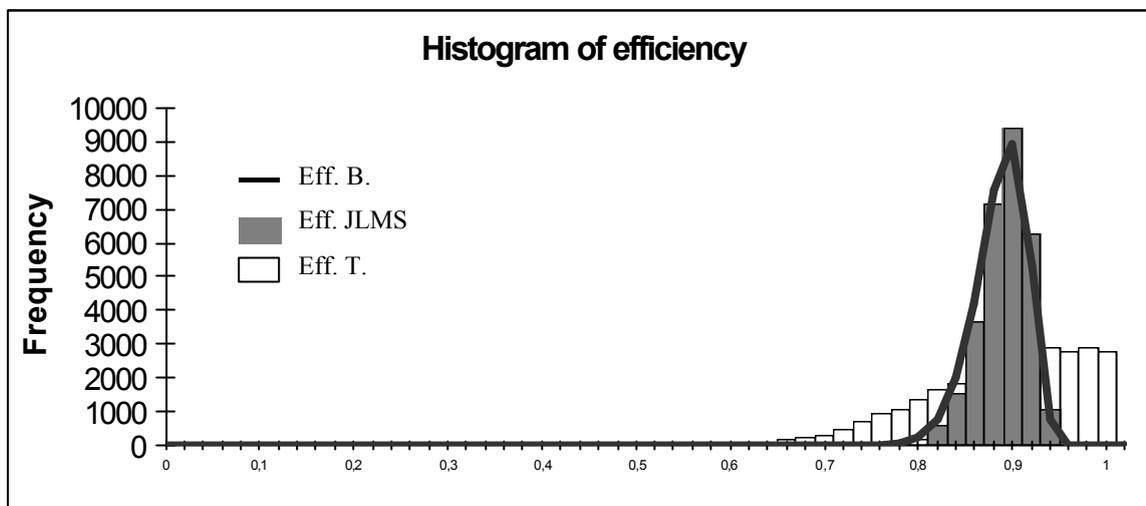


Figure 3.-Histogram of efficiency $s^2=1$; $g=0.1$; $s_u^2=0.1$

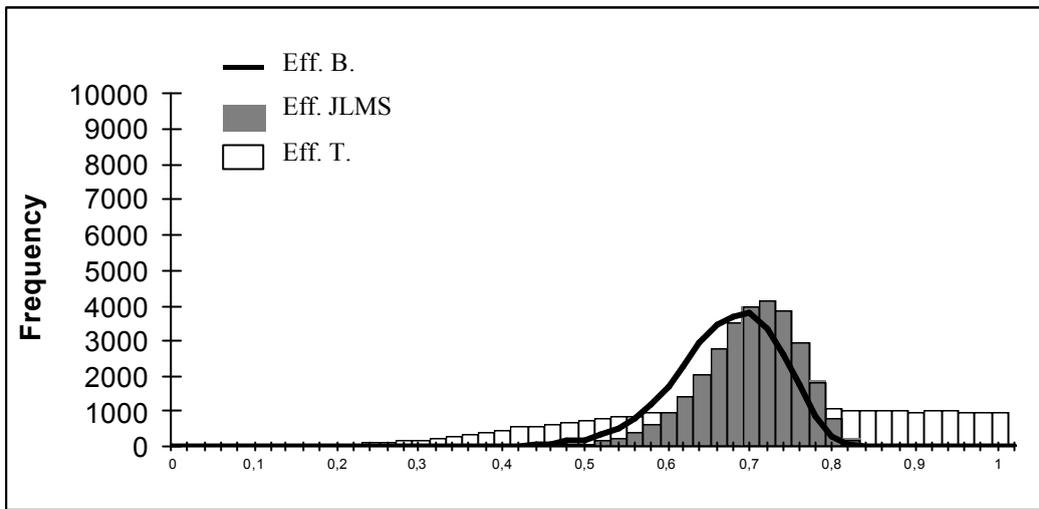


