SOBRAPO

CONDITIONAL FDH EFFICIENCY TO ASSESS PERFORMANCE FACTORS FOR BRAZILIAN AGRICULTURE

Geraldo da Silva e Souza, Eliane Gonçalves Gomes*
and Eliseu Roberto de Andrade Alves

ABSTRACT. In this article we assess the effect of market imperfections and income inequality on rural production efficiency. The analysis is carried out using the notion of stochastic conditional efficiency computed in terms of free disposal hull (FDH) efficiency measurements. Free disposal hull and conditional FDH are output oriented with variable returns to scale. They are evaluated for rural production at the county level, considering the rank of rural gross income as the output and the ranks of land expenses, labor expenses, and expenses on other technological factors as inputs. The conditional frontier is dependent on income inequality and other indicators related to market imperfections. The econometric approach is based on fractional regression models and the generalized method of moments (GMM). Overall, the market imperfection variables act to reduce performance, and income dispersion is positively associated with technical efficiency.

Keywords: conditional FDH, GMM, fractional regression, rural income inequality.

1   INTRODUCTION

Recent studies of Brazilian agriculture based on the Agricultural Census of 2006 suggest a positive association between the (income) Gini index and the production efficiency. Souza et al. (2015), using the Gini index as the dependent variable, found that the variable returns to scale (VRS) data envelopment analysis (DEA) score (Banker et al., 1984) is highly significant in a nonlinear fractional regression model relating income dispersion to DEA and other covariate determinants of market imperfections. The model was fitted to each of the five Brazilian regions (south, southeast, north, northeast, and center-west). The DEA variable was dominant in all the regions and the market imperfection variables varied in intensity by region. As Alves & Souza (2015) pointed out, due to market imperfections, small farmers sell their products at lower prices and buy inputs at higher prices. As a result, they are not able to adopt technologies at relatively

*Corresponding author.
Embrapa, Secretaria de Gestão e Desenvolvimento Institucional – SGI, PqEB Av. W3 Norte final, 70770-901 Brasília, DF, Brasil.
E-mails: geraldo.souza@embrapa.br; eliane.gomes@embrapa.br; eliseu.alves@embrapa.br

higher prices. The conditions leading to market imperfections and affecting agricultural production are schooling, access to credit, access to water and electricity, and the development level in general. Thus, it is important to identify and quantify the effects of market imperfections to guide public policies and to reduce income concentration in the rural areas.

In another regression study relating ranks of income concentration as a linear function of ranks to labor, land, and technological input expenses, Alves et al. (2013) found that, for Brazil, all the variables were statistically significant and that technological inputs dominated the relationship. They concluded that technology was responsible for the observed income concentration. Indeed, they reported that 11% of the farms were responsible for 87% of the value of the rural production, according to the 2006 agricultural census. In that article they emphasized the need for further studies quantifying the effect of technology in the presence of market imperfections. Other studies deserve mentioning. Ney & Hoffmann (2008) analyzed the contribution of agricultural and non-farm activities to the inequality of rural income distribution in Brazil, observing two pieces of evidence: the participation of each sector in the household earnings of different income strata delimited by percentiles and the decomposition of Gini coefficients. Their results show that agriculture and rural non-farm activities contribute, respectively, to reducing and to increasing the rural income inequalities in Brazil. Ney & Hoffmann (2009) assessed the effects of rural income determinants, in particular human capital and physical capital. Ferreira & Souza (2007) analyzed the participation and the contribution of the household income "retirements and pensions" for the inequality of the distribution of the per capita household income in Brazil and rural Brazil, in the period from 1981 to 2003. Neder & Silva (2004) estimated poverty indexes and the income distribution in rural areas based on the National Survey for Household Sampling for the period 1995-2001. They reported a drop in the rural income concentration in some Brazilian states. These facts were not confirmed by the Brazilian 2006 agricultural census, if we restrict our attention to rural income.

We contribute to this literature by evaluating production in the appropriate way, that is, by conditioning efficiency on the covariates of importance (market imperfection determinants). To the best of our knowledge this approach is new regarding the study of public policies effects in the Brazilian agricultural production. The output is the rural income and the inputs are labor, land, and capital expenses. The flavor of the analysis is stochastic, also a new approach in empirical work, and involves the conditional efficiency probability models proposed initially by Daraio & Simar (2007). The measure of efficiency considered is FDH. The statistical analysis that we carried out mimics the approach of Souza & Gomes (2015) in what concerns the use of GMM to deal with endogeneity. It is an extension of the work of Daraio & Simar (2007), which does not explore a proper model specification of the response, defined by the conditional FDH measurement, as we are suggesting here. Another contribution of our approach to this literature has to do with the specification of the multivariate kernel used in the analysis, which differs considerably of the proposal of Bădin et al. (2012).

