



A Stochastic Production Frontier Analysis of the Brazilian Agriculture in the Presence of an Endogenous Covariate

Geraldo da Silva e Souza^(✉) and Eliane Gonçalves Gomes

Secretaria de Inteligência e Relações Estratégicas – Embrapa,
Brasília, DF 70770-901, Brazil
{geraldodo.souza, eliane.gomes}@embrapa.br

Abstract. Production frontier analysis aims at the identification of best production practices and the importance of external factors, endogenous or not, that affect the production function and the technical efficiency component. In particular, in the context of the Brazilian agriculture, it is desirable for policy makers to identify the effect on production of variables related to market imperfections. Market imperfections occur when farmers are subjected to different market conditions depending on their income. In general, large scale farmers access lower input prices and may sell their production at lower prices, thereby making competition harder for small farmers. Market imperfections are typically associated with infrastructure, environment control requirements and the presence of technical assistance. In this article, at county level, and using agricultural census data, we estimate the elasticities of these variables on production by maximum likelihood methods. Technological inputs dominate the production response, followed by labor and land. Environment control has a positive net effect on production, as well as technical assistance. The indicator of infrastructure affects positively technical efficiency. There is no evidence of technical assistance endogeneity.

Keywords: Stochastic frontier · Endogeneity · Agriculture

1 Introduction

As pointed out in other sources [1–3], Brazilian agriculture is highly concentrated. Only five hundred thousand farmers, 11.4% of the total, produced 87% of the total production value in 2006 (last agricultural census). These data motivate studies that identify factors of importance for public policies leading to productive inclusion in agriculture in Brazil. Indeed, the major (state) agricultural research company in Brazil defines “productive insertion and poverty reduction” as one of the impact axes in its strategic planning map. Access to technology is the main cause of production concentration and, very likely, of poverty in the fields. We see, in this context, that the agricultural sector demands proper public policies in order to improve access to technology and to increase productive insertion and reduce rural poverty.

As emphasized in Souza and Gomes [3], market imperfections are the main cause of inhibition of the access of farmers to technology and, therefore, to productive

inclusion. Market imperfections are the result of asymmetry in credit for production, infrastructure, information availability, rural extension and technical assistance, among others [4].

The lack of physical infrastructure and education make it difficult for the rural extension to fulfill its role and, therefore, gain proper access to technology. Another point to be emphasized is related to the imperfection of the production markets. Souza et al. [5] highlight that small farmers sell their products at lower values and buy inputs at higher prices. The large scale producers are able to negotiate better input and output prices and the existence of these different prices characterizes a market imperfection. Unfavorable negotiation may lead to higher prices for the adoption of better technologies and thus lead to difficulties in achieving higher economic efficiency.

We contribute to this literature modeling production value as a function of several aggregates, reflecting, on a municipal level, the input usage, environment control, technical assistance and the effect of market imperfection variables on the technical efficiency of production. The modeling process postulates a Cobb-Douglas representation in a typical stochastic frontier approach and is carried out under the assumption of endogeneity of technical assistance. The models we used follow the basic lines of Karakaplan and Kutlu [6] and Karakaplan [7]. We extend Karakaplan's approach to the truncated normal and the exponential distributions. Alternatives to the approach are also suggested, considering non-linear models with the Murphy and Topel [8] variance correction. In this context we allow the use of fractional regressions [9, 10]. Our results extend Souza and Gomes' [3] findings.

2 Data

The data sources for this article are the Brazilian agricultural census of 2006, the Brazilian demographic census of 2010, and municipal databases on education and health.

We follow the approach of Souza et al. [2, 5] in the definition of production and contextual variables.

Production (inputs and output) is defined using monetary values. The source is the agricultural census of 2006 [11]. The output variable is the value of production and the inputs are expenses on labor (*labor*), land (*land*) and technological inputs (*techinputs*), which includes machinery, improvements in the farm, equipment rental, value of permanent crops, value of animals, value of forests in the establishment, value of seeds, value of salt and fodder, value of medication, fertilizers, manure, pesticides, expenses with fuel, electricity, storage, services provided, raw materials, incubation of eggs and other expenses. Value of permanent crops, forests, machinery, improvements on the farm, animals and equipment rental were depreciated at a rate of 6% a year (machines – 15 years, planted forests – 20 years, permanent cultures – 15 years, improvements – 50 years, animals – 5 years). Farm data from the agricultural census were aggregated to form totals for each county. A total of 4,965 counties (almost 90% of the total) provided valid data for our analysis.

