Assessing the Significance of Covariates in Output Oriented Data Envelopment Analysis with Two Stage Regression Models

GERALDO DA SILVA E SOUZA
ELIANE GONÇALVES GOMES
Brazilian Agricultural Research Corporation (Embrapa)
PqEB, Av. W3 Norte final, 70770-901, Brasília, DF
BRAZIL
geraldo.souza@embrapa.br
eliane.gomes@embrapa.br

Abstract - We propose alternative regression models to assess the effects of covariates in output oriented DEA scores. We use probability choice models combined with specifications related to the gamma and to the truncated normal families of distributions. These specifications imply different two stage regression models and alternative quasi maximum likelihood estimators. We apply these methods to assess the significance of technical effects – type of unit, processes improvement and technology impact – affecting DEA efficiency scores computed to agricultural research production in Brazil, measured through Embrapa. We favor models taking into account the whole sample of efficiency scores and not only the inefficient units. In our application we conclude that type of unit and processes improvement are significant effects, with the latter effect negatively associated to the efficiency of the units classified into the type of unit considered more efficient.

Key-words - Data envelopment analysis; Fractional regression models; Second stage DEA regressions; Bootstrap; Contextual variables; Agricultural research.

1. Introduction

In many applications of DEA one is faced with a setting where the main objective is to assess the significance of contextual variables in the efficiency measures. Similar settings will appear in Sytems and Control Theory. See [1] [2] [3] [4] [5] [6] [7] [8] [9]. Descriptive and parametric inferences are available for this purpose. An instance of the former is the work of [10]. Typical approaches for the latter are based on two stage regressions. Some criticism to the two stage regressions can be seen in the literature in the works of [11] [12], who establish the conditions under which it is valid. Basically it is assumed that the contextual variables are exogenous to the production process. Correlation between DEA measurements of different units and endogeneity of contextual variables may invalidate the statistical analysis. [13] [14] also discuss the problem.

Our objective in this article is the assessment of the influence of the contextual variables type of unit (product, thematic or ecoregional oriented), processes improvement and impact of technology on the DEA efficiency measurements, computed as performance measures for the Brazilian Agricultural Research Corporation (Embrapa) research centers. This is a DEA VRS model with a single pooled output measure and three inputs – personnel expenses, operational costs, and capital depreciation. For this purpose we use the fractional regression model, as proposed by [15][16].

Our discussion proceeds as follows. In Section 2, we present the state of the art on fractional regression models. In Section 3, we summarize the Embrapa production model. This is the case study in which the proposed approach will be applied. In Section 4, we describe the family of probability distributions we use in our regressions. Section 5 is on our statistical findings and discusses the proposed models regarding our case study. In Section 6, we summarize our conclusions.

2. State of the art

The fractional regression to assess the effect of covariates in a context where DEA scores are treated as descriptive measures of the relative technical efficiency of the sampled DMUs are proposed in [15] [16] These authors propose the use of flexible families of probability distributions to describe the response behavior of performance measures with values in the interval $(0,1]$ and apply their specifications to agricultural data. They consider one and two part models that can be used with quasi maximum likelihood, nonlinear least squares and maximum likelihood estimation. In the two part model, firstly a binary choice model is fit by maximum likelihood to all units. The contextual variables affect the expected response (choice of being efficient) through a distribution function evaluated on a linear construct. The second part fits a nonlinear mean with values in $(0,1)$ for the inefficient units. The dependence on the contextual variables in this instance is obtained via a monotonic function of another linear construct.

The two stage approach of [11] assumes that DEA scores measure efficiency relative to an estimated frontier, the true value of which is unobserved. This implies that estimates of efficiency from DEA models are subject to uncertainty because of sampling variation [15]. From an empirical point of view, the statistical analysis of the two part fractional regression and the two stage approach of [11] are similar. They assume a truncated normal distribution for the inefficient units and propose bootstrap estimation of the model parameters. The same idea can be used in the fractional regression. The difference is that the two stage regression does not take into account efficient DMUs. In this context, unity efficiency is viewed more as a natural consequence of the way DEA scores are defined than as informative as why units become efficient.

Motivated by the results of [11] [15] [16] [17], here we propose a model for the analysis of two stage regressions where the response is output oriented with values in $[1,+\infty)$. Our contribution to this literature is threefold. Firstly, our approach combines a choice model with probability distributions with support in $[1,+\infty)$. It extends the work of [17] on quasi maximum likelihood involving the Bernoulli log-likelihood function, allowing responses greater than one. Secondly, we use the gamma and the truncated normal families to define regression models to fit mean efficiency response encompassing the whole sample. We thus avoid the two part model. Finally, we compare the results of some of these distinct approaches in a context of interest in it involving the assessment of technical effects in a real case study.