Our discussion proceeds as follows. Section 2 introduces the variables used in the production process. Section 3 and Section 4 deal with the methodology. Section 5 presents and discusses the statistical results. Finally, Section 6 summarizes and concludes the article.

2   PRODUCTION VARIABLES AND COVARIATES

The main source for this work is the Brazilian Agricultural Census of 2006 (IBGE, 2012a). The other sources used are listed below on a case-by-case basis. For the inputs and the output, we worked with monetary values. The choice of monetary values as opposed to quantities arose from the fact that using values allows for the aggregation of all agricultural outputs and inputs in the production process.

Farm data were pooled to form averages for each county. A total of 4,965 counties provided valid data for our analysis. This figure represents 89.3% of the total number of Brazilian counties. The decision-making unit (DMU) for our production analysis is the county.

Table 1 provides a complete list of inputs and outputs used to construct the production variables used in the analysis. The production variables used are straightforward and do not require further explanation. They were measured on the farm level, as facilitated by the census, and aggregated by county.

The contextual variables considered are the Gini index, the proportion of farmers who received technical assistance, the total financing per farm, and the county performance indexes in the social, environmental, and demographic dimensions. These performance indexes require further comments. They have been considered in total or in part by Embrapa (2001), Monteiro & Rodrigues (2006), Rodrigues et al. (2010), and Souza et al. (2013). The idea was also used by the National Confederation of Agriculture (Confederação Nacional da Agricultura, 2013) to develop an overall indicator of rural development. Our version of these quantities presented here are similar to but not coincident with these sources. The technique of index construction is based on the work of Moreira et al. (2004).

Gini index

As a measure of income inequality, we used the county Gini index. This is defined as follows. If x_i is the rural income of farm i, the index is $g/2\bar{x}$, where $g = (1/n^2) \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|$ and $\bar{x}$ is the mean of x_i. The Gini index varies in the interval [0, 1) with values close to 1 indicating more intense income concentration.

Social dimension

The variables comprising the social dimension reflect the level of well-being, favored by factors such as the availability of water and electric energy in the rural residences. They also reflect indicators of the level of education, health, and poverty in the rural residences.

The data used in the social dimension were extracted from the Brazilian Demographic Census 2010 (IBGE, 2012b), the Brazilian Agricultural Census 2006 (IBGE, 2012a), and the databases of the National Institute of Research and Educational Studies (INEP), referring to education in 2009 (INEP, 2012), and the Ministry of Health 2011 data (Ministério da Saúde, 2011).

Table 1 – Description of the production variables.

Variable	Components	Unit	Notes
Y (output)	Value of production of cattle, swine, goats, equines, buffaloes, donkeys, mules, sheep, other birds, rabbits, apiculture, sericulture, raniculture, aquaculture, horticulture, flowers, forestry, agro industry, permanent crops, temporary crops, extractive activities	Reais	—
Land	4 percent of land expenses, the rent paid for the land	Reais	—
Labor	Salaries or other forms of compensation paid to family and hired laborers	Reais	—
Capital (technological inputs)	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	Reais	Value of permanent crops, forests, machinery, improvements on the farm, animals, and equipment rental were depreciated at a rate of 6 percent a year. Depreciation periods: machines – 15 years, planted forests – 20 years, permanent cultures – 15 years, improvements – 50 years, and animals – 5 years

Demographic dimension

The variables comprising the demographic dimension capture aspects of the population dynamics that relate to rural development. These are the proportion of the rural to the urban population, the average size of a rural family, the aging rate, the migration index, and the ratio of the inactive population (0 to 14 years and 60 years or more of age) to the active population (15 to 59 years of age). The data source is the Brazilian Demographic Census 2010 (IBGE, 2012b).

Environmental dimension

The variables comprising the environmental dimension are the proportion of farmers practicing the technique of vegetation fires, proportion of farmers who use agrochemicals, proportion of farmers practicing crop rotation, proportion of farmers practicing minimum tillage, proportion of farmers practicing no tillage, proportion of farmers planting in contour lines, proportion of farmers providing proper garbage disposal, proportion of forest and agro-forest areas, and proportion of degraded areas. The data source is the Brazilian Agricultural Census 2006 (IBGE, 2012a).

All the variables within each dimension were measured in such way as to correlate positively with the given dimension. They were rank transformed and normalized by the maximum value

in the dimension, which in this case is the number of counties. Each specific dimension index is a weighted average of the normalized variables comprising the dimension, with weights defined by the relative squared multiple correlation obtained in the regression of a variable with all the others; that is, if R_i^2 is the squared multiple correlation of the regression, considering the ith variable as the dependent variable in the dimension, its weight is $R_i^2 / \sum_j R_j^2$.

