A DEA APPROACH FOR AGRICULTURAL ECONOMIC EFFICIENCY AT STATE LEVEL IN BRAZIL

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

Abstract

We use Brazilian census data (1995/96 and 2006) to model agricultural production at state level in Brazil. Cost efficiency measurements were computed using data envelopment analysis techniques and the response was assessed via fractional regression. We studied the effects of time, geographic region, education and investment in agricultural research on economic efficiency. Investments in agricultural research as well as regional dummies have a significant effect on the efficiency measurements. South and Southeast states are more efficient than others on average. Increase in cost efficiency may be accomplished through investment in agricultural research.
Keywords: Economic efficiency, DEA frontiers, Quasi-maximum likelihood, Fractional regression.

Resumo

Os dados dos censos agropecuários brasileiros de 1995/96 e de 2006 foram usados para modelar a produção agrícola estadual. Medidas de eficiência custo foram calculadas com modelos de análise de envoltória de dados e a resposta foi avaliada via modelos de regressão proporcional. Foram estudados os efeitos de tempo, região geográfica, educação e investimentos em pesquisa agropecuária nas medidas de eficiência. A variável 'investimentos em pesquisa', assim como *dummies* de região, tiveram efeito estatístico significante. As regiões Sul e Sudeste são mais eficientes, em média, que as demais. Incrementos em eficiência custo podem ser obtidos via aumento do investimento em pesquisa agropecuária.
Palavras-chave: Eficiência econômica, Fronteiras DEA, Quase máxima verossimilhança, Regressão proporcional.

1. INTRODUCTION

Brazil is one of the most important countries in relation to agribusiness. In 2011, the agribusiness represented about 22% of the Brazilian GDP and 37% of its exports.

The states of the South and Southeast regions historically and, more recently, the Center-west use more technology, such as improved varieties of plants, fertilizers, irrigation, mechanization and chemicals. Brazilian agriculture differs regionally, due, primarily, to the differences in geographical areas, such as climate and natural resources, and thus production characteristics. For example, in the South region soybeans, maize, poultry and pork have particular significance, but in the Northern region rubber, nuts, wood extraction are important activities. These regional differences can cause different agricultural performances among the regions.

Since there are regional variations regarding the way the agribusiness is organized in Brazil, it seems to be plausible to expect that economic efficiency shall also differ from state to state. But some variation may also be expected from other factors, like education and investments in agricultural research. In this article we intend to investigate how these two variables affect economic efficiency.

We use Brazilian agricultural census data (1995/96 and 2006) to construct a cost frontier based on non parametric methods. Our approach for the specification of the frontier follows Banker and Natarajan (2004) and is robust relative to cost function specifications. It is not dependent on input prices. Input variables were chosen following Binswanger (1974) and Santos (1987) agricultural production model. The fractional regression approach proposed by Ramalho et al. (2010) is used to study the impact of covariates on efficiency scores.

Our discussion proceeds as follows. Section 2 is on material and methods, where we briefly discuss the approaches available for frontier analysis and present our choice of production model and statistical approach. Section 3 is on agricultural production and the type of data collected from the two censuses. Section 4 is on statistical results. Finally, in Section 5 we summarize the proposed approach and present some conclusions.

2. MATERIAL AND METHODS

Basically, two approaches are available in the literature on efficiency analysis: the stochastic efficiency frontier analysis and the deterministic frontier analysis. In the context of deterministic frontiers, Data Envelopment Analysis (DEA) is by far the most used technique.

With a single output, for the stochastic frontier one typically specifies a parametric log cost function $C(\ln p, \ln y, \theta)$ dependent on log factor input prices $\ln p$ and log output level $\ln y$, and postulates model (1), for cost data C_{it} available for a panel of N producing units in T time periods.

$$\ln C_{it} = C(\ln p_{it}, \ln y_{it}, \theta) + v_{it} + u_{it},\ i = 1...N, t = 1...T \quad (1)$$

In this formulation, θ is an unknown parameter, $C(\cdot)$ has a known functional form, and the stochastic components v_{it} and u_{it} represent random errors and inefficiency errors, respectively.

In our application the units will be the Brazilian 27 states, $N = 27$, and $T = 2$ representing two consecutive censuses (1995/96 and 2006). The specification of the distribution of the inefficiency error may include technical effects (contextual variables). A production model formulation is obtained changing prices by input quantities and changing the error term to $v_{it} - u_{it}$.