Other recent approaches on the two-stage regressions subject can be seen in [18] [19] [20] [21] [22], among others. Modelling in other contexts can be seen in [23] [24] [25] [26] [27] [28] [29] [30].

3. Embrapa's Production Model

Embrapa's research system comprises 42 research centers (DMUs) spread all over the country. Input and output variables have been defined from a set of performance indicators known to the company since 1991.

The set of production variables monitored by Embrapa, as considered here, comprises one output and a three dimensional input vector. The analysis is performed on a yearly basis. Here we restrict attention to 2009. Dynamic specifications are studied in [31] [32].

The input side of Embrapa's production process is composed of three factors: personnel, operational costs (consumption materials, travel and services less income from production projects), and capital measured by depreciation.

The output indicator is a pooled index of four categories: Scientific Production; Production of Technical Publications; Development of Technologies, Products, and Processes; Diffusion of Technologies and Image.

Inputs and output are indexes of complex computations that can be appreciated elsewhere. See [31] [32] [33] for more details.

Embrapa's production system is being monitored since 1996 for 37 research centers. Measures of efficiency and productivity are calculated and used for several managerial objectives. One of the most important is the negotiation of production goals with the individual research units. A proper management of the production system as a whole requires the identification of good practices and the implementation of actions with a view to improve overall performance and reduce variability in efficiency among research units.

Parallel to this endeavor is the identification of non-production variables that may affect positively or negatively the system. It is of managerial interest to detect controllable attributes causing the observed best practices.

Several attempts are in course in Embrapa to evaluate the effects of contextual variables in production efficiency. It is worth to mention [31] [32] [34] [35]. Here we analyze the effect of three exogenous covariates: process improvement (*PRO*), impact of technologies (*IMP*), and type of a research center. *PRO* and *IMP* are considered continuous scores. Type is a categorical variable. The construct *IMP* is a score computed by Embrapa's administration reflecting perceptions regarding the quality of the reports on impact of the technologies developed by the research centers; it's about form and contents of the reports, and not about the importance of the technologies under concern. On the other hand, *PRO* is a value intended to measure the successful implementation of changes on some administrative processes. These processes are selected by local Embrapa's administration. Type is an exogenous classification based on the research focus of each unit. There are three types: units or research centers that focus their research on agricultural products (*PRODUCT*), research centers focusing on agricultural specific themes (*THEMATIC*), and research centers focused on agricultural research pertaining to issues related to environment and ecological aspects (*ECOLOGICAL*). We assume that all contextual variables satisfy the separability assumption of [1].

The data on production (inputs – X1, X2, X3 – and output – Y), the DEA output oriented efficiencies under variable returns to scale (EFF), and contextual variables are shown in Table 1. The year of analysis is 2009.

4. DEA, Contextual Variables and Statistical Models

Consider a production process with *n* production units, the Decision Making Units (DMUs). Each DMU uses variable quantities of *s* inputs to produce a single output *y*. Denote by $Y=(y_1,...,y_n)$ the $1\times n$ output vector, and by $X=(x_1,...,x_n)$ the $s\times n$ input matrix. Notice that the element $y_j>0$ is the output of DMU *r* and $x_j\geq 0$, with at least one component strictly positive, is the $s\times 1$ vector of inputs used by DMU *j* to produce y_r.

For each DMU *j* the DEA measure of efficiency ϕ_r^* is the solution of the linear programming problem $\max_{\phi,\lambda}\phi$ subject to $\sum_r \lambda_r y_r \geq \phi\, y_j$ and $\sum_r \lambda_r x_r \leq x_j$, $\sum_r \lambda_r = 1$, $\lambda=(\lambda_1,...,\lambda_n)\geq 0$.

Our objective is to assess the effect of a vector of contextual variables on the DEA efficiency scores. In this context, in [5 15] it is considered a two part model for responses in the interval $(0,1]$. Basically, they assume that contextual variables may affect differently efficient and inefficient DMUs. They argue that a two part model should be used for modelling DEA scores. The first part of such model comprises a standard binary choice model that governs the probability of observing an efficient DMU. They suggest the use of the whole sample to estimate the model $\text{Prob}\left(\phi_j^*=1 \mid z_j\right)=F\left(z_j'\beta\right)$, where *z* is the vector of contextual variables, β is an unknown parameter vector, and F is a known probability distribution function. Typical choices for F are the logistic and the standard normal distributions. Other possibilities may be seen in [5]. For the second part of the model they assume the specification $E\left(\phi_j^* \mid z_j\right)=G\left(z_j'\theta\right)$, presented in [7 17], for the DEA scores in the interval $(0,1)$.