3 FDH UNCONDITIONAL AND CONDITIONAL MEASURES OF TECHNICAL EFFICIENCY

We begin with the notion of FDH efficiency and related measures. Daraio & Simar (2007) and Bădin et al. (2012, 2014) discussed a measure of conditional efficiency that is closely relates FDH efficiency measures to probability theory, and which we used to assess the influence of covariates in the production process. Convexity of the technology set is not required. Consider production observations (x_j, y_j), $j = 1, \ldots, n$, of n producing units (DMUs). In our case, the input vector $x_j = (x_j^1, x_j^2, x_j^3)$ is a vector in R^3 with non-negative components, representing the land, labor, and capital components with at least one strictly positive. The output vector y_j is a non-negative point in R. The technical efficiency FDH of DMU τ is taken relative to the frontier of free disposability (free disposal hull) of the set

$$\psi = \left\{ (x, y) \in R_+^{3+1}, y \leq \sum_{j=1}^n \gamma_j y_j, x \geq \sum_{j=1}^n \gamma_j x_j, \sum_{j=1}^n \gamma_j = 1, \gamma_j \in \{0, 1\}, j = 1 \ldots n \right\} \quad (1)$$

The input-oriented FDH efficiency measure $\hat{\theta}(x, y)$ is given by

$$\hat{\theta}(x_\tau, y_\tau) = \text{Min} \left\{ \theta; y_\tau \leq \sum_{j=1}^n \gamma_j y_j, \theta x_\tau \leq \sum_{j=1}^n \gamma_j x_j, \sum_{j=1}^n \gamma_j = 1, \gamma_j \in \{0, 1\} \right\} \quad (2)$$

The output-oriented FDH efficiency measure $\hat{\lambda}(x, y)$ is given by

$$\hat{\lambda}(x_\tau, y_\tau) = \text{Max} \left\{ \lambda; \lambda y_\tau \leq \sum_{j=1}^n \gamma_j y_j, x_\tau \geq \sum_{j=1}^n \gamma_j x_j, \sum_{j=1}^n \gamma_j = 1, \gamma_j \in \{0, 1\} \right\} \quad (3)$$

These quantities are similar to their DEA counterparts, including the interpretation of radial contraction (input orientation) and augmentation (output orientation). The main difference is that in DEA the γ_j are not restricted to be 0 or 1.

As stated in Dario & Simar (2007), the FDH estimator is computed in practice by a vector comparison procedure. In the case of three inputs and only one output, and assuming input orientation, we have

$$\hat{\theta}(x_\tau, y_\tau) = \text{Min}_{j=1\ldots n} \left\{ \text{Max}_{i=1,2,3} \left\{ \frac{x_j^i}{x_\tau^i} \right\} \right\}. \quad (4)$$

For the output oriented FDH model we then have

$$\hat{\lambda}(x_\tau, y_\tau) = \text{Max}_{j, \, x_j \leq x_\tau} \left\{ \frac{y_j}{y_\tau} \right\}. \tag{5}$$

In this article we will restrict our attention to output oriented measures of efficiency (5).

A very interesting interpretation of the FDH arises when the production process is described by a probability measure, defined, in our case, in the product space R_+^{3+1} by random variables (X, Y). For efficiency purposes, we were interested in the probability of dominance given by

$$H_{XY}(y, x) = P(Y \geq y, X \leq x). \tag{6}$$

Notice that

1. $H_{XY}(y, x)$ gives the probability that a unit operating at the input–output level (x, y) is dominated, that is, that another unit produces at least as much output while using no more of any input than the unit operating at (x, y);

2. $H_{XY}(y, x)$ is monotone non-decreasing in x and monotone non-increasing in y.

The support of $H_{XY}(x, y)$ is the technology set Ω given by

$$H_{XY}(x, y) = 0 \quad \forall (x, y) \notin \Omega, \tag{7}$$

where $\Omega = \{(x, y) : x \text{ can produce } y\}$.

We then have

$$H_{XY}(x, y) = P(X \leq x \,|\, Y \geq y) P(Y \geq y) = F_{X/Y}(x \,|\, y) S_Y(y) \tag{8}$$

$$H_{XY}(x, y) = P(Y \geq y \,|\, X \leq x) P(X \leq x) = S_{Y/X}(y \,|\, x) F_X(x). \tag{9}$$

New concepts of efficiency measures can be defined for the input-oriented and output-oriented cases, assuming $S_Y(y) > 0$ and $F_X(x) > 0$. For input and for output orientation we have, respectively, the following equations

$$\theta(x, y) = \inf\{\theta \,\big|\, F_{X/Y}(\theta x \,|\, y) > 0\} = \inf\{\theta \,|\, H_{XY}(\theta x, y) > 0\} \tag{10}$$

$$\lambda(x, y) = \sup\{\lambda \,\big|\, S_{Y/X}(\lambda y \,|\, x) > 0\} = \sup\{\lambda \,|\, H_{XY}(\lambda y, x) > 0\}. \tag{11}$$

Since the support of the joint distribution of (X, Y) is the technology set, boundaries of Ω can be defined in terms of conditional distributions.

