Tropentag 2009
University of Hamburg, October 6-8, 2009
Conference on International Research on Food Security, Natural Resource Management and Rural Development

Using Stochastic Frontier Approach to Assess Technical Efficiency in Brazilian Agriculture

Souza[a], Geraldo da Silva, Alcido Elenor Wander[b], Eliane Gonçalves Gomes[a] and Rosaura Gazzola[a]

a. Brazilian Agricultural Research Corporation (Embrapa), Secretariat of Management and Strategy (SGE), Parque Estação Biológica - PqEB s/n°. 70770-901 Brasília – DF, Brazil.
b. Brazilian Agricultural Research Corporation (Embrapa), National Rice and Bean Research Center (CNPAF), Rod. GO-462, km 12, 75375-000 Santo Antônio de Goiás – GO, Brazil. Email: awander@cnpaf.embrapa.br.

Introduction

Brazil is one of the most important countries in relation to agribusiness. Agribusiness represents about 25% of Brazilian GDP, 36% of 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 (Center West), mechanization and chemicals. Brazilian agriculture differs regionally, due, primarily, to the differences in geographical area, such as climate and natural resources, and thus production characteristics. For example, in South region soybeans, maize and poultry and pork have particular significance but in Northern region, rubber (hevea), nuts, wood extraction are important. These regional differences cause different technical efficiency among regions.
Thus, because of the peculiarities and differences of the states among the regions, further analyses were necessary. As found by Battese and Broca (1997) for the case of wheat in Pakistan, in the case of Brazil, the available data on agriculture and livestock are not suitable for some models of efficiency analysis.
Therefore production data were extracted from the agricultural census of 1995/96 and 2006. Together with production data, information on official credit used by farmers for investment and running costs in both mentioned years.
For the analysis, a stochastic frontier model was used to estimate the technical efficiency. The literature on technical efficiency measures has some examples of studies using stochastic frontier models to assess the efficiency of agricultural activities, considering regional aggregation levels. In this frame, studies like Chen and Song (2008), Onishi et al. (2008), Kaneko et al. (2004), Bhattacharayya and Parker (1999), and Hofler and Payne (1995) can be mentioned.
The objective of this study was to determine the technical efficiency of agriculture and livestock production of the 27 Brazilian states. Since there are regional variations regarding the way the agribusiness is organized in Brazil, it seems to be plausible that the technical efficiency shall differ from state to state. The topic is delicate, since there are considerable differences among states. The Human Development Index (HDI) in 2005 had considerable differences. Distrito Federal (0.874), Santa Catarina (0.840), São Paulo (0.833) and Rio de Janeiro (0.832) are states with higher HDI, while states like Alagoas (0.677) and Maranhão (0.683) have lower HDI (IDH, 2009).

Material and Methods

Stochastic Frontier Analysis

Basically, two approaches are available in the literature about 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, typically, one 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 the model (1), for cost data C_{it} for a panel of N producing units and T time periods.

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

For a production function one specifies (2), for log inputs $\ln x$.

$$\ln y_{it} = f(\ln x_{it}, \theta) + v_{it} - u_{it} \tag{2}$$

In these formulations, θ is an unknown parameter, $C(.)$ and $f(.)$ have known functional forms, and the stochastic components v_{it} and u_{it} represent random errors and inefficiency errors, respectively.

Typical parametric log cost families are provided by the Translog form (Coelli et al., 2005), the CES (Gallant, 1982), and the Fourier Flexible Form (Gallant, 1982). The latter endows the analysis with nonparametric properties. The random errors v_{it} are assumed to be uncorrelated across time and panel, and normally distributed with mean zero and variance $\sigma_v^{\ 2} > 0$. A flexible family of distributions to model the u_{it} (Kumbhakar and Lovell, 2000; Coelli et al., 2005) is provided by truncation of the normal. In this context one may postulate $u_{it} = z_{it}\delta + w_{it}$, where z_{it} is a vector of specific inefficiency variables (covariates), δ is a vector of unknown coefficients of the firm specific inefficiency variables, and w_{it} is the truncation at $-z_{it}\delta$ of the normal with mean zero and variance σ_u^2. Here we use the production function approach. Here we will follow the production approach using the Cobb-Douglas representation (3), which leads to (4), where $\alpha = \ln \theta_0$.