It is important to emphasize here that quasi maximum likelihood methods may not be used to compute standard errors for any of the previous approaches. The existing correlation among the DEA scores precludes the assumption of independent observations necessary for the validity of the asymptotic assumptions for quasi maximum likelihood and nonlinear least squares.

Motivated by the families of gamma and truncated normal distributions, we begin proposing two specifications for the mean efficiency with estimation based on the whole sample, when efficiency scores are output oriented and greater than or equal to one. Firstly, we specify the mean efficiency as the mean of a random variable of the form 1+H, where H has the gamma distribution with location parameter *p* and scale parameter $\lambda_j^{-1}=\exp\left(z_j'\theta\right)$. Here z_j, again, is the observation of the vector of contextual variables for the inefficient DMU *j*. The parameters θ are unknown. Secondly, another flexible family much used in stochastic frontier analysis is given by the truncated normal distribution. The corresponding mean in this case is given by $z_j'\theta+\sigma\phi\left(\left(1-z_j'\theta\right)/\sigma\right)/\left(1-\Phi\left(\left(1-z_j'\theta\right)/\sigma\right)\right)$, where $\phi(.)$ and $\Phi(.)$ are the density and the distribution functions of the standard normal. This expression is the mean of the random variable $z_j'\theta+u_j$, where u_j is the $N\left(0,\sigma^2\right)$ truncated at $1-z_j'\theta$.

As in [15], we see possible that contextual variables and corresponding parameters may differ for efficient and inefficient units. In both cases the mean response for the efficiency measure will be a monotonic function of the linear construct $z'\theta$. In a joint model, as we propose, this more general assumption will not be parsimonious, creating convergence problems in nonlinear estimation. The separate Bernoulli type regression of [15] will demand many efficient units for a reasonable assessment of the corresponding covariates.

Table 1. Production data, efficiency measurements and contextual variables.

	X1	X2	X3	Y	EFF	*PRO*	*IMP*	Type
DMU1	1.9491	2.31	2.7117	1.5779	1.1725	71.38	1.42	*THEMATIC*
DMU2	0.9475	0.7801	0.6516	0.8873	2.0851	45.88	4.27	*PRODUCT*
DMU3	0.6054	0.6833	0.7612	1.5432	1.1989	88.38	3.53	*THEMATIC*
DMU4	1.3058	1.1456	1.1190	0.5541	3.3389	72.79	4.20	*PRODUCT*
DMU5	1.0482	1.1079	1.1601	1.3029	1.4201	88.88	2.86	*THEMATIC*
DMU6	0.6746	0.8532	0.6409	0.7294	2.5368	58.50	3.86	*PRODUCT*
DMU7	0.4377	0.5439	1.0545	1.8501	1.0000	58.42	2.22	*THEMATIC*
DMU8	1.021	0.7785	0.7123	1.0453	1.7699	80.68	4.61	*PRODUCT*
DMU9	0.9175	0.9185	1.8102	0.7664	2.4143	80.92	3.94	*PRODUCT*
DMU10	1.3485	0.9039	1.5332	0.7837	2.3607	95.13	4.10	*PRODUCT*
DMU11	0.9720	1.0944	1.0455	0.7466	2.4777	85.88	3.75	*PRODUCT*
DMU12	1.0433	0.7983	1.0437	1.0598	1.7458	57.75	4.10	*THEMATIC*
DMU13	1.0481	1.0375	0.7269	1.2256	1.5094	70.04	4.91	*PRODUCT*
DMU14	1.4299	1.4462	1.4492	1.0583	1.7483	81.88	4.07	*PRODUCT*
DMU15	0.9104	0.7062	0.7744	1.0922	1.6938	73.63	3.32	*THEMATIC*
DMU16	0.8805	0.838	0.9973	0.6600	2.8027	79.48	4.54	*PRODUCT*
DMU17	1.3737	1.7809	1.5852	1.1443	1.6168	47.43	4.72	*PRODUCT*
DMU18	1.0264	0.9054	0.9540	0.9172	2.0169	76.50	4.47	*PRODUCT*
DMU19	0.5765	0.5647	0.6141	1.8501	1.0000	92.25	4.96	*THEMATIC*
DMU20	0.6892	0.9250	1.0699	0.7055	2.6226	76.38	4.02	*PRODUCT*
DMU21	1.2903	1.1155	0.8306	0.5272	3.5088	73.38	3.91	*ECOLOGICAL*
DMU22	1.7702	1.7286	1.5338	0.5682	3.2563	85.38	4.22	*ECOLOGICAL*
DMU23	1.6006	1.715	1.8198	1.1389	1.6244	84.08	3.16	*ECOLOGICAL*
DMU24	0.7749	1.194	0.673	0.6848	2.7012	85.40	3.84	*ECOLOGICAL*
DMU25	0.5078	0.4727	0.2901	0.4944	1.0000	73.00	1.41	*ECOLOGICAL*
DMU26	0.7037	0.5547	0.4159	1.1163	1.0000	0.00	3.10	*ECOLOGICAL*
DMU27	0.6122	0.5341	0.6379	1.4728	1.1955	50.17	3.73	*ECOLOGICAL*
DMU28	1.1706	1.0919	0.8334	0.5575	3.3190	2.50	1.77	*ECOLOGICAL*
DMU29	0.6368	0.7740	0.5731	0.6497	2.6490	73.88	4.51	*ECOLOGICAL*
DMU30	0.7758	0.5738	0.6142	0.8509	2.1744	86.54	4.85	*ECOLOGICAL*
DMU31	1.0206	0.9173	0.6094	1.2273	1.4954	95.50	2.71	*ECOLOGICAL*
DMU32	1.3446	1.3243	1.1444	0.6782	2.7278	65.14	4.32	*ECOLOGICAL*
DMU33	2.3904	2.1439	1.5218	0.8324	2.2227	83.13	4.32	*ECOLOGICAL*
DMU34	0.6753	0.6457	0.7747	0.9863	1.8758	83.80	4.35	*PRODUCT*
DMU35	0.4118	0.4548	0.5033	1.5013	1.0000	18.88	4.67	*PRODUCT*
DMU36	0.7590	0.8277	1.0633	1.8501	1.0000	90.25	3.04	*THEMATIC*
DMU37	0.350	0.8103	0.7465	0.7627	1.0000	0.00	4.26	*THEMATIC*

