INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 22 (2015) 329–337
DOI: 10.1111/itor.12055

Improving agricultural economic efficiency in Brazil

Geraldo da Silva e Souza and Eliane Gonçalves Gomes

Brazilian Agricultural Research Corporation (Embrapa), SGE, PqEB, W3 Norte final, Asa Norte, Brasília, 70770-901 DF, Brazil

E-mail: geraldo.souza@embrap.br [da Silva e Souza]; eliane.gomes@embrapa.br [Gomes]

Received 11 March 2013; received in revised form 13 September 2013; accepted 15 September 2013

Abstract

In this article, we use Brazilian census data (1995/1996 and 2006) to model agricultural production at the state level in Brazil. Cost-efficiency measurements are computed using data envelopment analysis techniques, and the response is assessed via fractional regression. We examine the effects of time, geographic region, education, and investment in agricultural research on economic efficiency. We found that investment in agricultural research and regional dummies have a significant effect on efficiency measurements. On average, South and Southeast states are more efficient than other states. An increase in cost efficiency can be accomplished through investment in agricultural research.

Keywords: economic efficiency; DEA frontiers; quasi-maximum likelihood; fractional regression

1. Introduction

Brazil is one of the most important countries in terms of agribusiness. In 2011, agribusiness represented about 22% of Brazilian GDP and 37% of its exports. The states of the South and Southeast regions, and more recently the Center-west, use more technology, such as improved varieties of plants, fertilizers, irrigation, mechanization, and chemicals. Brazilian agriculture differs regionally because of differences in geographical areas, such as climate and natural resources, and thus production characteristics also differ. For example, in the South region, soybeans, maize, poultry, and pork have particular significance, but in the Northern region, rubber, nuts, and wood extraction are important activities. These regional differences can cause different agricultural performances among regions.

Since there are regional variations in the way agribusiness is organized in Brazil, it can be expected that economic efficiency also differs from state to state. However, some variation can also be expected from other factors, such as education and investment in agricultural research. In this article, we investigate how these two variables affect economic efficiency. This study is useful for managers and decision makers as it examines factors that may cause or influence efficiency. In this sense, it can

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA02148, USA.

be seen as a decision support tool that can, for instance, identify sectors that need the allocation of resources for improved efficiency.

We use Brazilian agricultural census data (1995/1996 and 2006) to construct a cost frontier based on nonparametric methods. Our approach to 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 are chosen following the agricultural production models of Binswanger (1974) and Santos (1987). 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 materials and methods, where we briefly discuss the approaches available for frontier analysis and present our choice of production model and statistical approach. Section 3 discusses agricultural production and the type of data collected from the two censuses. Section 4 provides statistical results. Finally, in Section 5, we summarize the proposed approach and provide some conclusions.

2. Materials and methods

Basically two approaches are available in the literature on efficiency analysis: stochastic efficiency frontier analysis and deterministic frontier analysis. In the context of deterministic frontiers, data envelopment analysis (DEA) is by far the most used technique. Owing to the small sample size and our unsuccessful attempts to specify a meaningful parametric response function, our choice was the DEA approach. In our application, the sample comprises the 27 Brazilian states, which are the decision-making units (DMUs), and two time periods representing two consecutive agricultural censuses (1995/1996 and 2006).

DEA can easily deal with multiple outputs and assess economic efficiency without knowledge of factor input prices. This is another reason for its use in this article. Banker and Natarajan (2004) show how these measurements can be computed using only total expenditure data. In this context, if one is interested in the effects of contextual variables, such as education and investment in research in our case, the analysis is carried out in two stages. First, one computes DEA economic efficiency measures from the production model and then relates those to contextual variables via regression procedures. This approach is discussed in detail in Simar and Wilson (2007), Souza and Staub (2007), and Banker and Natarajan (2008). Assuming the exogeneity of the contextual variables, a 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 relate to the cross-sectional correlations induced by the way DEA measures are computed. Simar and Wilson (2007) suggest using the maximum likelihood estimation with bootstrap corrections, whereas Ramalho et al. (2010) suggest using quasi-maximum likelihood (QML) methods in addition to classical techniques, such as the nonlinear least squares and maximum likelihood estimations.