$$y_{it} = \theta_0 x_{1it}^{\theta_1} x_{2it}^{\theta_2} x_{3it}^{\theta_3} x_{4it}^{\theta_4} \exp(v_{it}) \exp(-u_{it}) \tag{3}$$

$$\ln y_{it} = \alpha + \theta_1 \ln x_{1it} + \theta_2 \ln x_{2it} + \theta_3 \ln x_{3it} + \theta_4 \ln x_{4it} + v_{it} - u_{it} \tag{4}$$

Production is measured by total value of agricultural production and the inputs are labor, capital, land and other inputs (running costs like fertilizers, seeds, pesticides, etc.). In the next section (data) we provide a more detailed explanation of the production variables. Let *r* denote the vector of log inputs, *q* the log output and θ the production function parameter vector.

DEA, on the other hand, assumes a deterministic frontier. Typical statistical models for which data envelopment analysis is optimal do not assume the presence of the stochastic component v_{it}.

As Coelli et al. (2005) put it; much of stochastic efficiency analysis is directed towards the prediction of inefficiency (efficiency) effects. The most common output-oriented measure of technical efficiency for firm *o* is estimated in the stochastic frontier case by (5),

$$E\left[\exp(-u_{it}) \mid \varepsilon_{it}\right] = \left\{ {[1 - \Phi(\sigma_* - \mu_{*it} / \sigma_*)]} \Big/ {\Phi(\mu_{*it} / \sigma_*)} \right\} \exp(-\mu_{*it} + 0.5\sigma_*^2) \tag{5}$$

where: $\varepsilon_{it} = q_{it} - w_{it}\theta, \mu_{*it} = (-\varepsilon_{it}\,\sigma_u^2 + \mu_{it}\sigma_v^2) / \sigma_S^2, \mu_{it} = z_{it}\delta, \sigma_S^2 = \sigma_u^2 + \sigma_v^2, \sigma_* = \sigma_u\sigma_v / \sigma_S$.

Assuming a normal-truncated normal specification the parameters are obtained maximizing the log likelihood function (Battese and Coelli, 1995), as in (6).

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ln(2\pi)+\ln(\sigma_S^2)\right\}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ln(2\pi)+\ln(\sigma_S^2)\right\}$$

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{[(q_{it}-r_{it}\theta)+z_{it}\delta]/\sigma_S^2\right\} \tag{6}$$

$$-\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ln\Phi(d_{it})-\ln\Phi(d_{it}^{*})\right\}$$

where: $d_{it}=\mu_{it}/\sigma_u\,, d_{*it}=\mu_{*it}/\sigma_*\,,\sigma_*=\sigma_u\sigma_v/\sigma_S$.

A convenient model re-parameterization, making $\gamma=\sigma_u^2/\sigma_S^2$, leads to the log likelihood as a function $L(\beta,\delta,\gamma,\sigma_S^2)$, where $\sigma_u^2=\gamma\sigma_S^2,\ \sigma_*^2=\gamma(1-\gamma)\sigma_S^2$. A classical production model is implied by $\gamma=0$.