Source: Author's calculations based on Embrapa's research units data.

Thus, imposing similar effects of the covariates in all DMUs, we have the two following nonlinear regressions to be estimated using the complete sample:

(1) Gamma assumption

$$E\left(\phi_j^* \mid z_j\right) = \left[1 + p\exp\left(z_j'\theta\right)\right]\left[1 - F\left(z_j'\beta\right)\right] + F\left(z_j'\beta\right)$$

(2) Truncated Normal assumption

$$E\left(\phi_j^* \mid z_j\right) = \left[z_j'\theta + \sigma\frac{\phi\left(\left(1 - z_j'\theta\right)/\sigma\right)}{1 - \Phi\left(\left(1 - z_j'\theta\right)/\sigma\right)}\right]\left[1 - F\left(z_j'\beta\right)\right] + F\left(z_j'\beta\right)$$

Another possibility to evaluate the effect of contextual variables z_j for technical efficiency measurements ϕ_j^* in the interval $[1, +\infty)$, mimicking the work of [17] for general truncated distributions, is to assume the following log likelihood functions for the efficiency response. For the gamma distribution we have

$$l(\beta, \theta, p) = \eta\left(\sum_{\phi_j^*=1}\log F\left(z_j'\beta\right)\right) + (1-\eta)\left(\sum_{\phi_j^*>1}\begin{matrix} p\log\left(\lambda_j\right) + (p-1)\log\left(\phi_j^* - 1\right) \\ -\lambda_j\left(\phi_j^* - 1\right) - \log\Gamma(p) \\ +\log\left(1 - F\left(z_j'\beta\right)\right)\end{matrix}\right),$$

where η is the indicator of an efficient DMU. For the truncated normal, we have

$$l(\beta, \theta, \sigma) = \eta\left(\sum_{\phi_j^*=1}\log F\left(z_j'\beta\right)\right) + (1-\eta)\left(\sum_{\phi_j^*>1}\begin{matrix} \log g\left(\phi_j^*, z_j'\theta, \sigma^2\right) \\ -\log\left(1 - Q\left(1, z_j'\theta, \sigma^2\right)\right) \\ +\log\left(1 - F\left(z_j'\beta\right)\right)\end{matrix}\right),$$

where g is the density function of the normal $N\left(z_j'\theta, \sigma^2\right)$ and Q *is* the corresponding distribution function.

In applications under exogeneity of the contextual variables all the models above may be estimated by quasi maximum likelihood or nonlinear least squares. Our choice is the nonlinear least squares, with standard errors and confidence intervals computed using nonparametric bootstrap based on centered residuals or not. These order of ideas represent new contributions and are clearly distinct from the proposals of [15], including the work of [17].