Although the formulation allows for multiple outputs, its main drawbacks in applications relate to the statistical fit of a proper flexible form, and the knowledge of input prices. In our case we only have reliable data on total input factor expenditures at state level. Our attempts to fit a Translog type production function with a normal-truncated normal specification with technical effects and using expenditures for proxies for input usage did not succeed. Therefore we were led to the DEA approach.

Data Envelopment Analysis is a technique easy to deal with multiple outputs and allows the assessment of economic efficiency without knowledge of factor input prices. This is the main reason for its use here. Banker and Natarajan (2004) show how these measurements can be computed only using total expenditures data. In this context if one is interested in the effects of contextual variables, like education and investment in research in our case, the analysis is carried out in two stages. Firstly one computes DEA economic efficiency measures from the production model, and then relates those to contextual variables, via regression procedures. The approach is discussed in detail in Simar and Wilson (2007), Souza and Staub (2007) and Banker and Natarajan (2008). Assuming exogeneity of the contextual variables, the two stage analysis is viable, as pointed out in Simar and Wilson (2007), Banker and Natarajan (2008) and Ramalho et al. (2010). The statistical problems in the two stage approach relates to cross-sectional correlations induced by the way the DEA measures are computed. Simar and Wilson (2007) suggest the use of maximum likelihood estimation with bootstrap corrections. Ramalho et al. (2010) suggest quasi-maximum likelihood methods besides the classical techniques, like nonlinear least squares and maximum likelihood estimation.

Motivated by these recent results in DEA we consider here two proposals set forth by Ramalho et al. (2010): the quasi-maximum likelihood proposed by Papke and Wooldridge (1996) and the nonlinear least squares with bootstrap standard errors. Both approaches are robust against the presence of cross-sectional correlations. In order to describe the methods, let y be the DEA score, x the vector of contextual variables and $G(\cdot)$ a nonlinear function with values in [0,1]. One postulates (2).

$$E(y \mid x) = G(x\theta) \tag{2}$$

Usual choices for $G(\cdot)$ are the logistic (3) and $G(\cdot) = \Phi(\cdot)$, where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Indeed, Papke e Wooldridge (1996) suggest the use of any distribution function adequate for binary data.

$$G(x\theta) = \frac{e^{x\theta}}{1 + e^{x\theta}} \tag{3}$$

The resulting statistical procedure is named fractional regression by Ramalho et al. (2010). The model specifies the expected value of the performance score as a monotone function of the linear construct $\mu = x\theta$. To estimate θ from observations (x_i, y_i) $i = 1, \ldots n$, we seek the vector $\hat{\theta}$ maximizing de quasi likelihood function (4).

$$\sum_{i=1}^{n}\left(y_i \log\left(G\left(x_i\theta\right)\right)+(1-y_i)\log\left(1-G\left(x_i\theta\right)\right)\right) \quad (4)$$

Papke and Wooldridge (1996) show that under correct specification for the mean function $\sqrt{n}\left(\hat{\theta}-\theta\right)\xrightarrow{d} N(0,V)$, where V is estimated using (5). As pointed out in Ramalho et al. (2010), the parameter θ may also be estimated by nonlinear least squares with corrected standard errors via bootstrap.

$$\begin{aligned}
&\hat{V}=\left(\hat{\mathrm{A}}\right)^{-1}\hat{\mathrm{B}}\hat{\mathrm{A}}\\
&\hat{\mathrm{A}}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\frac{\hat{g}_i^2}{\hat{G}_i(1-\hat{G}_i)}x_i'x_i\\
&\hat{\mathrm{B}}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\frac{\hat{u}_i^2\hat{g}_i^2}{\left(\hat{G}_i(1-\hat{G}_i)\right)^2}x_i'x_i\\
&\hat{G}_i=G(x_i\hat{\theta}),\ \hat{g}_i=G'(x_i\hat{\theta}),\ \hat{u}_i=y_i-\hat{G}_i
\end{aligned} \quad (5)$$

Economic efficiency is computed as suggested by Banker and Natarajan (2004). Let w_{it} denote aggregate agricultural output production for state *i* in period *t* and c_{it} its total factor input expenditures. Denote by $W_t=\left(w_{1t},...,w_{Nt}\right)$ the output vector for period *t* and by $C_t=\left(c_{1t},...,c_{Nt}\right)$ the factor input expenditures vector. The economic efficiency of state *i* in period *t* is simply the variable returns to scale solution to the one input one output DEA problem (6).

$$y_{it}=\min\{\theta; W_t\lambda\geq w_{it}, C_t\lambda\leq\theta c_{it}, \lambda 1=1, \lambda\geq 0\} \quad (6)$$

3. DATA

The agricultural variables we used to characterize the agricultural production model are the value of agricultural production (including livestock) on the output side, and expenditures on five factor inputs, following Binswanger (1974) and Santos (1987): land, labor, machinery, fertilizer and all other inputs.