The contextual variables we chose are a performance municipal index of social development (*social*), an index of demographic development (*demog*), the proportion of farmers who received technical assistance (*techassist*), the proportion of non-degraded areas (*ndareas*) and the proportion of forested areas (*forest*). The last two are proxies for environment control. Market imperfections are mainly associated with the social index.

The demographic index captures the population dynamics that tend to follow rural development. The variables considered in this dimension of development are the migration index (rural to urban areas), average number of farm dwellers, aging rate (total municipal rural population over rural population over 60), dependency rate (ratio of the rural population with age in the bracket 15–59 over the rural population with age in the bracket of 0–14 plus over 60), ratio of urban to rural population in the municipality. The source is the demographic census of 2010 [12] in general, and the 2000 and 2010 census for the migration index. The demographic score was computed using the ranks of these measurements, weighted by the relative multiple correlation coefficient.

The index of social development reflects the level of well-being, favored by factors such as the availability of water and electric energy in the rural residences, and level of education, health and poverty in the rural households. It was computed as a weighted average of normalized ranks of the following variables: education (illiteracy rate), poverty index, average gross per capita income of rural households, proportion of farms with access to electricity and water, index of basic education, index of performance of the public health system and vulnerability of children up to five years old. These indicators were obtained from the Brazilian demographic census 2010 [12], from the Brazilian agricultural census 2006 [11], and from the databases of the National Institute of Research and Educational Studies (INEP), referring to education in 2009 [13], and of the Ministry of Health 2011 data [14]. The social score was computed using the ranks of these measurements, weighted by the relative multiple correlation coefficient.

3 Methodology

Our approach to assess production and efficiency of production follows along the lines of Karakaplan and Kutlu [6] and Karakaplan [7]. The structural model for our application is defined by (1) for municipality i , where *techassist* is assumed endogenous and y_i is the log of gross income.

$$\begin{aligned}
 y_i &= b_0 + b_1 \log(\text{labor}_i) + b_2 \log(\text{land}_i) + b_3 \log(\text{techinputs}_i) + b_4(\text{forest}_i) \\
 &\quad + b_5(\text{ndareas}_i) + b_6(\text{techassist}_i) + v_i - u_i \\
 v_i, u_i &\text{ independent} \\
 v_i &\sim N(0, \sigma^2)
 \end{aligned} \tag{1}$$

The u_i are non-negative inefficiency components and the v_i are a random sample of an idiosyncratic error component. We assume three possible distributions for the inefficiency component: half-normal, exponential and truncated normal.

For the half-normal we have $u_i \sim N^+(0, \sigma_{u_i}^2)$ and $\sigma_{u_i}^2 = \exp\left(\begin{matrix} b_7 + b_8 \log(\text{labor}_i) + b_9 \log(\text{land}_i) + b_{10} \log(\text{techinputs}_i) + b_{11}(\text{forext}_i) + \\ b_{512}(\text{ndareas}_i) + b_{13}(\text{social}_i) + b_{14}(\text{demog}_i) \end{matrix}\right)$. For the exponential $u_i \sim \exp(\zeta_i)$, $\zeta > 0$, we assume the variance $\sigma_{u_i}^2 = \zeta^{-2}$ with the same representation as the half-normal. Finally, for the truncated normal $u_i \sim N^+(\mu_i, \sigma_u^2)$ and $\mu_i = b_7 + b_8 \log(\text{labor}_i) + b_9 \log(\text{land}_i) + b_{10} \log(\text{techinputs}_i) + b_{11}(\text{forext}_i) + b_{512}(\text{ndareas}_i) + b_{13}(\text{social}_i) + b_{14}(\text{demog}_i)$.

Endogeneity in Karakaplan and Kutlu [6] and Karakaplan [7] means correlation of a variable with v_i . This assumption invalidates the classic stochastic frontier analysis. A classic approach for handling this issue is to use two stage least squares or the general method of moments (GMM), as suggested in Amsler et al. [15]. On the other hand, Karakaplan and Kutlu [6] and Karakaplan [7] suggest the use of instrumental variables in a context of maximum likelihood estimation, resembling classical frontier analysis. In our application, we follow this approach and the instruments considered for *techassist* are the exogenous variables plus demographic and social indicators. The instrumental variable regression is assumed to be linear but the idea can be easily generalized to non-linear specifications $\text{techassist}_i = f(z_i, \delta) + \varepsilon_i$. In this formulation, z_i is a vector of instrumental variables and $\varepsilon' = (\varepsilon_1, \dots, \varepsilon_n)$ has mean zero and variance matrix $\sigma_\varepsilon^2 I$. Heteroskedastic formulations are possible assuming a general variance matrix Ω . This formulation also allows for the Bernoulli specification described in Papke and Wooldridge [9], which is particularly convenient if one is dealing with fractions. In this instance, the model can be estimated assuming $f(\cdot)$ to be a distribution function.