We believe that additional information on the contextual variables is gained considering the whole sample of efficiency scores. In our application the data do not support different parameterizations for efficient and inefficient units jointly, as we have a small sample. Therefore, we assume $\theta = \beta$. In other words, our assumption is that contextual variables affect the response equally on efficient and on inefficient DMUs. A difficulty that arises with the representations (1) e (2) is that the mean, in general, is no longer a monotone function of the linear construct $z'\theta$.

Restricting attention only to inefficient units, using maximum likelihood and assuming the truncated normal or the gamma distributions, the results will be equivalent to the two stage analysis proposed by [1].

5. Statistical Results

For the data of Table 1 we assume that the linear construct

$$\mu = \beta_0 + \beta_1 PRO + \beta_2 IMP + \beta_3 PRODUCT + \beta_4 THEMATIC$$

affects expected efficiency according to one of the models discussed above. *PRODUCT* and *THEMATIC* are dummy variables.

The gamma specification with nonlinear least squares estimation better fits the data (best correlation between predicted and observed efficiency scores). The separate mean specifications do not produce stable nonlinear least squares results. Maximum likelihood for the inefficient units under the gamma and the truncated normal are stable and leads to similar statistical results.

We begin our discussion with the two part analysis, as proposed by [15], using maximum likelihood methods and the bootstrap Algorithm #1 of [11], which assumes the truncated normal specification to compute standard errors and confidence intervals for the model, based on 5,000 replications. Table 2 shows results of the choice model with the probit assumption.

We see that only marginally the set of contextual variables is significant. *PRO* acts reducing the probability of being efficient and *THEMATIC* has an increasing effect. The analysis of [1] is shown in Table 3 (results with the gamma distribution are basically the same). Only type of unit is significant at the 5% level.

Table 4 shows the statistical results for the joint estimation of the specification (1) using nonlinear least squares. Correlation between predicted and observed values is 0.692. Nonlinear least squares did not converge for (2). The bootstrap (nonparametric) results are based on

5,000 replications. Fig.1 illustrates the bootstrap distributions.

One can see in Table 4 that Type and *PRO* are significant effects. The model is more informative regarding the effect of the contextual variables on efficiency. Fig.2 shows the derivative of the expected mean response as a function of the linear construct μ. We see that the mean response increases or decreases depending on the level of the contextual variables. The same behavior is observed with the marginal effect of *PRO*. The negative value of *THEMATIC* is in the direction of more efficiency. For *PRO* is more difficult to disentangle the marginal effect. For all thematic centers, which are the more efficient, *PRO* has a positive effect. For the other types the response will decrease with the level of μ.

We see that processes improvement has not been adequate for benchmark units and do not lead to overall improvement for the less efficient units. The impact of technologies on efficiency, as actually measured, is not statistically important. Appropriate presentation of reports will not lead to more efficiency in production. Managers should look more carefully into production and costs profiles and into processes actually carried out within benchmarks units that could indeed lead to the increase of overall performance.

Table 2. Results of the choice model with the probit specification.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	37.893	37.814
SC	39.504	45.868
-2 Log L	35.893	27.814

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.0798	4	0.0887
Score	7.4800	4	0.1126
Wald	6.4687	4	0.1668

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4009	1.6314	0.7373	0.3905
PRO	1	-0.0282	0.0179	2.4646	0.1164
IMP	1	-0.1006	0.2944	0.1168	0.7326
PRODUCT	1	-0.8279	0.8624	0.9214	0.3371
THEMATIC	1	0.9818	0.6213	2.4972	0.1140

Source: Author’s calculations.

Table 3. Results of the truncated normal specification for the inefficient units.