Motivated by these recent results in DEA, we consider the proposal in Ramalho et al. (2010): fractional regression with the QML of Papke and Wooldridge (1996) as the method of estimation. This approach is robust to the presence of cross-sectional correlations. Indeed, Ramalho et al. (2010) propose other alternative models based on what they call two-part models. These latter classes of models do not seem to be compatible with our data, given the small probabilities of efficient units.

In order to describe the fractional regression approach, 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

$$E(y|x) = G(x\theta). \tag{1}$$

The usual choices for $G(\cdot)$ are the logistic (2), $G(x\theta) = \Phi(x\theta)$, where $\Phi(\cdot)$ is the distribution function of the standard normal distribution and $G(x\theta) = 1 - \exp\{-\exp\{x\theta\}\}$. Indeed, Papke and Wooldridge (1996) suggest the use of any distribution function adequate for binary data:

$$G(x\theta) = \frac{e^{x\theta}}{1 + e^{x\theta}}. \tag{2}$$

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 the observations $(x_i, y_i)\ i = 1, \ldots, n$, we seek the vector $\hat{\theta}$ maximizing the QML function (3):

$$LL(\theta) = \sum_{i=1}^{n} (y_i \log(G(x_i\theta) + (1 - y_i) \log(1 - G(x_i\theta)))). \tag{3}$$

Papke and Wooldridge (1996) show that under the correct specification of the mean function $\sqrt{n}(\hat{\theta} - \theta)N(0,\ V)$, V is estimated using (4). Quoting Ramalho et al. (2010), the "QML estimator is efficient within the class of estimators containing all linear exponential family-based QML and weighted nonlinear least squares estimators." Although not efficient, the parameter θ may also be estimated by nonlinear least squares:

$$\begin{aligned}
\hat{V} &= (\hat{A})^{-1}\hat{B}\hat{A} \\
\hat{A} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}_i^2}{\hat{G}_i(1 - \hat{G}_i)} x_i' x_i \\
\hat{B} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{u}_i^2 \hat{g}_i^2}{(\hat{G}_i(1 - \hat{G}_i))^2} x_i' x_i \\
\hat{G}_i &= G(x_i\hat{\theta}),\ \hat{g}_i = G'(x_i\hat{\theta}),\ \hat{u}_i = y_i - \hat{G}_i.
\end{aligned} \tag{4}$$

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} is its total factor input expenditure. 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 expenditure 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:

$$y_{it} = \min\{\theta;\ W_t\lambda \geq w_{it}, C_t\lambda \leq \theta c_{it}, \lambda 1 = 1, \lambda \geq 0\}. \tag{5}$$

3. Data

The agricultural variables used to characterize the agricultural production model are the value of agricultural production (including livestock) on the output side and expenditure 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 1995/1996 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 (Center-west, North, Northeast, South, Southeast), the human development index (HDI) education component (Programa das Nações Unidas para o Desenvolvimento, 2004), and the number of researchers (research) working for Brazilian Agricultural Research Corporation (*Embrapa*) research centers and Brazilian agricultural state companies, called OEPAs (*Organizações Estaduais de Pesquisa Agropecuária*). Tables 1 and 2 provide all the data used in the article.

4. Results and discussion

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

Ramalho et al. (2010) suggest using Regression Equation Specification Error Tests (RESET), following Pagan and Vella (1989), and *P* tests for non-nested hypothesis, following Davidson and MacKinnon (1981), for the choice between different functional forms for the mean function. In our particular application, we did not obtain convergence with the Pagan and Vella (1989) approach or with the form of the test presented in Gallant (1987) and Asteriou and Hall (2007). The test of Davidson and MacKinnon (1981) also did not indicate a superior specification. Indeed, the three alternatives considered here (logistic, probit, and log–log) could not be rejected. The *p*-values for the comparisons involved for the pairs (null alternative) logistic versus probit, probit versus logistic, log–log versus probit, probit versus log–log, log–log versus logistic, and logistic versus log–log were 0.419, 0.491, 0.538, 0.274, 0.530, and 0.241, respectively. Further, the three functional forms provided similar fits. This is provided in Table 4 that shows the overall fits for each of them. The best choice seemed to be the log–log specification following the arguments of Asteriou and Hall (2007).