Karakaplan [7] in its ‘sfkk’ module in the Stata software makes use of the half-normal distribution and the linear instrumental variable regression.

Let ρ be the correlation between ε_i and v_i . Endogeneity means $\rho \neq 0$. We assume the bivariate normal distribution as in (2).

$$\begin{bmatrix} \tilde{\varepsilon}_i \\ v_i \end{bmatrix} = \begin{bmatrix} \varepsilon_i / \sigma_\varepsilon \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}\right) \quad (2)$$

Using a Cholesky decomposition we may write (3) and, therefore, we have (4).

$$\begin{bmatrix} \tilde{\varepsilon}_i \\ v_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho\sigma & \sigma\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} \tilde{\varepsilon}_i \\ \tilde{w}_i \end{bmatrix} \quad (3)$$

$$\begin{aligned} y_i &= b_0 + b_1 \log(\text{labor}_i) + b_2 \log(\text{land}_i) + b_3 \log(\text{techinputs}_i) + b_4(\text{forext}_i) \\ &\quad + b_5(\text{ndareas}_i) + b_6(\text{techassist}_i) + \eta\tilde{\varepsilon}_i + w_i - u_i \\ w_i &= \sigma\sqrt{1-\rho^2}\tilde{w}_i \\ \mu &= \rho\sigma \end{aligned} \quad (4)$$

Therefore, when the residual variance is constant, the component $\eta\tilde{e}_i$ is the correction term for bias. The test of $\eta = 0$ is an endogeneity test. The model is estimated by maximum likelihood.

For the half-normal distribution, the likelihood function is given by (5).

$$\begin{aligned} \log L(\theta) = & \sum_{i=1}^n \left\{ \frac{\ln(2/\pi) - \ln(\sigma_{Si}^2) - (e_i/\sigma_{Si}^2)}{2} + \ln \Phi\left(\frac{\lambda_i e_i}{\sigma_{Si}}\right) \right\} \\ & + \sum_{i=1}^n \left\{ \frac{\ln 2\pi - \ln \sigma_\varepsilon - \sum_{i=1}^n (\varepsilon_i^2/\sigma_\varepsilon^2)}{2} \right\} \end{aligned} \tag{5}$$

Here $\lambda_i = \sigma_{ui}/\sigma$ and $\sigma_{Si}^2 = \sigma_{ui}^2 + \sigma^2$. Notice that e_i is defined by (6).

$$e_i = y_i - \left(\begin{aligned} & b_0 + b_1 \log(\text{labor}_i) + b_2 \log(\text{land}_i) + b_3 \log(\text{techinputs}_i) + b_4(\text{forext}_i) \\ & + b_5(\text{ndareas}_i) + b_6(\text{techassist}_i) + \eta\tilde{e}_i \end{aligned} \right) \tag{6}$$

For the exponential model, the likelihood function becomes (7) and for the truncated normal it is defined as in (8), where $\gamma = \sigma_u^2/\sigma_S^2$, $\sigma_S^2 = \sigma_u^2 + \sigma^2$.

$$\begin{aligned} \log L(\theta) = & \sum_{i=1}^n \left\{ -\ln(\sigma_u) + \frac{\sigma^2}{2\sigma_u^2} + \ln \Phi\left(\frac{-e_i - \sigma^2/\sigma_u}{\sigma}\right) + \frac{e_i}{\sigma_u} \right\} \\ & + \sum_{i=1}^n \left\{ \frac{\ln 2\pi - \ln \sigma_\varepsilon - \sum_{i=1}^n (\varepsilon_i^2/\sigma_\varepsilon^2)}{2} \right\} \end{aligned} \tag{7}$$

$$\begin{aligned} \log L(\theta) = & \sum_{i=1}^n \left\{ \begin{aligned} & -\frac{\ln(2\pi)}{2} - \ln \sigma_S - \ln \Phi\left(\frac{\mu_i}{\sigma_S \sqrt{\gamma}}\right) + \ln \Phi\left(\frac{(1-\gamma)\mu_i - \lambda e_i}{\sigma_S \sqrt{\gamma(1-\gamma)}}\right) \\ & -\frac{1}{2} \left(\frac{e_i + \mu_i}{\sigma_S}\right)^2 \end{aligned} \right\} \\ & + \sum_{i=1}^n \left\{ \frac{\ln 2\pi - \ln \sigma_\varepsilon - \sum_{i=1}^n (\varepsilon_i^2/\sigma_\varepsilon^2)}{2} \right\} \end{aligned} \tag{8}$$