As we stated before, in our empirical work we used the output orientation. The output efficiency measure for given levels of input (x) and output (y) is given in (11) and it is nonparametrically estimated by

$$\hat{\lambda}_n(x, y) = \sup \left\{ \lambda \,\big|\, \hat{S}_{Y|X,n}(\lambda y \,|\, x) > 0 \right\}, \tag{12}$$

where

$$\hat{S}_{Y|X,n}(y|x) = \frac{\sum_{i=1}^{n} I(x_i \leq x, y_i \geq y)}{\sum_{i=1}^{n} I(x_i \leq x)}.$$

It can be shown that the following equation coincides with the FDH estimator.

$$\hat{\lambda}_n(x, y) = \text{Max}_{i:x_i \leq x} \left\{ \frac{y_i}{y} \right\}. \tag{13}$$

The estimated FDH production set is very sensitive to outliers, and consequently so are the estimated efficiency scores. Here we avoided the outlier problem by transforming the data into ranks before the analysis. The approach is well known in statistics to produce methods that are robust relative to outlying and heteroskedastic observations. See Wonnacott & Wonnacott (1990) for Anova applications & Conover (1999) for a general discussion. In our case we experienced difficulties in computing performance measures using the raw data. For our purposes the rank transformation was sufficient to obtain directions of better performance. If a ranked dependent variable responds positively to a ranked independent variable, it is likely that the same relationship will occur for raw data. This is typical in regression problems.

To assess the significance of a continuous contextual variable Z of dimension m on the output-oriented efficiency measurement, we conditioned on $Z = z$ to obtain

$$\lambda(x, y|z) = \sup\{\lambda | S_Y (\lambda_y | x, z) > 0\}, \quad \text{where}$$
$$S_Y(y|x, z) = \text{Prob}(Y \geq y | X \leq x, Z = z). \tag{14}$$

The nonparametric kernel estimate proposed by Daraio & Simar (2007) is defined by

$$\hat{S}_{Y,n}(y|x, z) = \frac{\sum_{i=1}^{n} I(x_i \leq x, y_i \geq y) K_{h_n}(Z_i - z)}{\sum_{i=1}^{n} I(x_i \leq x) K_{h_n}(Z_i - z)}, \tag{15}$$

where $K(\cdot)$ is the kernel and h_n is the bandwidth.

The bandwidth selection can be carried out using likelihood cross-validation as described by Silverman (1986), but other methods can also be used. For the multivariate case, we could use a bounded multivariate kernel or a product of univariate bounded kernels. An easy choice in the latter case would be the product of marginal kernels with support $[-1, 1]^k$. In our case $k = 6$, which is the number of covariates.

Substituting the estimator $\hat{S}_{Y,n}(y|x, z)$ into equation (14), we obtained the conditional FDH efficiency measure for the output-oriented case as below.

$$\hat{\lambda}_n(x, y|z) = \sup\{\lambda | \hat{S}_Y (\lambda y | x, z) > 0\}. \tag{16}$$

Therefore, for a multivariate bounded kernel we used the following estimator

$$\hat{\lambda}_n(x_j, y_j|z_j) = \max_{\{i; x_i \leq x_j, \|z_i - z_j\| \leq \tau_n \hat{\sigma}\}} \left\{ \frac{y_i}{y_j} \right\}, \tag{17}$$

where τ_n is the bandwidth and $\hat{\sigma}$ is the square root of the average variance of the covariates.

With a product of univariate bounded kernels, following Bădin et al. (2012), it is also possible to use the following formulation

$$\hat{\lambda}(x_j, y_j \,|z_j) = \max_{\left\{i; x_i \leq x_j, \left|z_i^1 - z_j^1\right| \leq \tau_{1n}, \ldots, \left|z_i^m - z_j^m\right| \leq \tau_{mn}\right\}} \left\{\frac{y_i}{y_j}\right\}, \tag{18}$$

where $(\tau_{1n}, \ldots, \tau_{mn})$ is the vector of marginal univariate bandwidths of the covariates.

The statistical inference will be dependent on how the kernels and bandwidths are chosen. Typically, nonparametric estimates are not dependent on the kernel choice and we proceed with the suggestion of Silverman (1986) and Daraio & Simar (2007), which is the Epanechnikov kernel. In our approach we obtained a better fit using a single bandwidth, also following Silverman (1986) for the bandwidth choice. The multivariate Epanechnikov kernel is given by

$$K(u) = \begin{cases} (1/(2c_d)) (d + 2)(1 - u'u) & \text{if } u'u < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{19}$$

where u is a point in a d-dimensional space and c_d is the volume of the unit sphere. In our application $d = 6$ and $c_6 = (1/6)\pi^3$.

