A NONPAPRAMETRIC APPROACH TO EVALUATE THE IMPACT OF CONTEXTUAL VARIABLES ON THE AGRICULTURAL RESEARCH EFFICIENCY IN BRAZIL

Geraldo da Silva e Souza
Eliane Gonçalves Gomes
Brazilian Agricultural Research Corporation (Embrapa) – SGE
Parque Estação Biológica, Av. W3 Norte final, Asa Norte, 70770-901, Brasília, DF, Brazil
{geraldo.souza, eliane.gomes}@embrapa.br

RESUMO

Neste artigo são calculadas medidas de eficiência técnica para cada um dos centros de pesquisa da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). As medidas de eficiência DEA são modeladas como função de variáveis contextuais: capacidade de geração de receita, intensidade de parcerias, melhoria de processos administrativos, e impacto das tecnologias geradas pelos centros de pesquisa. A produção é modelada com erros aleatórios e de ineficiência, de forma semelhante às fronteiras estocásticas. A avaliação da significância para o conjunto de variáveis contextuais é realizada por meio de modelos de programação linear e testes de adequabilidade de distribuições e tem base não-paramétrica. Conclui-se que há significância conjunta de todas as variáveis contextuais.

PALAVRAS-CHAVE: Variáveis contextuais, Eficiência, Pesquisa Agropecuária

Área principal: DEA Análise Envoltória de Dados

ABSTRACT

In this paper we measure technical efficiency for each of Embrapa (Brazilian Agricultural Research Corporation) research centers. We model DEA efficiency as a function of contextual variables: revenue generation capacity, partnership intensity, improvement of administrative processes, and impact of technologies generated by the research centers. Production is modeled with random and inefficient errors, in a manner similar to stochastic frontiers. The assessment of significance for the set of contextual variables is carried out by means of linear programming and goodness of fit tests and has a nonparametric basis. We conclude that there is joint significance of all contextual variables.

KEYWORDS: Contextual Variables, Efficiency, Agricultural Research.

Main area: DEA Data Envelopment Analysis

1. Introduction

The Brazilian Agricultural Research Corporation (Embrapa) monitors, since 1996, the production process of its 37 research centers, using a nonparametric DEA (Data Envelopment Analysis) production model. This model provides a measure of technical efficiency of production for each research center. For details see Souza et al. (1997, 1999, 2007) and Souza & Avila (2000).

The measure of technical efficiency proposed here assesses the performance of Embrapa research centers using a single output and a three dimensional input vector. Inefficiency errors are stochastic and further assumed to be a monotonic concave function of the contextual variables.

The use of technical efficiency measures as a performance indicator raises some questions within the organization. An important one is whether or not the process generates unwanted competition among the research centers. A typical criticism is that the evaluation system may inhibit partnerships.

This article is concerned with the identification of contextual variables external to the production process that may be affecting or causing efficiency. Typically these variables are in the control of the institution. The assessment of their effect is of managerial importance, since they may serve as a tuning device to improve management practices leading to efficient units. Here we are interested in studying the effects on technical efficiency of revenue generation capacity, partnership intensity, improvement of administrative processes, and impact of the technologies generated by the research centers.

The identification of causal factors of efficiency demands appropriate statistical modeling. The literature is rich in parametric and semi parametric statistical models to assess the significance of covariates in efficiency models. Typical semi parametric approaches can be seen in a DEA context in Souza and Staub (2007) and Simar and Wilson (2007). Recently, Souza (2006) and Souza et al. (2007) assessed the influence of covariates on the DEA efficiency measurements using analysis of variance, dynamic panel data, generalized method of moments and maximum likelihood methods. The typical approach followed in all those cases is based on a two stage DEA. Efficiency measurements are computed and then regressed on a set of covariates. To lessen the problem of interference of the covariates on the production frontier, Daraio and Simar (2007) proposed a measure based on the conditional FDH to obtain insights on the effects of covariates. Souza et al. (2010) explores these ideas and, for the Embrapa application described here, concluded via generalized method of moments that the set of contextual variables is statistically significant. Their analysis is dynamic and they pinpoint efficiency persistence in the process and marginal significance of processes improvements, revenue generation capacity and changing in administration.