Figure 4.-Histogram of efficiency $s^2=1$; $g=0.9$; $s_u^2=0.9$

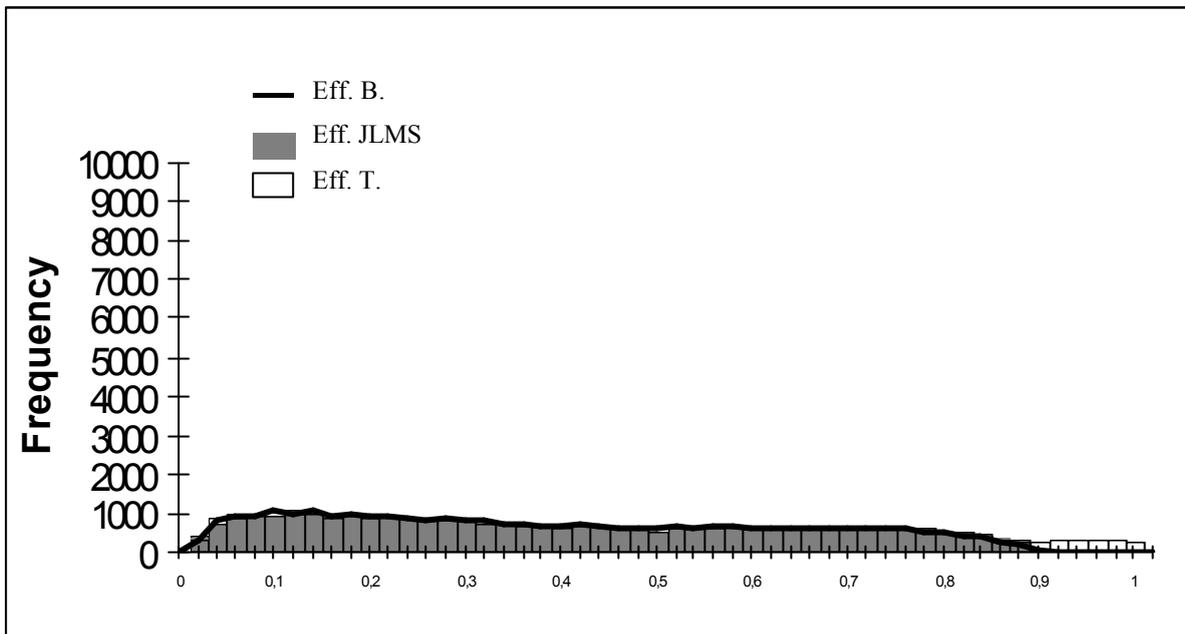


Figure 5: Rank $s^2 = 0.1$

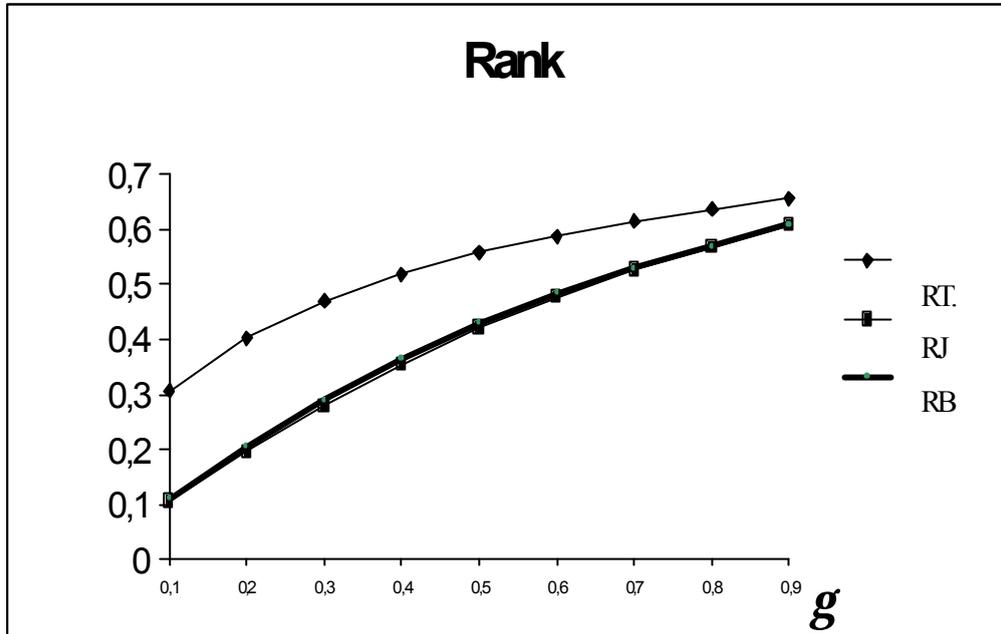