Karakaplan and Kutlu [6] suggest an alternative to estimation easier to implement, which can be extended to accommodate fractional regression models in the instrumental regression. The idea is to perform the estimation in two steps. Firstly, one fits the instrumental variable regression and computes residuals $\hat{e}_i = \text{techassist} - f(z_i, \hat{\delta})$ and then runs the standard stochastic frontier model (9).

$$y_i = b_0 + b_1 \log(\text{labor}_i) + b_2 \log(\text{land}_i) + b_3 \log(\text{techinputs}_i) + b_4(\text{forext}_i) + b_5(\text{ndareas}_i) + b_6(\text{techassist}_i) + \eta \hat{\epsilon}_i + w_i - u_i \quad (9)$$

The process will not produce the same results as the full maximum likelihood estimation. Greene [4] names it limited information maximum likelihood. The variance matrix of the estimator requires the Murphy and Topel [8] correction. Let $\hat{\delta}$ be the maximum likelihood estimate obtained from the instrumental variable regression with variance matrix \hat{V}_1 . The likelihood function is $\ln(\text{techassist}, z, \delta)$. Let $\hat{\theta}$ be the maximum likelihood of the resulting frontier model obtained when $\delta = \hat{\delta}$. The variance matrix is \hat{V}_2 and the likelihood function is $\ln f_2(y, x, \hat{\delta}, \theta)$, where the vector x includes inputs, technical assistance, non-degraded areas, forests, and the residual from the instrumental variable regression. Following Greene [16], we may define the matrices (10) and (11).

$$\hat{C} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{2i}}{\partial \hat{\theta}} \right) \left(\frac{\partial \ln f_{2i}}{\partial \hat{\delta}'} \right) \quad (10)$$

$$\hat{R} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{2i}}{\partial \hat{\theta}} \right) \left(\frac{\partial \ln f_{1i}}{\partial \hat{\delta}'} \right) \quad (11)$$

The estimated variance matrix of the limited information maximum likelihood estimator is defined as in (12).

$$\hat{V} = \frac{1}{n} (\hat{V}_2 + \hat{V}_2 (\hat{C} \hat{V}_1 \hat{C}' - \hat{R} \hat{V}_1 \hat{C}' - \hat{C} \hat{V}_1 \hat{R}') \hat{V}_2) \quad (12)$$

In our exercise we used both methods, that is, full likelihood estimation as well as the two-step procedure. Regression in the first step used both the fractional approach of Papke and Wooldridge [9] and linear regression.

4 Statistical Results

Following the standard literature in stochastic frontier analysis we fitted 11 models to the data described in Sect. 2, using the approaches of Sect. 3. The models considered are: Case 1 – The full information maximum likelihood approach under the half-normal and exponential inefficiency distributions, and the correspondent limited information maximum likelihood for the best model under linear and fractional instrumental variables regressions. The only inefficiency effect considered is the social indicator; Case 2 – The limited information maximum likelihood assuming both instrumental variables' regression assumptions, including as efficiency effects all independent factors for the half-normal and truncated normal. Tables 1 and 2 show the goodness of fit measures considered for model choice.

Table 1. Fit statistics: Case 1 – Social indicator is the only inefficiency effect.

| Model | Inefficiency distribution | Likelihood |
|-------------------------------------|---------------------------|-----------------|
| Full information ML | Half-normal | -3041.3 |
| Full information ML | Truncated normal | Do not converge |
| Full information ML | Exponential | -3042.8 |
| Limited information ML – linear | Half-normal | -4663.3 |
| Limited information ML – fractional | Half-normal | -4661.8 |

Table 2. Fit statistics: Case 2 – All exogenous variables are inefficiency effects.