An elasticity for firm *i* in period *t* relative to a contextual variable measured in logs with parameter estimate *b* is computed using the formula (7).

$$b\left(1-\frac{d_{it}\phi(d_{it})\Phi(d_{it})+[\phi(d_{it})]^2}{[\Phi(d_{it})]^2}\right) \tag{7}$$

If X^t and Y^t represent the input and output matrices, the efficiency estimate of firm *i* in period *t*, under constant returns to scale and output orientation, is given by the solution of the linear programming problem Min φ, subject to $X^t\lambda\le x_i^t$, $Y\lambda\ge\varphi y_i^t$, $\lambda\ge 0$ (Charnes et al., 1978). Here x_i^t and y_i^t represent the input vector and the output used by firm *i* in period *t*. The effect of contextual variables may be studied in a second stage regression using the efficiencies computed in the first stage as proposed in Simar and Wilson (2007), Souza and Staub (2007) and Banker and Natarajan (2008). The stochastic and deterministic specifications underlying these approaches did not provide a good fit for our data. Thus the classical stochastic frontier with technical effects was our choice of model.

Data

In this model we used the value of agricultural production as dependent and land, labor, capital and running costs as independent variables. The data on value of agricultural and livestock production of all 27 Brazilian states in the years 1995/96 and 2006 was used. The two years correspond to the two last available agricultural census data in Brazil. The output variable used was the total value of Brazilian agricultural and livestock production in the years 1995/96 and 2006 (total value of production; R$). The inputs of the model were: total land area used (planted area; hectares), labor force (employment in agriculture and livestock; number of persons), investment and running costs (in monetary value, R$).

The data on area, labor force and value of production were obtained from the agricultural census (IPEA, 2008). Credit data on investments or capital and running costs or other inputs were extracted from the "Anuários Estatísticos do Crédito Rural" (BACEN, 1995, 2008), representing all official credit taken by farmers in all 27 states in the two years of consideration.

Results and Discussion

For the analysis all variables are measured in log form. The actual model used postulates a linear relationship between the log agricultural production *y* and the log inputs *l*, *k*, *t*, and *c* denoting labor, capital, land and other inputs, respectively.

Table 1 shows the statistical results of maximum likelihood estimation of the stochastic frontier model using Stata 10.1 software (Stata, 2007).

Table 1. Stochastic frontier estimation.

	Coefficient	Standard error	z	P>|z|	[95% Confidence interval]	
Production (*y*)						
Labor (*l*)	0.3238	0.0442	7.33	0.000	0.2373	0.4103
Capital (*k*)	0.1413	0.0529	2.67	0.008	0.0376	0.2449
Land (*t*)	0.2672	0.0286	9.34	0.000	0.2111	0.3232
Other Inputs (*c*)*	0.2633	0.0470	5.60	0.000	0.1712	0.3554
Constant	-0.9182	0.4623	-1.99	0.047	-1.8243	-0.0120
Technical effect						
l (GDP pc)	-0.7699	0.1114	-6.91	0.000	-0.9881	-0.5516
contant	1.8278	0.2726	6.71	0.000	1.2935	2.3620
sigma_S2	0.0604				0.0396	0.0921
gamma	0.4627				0.0319	0.9574

*Other inputs are running costs for expenditures on annual and perennial cultivates and livestock.

The likelihood ratio test statistic for the joint hypothesis implying the presence of HDI (Human Development Index), time, and regional effects has a value of 7.15 with a p-value of about 31%. For this reason we dropped log HDI and all the other categorical variables, and use the more parsimonious model presented in Table 1. Table 1 also shows that a 1% increase in different inputs would have different impacts on production: in capital the gains in production would be 0.14% *ceteris paribus*, in labor 0.32%, in land, 0.27%, and in other inputs, 0.26%.
We see in Table 2 that the Pearson correlation between the two measures from Tocantins (last in left Table) is only 57% indicating strong differences between the two approaches.

Table 2. Efficiency DEA CCR-O estimation, Years: 1995/96 and 2006 (Data are in efficiency order).