Figure 6: Coverage $s^2 = 0.1$

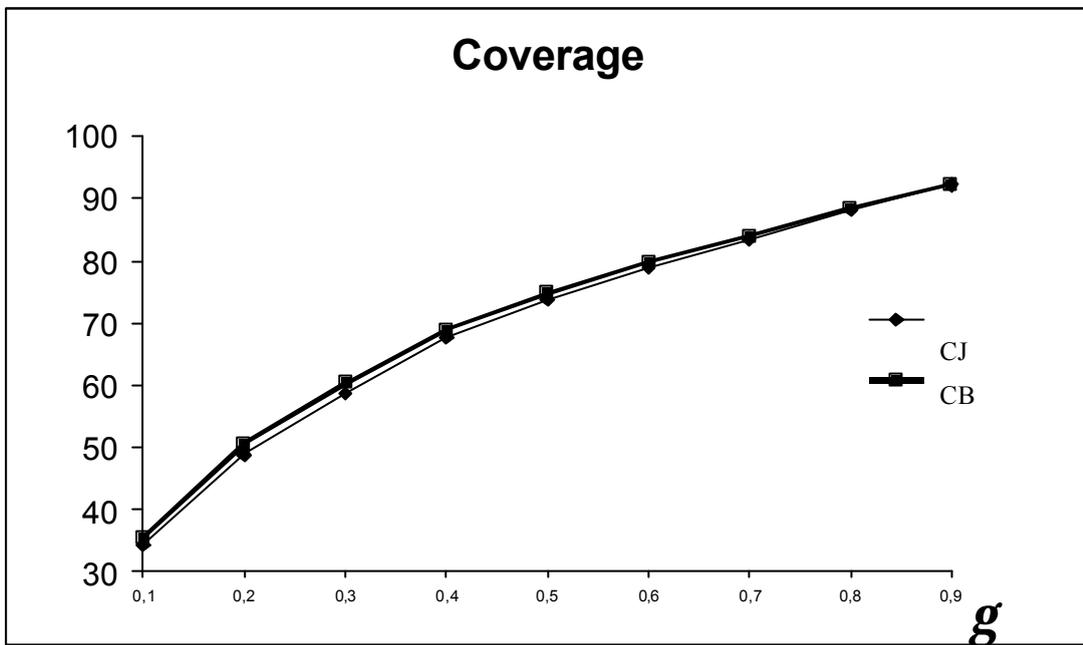


Figure 7: Comparative representation of the five distributions

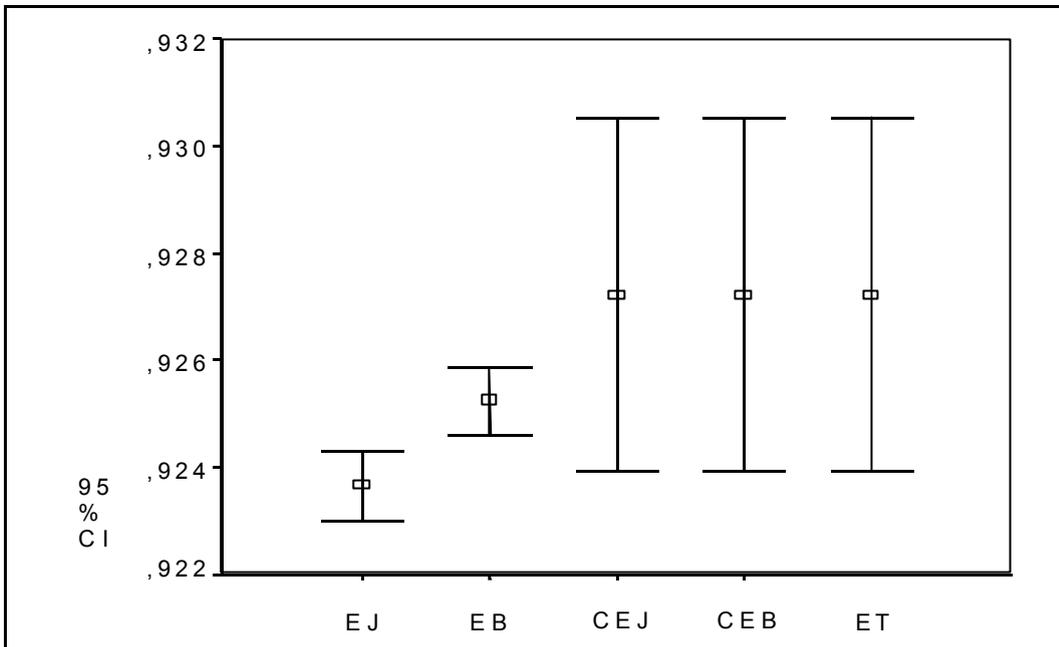