| Model | Inefficiency distribution | Likelihood |
|-------------------------------------|---------------------------|-----------------|
| Limited information ML – fractional | Half-normal | -4603.4 |
| Limited information ML – fractional | Truncated normal | -4607.4 |
| Limited information ML – fractional | Exponential | Do not converge |
| Limited information ML – linear | Half-normal | -4603.4 |
| Limited information ML – linear | Truncated normal | -4612.0 |
| Limited information ML – linear | Exponential | Do not converge |

We experienced convergence problems with some of the assumptions for the inefficiency distribution, depending on the assumption itself and on the number of variables included in the efficiency effect. The full information maximum likelihood with all exogenous variable included in the effect did not converge, inhibiting the application of the standard likelihood approach to test nested hypothesis. We see from Tables 1 and 2 that the best fit is the full information estimator under the half-normal distribution, reducing the set of inefficiency factor effects to the social indicator. The models fitted in two stages using the linear and the non-linear binomial Papke and Woodridge [12] assumptions indicate similar results, with a slight superiority for the fractional regression. Correlations between actual and estimated values for the instrumental regressions are, respectively, 80.1% and 80.4%. Programming was carried out using Stata 14 and SAS 9.2 software.

Table 3 shows statistical estimation for full information half-normal model including a social effect for the inefficiency component. Table 4 shows the fractional regression for technical assistance. Table 5 shows the limited information maximum likelihood with the Murphy-Topel variance correction [8], under the binomial specification for the instrumental variable regression.

In the context of the full information maximum likelihood estimation correlation between actual and predicted values of the frontier model, including efficiency effects, is 88.6%. The component technical assistance affects significantly and positively the response variable (log income). There is no evidence of endogeneity (p -value = 0.1858).

Table 6 summarizes the relative importance of production factors, including returns to scale. We see that technology dominates, followed by labor and land. The technology shows decreasing returns to scale. These results fairly agree with Souza et al. [5].

Table 3. Full information maximum likelihood estimation. Half-normal stochastic frontier under endogeneity of technical assistance. Stata output.

| | Coefficient | Std error | z | P > z | [95% confidence interval] | |
|------------------------------------|-------------|-----------|--------|--------|---------------------------|----------|
| Frontier | | | | | | |
| <i>labor</i> | 0.231137 | 0.011531 | 20.04 | 0.000 | 0.208536 | 0.253738 |
| <i>land</i> | 0.09003 | 0.013968 | 6.45 | 0.000 | 0.062653 | 0.117406 |
| <i>techinputs</i> | 0.45581 | 0.021104 | 21.6 | 0.000 | 0.414446 | 0.497173 |
| <i>forest</i> | -0.12398 | 0.032878 | -3.77 | 0.000 | -0.18842 | -0.05954 |
| <i>ndareas</i> | 0.250139 | 0.036281 | 6.89 | 0.000 | 0.17903 | 0.321249 |
| <i>techassist</i> | 0.567809 | 0.140459 | 4.04 | 0.000 | 0.292514 | 0.843105 |
| constant | 2.736811 | 0.104023 | 26.31 | 0.000 | 2.53293 | 2.940691 |
| Instrumental regression | | | | | | |
| <i>labor</i> | -0.02131 | 0.003139 | -6.79 | 0.000 | -0.02747 | -0.01516 |
| <i>land</i> | 0.007906 | 0.003929 | 2.01 | 0.044 | 0.000207 | 0.015606 |
| <i>techinputs</i> | 0.077737 | 0.004742 | 16.39 | 0.000 | 0.068443 | 0.087031 |
| <i>forest</i> | 0.020425 | 0.009285 | 2.2 | 0.028 | 0.002227 | 0.038624 |
| <i>ndareas</i> | 0.086496 | 0.008944 | 9.67 | 0.000 | 0.068967 | 0.104026 |
| <i>social</i> | 0.659066 | 0.015642 | 42.14 | 0.000 | 0.628409 | 0.689723 |
| <i>demog</i> | -0.12634 | 0.028992 | -4.36 | 0.000 | -0.18316 | -0.06952 |
| constant | -0.44813 | 0.023053 | -19.44 | 0.000 | -0.49331 | -0.40294 |
| η | | | | | | |
| constant | -0.1976 | 0.149364 | -1.32 | 0.186 | -0.49035 | 0.095144 |
| $\ln \sigma_u^2$ | | | | | | |
| <i>social</i> | -2.17789 | 0.737983 | -2.95 | 0.003 | -3.62432 | -0.73147 |
| constant | -2.47837 | 0.762352 | -3.25 | 0.001 | -3.97255 | -0.98419 |
| $\ln \sigma^2$ | | | | | | |
| constant | -0.9899 | 0.027306 | -36.25 | 0.000 | -1.04341 | -0.93638 |

η Endogeneity Test: Ho: Correction for endogeneity is not necessary; Ha: There is endogeneity in the model and correction is needed.