The data were obtained from the agricultural censuses of 1995/96 and 2006 (Instituto Brasileiro de Geografia e Estatística, 2009), for each of the 27 Brazilian states. The contextual variables of interest are time dummy (Year), regional dummies, the Human Development Index (HDI) Education component (Programa das Nações Unidas para o Desenvolvimento, 2004) (Education) and the number of researchers (Research) working for the Brazilian Agricultural Research Corporation (Embrapa) research centers and for the Brazilian agricultural state companies, called OEPAs (Organizações Estaduais de Pesquisa Agropecuária). Tables 1 and 2 provide all the data information used in the article.

Table 1 – Input, output, contextual variables and economic efficiency data for Year = 1995/96.

State	Region	Land	Labor	Other costs	Fertilizers	Capital	Output	HDI-Education	Investment Research	Efficiency
Acre	North	63,596	15,650	53,131	359	2,576	276,100	0.698	23	1.0000
Alagoas	Northeast	388,333	254,135	459,091	109,586	10,753	1,686,143	0.634	42	0.4551
Amapá	Northeast	33,193	21,063	57,934	4,224	7,129	177,382	0.856	19	1.0000
Amazonas	North	126,916	28,955	1,817,801	2,745	4,711	943,931	0.772	61	0.1689
Bahia	Northeast	4,559,462	795,835	144,146	326,446	54,534	5,414,449	0.701	143	0.2837
Ceará	Northeast	680,289	262,719	717,229	32,949	19,077	2,367,382	0.664	123	0.4435
Distrito Federal	Center-West	32,093	39,815	146,002	35,670	7,108	348,587	0.902	271	0.6020
Espírito Santo	Southeast	1,255,774	345,196	778,117	151,072	23,240	2,788,048	0.811	66	0.3465
Goiás	Center-West	3,435,597	713,795	2,524,097	506,800	99,015	6,652,280	0.799	118	0.2799
Maranhão	Northeast	839,310	141,068	371,889	35,645	14,979	1,798,160	0.656	49	0.4202
Mato Grosso	Center-West	3,887,637	566,202	2,396,541	624,719	99,081	5,112,096	0.811	39	0.2084
Mato Grosso do Sul	Center-West	3,630,812	572,055	2,350,199	309,276	86,292	5,619,410	0.811	133	0.2489
Minas Gerais	Southeast	5,593,134	2,437,773	6,053,015	1,122,986	228,376	16,506,998	0.813	292	0.4119
Pará	North	1,165,733	229,711	740,164	27,977	17,394	2,644,358	0.756	135	0.3860
Paraíba	Northeast	465,307	157,907	286,449	27,954	4,958	1,206,259	0.679	113	0.4380
Paraná	South	3,696,631	1,018,028	5,410,523	926,809	287,298	14,327,529	0.828	268	0.3817
Pernambuco	Northeast	795,203	517,332	918,542	120,962	21,695	3,166,633	0.719	193	0.4203
Piauí	Northeast	447,408	93,157	242,311	14,431	15,945	881,507	0.663	54	0.3885
Rio de Janeiro	Southeast	855,652	225,011	559,423	62,444	11,306	1,623,740	0.874	220	0.3136
Rio Grande do Norte	Northeast	467,797	158,653	5,606,880	44,240	7,466	916,720	0.712	48	0.0519
Rio Grande do Sul	South	2,018,147	822,716	623,414	1,018,621	311,546	15,890,978	0.867	313	1.0000
Rondônia	North	610,954	64,126	301,905	4,811	8,444	860,781	0.802	25	0.3128
Roraima	North	211,305	16,683	52,633	6,384	2,923	159,904	0.837	22	0.4261
Santa Catarina	South	1,151,200	419,741	4,013,283	393,406	128,088	8,423,301	0.860	212	0.4203
São Paulo	Southeast	3,927,011	3,878,861	9,764,660	1,533,964	340,724	21,666,578	0.882	794	1.0000
Sergipe	Northeast	523,099	89,668	176,919	32,455	5,155	704,483	0.737	41	0.3181
Tocantins	North	1,088,050	124,748	390,651	24,733	14,896	917,843	0.758	0	0.1989

Table 2 – Input, output, contextual variables and economic efficiency data for Year = 2006.