The model we propose here to assess the statistical significance of contextual variables is non dynamic, based on DEA, and follows the production model of Banker (1993), Banker and Natarajan (2004, 2008) and Souza and Staub (2007). It is a two stage approach, where efficiencies computed in the first stage are assumed to follow a production model defined by a nonnegative monotone concave function of the covariates. The two stage approach is robust against stochastic models, since it allows for a random inefficient component and a two sided random error. The use of this approach is new in the literature.

Our exposition proceeds as follows. In Section 2 we review the DEA models and the production model relative to which DEA production functions may produce consistent and non parametric maximum likelihood estimates. These results are basic for the assessment of the significance of covariates and to test for scale of operation. In this section we also describe our fully nonparametric approach to study significance of contextual variables based on Banker and Natarajan (2004, 2008) results. In Section 3 we review Embrapa production process and the production variables used in the analysis including contextual variables. Section 4 is on statistical results. Finally Section 5 summarizes our findings.

2. DEA, Production Functions, Statistical Models and Contextual Variables

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 $Y=(y_1,\ldots,y_n)$ the $1\times n$ output vector, and by $X=(x_1,\ldots,x_n)$ the $s\times n$ input matrix. Notice that the element $y_r>0$ is the output of DMU r and $x_r\geq 0$, with at least one component strictly positive, is the $s\times 1$ vector of inputs used by DMU r to produce y_r.

Let K be compact and convex in the nonnegative orthant of R^s. The maximum output (frontier output) achievable from $x\in K$ is given by the production function $y=g(x)$. We assume $g(x)$ to be continuous and, additionally,

1. Monotonicity: If $x\geq w$ are points in K, then $g(x)\geq g(w)$.
2. Concavity: If x and w are points in K, then $g(tx+(1-t)w)\geq tg(x)+(1-t)g(w)$, for $t\in[0;1]$.
3. For each $j=1,\ldots,n,\ g(x_j)\geq y_j$.

One can use the observations (x_j,y_j), with $x_j\in K$, and DEA to estimate $g(x)$ only in the set (1).

$$K^*=\left\{x\in K; x\geq\sum_{j=1}^{n}\lambda_j x_j, \text{for some } (\lambda_1,\ldots,\lambda_n)\geq 0, \sum_{j=1}^{n}\lambda_j=1\right\} \quad (1)$$

For $x\in K^*$ the DEA production function is defined by (2).

$$g_n^*(x)=\sup_{\lambda_1,\ldots,\lambda_n}\left\{\sum_j\lambda_j y_j; \sum_j\lambda_j x_j\leq x, \lambda_j\geq 0, \sum_{j=1}^{n}\lambda_j=1\right\} \quad (2)$$

This formulation imposes variable returns to scale. If the technology defined by $g(x)$ shows constant returns to scale only non negativity is imposed on the weights λ_j.

The subset K^* is convex and closed in K. For each r, $g_n^*(x_r)=\phi_r^* y_r$, where ϕ_r^* is the solution of the LP problem $max_{\phi,\lambda}\phi$ subject to $\sum_j\lambda_j y_j\geq\phi y_r$ and $\sum_j\lambda_j x_j\leq x_r$, $\lambda=(\lambda_1,\ldots,\lambda_n)\geq 0$, $\sum_j\lambda_j=1$. The function $g_n^*(x)$ satisfies conditions 1-3 and has the property of minimum extrapolation, that is, $g(x)\geq g_n^*(x), x\in K^*$.

If one assumes that the production observations (x_j,y_j) satisfy the deterministic statistical model $y_j=g(x_j)-\varepsilon_j$, where the technical inefficiencies ε_j are nonnegative random variables with probability density functions $f_j(\varepsilon)$ concentrated on R^+, and the inputs x_j are a random sample drawn independently with density functions $h_j(x)$ with support set contained in K, one can show that if x_0 is a point in K^* interior to K, then $g_n^*(x_0)$ converges almost surely to $g(x_0)$. See Souza and Staub (2007).