States	Region	1995/1996		States	Region	2006
Acre	North	1		Alagoas	North East	1
Amazonas	North	1		Amazonas	North	1
Amapá	North	1		Amapá	North	1
Distrito Federal	Center West	1		Ceará	North East	1
Espírito Santo	South East	1		Mato Grosso	Center West	1
Mato Grosso do Sul	Center West	1		Pará	North	1
Rio de Janeiro	South East	1		Pernambuco	North East	1
Santa catarina	South	1		Rio de Janeiro	South East	1
São Paulo	South East	1		São Paulo	South East	1
Paraíba	North East	0.9923		Espírito Santo	South East	0.9504
Alagoas	North East	0.9855		Paraíba	North East	0.9484
Pernambuco	North East	0.9723		Rondônia	North	0.9301
Rio Grande do Sul	South	0.8218		Mato Grosso do Sul	Center West	0.9278
Minas Gerais	South East	0.7916		Rio Grande do Norte	North East	0.902
Ceará	North East	0.7875		Roraima	North	0.893
Paraná	South	0.7746		Distrito Federal	Center West	0.8439
Rio Grande do Norte	North East	0.7352		Acre	North	0.8074
Goiás	Center West	0.681		Sergipe	North East	0.7921
Mato grosso	Center West	0.6611		Bahia	North East	0.7884
Maranhão	North East	0.6231		Santa Catarina	South	0.7872
Piauí	North East	0.5085		Paraná	South	0.7541
Rondônia	North	0.5082		Goiás	Center West	0.7188
Bahia	North East	0.4877		Minas Gerais	South East	0.6522

Pará	North	0.4651		Rio Grande do Sul	South	0.6501
Roraima	North	0.4318		Tocantins	North	0.5691
Sergipe	North East	0.4116		Maranhão	North East	0.4597
Tocantins	North	0.335		Piauí	North East	0.3906

Table 2 shows results from DEA analysis, which are does not represent Brazilian reality. Knowing the reality of agriculture in the states and the data used on capital and running costs, the obtained result with DEA is far away from agricultural reality in the states. This shows that, even with excellent data quality, the model may lead to incorrect conclusions. These results show the states of the Northern region with a similar efficiency than states like São Paulo, what would very hard to explain.
Table 3 shows statistics computed as a function of model parameter estimates. These are stochastic efficiency estimates. The Pearson correlation in Table 3 (stochastic efficiencies) between observed and predicted values is about 99% indicating a good fit for the frontier model. The 95% confidence interval for the parameter γ suggests a technical components model.

Table 3. Rank of stochastic technical efficiencies, Years: 1995/96 and 2006.

States	**Region**	**1995/1996**		**States**	**Region**	**2006**
São Paulo	South East	0.9040		Distrito Federal	Center West	0.9529
Distrito Federal	Center West	0.8898		São Paulo	South East	0.9467
Amapá	North	0.8142		Rio de Janeiro	South East	0.8773
Santa Catarina	South	0.8125		Espírito Santo	South East	0.8571
Rio de Janeiro	South East	0.8018		Santa Catarina	South	0.8426
Rio Grande do Sul	South	0.7538		Paraná	South	0.8113
Espírito Santo	South East	0.7387		Rio Grande do Sul	South	0.7971
Paraná	South	0.6857		Mato Grosso	Center West	0.7937
Amazonas	North	0.6814		Minas Gerais	South East	0.6720
Minas Gerais	South East	0.5947		Goiás	Center West	0.6632
Mato Grosso do Sul	Center West	0.5900		Mato Grosso do Sul	Center West	0.6599
Pernambuco	North East	0.4811		Amazonas	North	0.6004
Goiás	Center West	0.4716		Rondônia	North	0.5953
Acre	North	0.4538		Amapá	North	0.5454
Mato Grosso	Center West	0.4445		Pernambuco	North East	0.5075
Alagoas	North East	0.4261		Roraima	North	0.5021
Pará	North	0.3821		Acre	North	0.4929
Rondônia	North	0.3804		Rio Grande do Norte	North East	0.4877
Rio Grande do Norte	North East	0.3748		Bahia	North East	0.4869
Ceará	North East	0.3743		Pará	North	0.4732
Paraíba	North East	0.3634		Sergipe	North East	0.4661
Bahia	North East	0.3412		Alagoas	North East	0.4408
Sergipe	North East	0.3351		Ceará	North East	0.4373
Roraima	North	0.3206		Paraíba	North East	0.4282
Piauí	North East	0.2443		Tocantins	North	0.4247
Maranhão	North East	0.2441		Maranhão	North East	0.3098
Tocantins	North	0.2215		Piauí	North East	0.2619