(1) $[\eta_techassist]_{constant} = 0$

$\chi^2(1) = 1.75$

Prob > $\chi^2 = 0.1858$

Result: Cannot reject Ho at 10% level.

Technical assistance, non-degraded areas and the proportion of forested areas are all statistically significant (Table 4). The former act favoring production and the latter have a negative effect.

Table 7 shows 5-number summaries for technical efficiency. Figure 1 shows box plots for the normalized ranks of the efficiency measurements. Efficiency differs significantly by regional classification. There is a clear domination of South, Southeast, and Center-West.

Table 4. Instrumental variable fractional regression. SAS output.

| Parameter | Estimate | Error | DF | t Value | Pr > t | [95% confidence interval] | |
|-------------------|----------|--------|------|---------|---------|---------------------------|---------|
| <i>Labor</i> | -0.0579 | 0.0244 | 4965 | -2.37 | 0.0178 | -0.1058 | -0.0100 |
| <i>land</i> | 0.0160 | 0.0306 | 4965 | 0.52 | 0.6008 | -0.0440 | 0.0761 |
| <i>techinputs</i> | 0.2375 | 0.0374 | 4965 | 6.35 | <.0001 | 0.1642 | 0.3108 |
| <i>forest</i> | 0.0555 | 0.0718 | 4965 | 0.77 | 0.4394 | -0.0853 | 0.1964 |
| <i>ndareas</i> | 0.2609 | 0.0692 | 4965 | 3.77 | 0.0002 | 0.1253 | 0.3965 |
| <i>social</i> | 1.8260 | 0.1217 | 4965 | 15.01 | <0.0001 | 1.5874 | 2.0645 |
| <i>demog</i> | -0.4441 | 0.2226 | 4965 | -2.00 | 0.0461 | -0.8805 | -0.0078 |
| constant | -2.7670 | 0.1840 | 4965 | -15.04 | <0.0001 | -3.1277 | -2.4063 |

Table 5. Limited Information maximum likelihood estimation. Half-normal stochastic frontier under endogeneity of technical assistance and fractional regression. SAS output.

| Parameter | Estimate | Std error frontier model | Std error Murphy-Topel | Murphy-Topel 95% confidence interval | P > z | |
|-------------------|----------|--------------------------|------------------------|--------------------------------------|----------|----------|
| Frontier | | | | | | |
| <i>labor</i> | 0.2320 | 0.01139 | 0.01023 | 0.21196 | 0.25204 | 0.00000 |
| <i>land</i> | 0.0880 | 0.01394 | 0.01108 | 0.06623 | 0.10967 | 0.00000v |
| <i>techinputs</i> | 0.4522 | 0.02027 | 0.01829 | 0.41635 | 0.48805 | 0.00000 |
| <i>techassist</i> | 0.6050 | 0.12580 | 0.16852 | 0.27470 | 0.93530 | 0.00033 |
| <i>forest</i> | -0.1230 | 0.03280 | 0.03011 | -0.18202 | -0.06398 | 0.00004 |
| <i>ndareas</i> | 0.2424 | 0.03558 | 0.03517 | 0.17346 | 0.31134 | 0.00000 |
| residual | -0.2529 | 0.13590 | 0.17634 | -0.59853 | 0.09273 | 0.15153 |
| constant | 2.7732 | 0.10250 | 0.12101 | 2.53602 | 3.01038 | 0.00000 |
| $\ln \sigma_u^2$ | | | | | | |
| <i>social</i> | -2.0794 | 0.6869 | 2.47472 | -6.92985 | 2.77105 | 0.40076 |
| constant | -2.4200 | 0.62090 | 0.83613 | -4.05881 | -0.78119 | 0.0038 |
| $\ln \sigma^2$ | | | | | | |
| constant | 0.3702 | 0.00955 | 0.01528 | 0.34026 | 0.40014 | 0.00000 |

The social indicator positively affects technical efficiency, as reported in Table 4. Regions that are to benefit the most from improvements in the social indicators are the North and Northeast. The instrumental variable regression indicates a strong dependence of technical assistance on the environment, demographics and the social conditions. The increased population dynamics makes the presence of technical assistance unnecessary, implying, therefore, a negative effect of the demographic index. The other indices are positively related to technical assistance.