Table 5 shows the statistical results of the QML estimation for fractional regression with the log–log specification. We used SAS 9.2 software—Proc Nlmixed (SAS, 2012) for the computation of the QML estimator and Proc IML to estimate the variance–covariance matrix.

The joint test of $\theta = 0$ produced a significant chi-square statistic of 40.6 with 8 df (p-value $<$ 0.001). The joint regional effect produced a significant chi-square statistic of 33.1 with 4 df (p-value < 0.001). The performance suggested by the QML, in an increasing order, was Center-West, Southeast, Northeast, North, and South. The first and last positions are in agreement with Table 3. The marginal tests, however, provided some indication that only Center-West differs significantly from South (p-value $= 0.077$). Investment in agricultural research was statistically significant. For each additional 100 researchers hired, we can expect a significant 0.314 increase in economic efficiency, where the effect is computed considering the average value of $\mu = x\theta$ over both censuses. The marginal effect of a variable is given by $effect_j = \theta_j \exp(\mu - \exp(\mu))$. Figure 1 depicts the marginal effects for investment in research as a function of $\mu = x\theta$. The maximum response is always obtained when expected cost efficiency is 0. For the average $\mu = x\theta$, the corresponding expected efficiency was 0.430 for 1995/1996 and 0.639 for 2006, suggesting a significant increase in efficiency over time.

Table 1
Input, output, and contextual variables and economic efficiency data for year 1995/1996

State	Region	Land	Labor	Other costs	Fertilizers	Capital	Output	HDI-education	Investment research	Efficiency
Acre	North	63,596	15,650	53,131	359	2576	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	4224	7129	177,382	0.856	19	1.0000
Amazonas	North	126,916	28,955	1,817,801	2745	4711	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	7108	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	4958	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	7466	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	4811	8444	860,781	0.802	25	0.3128
Roraima	North	211,305	16,683	52,633	6384	2923	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	5155	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, and 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	2016	6117	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	6214	8575	4216	1572	100,228	0.919	19	1.0000
Amazonas	North	377,487	63,432	126,618	6613	5377	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	3136	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

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
Fit statistics

Functional form	−2*LL	AIC (smaller is better)	BIC (smaller is better)
Logistic	67.3	83.3	99.2
Log–log	66.8	82.8	98.7
Probit	67.1	83.1	99.1

AIC, Akaike information criterion; BIC, Bayesian information criterion.

Table 5
Statistical results of the QML estimation for fractional regression with the log–log functional form

Parameter	Coefficient	Standard deviation	z	p-Value
Constant	– 1.325	5.547	0.239	0.811
Research	0.003	0.001	2.681	0.007
Center-West	– 1.003	0.567	1.770	0.077
Northeast	– 0.333	1.031	0.323	0.747
North	– 0.093	0.681	0.137	0.891
Southeast	– 0.722	0.550	1.314	0.189
Education	0.898	6.710	0.134	0.894
Year	0.510	0.683	0.746	0.455

Agricultural research has played an important role in agricultural production and productivity growth as well as overall economic development in the world. In Brazil, high-yielding varieties and other modern technologies released and developed by national agricultural research systems have substantially increased crop and animal yields, the productivity of labor, land, and capital as well as agriculture's diversification and competitiveness in the world market.

Rapid agricultural growth driven by such productivity increases and competitiveness has played a crucial role in Brazil's economic transformation and development process. Both rural and urban populations have benefited from such development either through the direct rural income effect or indirect food price effect on the cost of living. Moreover, opportunities for rural-to-urban migration increase income and thereby the tax revenue of the government that has financed the direct transfer program to the poor.

Our results are in agreement with Gasques et al. (2012), who emphasize the importance of technological factors in agricultural productivity. At the state level, we see that the implication of our study, in terms of the decision-making process, is that production efficiency may be achieved