The mean stochastic technical efficiencies for each state are shown in Table 3. Considering an average of both analyzed years, the most efficient state is São Paulo, SP (0.93) and the least efficient is Piauí, PI (0.25). The South-East (SE) dominates, followed by South (S), Center-West (CW), North (N) and North-East (NE). The dominance of the South-East and South over the other regions is strong. These results are somewhat expected and serves the purpose to further validation of our model.

As shown in Tables 4 and 5, the chi-square test for constant returns to scale ($l+k+t+c=1$) has a p-value of 90%, non significant. Although the confidence intervals for all input variables do intercept, the pair wise Wald test of equality indicates that the labor (*l*) elasticity is stronger than the capital (*k*) elasticity (p-value 0.02), and the capital elasticity is weaker then the land elasticity (*t*) (p-value 0.03). The difference between labor and land elasticities is marginal (p-value 0.10). No other pair wise comparison was significant.

Average income *per capita* elasticity over all states and years is 0.67 with a standard error of 0.17. The minimum *per capita* income elasticity is 0.05, and the maximum 0.77. States highly efficient have smaller income elasticities.

This means, that a 1% increase in *per capita* income will increase the agricultural production in wealthier states like Distrito Federal by only 0.14%, but can achieve up to 0.77% in those states with lower technical efficiency.

Table 4. Average income *per capita* elasticities by state, in ascending order.

States	Average elasticity
Distrito Federal	0.1383
São Paulo	0.2701
Rio de Janeiro	0.4020
Rio Grande do Sul	0.5043
Santa Catarina	0.5253
Espírito Santo	0.5589
Paraná	0.6479
Amazonas	0.6689
Mato Grosso	0.7069
Minas Gerais	0.7359
Mato Grosso do Sul	0.7470
Goiás	0.7601
Amapá	0.7618
Roraima	0.7668
Rondônia	0.7689
Sergipe	0.7697
Tocantins	0.7698
Acre	0.7699
Alagoas	0.7699
Bahia	0.7699
Ceará	0.7699
Maranhão	0.7699
Pará	0.7699
Paraíba	0.7699
Pernambuco	0.7699
Piauí	0.7699
Rio Grande do Norte	0.7699

Table 5. Average income *per capita* elasticities per region, in ascending order.

Region	Frequency	Average elasticity
South East	8	0.4917
South	6	0.5592
Center West	8	0.5881
North	14	0.7537
North East	18	0.7699

Distrito Federal was the most efficient state in 2006 and second most efficient in 1995/96. The state was created with the foundation of Brasilia as federal capital. The state is small, if compared to other states, but with a strong market-oriented agriculture producing mainly cotton, soybeans, common beans and poultry production. Additionally, its population is small, since an important part lives outside Brasilia, belonging to Goiás state and the high per capita GDP is strongly related to the higher salaries of federal government staff.

With the study, we confirmed the strong presence of agriculture and livestock production in the states of South, South East and Center West regions. These states have export oriented agriculture. Whereas, the states with lower efficiency represent dryer regions, like in the Northeast, with rainfall between 200 and 600 mm per year, high temperatures, with limited infrastructure for irrigation and processing of agricultural production. In the North region, the lower efficiency is explained by large areas and a more subsistence oriented agriculture, exploring mostly native species like hevea, brazil nuts and timber extraction. Those activities represent the main vocation of the North region and could be further explored in a study focusing only on the states of the Amazon region.

Summary and Conclusions

We fitted a DEA Model (CCR-O) and a stochastic frontier model to state agricultural production data in Brazil. The second fit was very good as measured by a correlation of about 99% between observed and predicted values. The technology seems to show constant returns to scale.