Let M be a subset of the DMUs included in the sample that generates the n production observations. The asymptotic joint distribution of the technical inefficiencies $\varepsilon_{nj}^*=g_n^*(x_j)-y_j,\ j\in M$, coincides with the product distribution of the $\varepsilon_j,\ j\in M$. For these results to hold is sufficient that the sequence of input densities $h_j(x)$ satisfies (3).

$$0<l(x)\leq\inf{}_j h_j(x)\leq\sup{}_j h_j(x)\leq L(x) \quad (3)$$

for integrable functions $l(x)$ and $L(x)$ and x interior to K and that the inefficiency densities $f_j(\varepsilon)$ are such that (4) is true, where $F_j(u) = \int_0^u f_j(\varepsilon) d\varepsilon$.

$$F(u) = inf_j F_j(u) > 0, \ u > 0 \tag{4}$$

The importance of these results, whose proof one can see in Souza and Staub (2007), is that the statistical model allows for inefficiency variables not equally distributed as in Banker (1993). This is precisely the environment necessary for contextual variables when they are exogenous to the production frontier.

Here, following Banker and Natarajan (2004,2008), we assume that inefficiency errors satisfy $\varepsilon_j = h(z_j) + u_j$, where $h(z)$ is a nonnegative, monotone and concave function of the vector of contextual variables z. Production is assumed to follow the statistical model (5).

$$y_j = g(x_j) + v_j - \varepsilon_j \tag{5}$$

The component v_j is a random error with density function concentrated in $(-V^M; V^M)$, $V^M > 0$. The disturbance u_j has a density with support in $(0; B)$, $B > 0$. The function $g(x_j)$ is nonnegative, monotone and concave. Adding and subtracting V^M one obtains (6).

$$\begin{aligned} y_j &= g(x_j) + V^M - \left[V^M - v_j + \varepsilon_j\right] \\ &= g^*(x_j) - \delta_j \end{aligned} \tag{6}$$

The function $g^*(x_j)$ is also nonnegative, monotone and concave. It follows that the error component $\delta_j > 0$ can be estimated by DEA methods and satisfies the assumptions of the one sided inefficiency model. Also (7).

$$\begin{aligned} \delta_j &= h(z_j) + V^M - v_j + u_j \\ &= h(z_j) - \left[v_j - V^M - u_j\right] \end{aligned} \tag{7}$$

One may add and subtract $C = 2V^M + B$ to obtain (8), where the function $h^*(z_j)$ is nonnegative, monotone and concave.

$$\begin{aligned} \delta_j &= h(z_j) + C - \left[C + v_j - V^M - u_j\right] \\ &= h^*(z_j) - l_j, \ \ l_j > 0. \end{aligned} \tag{8}$$

Thus the assumptions of the (deterministic) statistical production model with one sided inefficiency errors also hold for this latter model. One may then assess the significance of the set of contextual variables by nonparametric methods comparing the DEA estimates of two models. Firstly one computes DEA (output oriented) residuals $\delta^*_{nj} = g^*_n(x_j) - y_j$ and uses these residuals as response variables in a new DEA (output oriented) model, having for response the δ^*_{nj} and for inputs the z_j. For generality we impose variable returns to scale in all stages. Significance of the whole set of contextual variables is assessed comparing the distribution of the inefficiency errors in the first stage with $l^*_{nj} = (\phi^2_{nj} - 1)\delta^*_{nj}$. Here ϕ^2_{nj} is the

DEA measure of efficiency in the second stage. Under the null hypothesis of no contextual variables effect, we would expect the two distributions to be coincidental. Any marginal contextual variable effect, following overall significance, is assessed comparing the l_{nj}^{*} with a third stage DEA residual, computed as $\eta_{nj}^{**} = (\phi_j^3 - 1)\delta_{nj}^{*}$, where the contextual variable(s) z^s is (are) omitted.

3. Embrapa Production Model

Embrapa research system comprises 37 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 company uses routinely some of these indicators to monitor performance through annual work plans. With the active participation of the board of directors of Embrapa, as well as the administration of each of its research units, we selected 28 output and 3 input indicators as representative of production actions in the company.

The output indicators were classified into four categories: Scientific Production; Production of technical publications; Development of Technologies, Products, and Processes; and Diffusion of Technologies and Image.