Figure 8: Original and corrected Jondrow et al. indices

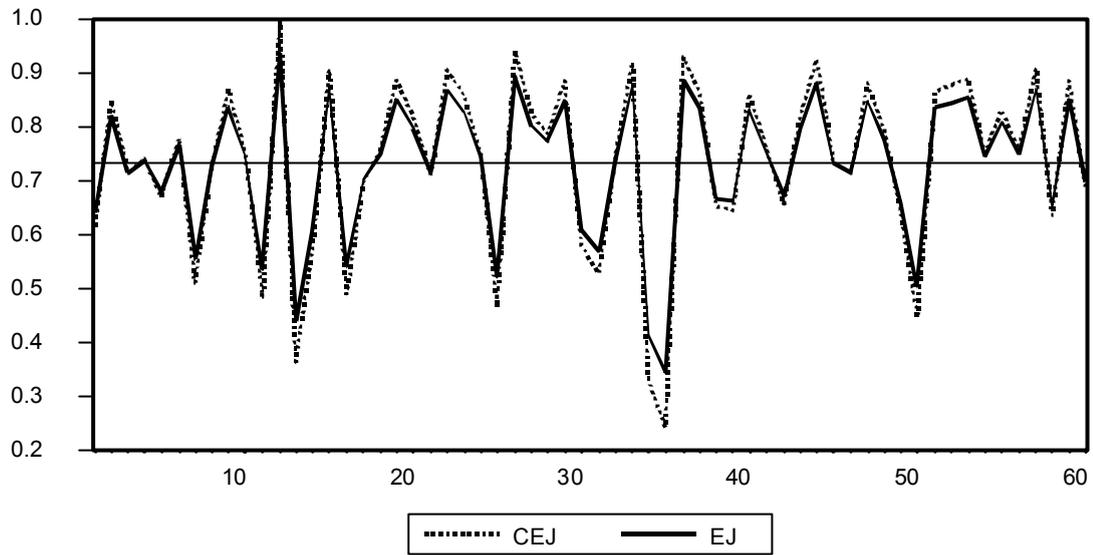


Figure 9: Original and corrected Battese and Coelli indices