The model also includes a statistically significant contextual inefficiency effect defined by *per capita* income. The average *per capita* income elasticity is 0.67, with a standard error of 0.17. The income variable is used as a proxy for infrastructure and technology assessment. We find stronger elasticities results for labor, other inputs (running costs) and land.

South-East and South states are significantly more efficient than other states on average. São Paulo (SP), Distrito Federal (DF), Rio de Janeiro (RJ), Santa Catarina (SC), Espírito Santo (ES), Rio Grande do Sul (RS) and Paraná (PR) are the most efficient states with product oriented technical efficiencies over 70% well above the other states.

These empirical results suggest one important findings. There are significant possibilities to increase efficiency levels in Brazil agriculture production, especially in the Northeast and North Region. Finally the results indicate the diversity of the scores of efficiency among regions. This suggests that the considerable variability of regions in climate, natural resources, irrigation, etc. (infrastructure, agro industries), can have different impacts on efficiency in Brazil agricultural production in different regions.

References

BANCO CENTRAL DO BRASIL (BACEN). (1995). *Anuário Estatístico do Crédito Rural – 1995*; Departamento de Cadastro e Informação, Brasília.

BANCO CENTRAL DO BRASIL (BACEN). (2008). *Anuário Estatístico do Crédito Rural - 2006*. Available at: <http://www.bcb.gov.br/default.asp?id=relrural&ano=2006>. Accessed: 02 Oct. 2008.

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

BATTESE, G.E. and BROCA, S.S. (1997). Functional forms of stochastic frontier production functions and models for technical inefficiency effects: a comparative study for wheat farmers in Pakistan. *Journal of Productivity Analysis*; 8:395-414.

BATTESE, G.E. and COELLI, T.J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*; 20:325-332.

BHATTACHARAYYA, A. and PARKER, E. (1999). Labor productivity and migration in Chinese agricultures: a stochastic frontier approach. *China Economic Review*; 10:59-74.

CHARNES, A., COOPER, W.W. and RHODES, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*; 2:429-444.

CHEN, Z. and SONG, S. (2008). Efficiency and technology gap in China's agriculture: A regional meta-frontier analysis. *China Economic Review*; 19:287-296.

COELLI, T.J., PRASADA RAO, D.S., O'DONNELL, C.J. and BATTESE, G.E. (2005). *An Introduction to Efficiency and Productivity Analysis*; Springer, New York.

GALLANT, A.R. (1982). Unbiased determination of production technologies. *Journal of Econometrics*; 20:285-323.

HOFLER, R.A. and PAYNE, J.E. (1995). Regional efficiency differences and development policy of agriculture in the Yugoslav republics: estimates from panel data. *Review of Regional Studies*; 25:287-300.

ÍNDICE DE DESENVOLVIMENTO HUMANO (IDH). (2009). *Lista de estados do Brasil por IDH*. Available at: <http://pt.wikipedia.org/wiki/Lista_de_estados_do_Brasil_por_IDH>. Accessed: 18 Mar. 2009.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). (2008). *Ipeadata*. Available at: <http://www.ipeadata.gov.br>. Accessed: 02 Oct. 2008.

KANEKO, S., TANAKA, K., TOYOTA, T. and MANAGI, S. (2004). Water efficiency of agricultural production in China: regional comparison from 1999 to 2002. *International Journal of Agricultural Resources, Governance and Ecology*; 3:231-251.

KUMBHAKAR, S. and LOVELL, C.A.K. (2000). *Stochastic Frontier Analysis*; Cambridge University Press, New York.

ONISHI, A., MORISUGI, M., IMURA, H., SHI, F., WATANABE, T. and FUKUSHIMA, Y. (2008). Study on the efficiency of agricultural water use in the Yellow River Basin. *Journal of Global Environment Engineering*; 13:51-67.

SIMAR, L. and 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. 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.

STATA. (2007). *STATA v. 10.1*; StataCorp LP, College Station, TX, USA.