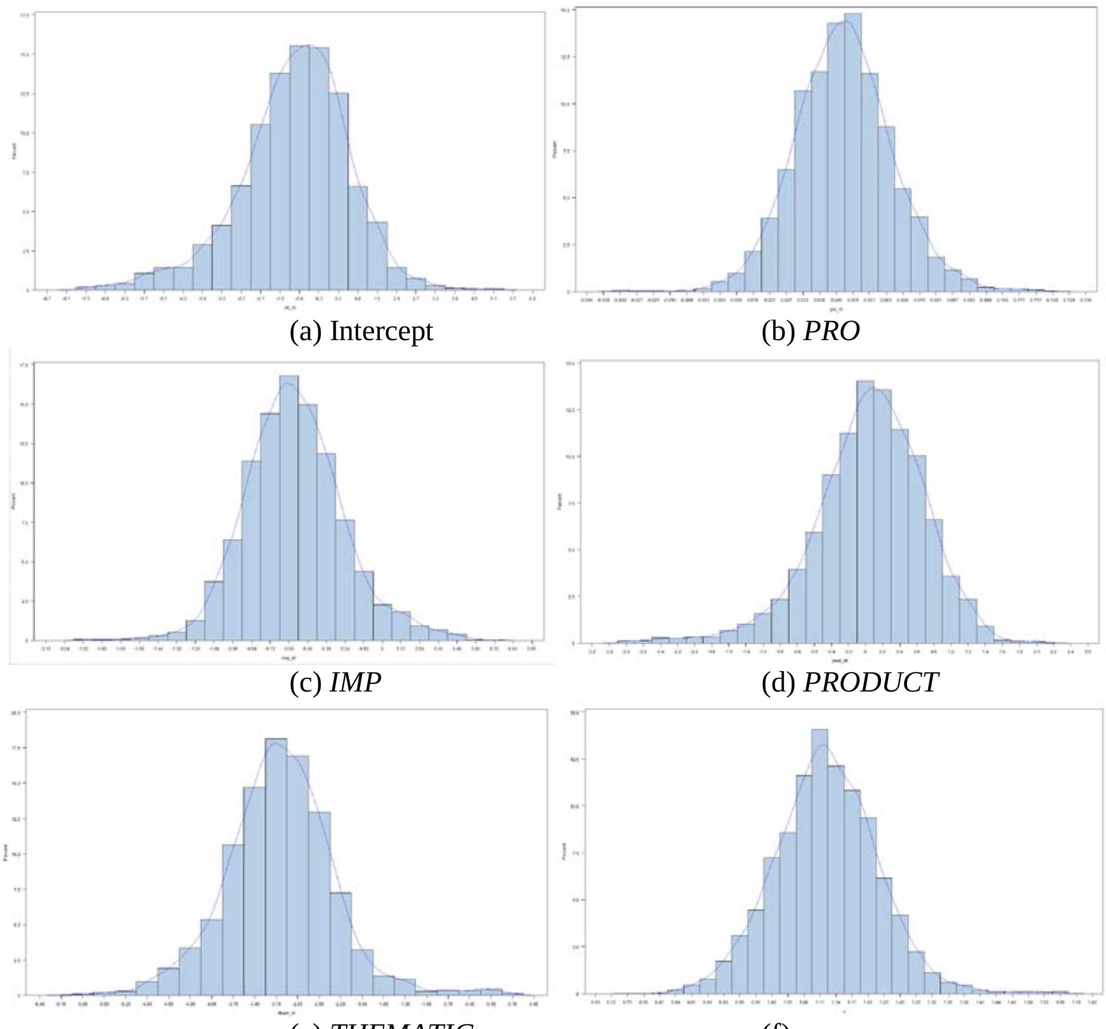
Parameter	Estimate	Bootstrap Standard Error	DF	t Value	95% Bootstrap Percentile Confidence Interval	
Intercept	2.4618	1.1369	30	2,1654	0.0308	4.5015
PRO	0.0008	0.0101	30	0,0816	-0.0181	0.0225
IMP	-0.0278	0.2076	30	-0,1340	-0.4411	0.4126
PRODUCT	-0.2265	0.2917	30	-0,7765	-0.7991	0.3632
THEMATIC	-1.6402	0.6780	30	-2,4192	-3.3985	-0.6399
sigma	0.6423	0.1067	30	6,0197	0.3968	0.8267

Source: Author’s calculations

Table 4. Nonlinear least squares for $E\left(\phi_j^* \mid z_j\right) = \left[1 + p\exp\left(z_j'\theta\right)\right]\left[1 - F\left(z_j'\beta\right)\right] + F\left(z_j'\beta\right)$.

Parameter	Estimate	Bootstrap Standard Error	95% Bootstrap Percentile Confidence Interval	
Intercept	-1.0188	0.7098	-5.2688	1.9396
PRO	0.0528	0.0176	0.0137	0.0852
IMP	-0.6738	0.3174	-1.1485	0.1667
PRODUCT	0.0770	0.6443	-1.4742	1.2034
THEMATIC	-4.1007	0.7654	-4.6618	-1.4163
v (p=exp(v))	1.0673	0.0968	0.9292	1.3139

Source: Author's calculations



(a) Intercept (b) *PRO*

(c) *IMP* (d) *PRODUCT*

(e) *THEMATIC* (f) v

Fig.1. Bootstrap distributions (Source: Author's calculations).