The optimal window width for the smoothing of normally distributed data with unit variance is given by

$$h_{opt} = A(K)n^{-1/(d+4)}$$

$$A(K) = \left\{8c_d^{-1}(d + 4) \left(2\sqrt{\pi}\right)^d\right\}^{1/(d+4)} \tag{20}$$

If the data are not transformed for unit variances, the indicated procedure is to choose a single-scale parameter $\hat{\sigma}$ and to use the value $\hat{\sigma} h_{opt}$ for the window width. A possible choice is $\hat{\sigma}^2 = \frac{1}{d} \sum_{i=1}^d \hat{\sigma}_i^2$ as above, where $(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_d^2)$ is the vector of variances of the covariates.

4   STATISTICAL INFERENCE AND COVARIATE ASSESSMENT

For the assessment of the influence of the covariates on the efficiency measurements, Daraio & Simar (2007) suggested a nonparametric statistical analysis using the ratio

$$q_n(x_j, y_j \,|z_j) = \frac{\hat{\lambda}_n(x_j, y_j \,|z_j)}{\hat{\lambda}_n(x_j, y_j)}$$

as the response variable. The underlying suggested statistical model (Daraio & Simar, 2007) is $E\left(q(x_j, y_j \,|z_j)\right) = G(\mu_j)$, $\mu_j = z_j'\theta$, assuming the vector of covariates to be exogenous. If the mean specification is correct and the errors are uncorrelated, standard quasi-maximum likelihood inference is valid even for efficiency distributions concentrating the probability mass on 1 (Ramalho et al., 2010). When the independent variables are endogenous, the statistical inference may not be valid. In statistics an endogeneity problem occurs when an explanatory

variable is correlated with the error term. This problem seems to be more serious than the cross-sectional correlations. A similar condition is "separability," which was discussed by Simar & Wilson (2007). The typical approach to overcome the problem of endogeneity is the instrumental variable estimation.

A natural parametric version of the model above is to consider the flexible fractional regression specification of Ramalho et al. (2010). We assumed that the observations are correlated and that some of the covariates are endogenous (the Gini index, financing, technical assistance, and social and environment indicators). The correlation is induced by the FDH computations.

A common choice for $G(\cdot)$ is $G(\mu) = \Phi(\mu)$, where $\Phi(\cdot)$ denotes the distribution function of the standard normal. Other specifications could be the logistic $G(\mu) = e^{\mu}/(1 + e^{\mu})$ or the complementary log-log $G(\mu) = 1 - e^{-e^{\mu}}$, but frequently they lead to similar conclusions. In our case we obtained the best response for the standard normal, followed by the logistic. The complementary log-log did not converge. Our preferred method of estimation was the generalized method of moments – GMM (Gallant, 1987; Greene, 2011), with robust estimation of the parameter vector variance-covariance matrix. The GMM handles the problems of endogeneity and nonlinearity of the expected mean and, in many instances, provides robustness relative to the correlated observations as well (Conley, 1999; Stata, 2015).

Let h_{inst} be a vector of strictly exogenous variables available for use as instruments. If

$$u_{jn}(\theta, z_j) = \frac{\hat{\lambda}_n(x_j, y_j \,|\, z_j)}{\hat{\lambda}_n(x_j, y_j)} - G(z'_j \theta)$$

and assuming the moment condition $E\left\{h \otimes u_{jn}(\theta, z_j)\right\} = 0$ to be true, we may estimate the parameter θ using the generalized method of moments. An endogeneity and goodness-of-fit test is performed, testing for overidentifying restrictions (Hansen's J test; Hansen, 1982).

To obtain bias-corrected confidence intervals and standard errors adjusted for all potential mis-specifications (endogeneity, cross-sectional correlation, and heteroskedasticity), we performed nonparametric bootstrapping (Stata, 2015) with 5,000 replications. The instruments that we chose here are regional dummies and variables measured after the agricultural census of 2006, available in the demographic census of 2010.

5   STATISTICAL FINDINGS

Table 2 shows the GMM results for the fitting of the model defined by $G(\mu) = \Phi(\mu)$. The other parametric alternatives given in Section 4 are worse. The criterion GMM function (the less the better) is 2.6167×10^{-4} for the standard normal and 3.4668×10^{-4} for the logistic response function. As we stated in Section 4, the complementary log-log function did not converge. The only covariate considered to be exogenous is the demographic index, as all the others were measured by or are somehow related to the agricultural census. The instruments used, that is, exogenous and or predetermined variables assumed to be independent of the error term (Davidson & MacKinnon, 1993), are regional dummies, the proportion of farms with a power supply

service in 2010, the demographic index, and demographic $\wedge 2$. A standard procedure in GMM estimation is to also include as instruments low order monomials in the in the instrument set (Gallant, 1987). Standard errors were computed from the bootstrap replications. The Hansen's J test statistic for the model correctness and instruments specification is 1.299, with 1 degree of freedom and p-value $= 0.254$. There is no evidence against the model specification and the instruments considered.

