STOCHASTIC FRONTIER APPROACH TO AGRICULTURAL PRODUCTION INCLUDING TECHNICAL EFFECTS: THE BRAZILIAN AGRICULTURAL CENSUSES OF 1995-1996 AND 2006

Geraldo da Silva e Souza
Eliseu Roberto de Andrade Alves
Eliane Gonçalves Gomes
Empresa Brasileira de Pesquisa Agropecuária – Embrapa
Parque Estação Bilógica, Av. W3 Norte final, 70770-901, Brasília, DF, Brazil
geraldo.souza@embrapa.br
eliseu.alves@embrapa.br
eliane.gomes@embrapa.br

Eduardo Castelo Magalhães
Datalyze Consulting Corporation
1 Rideau Street, K1N 8S7, Ottawa, Ontario, Canada
emagalhaes@gmail.com

Daniela de Paula Rocha
Fundação Getúlio Vargas – IBRE/RJ
Rua Barão de Itambi60, 8º andar, 22231-000, Rio de Janeiro, RJ, Brazil
daniela@fgv.br

ABSTRACT

In this paper we analyze the impact of Embrapa's research on the technical efficiencies of Brazilian farmers, particularly the poor. Using a stochastic frontier approach, we estimate a production function to obtain not only the technical efficiencies but also the elasticities of the key inputs. Our results indicate that Embrapa has generally had a positive effect on technical efficiencies, but there are still challenges ahead for Embrapa. Key among these challenges is the dissemination of its technology, which seems to be predominantly adopted by wealthier farmers. The low or absent use of modern inputs by poor farmers is highly detrimental of their ability to succeed.

Key-words: Stochastic frontiers, Brazilian agriculture, Rural poverty, Embrapa's research.

RESUMO

Neste artigo analisa-se o impacto da pesquisa da Embrapa sobre a eficiência técnica dos agricultores brasileiros, especialmente os mais pobres. Usando uma abordagem de fronteira estocástica, estimou-se uma função de produção para obter não só as eficiências técnicas, mas também as elasticidades dos principais insumos. Os resultados indicam que a Embrapa, em geral, tem tido um efeito positivo na eficiência técnica, mas ainda há desafios à frente da empresa. Entre esses desafios está a difusão de sua tecnologia, que parece ser predominantemente adotada pelos agricultores mais ricos. A ausência ou o baixo uso de insumos modernos pelos agricultores mais pobres é um fator altamente prejudicial à capacidade de sucesso de seus empreendimentos rurais.

Palavras-chave: Agricultura brasileira, Pobreza rural, Pesquisa da Embrapa, Fronteiras estocásticas.

1. Introduction

Brazil's relevance and emergence in the current world economy need not be described. It is a tour-de-force acknowledged globally. Not as much has been published, however, to take account of its success in combating poverty, and in particular how its primary source of public agricultural research – the Brazilian Agricultural Research Corporation, known as Embrapa – has contributed to poverty reduction in light of the large body of literature that promotes agricultural research as a poverty-reduction tool.

The current literature expounds the process through which the growth of agriculture leads to the development of other sectors and in doing so promotes gains in income and welfare. In order to have a strong agricultural sector, however, growth in productivity and strong agricultural research efforts are needed, both of which play key roles in poverty reduction in addition to propelling the agricultural sector and the economy as a whole (Christiansen and Demery, 2007; Thirtle et al., 2003; Fan et al., 2008; Fan and Zhang, 2008).

The evidence behind the importance of agricultural research as a poverty reducing mechanism is vast and ranges from its direct effects on the poor to a tool to more effectively use public expenditure as a poverty reducing mechanism (Fan et al., 2008). Much of the work, however, done to assess the impact of agricultural research has taken place in Africa and Asia (see for instance, Fan and Zhang, 2008). In addition, many of the economies for which these findings hold true were in an early stage of development by which is meant that agriculture still plays a very relevant role in the economy as a whole, vis-à-vis manufacturing and services.

Brazil has long left that stage. While agriculture is very strong as a sector as displayed by record production levels and the fact that Brazil is now a major agricultural exporter, agriculture accounts for only 5-6 percent of the gross domestic product. Brazil's success owes much to the successful economic policies put in place in the mid 1990s to stabilize inflation, reduce subsidies, price controls and monopolies, and virtually eliminate taxes on primary and semi-manufactured export products (Almeida, 2009).

These policies were further complemented by two key factors: the systematic presence of the Brazilian Agricultural Research Corporation (Embrapa) since the early 70s and the pro-poor policies put in place in early 2000s. For agriculture, the role of the PRONAF (National Program on Family Agriculture), a program aimed to provide small farmers with technical assistance and ease of credit, has been particularly prominent.