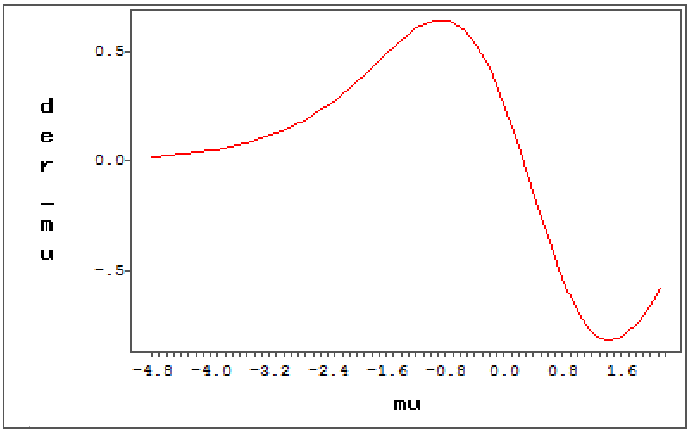


Fig.2. Derivative of the expected mean response as a function of the linear construct μ (Source: Author's calculations).

6. Summary and Conclusions

We propose and fit a new family of probability distributions for DEA output oriented measures of efficiency as an extension of fractional regression models for responses outside the $(0,1)$ interval. The objective of the analysis was to study the effect of contextual variables in the DEA performance measure computed for Embrapa's research centers.

The models considered allow for the dependence on a vector of contextual variables via a linear construct. As in fractional regression, the models combine two parts. These are a choice model explaining the expectation of being efficient, and a flexible family describing the expected mean behavior of the inefficient firms. These are motivated by the gamma and the truncated normal distributions.

The separate analysis for efficient and inefficient units seem to provide support for the two stage regression of [11], since we do not detect significant effects in the choice model related to the efficient units. For the inefficient units both approaches lead to the same results when we fit the data using maximum likelihood under the assumptions of the gamma or the truncated normal.

Combination of the two part models estimated via nonlinear least squares and bootstrap leads to different conclusions regarding the marginal effects of the contextual variables. The impression is that the inclusion of efficient units adds relevant information to the statistical analysis. This point is particularly important for the instrumentalist approach, where we look at efficiency measurements more as measures of performance than as realizations associated with a true unknown production process.

We conclude that the joint regression approach is more informative. The set of contextual variables studied in Embrapa's application is defined by PRO – Process improvement, IMP - impact of technologies and Type (Product, Thematic and Ecoregional). Type and PRO are statistically significant. The type category Thematic includes the more efficient units. The response effect to PRO varies with the expected mean efficiency level and is positively associated with performance for the Thematic research centers.

7. Acknowledgment

To the National Council for Scientific and Technological Development (CNPq), for the financial support.

References

[1] Doroshin, A. V., Neri, F. (2014) Open research issues on Nonlinear Dynamics, Dynamical Systems and Processes. WSEAS Transactions on Systems, 13,in press.

[2] Ciufudean, C., Neri, F. (2014) Open research issues on Multi-Models for Complex Technological Systems. WSEAS Transactions on Systems, 13, in press.

[3] Neri, F. (2014) Open research issues on Computational Techniques for Financial Applications. WSEAS Transactions on Systems, 13, in press.

[4] Karthikeyan, P., Neri, F. (2014) Open research issues on Deregulated Electricity Market:

Investigation and Solution Methodologies. WSEAS Transactions on Systems, 13, in press.

[5] Neri, F. (2014) Open research issues on Advanced Control Methods: Theory and Application. WSEAS Transactions on Systems, 13, in press.

[6] Pekař, L., Neri, F. (2013) An introduction to the special issue on advanced control methods: Theory and application (2013) WSEAS Transactions on Systems, 12 (6), pp. 301-303.

[7] Azzouzi, M., Neri, F. (2013) An introduction to the special issue on advanced control of energy systems (2013) WSEAS Transactions on Power Systems, 8 (3), p. 103.

[8] Muntean, M., Neri, F. (2012) Foreword to the special issue on collaborative systems WSEAS Transactions on Systems, 11 (11), p. 617.

[9] Volos, C., Neri, F. (2012) An introduction to the special issue: Recent advances in defense systems: Applications, methodology, technology WSEAS Transactions on Systems, 11 (9), pp. 477-478.

[10] Lima, V.S., Soares de Mello, J.C.C.B., Angulo Meza, L. Cost-benefit analysis in selected air trips using a non parametric method, African Journal of Business Management, Vol. 5, 2011, pp. 9678-9685.

[11] Simar, L., Wilson, P.W. Estimation and inference in two-stage, semi-parametric models of production processes, Journal of Econometrics, Vol. 136, No. 1, 2007, pp. 31-64.

[12] Simar, L., Wilson, P.W. Two-stage DEA: caveat emptor, Journal of Productivity Analysis, Vol. 36 No. 2, 2011, pp. 205-218.

[13] Banker, R.D., Natarajan, R. Evaluating contextual variables affecting productivity using data envelopment analysis, Operations Research, Vol. 56, 2008, pp. 48-58.