Ratkowsky (1983) considered the bound 1% for the relative bias as the threshold for linear normal behavior of a parameter estimator. Relative bias is defined as the ratio of the bias to the parameter estimate in absolute values. It is necessary to access the bootstrap distribution in order to estimate the relative bias. The 1% relative bias is achieved for the variables financing, social, demographic, and technical assistance. We failed to accept the normal assumption for the estimators of these parameters considering the bootstrap distribution. Only the Gini variable showed normal behavior. The Kolmogorov-Smirnov statistics of normality for financing, environment, social, demographic, technical assistance, and the Gini index are 0.014 (p-value $= 0.0184$), 0.028 (p-value < 0.01), 0.014 (p-value $= 0.0186$), 0.030 (p-value < 0.01), 0.020 (p-value < 0.01), 0.038 (p-value < 0.01), and 0.007 (p-value > 0.15), respectively. In general, the distributions are skewed, and bias-corrected confidence intervals should be taken into account to assess the statistical significance. In this context, the environment and demographic variables are not significant.

The interpretation of the sign of a covariate (Z) coefficient is taken from Daraio & Simar (2007, 115): "an increasing regression corresponds to a favorable environmental factor and a decreasing regression indicates an unfavorable factor." We see from Table 2 that financing, social, and Gini are favorable. These variables show the correct signs. The influence of the Gini index points to a strong association with production efficiency, confirming the results of previous studies reporting that technology is responsible for the rural income concentration in the country. We see the negative sign of technical assistance as an indication that the rural extension in the country does not lead to higher efficiency levels, which is very likely due to market imperfections that inhibit the use of proper technologies to increase rural productivity. Market imperfections are related to the markets of products, exports, and inputs, markets of water and power supply, markets of land, equipment, and machinery rentals, access to information, and education. As we have already pointed out, the small rural producers sell their products at lower prices and buy inputs at higher prices; therefore, they are not able to adopt the optimum technologies. These will be expensive and the cost will not be covered by the output revenue. Two other points inhibiting technical assistance for small farmers are the low educational level and the difficulties in accessing information. Rural production efficiency in Brazil is achieved mainly by the large producers. They may be owners of small pieces of land and adopt modern technologies, such as irrigation. Since they trade with large volumes, they gain lower input prices and easier access to technology.

Table 2 – GMM estimation.

Variable	Estimate	Relative bias	Bootstrap standard error	Robust standard error	Robust SE/ Boot SE	95% bootstrap confidence interval	
Constant	−2.7224	3.9017	1.0081	0.9903	0.9823	−4.6988	−0.6540
Financing	1.6870	0.5719	0.4904	0.5131	1.0463	0.8489	2.8144
Environment	1.3131	7.1068	1.1484	1.1800	1.0276	−0.8677	3.6785
Social	1.4740	0.9230	0.5705	0.5481	0.9608	0.4955	2.7652
Demographic	0.5780	0.0194	0.3407	0.3324	0.9758	−0.0659	1.2892
Tech. Assist.	−2.7881	0.0087	0.8831	0.8507	0.9633	−4.8169	−1.3202
Gini	3.1425	3.2974	1.0285	0.9865	0.9592	1.0724	5.1064

The marginal effects were computed following the equation

$$\frac{dG(\mu)}{d\mu}\theta_j \,|\mu = z'\theta, \qquad \frac{dG(\mu)}{d\mu} = \phi(\mu), \qquad (21)$$

where $\phi(\cdot)$ is the standard normal density function. The intensity of the marginal effects on the response may be appreciated by the median values of the factor $\phi(\mu)$. Table 3 shows five-number summaries for each region, and the corresponding box plots are shown in Figure 1. The region that will benefit the most from public policies reducing market imperfections, *ceteris paribus*, is the northeast, closely followed by the northern, southern, and southeastern regions.

Table 3 – Five-number summary statistics for the regional marginal effects.

Region	Minimum	Twenty-fifth percentile	Median	Seventy-fifth percentile	Maximum	Number of observations
North	0.0065	0.1771	0.2786	0.3610	0.3989	405
Northeast	0.0027	0.2096	0.2936	0.3615	0.3989	1666
Southeast	0.0034	0.1244	0.2124	0.3133	0.3989	1531
South	0.0021	0.1596	0.2659	0.3510	0.3989	1143
Center-west	0.0035	0.0684	0.1258	0.2105	0.3952	220

