ECONOMIC EFFICIENCY AND THE EFFECTS OF EDUCATION AND INVESTMENT IN RESEARCH IN THE BRAZILIAN AGRICULTURAL SECTOR

Geraldo da Silva e Souza
Eliane Gonçalves Gomes
Rosaura Gazzola
Antonio Flavio Dias Avila
Brazilian Agricultural Research Corporation (Embrapa)
Parque Estação Biológica – PqEB, Av. W3 Norte final, 70770-901, Brasília, DF
{geraldo.souza; eliane.gomes; rosaura.gazzola; flavio.avila}@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 censored regressions. We studied the effects of region, education and investment in agricultural research on economic efficiency. Education had a strong significant effect, as well investment in research. The intensity of the effects varied over regions for education and was statistically the same for investment in research. The South region responded better to stimulus in education, followed closely by the Southeast. The highest efficiency levels were observed for the same regions.

KEYWORDS. Economic efficiency. DEA frontiers. Panel regressions. Tobit regressions. Agriculture.

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 com dados censurados. Foram estudados os efeitos de região, educação e investimentos em pesquisa agropecuária nas medidas de eficiência. A variável educação teve forte efeito positivo, assim como investimentos em pesquisa. A intensidade dos efeitos varia por região para a covariável educação e não há diferença estatística para investimentos em pesquisa. A região sul responde melhor a estímulos em educação, seguida imediatamente pela região sudeste. As maiores medidas de eficiência são também observadas para estas regiões.

PALAVRAS-CHAVE. Eficiência econômica. Fronteiras DEA. Regressões em painel. Regressão Tobit. Agricultura.

1. Introduction

Brazil is one of the most important countries in relation to agribusiness. Agribusiness represents about 25% of Brazilian GDP, 36% of its exports in 2008 and 37% of jobs in 2008.

The states of the South and Southeast 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 investment in 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.

Our discussions in the article 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.

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

In this formulation, θ is an unknown parameter, $C(.)$ 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, 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 Simmar and Wilson (2007), Souza and Staub (2007) and Banker and Natarajan (2008). Assuming exogeneity of the contextual variables, the two stage analysis is viable. Motivated by these recent results in DEA we consider here an extension to panel data, following Staub et al. (2010). We assume the Tobit specification (2).

$$y_{it}^{o} = \sum_{k=1}^{K} x_{itk}\beta_k + u_{it} \tag{2}$$

In (2), the x_{itk} are observations on k covariates of interest, the quantities β_k are unknown parameters, and the component u_{it} is the independent sum of a state specific random effect and a random error, that is $u_{it} = \nu_i + \varepsilon_{i,t}$. The random effects ν_i are assumed iid $N(0,\sigma^2)$ and $\varepsilon_{i,t}$ are iid $N(0,\sigma_\varepsilon^2)$. The responses y_{it}^0 represent the censored values of the economic efficiencies y_{it}. For an efficiency measurement in (0,1), $y_{it} = y_{it}^0$. If $y_{it} = 1$, then $y_{it} \le y_{it}^0$. In this context one allows for over efficiency. See McCarty and Yaisawarng (1993) and Souza et al. (2006) for more details on the Tobit formulation.

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 $W_t = (w_{1t}, \ldots, w_{Nt})$ the output vector for period *t* and by $C_t = (c_{1t}, \ldots, c_{Nt})$ 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 (3).

$$y_{it} = \min\left\{\theta; W_t\,\lambda \ge w_i,\, C_t\,\lambda \le \theta c_{it}, \lambda 1 = 1, \lambda \ge 0\right\} \tag{3}$$

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 regional dummies (reg), the Human Development Index (HDI) Education component (Programa das Nações Unidas para o Desenvolvimento, 2004) (hdi-e) 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 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 maximum likelihood estimation for the Tobit model. We used the Stata 10.1 software (Stata, 2007). The covariates of interest are the regional dummies reg1-reg4, representing the regions Center-West, Northeast, North and Southeast, respectively (South = reg5 was dropped from the model), HDI Education – hdi-e, investment in agricultural research – research, and the interaction products x1 = reg1*hdi-e, x2 = reg_2*hdi-e, x3 = reg3*hdi-e and x4 = reg4*hdi-e.