Table 6. Relative elasticities and returns to scale.

| Production factor | Relative elasticity | Standard error |
|-------------------|---------------------|----------------|
| Labor | 0.297 | 0.016 |
| Land | 0.116 | 0.018 |
| Technology | 0.587 | 0.022 |
| Returns to Scale | 0.777 | 0.014 |

Table 7. Normalized rank of technical efficiency – 5-number summary.

| Region | Minimum | Q1 | Median | Q3 | Maximum |
|-------------|---------|-------|--------|-------|---------|
| North | 0.000 | 0.192 | 0.313 | 0.425 | 0.957 |
| Northeast | 0.000 | 0.098 | 0.206 | 0.344 | 0.999 |
| Southeast | 0.004 | 0.502 | 0.695 | 0.854 | 1.000 |
| South | 0.055 | 0.618 | 0.744 | 0.866 | 1.000 |
| Center-west | 0.012 | 0.405 | 0.526 | 0.543 | 0.990 |

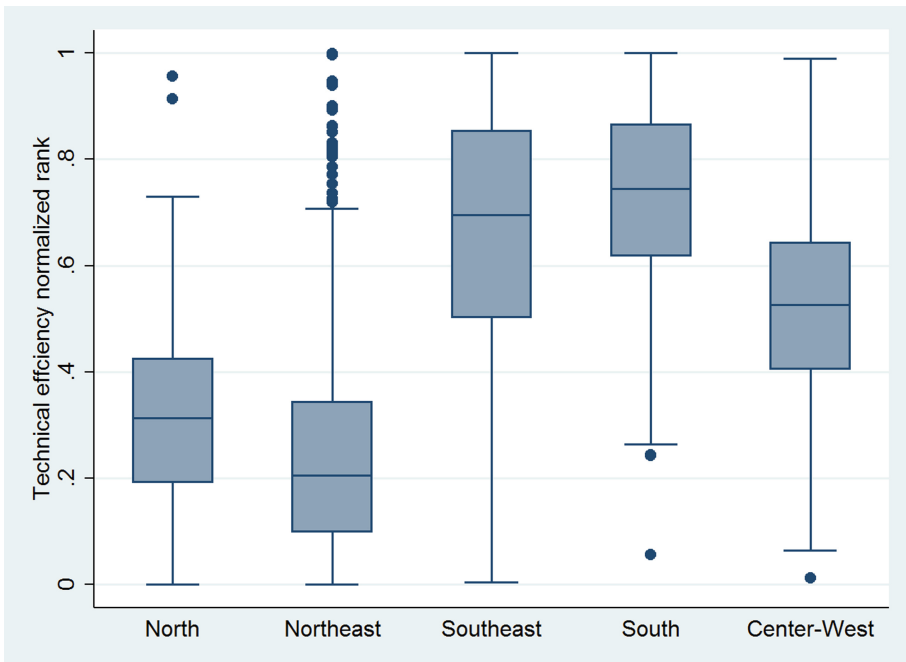


Fig. 1. Box plots of technical efficiency by region.

The limited information maximum likelihood estimation agrees, in general, with the full information maximum likelihood results. There is no evidence of technical assistance endogeneity. See Table 5. The main difference regards the standard error of the estimated coefficient of the social indicator in the inefficiency variance (Table 5). The Murphy-Topel correction inflates the variance, forcing non-significance. However, the coefficient values are similar. The fractional instrumental variables regression indicates positive relation to the social indicator and to non-degraded areas (Table 4). The demographic index is negatively related to the response and the proportion of forested areas is not significant.

Limited information estimation, including all instrumental variables as technical efficiency effects, is clearly inferior to the full information model estimated, including only the social indicator (Tables 1 and 2). The interesting feature of these models is the similarity of the results obtained with the linear and non-linear instrumental regression, suggesting robustness of the linear instrumental regression.

5 Concluding Remarks

We fitted a stochastic frontier under endogeneity to municipal data using the Brazilian agricultural census of 2006 – the last one available. The objective of this study, besides assessing input elasticities, was to investigate effects of market imperfection variables on production. Market imperfections come from different realities in production experienced by small and large farmers. They relate to infrastructure, level of education and access to credit, implying in different input and output prices for small and large farmers. The presence of market imperfection makes it harder for rural extension and technical assistance to promote productive inclusion.