State	Region	Land	Labor	Other costs	Fertilizers	Capital	Output	HDI-Education	Investment Research	Efficiency
Acre	North	140,714	24,766	84,433	2,016	6,117	347,876	0.844	33	0.6621
Alagoas	Northeast	429,693	399,694	459,825	480,789	21,578	3,273,161	0.759	14	0.7819
Amapá	Northeast	46,144	6,214	8,575	4,216	1,572	100,228	0.919	19	1.0000
Amazonas	North	377,487	63,432	126,618	6,613	5,377	650,508	0.925	56	0.5144
Bahia	Northeast	3,880,293	1,399,411	3,634,427	1,444,147	255,219	8,415,197	0.830	122	0.4492
Ceará	Northeast	642,920	289,346	635,787	56,338	18,256	3,848,241	0.808	99	1.0000
Distrito Federal	Center-West	36,701	69,604	144,096	49,161	10,402	432,828	0.962	269	0.6664
Espírito Santo	Southeast	723,982	460,140	650,029	219,679	57,704	2,343,280	0.887	48	0.4783
Goiás	Center-West	3,615,340	1,007,670	3,349,043	1,133,859	193,319	6,242,251	0.891	122	0.3528
Maranhão	Northeast	1,364,820	373,944	590,268	248,556	51,201	3,121,509	0.784	0	0.5086
Mato Grosso	Center-West	4,058,945	1,400,245	6,544,208	3,784,176	271,954	9,601,893	0.898	33	0.3632
Mato Grosso do Sul	Center-West	3,769,700	1,012,603	2,898,968	1,046,368	227,289	3,563,155	0.894	149	0.1700
Minas Gerais	Southeast	3,980,624	3,665,154	8,111,462	2,822,284	471,513	18,839,267	0.878	282	0.8279
Pará	North	1,729,312	551,154	884,905	83,138	186,822	3,335,581	0.861	122	0.4154
Paraíba	Northeast	387,392	117,039	392,442	61,662	11,203	1,422,049	0.793	132	0.6419
Paraná	South	3,445,894	1,680,067	5,618,565	2,243,063	473,674	15,897,868	0.913	246	0.9160
Pernambuco	Northeast	1,214,473	447,818	1,506,802	246,877	20,626	4,819,188	0.811	156	0.6713
Piauí	Northeast	866,584	115,605	460,547	99,813	120,287	1,327,899	0.779	56	0.3506
Rio de Janeiro	Southeast	501,228	265,171	357,608	57,273	19,021	1,247,884	0.945	207	0.4576
Rio Grande do Norte	Northeast	409,922	178,113	251,882	131,091	13,008	1,121,001	0.81	55	0.5040
Rio Grande do Sul	South	2,543,379	1,347,273	5,476,487	3,199,655	584,744	16,693,595	0.921	317	1.0000
Rondônia	North	511,531	103,619	519,338	33,988	38,278	850,749	0.885	26	0.3168
Roraima	North	111,226	12,403	33,947	11,685	3,136	98,916	0.885	26	1.0000
Santa Catarina	South	1,028,090	598,088	2,517,056	616,543	320,924	8,873,639	0.934	182	1.0000
São Paulo	Southeast	4,195,518	5,773,992	9,717,815	3,494,639	756,122	25,523,374	0.921	1013	1.0000
Sergipe	Northeast	328,647	272,287	679,459	121,174	10,432	1,065,216	0.827	63	0.3346
Tocantins	North	743,998	216,638	477,682	458,827	56,092	764,955	0.86	0	0.1773

4. RESULTS AND DISCUSSION

Average cost efficiency statistics are shown on Table 3. We see that the South and Southeast regions are considerably more economic efficient than the other regions.

Table 3 – Economic (cost) efficiency by regions.