[14] Banker, R.D., Natarajan, R. Statistical tests based on DEA efficiency scores. In: Cooper, W.W., Seiford, L.M., Zhu, J. (Eds.). Handbook on Data Envelopment Analysis (pp. 273-295). Boston: Kluwer International Series, 2011.

[15] Ramalho, E.A., Ramalho, J.J.S., Henriques, P.D. Fractional regression models for second stage DEA efficiency analyses, Journal of Productivity Analysis, Vol. 34, 2010, pp. 239-255.

[16] Ramalho, E.A., Ramalho, J.J.S., Murteira, J.M.R. Alternative estimating and testing empirical strategies for fractional regression models, Journal of Economic Surveys, Vol. 25, No. 1, 2011, pp. 19-68.

[17] Papke, L.E., Wooldridge, J.M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates, Journal of Applied Econometrics, Vol. 11, 1996, pp. 619-632.

[18] Bădin, L., Daraio, C., Simar, L. How to Measure the Impact of Environmental Factors in a Nonparametric Production Model, European Journal of Operational Research, Vol. 223, 2012, pp. 818-833.

[19] Daraio, C., Simar, L. Advanced Robust and Nonparametric Methods in Efficiency Analysis. New York: Springer. 2007.

[20] Hoff, A. Second stage DEA: Comparison of Approaches for Modelling the DEA Score, European Journal of Operational Research, Vol. 181, 2007, pp. 425-435.

[21] Johnson, A.L., Kuosmanen. T. One-Stage and Two-Stage DEA Estimation of the Effects of Contextual Variables, European Journal of Operational Research, 220, 2012, pp. 559-570.

[22] McDonald, J. Using Least Squares and Tobit in Second Stage DEA Efficiency Analyses, European Journal of Operational Research, Vol. 197, 2009, pp. 792-798

[23] Panoiu, M., Neri, F. (2014) Open research issues on Modeling, Simulation and Optimization in Electrical Systems. WSEAS Transactions on Systems, 13, in press.

[24] Hájek, P., Neri, F. (2013) An introduction to the special issue on computational techniques for trading systems, time series forecasting, stock market modeling, financial assets modeling WSEAS Transactions on Business and Economics, 10 (4), pp. 201-292.

[25] Bojkovic, Z., Neri, F.(2013) An introduction to the special issue on advances on interactive multimedia systems WSEAS Transactions on Systems, 12 (7), pp. 337-338.

[26] Pekař, L., Neri, F. (2013) An introduction to the special issue on advanced control methods: Theory and application (2013) WSEAS Transactions on Systems, 12 (6), pp. 301-303.

[27] Guarnaccia, C., Neri, F. (2013) An introduction to the special issue on recent methods on physical polluting agents and environment modeling and simulation WSEAS Transactions on Systems, 12 (2), pp. 53-54.

[28] Neri, F. (2012) An introduction to the special issue on computational techniques for trading systems, time series forecasting, stock market modeling, and financial assets modeling WSEAS Transactions on Systems, 11 (12), pp. 659-660.

[29] Pekař, L., Neri, F. (2012) An introduction to the special issue on time delay systems: Modelling, identification, stability, control and applications WSEAS Transactions on Systems, 11 (10), pp. 539-540.

[30] Hájek, P., Neri, F. (2013) An introduction to the special issue on computational techniques for trading systems, time series forecasting, stock market modeling, financial assets modeling WSEAS Transactions on Business and Economics, 10 (4), pp. 201-292.

[31] Souza, G.S., Gomes, E.G., Magalhães, M.C., Avila, AF.D. Economic efficiency of Embrapa's research centers and the influence of contextual variables, Pesquisa Operacional, Vol. 27, 2007, pp. 15-26.

[32] Souza, G.S., Gomes, E.G., Staub, R.B. Probabilistic measures of efficiency and the influence of contextual variables in nonparametric production models: an application to agricultural research in Brazil, International Transactions in Operational Research, Vol. 17, 2010, pp. 351-363.

[33] Souza, G.S., Gomes, E.G. Scale of operation, allocative inefficiencies and separability of inputs and outputs in agricultural research, Pesquisa Operacional, Vol. 33, 2013, pp. 399-415.

[34] Souza, G.S. Significância de efeitos técnicos na eficiência de produção da pesquisa agropecuária brasileira, Revista Brasileira de Economia, Vol. 60, No. 1, 2006, pp. 91-117.

[35] Souza, G.S., Alves, E., Avila, A.F.D. Technical efficiency in agricultural research, Scientometrics, Vol. 46, 1999, pp. 141-160.