By Scientific Production we mean the publication of articles and book chapters. We require that each item be specified with complete bibliographical reference.

The category of Technical Publications groups publications produced by research centers aiming, primarily, agricultural businesses and agricultural production.

The category of Development of Technologies, Products, and Processes groups indicators related to the effort made by a research unit to make its production available to the society in the form of a final product. We include here only new technologies, products and processes. These must be already tested at the client's level in the form of prototypes or through demonstration units, or be already patented.

Finally, the category of Diffusion of Technologies and Image encompasses production actions related with Embrapa effort to make its products known to the public and to market its image.

The input side of Embrapa 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.

A single production indicator, weather output or input, is defined by the quantity observed for the item divided by the company´s mean. In principle it is possible to work with a separate four dimensional output vector. However, to make the research centers more comparable, we reduced the response to a single output using a weighting system variable for each unit.

The weights, in principle, are supposed to reflect the administration perception of the relative importance of each variable to each DMU. Defining weights is a hard and questionable task. In our application in Embrapa we followed an approach based on the law of categorical judgment of Thurstone. See Torgerson (1958) and Kotz and Johnson (1989). The model is competitive with the AHP method of Saaty (1994) and is well suited when several judges are involved in the evaluation process. Basically we sent out about 500 questionnaires to researchers and administrators and asked them to rank in importance – scale from 1 to 5 – each production category and each production variable within the corresponding production category. A set of weights was determined under the assumption that the psychological continuum of the responses projects onto a lognormal distribution.

To further improve the DEA assumptions of homogeneity and to reduce variability, the production variables were corrected for outliers and further normalized by a personnel quantity index. Inputs were not corrected for outliers. The outlier corrections are performed using the Box-Plots superior fence. Any output variable with an observation greater than the third quartile plus 1.5 times the inter-quartile range is reduced to this mark.

We therefore see that all production variables are measured on a per capita basis. This fact calls for a variable returns to scale production function (Hollingsworth and Smith, 2003).

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. Dynamic specifications are considered elsewhere (Souza et al., 2010).

Embrapa production system is being monitored since 1996. 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 Souza (2006) and Souza et al. (1999, 2007). These studies consider DEA and FDH measures of efficiency, and have evaluated, for distinct periods, the effects of rationalization of costs, processes improvement, intensity of partnerships, type and size.

We now use the information of 2002 to 2009 and analyze the effect of these variables on Embrapa production model following the procedures laid out in the previous section. In this context we consider a vector of 4 covariates, corresponding to process improvement (PROC), financial resources generation capacity (REV), partnership intensity (PART), and impact of technologies (IMP). These are considered continuous covariates. Process improvement and intensity of partnerships are indexes. All continuous covariates are normalized by the maximum. The definition of these scores can be seen in detail in Embrapa (2006). The vector of contextual variables is assumed to be exogenous to the production process. Contextual variables are inverted to produce a positive effect on the inefficiency component.

4. Statistical Results

Table 1 presents the data base used in our work. Table 2 shows the residuals computed as in Section 2, assuming variable returns to scale. Only the year 2009 is shown.

The distribution of these residuals is compared by means of the Kolmogorov-Smirnov D test statistic (Conover, 1998). Notice that only inefficient units are considered in subsequent analysis, since DEA cannot handle zero outputs. This procedure is not strange to the literature. Simar and Wilson (2007) adopt the same approach to compute confidence intervals and bias corrected efficiency estimates for DEA measures. In this second stage, we also removed the units with zero values for PROC, IMP, PART and REV. Although Tables 1 and 2 show data for only 2009, the analysis was repeated for each year in the period 2002-2009.

Table 3 shows the results of the statistical tests. Only the statistical significant effects are shown. We used SAS 9.2 software (Proc Npar1way) in our analyses. Probability values are for the Kolmogorov-Smirnov two sample test. The hypothesis that covariates jointly matter in production is true for all years. The statistics do not indicate statistical significance for the marginal analysis. Combinations of effects are significant in particular years. We do not see any particular trend in the tests, indicating the isolated importance of a particular effect. It seems that the units should concentrate on improving all of them together to improve production efficiency. It is worth to mention that for 2009 all pair wise combinations are effective.