6   SUMMARY AND CONCLUSIONS

Using the Brazilian 2006 agricultural census data and other official sources, we examined the influence of market imperfection variables and income concentration on the efficiency of rural production in Brazil at the county level. As a measure of production efficiency, we took the FDH performance score. The analysis was based on the conditional frontier, assuming a random data-generating process for the frontier. The statistical analysis was carried out using fractional regression and GMM and postulated a nonlinear relationship between the ratio of the conditional to unconditional FDH frontiers and the set of conditioning variables. The statistical inference
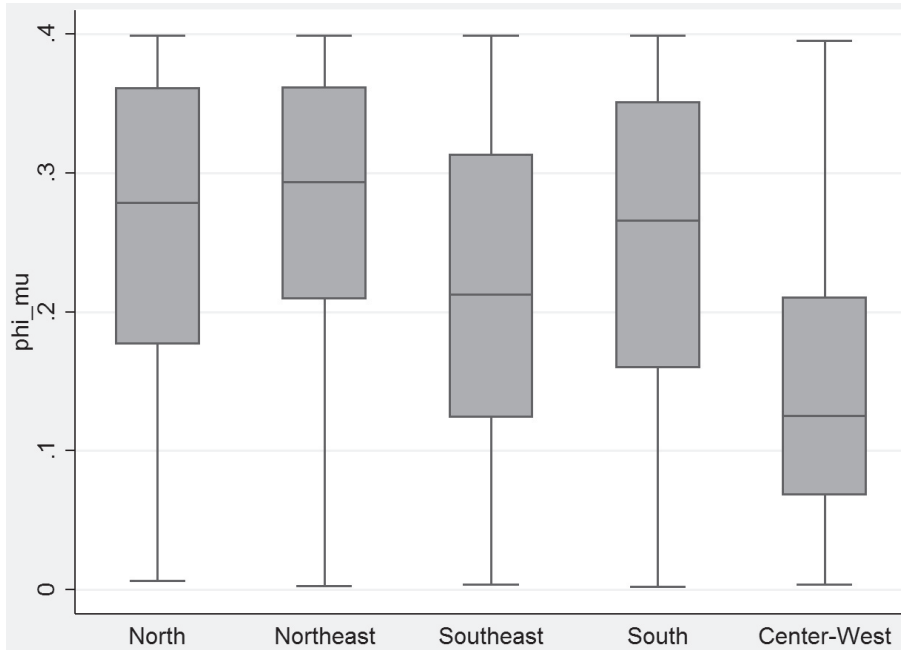
Figure 1 – Regional marginal effects box plots.

was performed with instrumental variables, to handle the eventual endogeneity and separability problems, and bootstrapping, to address the cross-sectional correlation and heteroskedasticity properly.

We conclude that the market imperfections discriminate against the small rural producers. The public policies are directed to the rich farmers, since the interest rates are higher for the small producers, the leasing of equipment and machinery is also more expensive, access to health and education services is compromised, and the infrastructure drawbacks are more difficult to overcome. These conditions are unfavorable for production, making it impossible for small farms to adopt technologically intensive inputs. We also found a strong positive association between rural production efficiency and income concentration, emphasizing the difficulties faced by poor farmers in accessing proper technology at competitive prices due to market imperfections.

In a recent study, Alves et al. (2015) compared the behavior of market imperfections in the south of Brazil (state of Rio Grande do Sul) and in the northeastern region (state of Bahia). We believe that their observations extrapolate for the other regions and suggest a comparison between the north and northeast and the other regions from the point of view of efficiency in production. They pointed out that in the south the better values of gross production are due to better control of market imperfections, by actions carried out by the county's administrations, by the rural and urban leadership, and by the rural cooperatives and farmers' associations. The variable that they used to compute the Gini index is the value of gross production and should

function as a proxy for the Gini quantity computed here. If the purpose of public policies is to reduce income concentration, one should invest in technologies that will increase land and labor productivity, and this cannot be achieved in the presence of market imperfections. In our study, we found that the northern and the northeastern regions in Brazil are likely to be the most responsive to marginal effects in income concentration and market imperfections. Therefore, public policies should be aimed at reducing the income concentration and creating conditions that allow productive inclusion and rural poverty reduction.

REFERENCES

[1] Alves E & Souza GS. 2015. Pequenos estabelecimento em termos de área também enriquecem? Pedras e tropeços. *Revista de Política Agrícola*, **24**(3): 7–21.

[2] Alves E, Souza GS & Rocha DP. 2013. Desigualdade nos campos sob a ótica do censo agropecuário 2006. *Revista de Política Agrícola*, **22**(2): 67–75.

[3] Alves E, Souza GS, Garagorry FL & Mello PF. 2015. O sonho de produzir: assentados da reforma agrária da Bahia e do Rio Grande do Sul. *Revista de Política Agrícola*, **24**(3): 114–133.

[4] Badin L, Daraio C & Simar L. 2012. How to Measure the Impact of Environmental Factors in a Nonparametric Production. *European Journal of Operations Research*, **223**: 818–833.

[5] Badin L, Daraio C & Simar L. 2014. Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, **214**(1): 5–30.

[6] Banker RD, Charnes A & Cooper WW. 1984. Some models for estimating technical scale inefficiencies in Data Envelopment Analysis. *Management Science*, **30**(9): 1078–1092.