Region	Mean	Standard Error	[95% Confidence	Interval]
South	0.7863	0.1227	0.5403	1.0324
Southeast	0.6045	0.1026	0.3988	0.8101
North	0.5413	0.0878	0.3653	0.7173
Northeast	0.4701	0.0488	0.3721	0.5680
Center-West	0.3615	0.0641	0.2328	0.4901

Table 4 shows the statistical results of quasi-maximum likelihood estimation for fractional regression with the logistic specification. The probability normal assumption leads to almost equivalent results. We used SAS 9.2 software – Proc Nlmixed (SAS, 2012) for these computations. The covariates of interest are a census dummy (time effect), regional dummies – Center-West, Northeast, North and Southeast (South was dropped from the model), HDI Education, and investment in agricultural research.

Table 4 – Statistical results of quasi maximum likelihood estimation for fractional regression with the logistic.

Parameter	Coefficient	Standard deviation	z	p-value
Constant	-0.58180	4.82681	0.12054	0.90406
Center-West	-1.42540	0.60677	2.34916	0.01882
Northeast	-0.61690	0.96332	0.64039	0.52192
North	-0.15470	0.70291	0.22009	0.82580
Southeast	-1.06780	0.59616	1.79113	0.07327
Education	0.41210	5.89379	0.06992	0.94426
Research	0.00464	0.00124	3.74113	0.00018
Year	0.75950	0.62206	1.22094	0.22211

Statistically significant contextual variables are regional dummies and investments in agricultural research. Regional effects significant in $\mu = x\theta$ are Center-West and Southeast, the later being marginal. The order induced by the estimates is given in Table 5, which mimics Table 3.

Table 5 – Value of $\mu = x\theta$ by regions.

Region	Mean	Standard Error	[95% Confidence	Interval]
South	1.3528	0.4624	0.8676	1.8380
Southeast	0.7855	1.7056	-0.6404	2.2114
North	0.1737	0.4553	-0.0891	0.4366
Northeast	-0.1257	0.4597	-0.3543	0.1029
Center-West	-0.6111	0.6081	-1.1195	-0.1027

For each additional 100 researchers hired, we would expect a significant 0.115 increase in economic efficiency, where the effect is computed considering the average value of $\mu = x\theta$ over both censuses. The marginal effect is given by $effect_j = \theta_j \frac{e^{\mu}}{(1+e^{\mu})^2}$. Figure 1 depicts marginal effects as a function of $\mu = x\theta$. The maximum response is obtained when expected cost efficiency is 0.5. For the average $\mu = x\theta$, the corresponding expected efficiency is 0.7805.

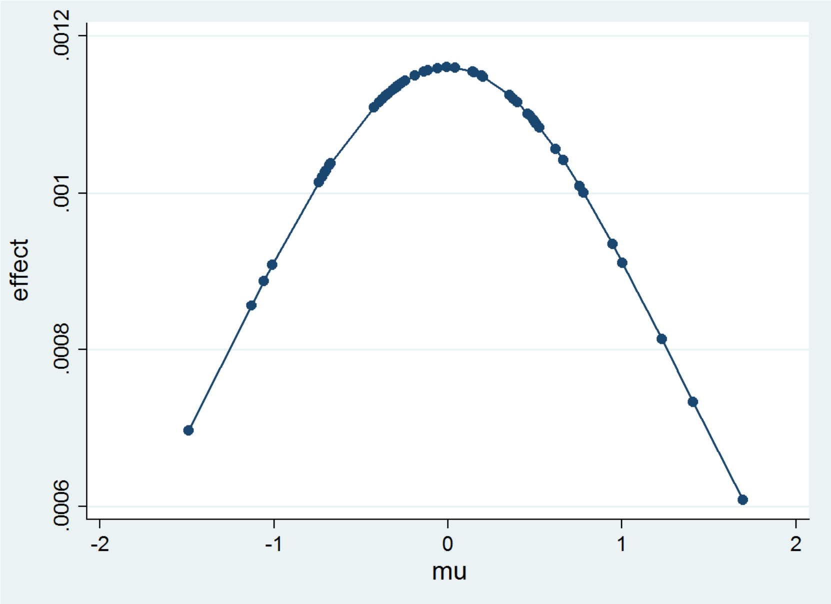

Figure 1 – Marginal effect of research as a function of the linear construct $\mu = x\theta$.

Overall correlation between observed and predicted values is 0.623, indicating a reasonable fit. Within 1995/96 the correlation is 0.529, and within 2006, 0.616.