Table 1. Production data and contextual variables. Inputs are X1, X2, X3. Output is Y. Contextual variables are Processes Improvement (PROC), Impact (IMP), Partnership Intensity (PART) and Revenue generation capacity (REV). Year = 2009.

	X1	X2	X3	Y	PROC	IMP	PART	REV
DMU1	1.9491	2.3100	2.7117	1.5779	71.38	1.42	3.45	71.30
DMU2	0.9475	0.7801	0.6516	0.8873	45.88	4.27	3.64	36.50
DMU3	0.6054	0.6833	0.7612	1.5432	88.38	3.53	5.65	61.20
DMU4	1.3058	1.1456	1.1190	0.5541	72.79	4.20	7.65	17.70
DMU5	1.0482	1.1079	1.1601	1.3029	88.88	2.86	3.09	21.00
DMU6	0.6746	0.8532	0.6409	0.7294	58.50	3.86	8.15	21.40
DMU7	0.4377	0.5439	1.0545	1.8501	58.42	2.22	3.81	142.60
DMU8	1.0210	0.7785	0.7123	1.0453	80.68	4.61	3.91	44.40
DMU9	0.9175	0.9185	1.8102	0.7664	80.92	3.94	4.09	95.10
DMU10	1.3485	0.9039	1.5332	0.7837	95.13	4.10	4.15	262.70
DMU11	0.9720	1.0944	1.0455	0.7466	85.88	3.75	5.17	51.50
DMU12	1.0433	0.7983	1.0437	1.0598	57.75	4.10	6.17	7.30
DMU13	1.0481	1.0375	0.7269	1.2256	70.04	4.91	4.42	29.30
DMU14	1.4299	1.4462	1.4492	1.0583	81.88	4.07	7.40	74.80
DMU15	0.9104	0.7062	0.7744	1.0922	73.63	3.32	2.51	56.40
DMU16	0.8805	0.8380	0.9973	0.6600	79.48	4.54	3.18	76.30
DMU17	1.3737	1.7809	1.5852	1.1443	47.43	4.72	5.43	195.40
DMU18	1.0264	0.9054	0.9540	0.9172	76.50	4.47	5.75	72.90
DMU19	0.5765	0.5647	0.6141	1.8501	92.25	4.96	4.55	21.00
DMU20	0.6892	0.9250	1.0699	0.7055	76.38	4.02	6.10	31.30
DMU21	1.2903	1.1155	0.8306	0.5272	73.38	3.91	6.44	14.70
DMU22	1.7702	1.7286	1.5338	0.5682	85.38	4.22	5.16	56.70
DMU23	1.6006	1.7150	1.8198	1.1389	84.08	3.16	7.46	91.00
DMU24	0.7749	1.1940	0.6730	0.6848	85.40	3.84	6.87	19.10
DMU25	0.5078	0.4727	0.2901	0.4944	73.00	1.41	13.71	16.70
DMU26	0.7037	0.5547	0.4159	1.1163	0.00	3.10	15.39	34.00
DMU27	0.6122	0.5341	0.6379	1.4728	50.17	3.73	12.89	15.90
DMU28	1.1706	1.0919	0.8334	0.5575	2.50	1.77	7.03	95.10
DMU29	0.6368	0.7740	0.5731	0.6497	73.88	4.51	16.09	24.40
DMU30	0.7758	0.5738	0.6142	0.8509	86.54	4.85	4.90	9.70
DMU31	1.0206	0.9173	0.6094	1.2273	95.50	2.71	9.53	21.50
DMU32	1.3446	1.3243	1.1444	0.6782	65.14	4.32	5.70	49.60
DMU33	2.3904	2.1439	1.5218	0.8324	83.13	4.32	7.11	40.30
DMU34	0.6753	0.6457	0.7747	0.9863	83.80	4.35	5.55	61.50
DMU35	0.4118	0.4548	0.5033	1.5013	18.88	4.67	11.40	21.60
DMU36	0.7590	0.8277	1.0633	1.8501	90.25	3.04	4.45	46.70
DMU37	0.3500	0.8103	0.7465	0.7627	0.00	4.26	2.42	121.80

Table 2. Residuals for joint and marginal tests of significance. VRS is the original DEA residual. ALL, PROC, IMP, PART and REV are residuals for joint and marginal effects respectively. Technology is assumed to have variable returns to scale (VRS). Year = 2009.