For public policy decision-making, the identification of production function components elasticities is of importance to guide rural governmental assistance. This is critical for reducing poverty in the fields and for increasing production. We conclude that technology is the main input factor for increasing income in rural Brazil. The social indicator is the key variable to reducing inefficiency. The indicator is relatively too low for the Northern and Northeastern regions. Values are less than half of the corresponding values of other regions. Public policies should be oriented to improve this indicator particularly in these regions.

Technical assistance is an important part of rural extension and has a direct positive effect on income. Improvement of the social indicator will tend to facilitate the access of technical assistance creating, in this way, a synergic positive effect on income.

Environment in our study was measured in two ways: non-degraded areas and the proportion of forested areas. Keeping non-degraded areas relates to technology and has a positive impact on production. On the other hand, keeping a relative large area of uncultivated land in the farm will have a negative effect on income. Extension and technical assistance may be the key factor to extract value from forests and properly preserve these areas.

Finally, we emphasize the fact that the use of limited information maximum likelihood estimation indeed eases convergence in the stochastic frontier models. The linear instrumental regression seems to be robust, but it produces inferior fits when compared with fractional regressions. The Murphy-Topel variance matrix correction may change the significance of important variables relative to the full information maximum likelihood estimation.

References

1. Alves, E., Souza, G.S., Rocha, D.P.: Desigualdade nos campos sob a ótica do censo agropecuário 2006. *Revista de Política Agrícola* **22**, 67–75 (2013)
2. Souza, G.S., Gomes, E.G., Alves, E.R.A., Magalhães, E., Rocha, D.P.: Um modelo de produção para a agricultura brasileira e a importância da pesquisa da Embrapa. In: Alves, E. R.A., Souza, G.S., Gomes, E.G. (eds.) *Contribuição da Embrapa para o desenvolvimento da agricultura no Brasil*, pp. 49–86. Embrapa, Brasília (2013)
3. Souza, G.S., Gomes, E.G.: The effect of marketing imperfection variables on production in the context of Brazilian agriculture. In: *Proceedings of the 7th International Conference on Operations Research and Enterprise Systems (ICORES 2018)*, pp. 15–20. Scitepress, Setúbal (2018)
4. Alves, E., Souza, G.S.: Pequenos estabelecimentos em termos de área também enriquecem? Pedras e tropeços. *Revista de Política Agrícola* **24**, 7–21 (2015)
5. Souza, G.S., Gomes, E.G., Alves, E.R.A.: Conditional FDH efficiency to assess performance factors for Brazilian agriculture. *Pesquisa Operacional* **37**, 93–106 (2017)
6. Karakaplan, M.U., Kutlu, L.: Handling endogeneity in stochastic frontier analysis (2013). <http://www.mukarakaplan.com/Karakaplan%20-%20EndoSFA.pdf>. Accessed 10 Mar 2017
7. Karakaplan, M.U.: Fitting endogenous stochastic frontier models in Stata. *Stata J.* **17**(1), 39–55 (2017)
8. Murphy, K.M., Topel, R.H.: Estimation and inference in two step econometric models. *J. Bus. Econ. Stat.* **3**, 370–379 (1985)
9. Papke, L.E., Wooldridge, J.M.: Econometric methods goes fractional response variables with an application to 401(k) plan participation rates. *J. Appl. Econ.* **11**(6), 619–632 (1996)
10. Ramalho, E.A., Ramalho, J.J.S., Henriques, P.D.: Fractional regression models for second stage DEA efficiency analyses. *J. Prod. Anal.* **34**, 239–255 (2010)
11. IBGE Homepage. Censo Agropecuário (2006). <http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/>. Accessed 24 Jan 2012
12. IBGE Homepage. Censo Demográfico (2010). <http://censo2010.ibge.gov.br/>. Accessed 24 Jan 2012
13. INEP Homepage. Nota Técnica do Índice de Desenvolvimento da Educação Básica (2012) <http://ideb.inep.gov.br/resultado/>. Accessed 24 Jan 2012
14. Ministério da Saúde Homepage. IDSUS – Índice de Desempenho do SUS (2011). <http://portal.saude.gov.br/>. Accessed 02 Mar 2012
15. Amsler, C., Prokhorov, A., Schmidt, P.: Endogeneity in stochastic frontier models. *J. Econometrics* **190**, 280–288 (2016)
16. Greene, W.H.: *Econometric Analysis*, 6th edn. Prentice Hall, Englewood Cliffs (2008)