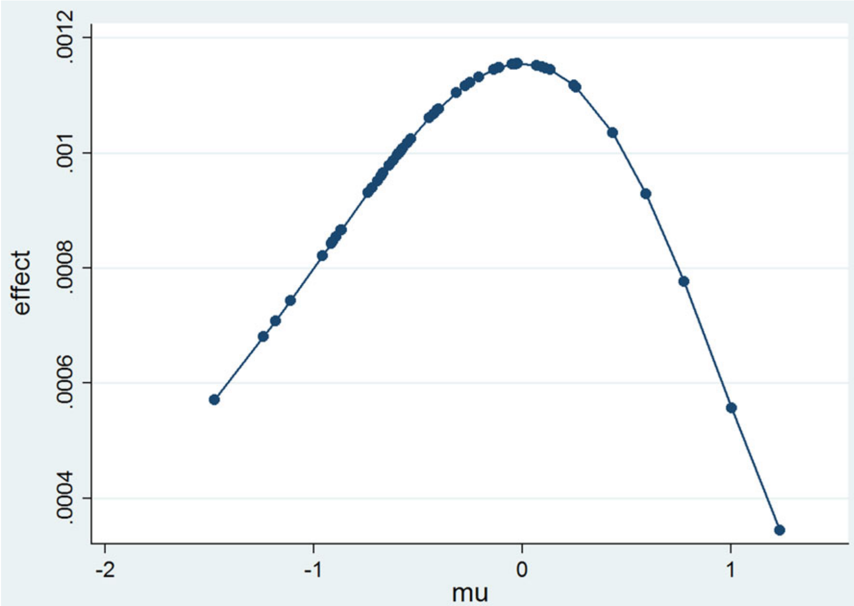

Fig. 1. Marginal effect of research as a function of the linear construct $\mu = x\theta$.

through the incorporation of technology via rural extension. The agricultural census data of 2006 also indicate a strong technical assistance effect on the production frontier (Alves et al., 2012). Effective rural extension is further associated with the number of qualified researchers due to spillover effects.

5. Summary and conclusions

We use DEA and Brazilian agricultural census data (1995/1996 and 2006) to assess the effect of contextual variables on cost efficiency. These variables were education and investment in agricultural research. This kind of approach is useful to support decision-making processes, as it can identify factors that may cause or influence efficiency. The production model proposed here uses the value of total agricultural output as the output variable and aggregate expenditure on land, fertilizers, labor, machinery, and other inputs as the input variable.

We conclude that investment in agricultural research and regional dummies have a significant effect on efficiency measurements. Overall, the economic efficiency of the agricultural sector increased (by 39%) from 0.442 in 1995/1996 to 0.613 in 2006, whereas HDI-education increased (by 12%) from 0.774 to 0.868. Investment in research was stable in the period. Expected efficiency agreed with these data, increasing by 49% from 0.430 in 1995/1996 to 0.639 in 2006.

South and Southeast states are more efficient than other states on average in the original cost-efficiency score over censuses. These empirical results suggest that there are significant possibilities to increase cost-efficiency levels in Brazilian agriculture, especially in the Center-West, Northeast, and North regions. Such an increase in efficiency can be accomplished by means of investment in agricultural research. The incorporation of technology at the farm level and resulting increase in economic performance are understood as a spillover effect.

References

Alves, E.R.A., Souza, G.S., Gomes, E.G., Magalhães, E., Rocha, D.P., 2012. Um modelo de produção para a agricultura brasileira e a importância da pesquisa da Embrapa. *Revista de Politica Agrícola* XII, 35–59.

Asteriou, D., Hall, S.G., 2007. *Applied Econometrics: A Modern Approach*. Palgrave Macmillan, New York.

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

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

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.

Davidson, R., MacKinnon, J.G., 1981. Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49, 3, 781–793.

Gallant, A.R., 1987. *Nonlinear Statistical Models*. John Wiley and Sons, New York.

Gasques, J., Bastos, E.T., Valdes, C., Bacchi, M., 2012. Produtividade da agricultura brasileira e os efeitos de algumas políticas. *Revista de Política Agrícola* XXI, 3, 83–92.

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 18 November 2009).

Pagan, A., Vella, F., 1989. Diagnostic test for models based on individual data: a survey. *Journal of Applied Economics* 49, S29–S59.

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

Programa das Nações Unidas para o Desenvolvimento, 2004. Desenvolvimento Humano e IDH. Available at http://www.pnud.org.br/idh/ (accessed 18 November 2009).

Ramalho, E.A., Ramalho, J.J.S., Henriques, P.D., 2010. Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis* 34, 239–255.

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

SAS, 2012. *SAS/STAT® 12.1 User's Guide*. SAS Institute, Cary, NC.

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

Souza, G.S., 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.