	Residuals					
	VRS	**ALL**	**PROC**	**IMP**	**PART**	**REV**
DMU1	0.2722	0.9542	1.0223	0.9542	0.9542	1.0507
DMU2	0.9628	0.3060	0.3060	0.3274	0.3060	0.3100
DMU3	0.3069	0.6798	0.9061	0.6798	0.9036	0.8257
DMU4	1.2960	0.0000	0.0000	0.0000	0.0000	0.0000
DMU5	0.5472	0.6518	0.7641	0.6518	0.6518	0.6518
DMU6	1.1207	0.1174	0.1279	0.1174	0.1850	0.1490
DMU7	0.0000	.	.	.	.	.
DMU8	0.8048	0.0000	0.2471	0.4811	0.0000	0.3263
DMU9	1.0837	0.0876	0.0922	0.0876	0.0876	0.2120
DMU10	1.0664	0.0000	0.0000	0.0000	0.0000	0.0000
DMU11	1.1035	0.1588	0.1813	0.1588	0.1661	0.1588
DMU12	0.7903	0.5145	0.5145	0.5326	0.5145	0.5145
DMU13	0.6245	0.0000	0.0000	0.6709	0.0000	0.0000
DMU14	0.7918	0.0000	0.0000	0.0000	0.4239	0.3796
DMU15	0.7578	0.5242	0.5273	0.5242	0.5242	0.5640
DMU16	1.1901	0.0000	0.0000	0.0216	0.0000	0.0000
DMU17	0.7058	0.0000	0.0000	0.0000	0.0000	0.3415
DMU18	0.9329	0.0000	0.0310	0.1014	0.2709	0.2336
DMU19	0.0000	.	.	.	.	.
DMU20	1.1446	0.1272	0.1329	0.1272	0.1489	0.1640
DMU21	1.3229	0.0000	0.0000	0.0000	0.0000	0.0000
DMU22	1.2818	0.0000	0.0000	0.0000	0.0000	0.0000
DMU23	0.7111	0.0000	0.0000	0.0000	0.4674	0.4248
DMU24	1.1653	0.0000	0.1360	0.0000	0.1161	0.0000
DMU25	0.0000	.	.	.	.	.
DMU26	0.0000	.	.	.	.	.
DMU27	0.2879	0.8339	0.8339	0.8339	1.0311	0.8339
DMU28	1.2926	0.0000	0.0000	0.0000	0.0000	0.0160
DMU29	1.0712	0.0000	0.0000	0.0000	0.1275	0.0000
DMU30	0.9992	0.0000	0.0000	0.2545	0.0000	0.0000
DMU31	0.6081	0.0000	0.5998	0.0000	0.0000	0.0000
DMU32	1.1719	0.0176	0.0176	0.0655	0.0806	0.0750
DMU33	1.0177	0.0000	0.1107	0.0163	0.2341	0.0000
DMU34	0.8638	0.0000	0.2508	0.2925	0.3454	0.3458
DMU35	0.0000	.	.	.	.	.
DMU36	0.0000	.	.	.	.	.
DMU37	0.0000	.	.	.	.	.

Table 3: Results of the statistical tests.

Effect	Year	p-value
ALL	2002	<0.0001
	2003	<0.0001
	2004	0.0257
	2005	0.0005
	2006	0.0002
	2007	0.0028
	2008	<0.0001
	2009	<0.0001
PROC, IMP	2009	0.0165
PROC, PART	2004	0.0281
	2005	0.0002
	2009	0.0354
PROC, REV	2005	0.0158
	2006	0.0562
	2009	0.0072

Effect	Year	p-value
IMP, PART	2002	0.0366
	2005	0.0366
	2009	0.0072
IMP, REV	2009	0.0072
PART, REV	2002	0.0783
	2005	0.0783
	2006	0.0266
	2009	0.0354
IMP, PART, REV	2002	<0.0001
	2003	0.0063
	2005	0.0008
	2006	0.0266
	2008	0.0165
	2009	0.0011
PROC, PART, REV	2002	0.0063
PROC, IMP, REV	2002	0.0366
PROC, IMP, PART	2002	0.0023
	2004	0.0456