Table 5 – Tobit model estimation statistics.

	Coefficient	Standard Error	z	P>|z|	[95% Confidence Interval]	
x1	-7.2271	2.5469	-2.84	0.005	-12.2188	-2.2353
x2	-6.2136	2.2851	-2.72	0.007	-10.6923	-1.7349
x3	-7.5041	2.3222	-3.23	0.001	-12.0556	-2.9527
x4	-5.6264	2.8004	-2.01	0.045	-11.1152	-0.1376
reg1	5.9876	2.2390	2.67	0.007	1.5991	10.37601
reg2	5.5236	1.9884	2.78	0.005	1.6263	9.420846
reg3	6.5665	2.0355	3.23	0.001	2.5770	10.5561
reg4	4.6942	2.4512	1.92	0.055	-0.1101	9.4984
research	0.00104	0.00046	2.26	0.024	0.00013	0.00195
hdi-e	8.1187	2.1897	3.71	0.000	3.8269	12.4104
constant	-6.5506	1.9324	-3.39	0.001	-10.3381	-2.76321
sigma_*v*	0.1895	0.0412	4.59	0.000	0.1087	0.27031
sigma_*u*	0.1601	0.0254	6.30	0.000	0.1103	0.21001
rho	0.5833	0.1463			0.2999	0.8278

All contextual variables are statistically significant and the evidence is in the direction of strong associations of economic efficiency and HDI-Education and investment in research. The significance of the interaction effects provide indication that education effects vary over regions. On the other hand, investment in research has an uniform effect over all regions. For each additional 100 researchers hired, we would expect a significant 0.1 increase in economic efficiency.

The slope coefficients for HDI-Education in regions Center-West, Northeast, North, Southeast and South are 0.892, 1.905, 0.615, 2.492, and 8.119, respectively, indicating dominance of the South and Southeast regions.

Overall correlation between observed and predicted values is 0.637, indicating a reasonable fit. Within 1995/96 the correlation is 0.575, and within 2006, 0.614.

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 adopted 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 both contextual variables 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. Response to education follows the same pattern. Response to investment in research is uniform over regions.

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

References

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

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

Binswanger, H.P. (1974), A cost function approach to the measurement of elasticities of factor demand and elasticities of substitution, *American Journal of Agricultural Economics*, 56 (2), 377-386.

Instituto Brasileiro de Geografia e Estatística (2009), *Banco de Dados Agregados. Sistema IBGE de Recuperação Automática - SIDRA*. Available at: <http://www.sidra.ibge.gov.br>. Accessed in: 18 Nov. 2009.

McCarty, T.A. e Yaisawarng, S. Technical efficiency in New Jersey school districts, Fried, H., Lovell, C.A.K. e Schmidt, S.S. (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford, 271-287, 1993.

Programa das Nações Unidas para o Desenvolvimento (2004), *Desenvolvimento Humano e IDH*. Available at: <http://www.pnud.org.br/idh/>. Accessed in: 18 Nov. 2009.

Santos, R.F. dos. *Presença de Vieses de Mudança Técnica na Agricultura Brasileira*, IPE/USP, São Paulo, 1987.

Simar, L. e Wilson, P.W. (2007), Estimation and inference in two-stage: semi-parametric models of production processes. *Journal of Econometrics*, 136, 31-64.

Souza, G. da S., Staub, R.B. e Tabak, B.M. (2006), Assessing the significance of factors effects in output oriented DEA measures of efficiency: an application to Brazilian banks, *Revista Brasileira de Economia de Empresas*, 6 (1), 7-20.

Souza, G.S. e 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.
Stata. *STATA v. 10.1*. StataCorp LP, College Station, TX, USA, 2007.
Staub, R.B., Souza, G. da S. e Tabak, B.M. (2010), Evolution of bank efficiency in Brazil: A DEA approach, European Journal of Operational Research, 202 (1), 204-213.