[7] Confederação Nacional da Agricultura. 2013. Índice de Desenvolvimento Rural CNA. Brasília: CNA (unpublished document).

[8] Conley TG. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics*, **92**: 1–45.

[9] Conover WJ. 1999. Practical Nonparametric Statistics. 3rd ed. New York: Wiley.

[10] Daraio C & Simar L. 2007. Advanced Robust and Nonparametric Methods in Efficiency Analysis. New York: Springer.

[11] Davidson R & MacKinnon JG. 1993. Estimation and Inference in Econometrics. New York: Oxford University Press.

[12] Embrapa. 2001. Balanço Social da Pesquisa Agropecuária Brasileira. Brasília: Secretaria de Comunicação/Secretaria de Gestão Estratégica.

[13] Ferreira CR & Souza SCI. 2007. As aposentadorias e pensões e a concentração dos rendimentos domiciliares per capita no Brasil e na sua área rural: 1981 a 2003. *Revista de Economia e Sociologia Rural*, **45**(4): 985–1011.

[14] Gallant AR. 1987. Nonlinear Statistical Models. New York: Wiley.

[15] Greene WH. 2011. Econometric Analysis. 7th ed. New Jersey: Prentice Hall.

[16] Hansen LP. 1982. Large sample properties of generalized method of moments estimators. *Econometrica*, **50**: 1029–1054.

[17] IBGE. 2012. Censo Agropecuário 2006. 2012a. Disponível em: `http://www.ibge.gov.br/home/estatistica/economia/agropecuaria/censoagro/`. Acesso em: 24 jan. 2012.

[18] IBGE. 2012. Censo Demográfico 2010. 2012b. Disponível em: `http://censo2010.ibge.gov.br/`. Acesso em: 24 jan. 2012.

[19] INEP. 2012. Nota Técnica do Índice de Desenvolvimento da Educação Básica. Disponível em: `http://ideb.inep.gov.br/resultado/`. Acesso em: 24 jan. 2012.

[20] MINISTÉRIO DA SAÚDE. 2011. IDSUS – Índice de Desempenho do SUS. Ano 1. Disponível em: `http://portal.saude.gov.br/`. Acesso em: 02 mar. 2012.

[21] MONTEIRO RC & RODRIGUES GS. 2006. A system of integrated indicators for socio-environmental assessment and eco-certification in agriculture – Ambitec-agro. *Journal Technology Management & Innovation*, **1**(3): 47–59.

[22] MOREIRA TB, PINTO MB & SOUZA GS. 2004. Uma metodologia alternativa para mensuração de pressão sobre o mercado de câmbio. *Estudos Econômicos*, **34**: 73–99.

[23] NEDER HD & SILVA JLM. 2004. Pobreza e distribuição de renda em áreas rurais: uma abordagem de inferência. *Revista de Economia e Sociologia Rural*, **42**(3): 469–486.

[24] NEY MG & HOFFMANN R. 2008. A contribuição das atividades agrícolas e não-agrícolas para a desigualdade de renda no Brasil rural. *Economia Aplicada*, **12**(3): 365–393.

[25] NEY MG & HOFFMANN R. 2009. Educação, concentração fundiária e desigualdade de rendimentos no meio rural brasileiro. *Revista de Economia e Sociologia Rural*, **47**(1): 147–181.

[26] RAMALHO EA, RAMALHO JJS & HENRIQUES PD. 2010. Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis*, **34**: 239–255.

[27] RATKOWSKY DA. 1983. Nonlinear Regression Modeling, a unified practical approach. New York: Marcel Dekker.

[28] RODRIGUES GS, BUSCHINELLI CCA & AVILA AFD. 2010. An environmental impact assessment system for agricultural research and development II: institutional learning experience at Embrapa. *Journal Technology Management & Innovation*, **5**(4): 38–56.

[29] SILVERMAN BW. 1986. Density Estimation. New York: Chapman and Hall.

[30] SIMAR L & WILSON PW. 2007. Estimation and Inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, **136**(1): 31–64.

[31] SOUZA GS & GOMES EG. 2015. Management of agricultural research centers in Brazil: a DEA application using a dynamic GMM approach. *European Journal of Operational Research*, **240**: 81–824.

[32] SOUZA GS, GOMES EG & ALVES ERA. 2015. Assessing the determinants of rural income dispersion in Brazil. *Anais do XLVII Simpósio Brasileiro de Pesquisa Operacional*, 582–591.

[33] SOUZA MO, MARQUES DV & SOUZA GS. 2013. Uma metodologia alternativa para cálculo dos indices de impactos sociais e ambientais das tecnologias da Embrapa. *Cadernos do IME-Série Estatística*, **34**: 1–15.

[34] STATA. 2015. Base Reference Manual Release 14. Texas: Stata Press.

[35] WONNACOTT TH & WONNACOTT RJ. 1990. Introductory Statistics. 5th. ed. New York: Wiley.