5. Summary and Conclusions

We fit a non parametric model for production data generated by Embrapa research centers for 2002 to 2009. A single output combined variables in the categories of scientific publications, technical publications, development of technologies, products and processes, and diffusion of technologies and image, to model production as a function of inputs personnel expenses, capital expenses and other expenses. The data per research center is outlier corrected and normalized by a quantity index of personnel. Residuals were computed under the assumption of variable returns to scale.

Assuming a variable returns technology we proceed to investigate the joint effects of contextual variables process improvement, financial resources generation capacity, partnership intensity, and impact of generated technologies. The assumption behind this analysis is that these variables jointly positively affect the technology through a non negative, monotone, concave function. We found the covariates jointly, but not marginally significant, indicating an effect similar to that of multicollinearity (Souza, 1998).

6. Acknowledgment

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

7. References

Banker, R.D. (1993), Maximum likelihood, consistency and DEA: a statistical foundation, *Management Science*, 39(10), 1265-1273.

Banker, R.D. and Natarajan, R. (2008), Evaluating contextual variables affecting productivity using data envelopment analysis, *Operations Research*, 56, 48-58.

Banker, R.D. and Natarajan, R., Statistical tests based on DEA efficiency scores, in Cooper, W.W., Seiford, L.M. and Zhu, J. (Eds.), *Handbook on Data Envelopment Analysis*, Kluwer International Series, Boston, 299-321, 2004.

Conover, W.J. *Practical Nonparametric Statistics*. Wiley, New York, 1998.

Daraio, C. and Simar, L. *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, Springer, New York, 2007.
Embrapa. *Manual dos indicadores de avaliação de desempenho das unidades descentralizadas da Embrapa: Metas quantitativas - Versão para ano base 2007*, Superintendência de Pesquisa e Desenvolvimento, Brasília, 2006.
Hollingsworth, B. and Smith, P. (2003), Use of ratios in data envelopment analysis, *Applied Economics Letters*, 10, 733-735.
Kotz, N. and Johnson, L. (1989), Thurstone's theory of comparative judgment, *Encyclopedia of Statistical Sciences*, 9, 237-239.
Saaty, T.L. *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*, RWS Publication, Pittsburgh, 1994.
Simar, L. and Wilson, P.W. (2007), Estimation and inference in two-stage, semi-parametric models of production processes, *Journal of Econometrics*, 136 (1), 31-64.
Souza, G.S. and Avila, A.F.D. (2000), A psicometria linear da escalagem ordinal: uma aplicação na caracterização da importância relativa de atividades de produção em ciência e tecnologia, *Cadernos de Ciência e Tecnologia*, 17 (3), 11-27.
Souza, G.S. Introdução aos Modelos de Regressão Linear e Não-Linear, Embrapa-SCT, Brasília, 1998.
Souza, G.S. and Staub, R.B. (2007), Two-stage inference using data envelopment analysis efficiency measurements in univariate production models, *International Transactions in Operational Research*, 14, 245-258.
Souza, G.S. (2006), Significância de efeitos técnicos na eficiência de produção da pesquisa agropecuária brasileira, *Revista Brasileira de Economia* 60 (1), 91-117.
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, v. 17, p. 351-363, 2010.
Souza, G.S., Alves, E. and Avila, A.F.D. (1999), Technical efficiency in agricultural research, *Scientometrics*, 46, 141-160.
Souza, G.S., Alves, E., Ávila, A.F.D. and Cruz, E.R. (1997), Produtividade e eficiência relativa de produção em sistemas de produção de pesquisa agropecuária, *Revista Brasileira de Economia*, 51 (3), 281-307.
Souza, G.S., Gomes, E.G., Magalhães, M.C. and Avila, A.F.D. (2007), Economic efficiency of Embrapa s research centers and the influence of contextual variables, *Pesquisa Operacional*, 27, 15-26.
Torgerson, W.S. *Theory and Methods of Scaling*, Wiley, New York, 1958.