Estimation results using nonlinear least squares with bootstrap corrections (5,000 replicates) are given in Table 6. We used Stata 12 software (STATA, 2011) for these computations. Estimation is similar, in overall fit, to fractional regression, but the marginal regional effects are not significant. Our choice for the fractional model relates to quasi-maximum likelihood being more efficient (Ramalho et al., 2010), i.e., it has smaller standard deviations. In addition, the likelihood function of the quasi-maximum likelihood approach seems to be more adequate to describe efficient units. Also, intuition leads one to expect the South region as significantly more efficient, which is in agreement with the fractional regression results.

Table 6 – Statistical results of nonlinear least squares with bootstrap corrections.

Parameter	Observed Coefficient	Bootstrap Standard Error	z	P>|z|	Normal-based [95% Confidence Interval]	
Constant	-0.89363	9.53318	-0.09	0.925	-19.57832	17.79106
Center-West	-1.33232	7.43970	-0.18	0.858	-15.91387	13.24923
Northeast	-0.50871	7.49612	-0.07	0.946	-15.20083	14.18341
North	-0.10218	7.44547	-0.01	0.989	-14.69504	14.49068
Southeast	-0.98283	7.48110	-0.13	0.895	-15.64551	13.67985
Education	0.81148	7.39101	0.11	0.913	-13.67462	15.29758
Research	0.00394	0.00174	2.27	0.023	0.00053	0.00735
Year	0.69039	0.78462	0.88	0.379	-0.84742	2.22821

5. SUMMARY AND CONCLUSIONS

We use DEA and Brazilian agricultural censuses data (1995/96 and 2006) to assess the effect of contextual variables on cost efficiency. These variables were education, measured by Human Development Index (HDI) Education indicator, and investment in agricultural research, measured by number of researchers.

The production model here proposed uses the value of total agricultural output as the output variable and aggregate expenditures on land, fertilizers, labor, machinery and other inputs as the input variable.

We conclude that investments in agricultural research as well as regional dummies have a significant effect on the efficiency measurements. Overall economic efficiency of the agricultural sector increased (39%) from 0.442 in 1995/96 to 0.613 in 2006, while the HDI-Education increased (12%) from 0.774 to 0.868. Investment in research was stable in the period.

South and Southeast states are significantly more efficient than other states on average.

These empirical results suggest that there are significant possibilities to increase cost efficiency levels in the Brazilian agriculture, especially in the Center-West, Northeast and North regions. Increase in efficiency may be accomplished through investments in agricultural in research.

REFERENCES

[1] BANKER, R.D.; NATARAJAN, R. Evaluating contextual variables affecting productivity using data envelopment analysis. **Operations Research**, v. 56, p. 48-58, 2008.

[2] 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**. Boston: Kluwer International Series, 2004. p. 299-321.

[3] BINSWANGER, H.P. A cost function approach to the measurement of elasticities of factor demand and elasticities of substitution. **American Journal of Agricultural Economics**, v. 56, n. 2, p. 377-386, 1974.

[4] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Banco de Dados Agregados. Sistema IBGE de Recuperação Automática - SIDRA**. 2009. Disponível em: http://www.sidra.ibge.gov.br. Accesso em: 18 Nov 2009.

[5] 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**, v. 11, p. 619-632, 1996.

[6] PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Desenvolvimento Humano e IDH**. 2004. Disponível em: http://www.pnud.org.br/idh/. Acesso em: 18 Nov 2009.

[7] RAMALHO, E.A.; RAMALHO, J.J.S.; HENRIQUES, P.D. Fractional regression models for second stage DEA efficiency analyses. **Journal of Productivity Analysis**, v. 34, p. 239-255, 2010.

[8] SAS. SAS/STAT® 12.1 User's Guide. North Carolina: SAS Institute Inc. 2012.

[9] SANTOS, R.F. **Presença de Vieses de Mudança Técnica na Agricultura Brasileira**. São Paulo: IPE/USP. 1987.

[10] SIMAR, L.; WILSON, P.W. Estimation and inference in two-stage: semi-parametric models of production processes. **Journal of Econometrics**, v. 136, p. 31-64, 2007.

[11] SOUZA, G.S.; STAUB, R.B. Two-stage inference using data envelopment analysis efficiency measurements in univariate production models. **International Transactions in Operational Research**, v. 14, p. 245-258, 2007.

[12] STATA. **STATA Base Reference Manual Release 12**. College Station: Stata Press. 2011.