In this context, it becomes particularly relevant to analyze the impact of Embrapa's research and of agricultural research in general on the welfare of farmers across the country, but particularly less-favored ones. The question we ask is: how effective has Embrapa – through its broad suite of technologies – been in helping the poor improve their wellbeing, and, in particular, their levels of productivity?

The motivation of this work is four-fold: first, Alston et al. (2000) has shown that the impacts of public agricultural research are usually very large from a primarily cost-benefit point of view. So we would expect that Embrapa, as a public entity, is likely to have a large return in terms of the money invested in it. In fact, Pardey et al. (2006) have looked at only the crop breeding side of Embrapa's research for three commodities and found an astounding rate of return of 16:1 for each dollar invested, even when using the most strict benefit attribution rule. Thus, we know that Embrapa is powerful in its ability to generate impact. To our knowledge though, Embrapa's direct impact on the poor has not been measured. Second, the assessment done by Pardey looks at only one technology – crop breeding – and here we attempt to quantify the impacts of all technologies. Thirdly, as we have indicated above much of the literature on the impacts of agricultural research has not been done in a context that resembles or emulates the Brazilian reality. Lastly, in 2011, the Brazilian Bureau of Geography and Statistics (IBGE) released the final dataset of the agricultural census of 2006, which allowed us to measure the impact at the farmer's level. In addition, we were able to assess whether progress has been made between 1995-1996 and 2006 (the two last agricultural censuses).

We have chosen to conduct the analysis using a stochastic frontier approach. This allowed us to use the rich dataset available at IBGE to frame the impact discussion on the poor in a productivity context. Doing so has several advantages: First, Embrapa's technology and mandate is primarily aimed at improving productivity and as such it is only logical to analyze its impact using a productivity approach. Second, since this approach requires a production function, it allows us to analyze the performance of different production inputs while controlling for different regions and different strata of output, thus controlling for the vast heterogeneity that exists in Brazil. Third, it provides us with the ability to measure and explain technical efficiency. Technical efficiency is at the core of productivity growth (along with technical change) and is the component that will be most affected by Embrapa, especially for the poor. This is so because the poor are not at the frontier of technology (as we will see, most do not even adopt technology) and as such the best way to make them better off is to enable them to become more efficient, to produce more with the same or less.

The remainder of this paper is organized as follows. Section 2 provides support to the use of stochastic frontier as an impact assessment tool and outlines technical aspects of the approach utilized in the paper. This is followed in section 3 by a discussion on the data and the sampling strategy, followed by the analyses of the descriptive statistics. Section 4 discusses the results and is followed by a conclusion and implications for Embrapa.

2. Methodological Discussion

The primary use of stochastic frontier analysis is to determine not only input elasticities but also levels of technical efficiency, both in and outside of the agricultural field. And for this purpose, much has been written and published; see for instance Vicente (1999) and Belloumi and Matoussi (2006). Much can be learned and gained from simply looking at elasticities and technical efficiencies, particularly if it is applied within an impact assessment framework.

Indeed, a considerable body of work has used this approach to measure the impact of selected shocks on technical efficiency. Khumbakar et al. (2012) looked at the impact of corporate research on the technical efficiencies of R&D investors in Europe. In a more agriculture-related theme, Zhang et al. (2011) analyzed the impacts of land reallocation in China and found that policy had significant impacts on technical efficiency. Nin-Pratt and Magalhães (in preparation) looked at the impact of seed programs in Ethiopia.

This non-exhaustive set of studies bring out two important points: first, they validate the stochastic frontier approach as a tool to conduct impact assessment. Second, they highlight the relevance of the technical efficiency component in an impact assessment study. This, as we indicated in the introduction, is not surprising given that the other component of productivity – technical change – is often more of a long-term effect, which directly affects only those that are at the frontier of knowledge. Keeping up and advancing towards the frontier producers is therefore the best the poor or less resourceful producers can do. It is in this context, i.e. of advancing poor farmers towards the frontier that we expect to see Embrapa's effort in alleviating poverty.

2.1. Theoretical Underpinnings

The discussion that follows in this section draws from key references in the productivity literature, both theoretical and applied, including Khumbakar and Lovell (2000), Coelli et al. (2005), Greene (2011) and Stata (2011). All these, in turn, are evolutions of the work done by Aigner et al. (1977).

The basic set up of a stochastic production frontier analysis starts with the definition of a production function $f(x, z, \theta)$, which is a function of *x* inputs of *k* dimension, a vector of *z* explanatory variables with a *g* dimension and a parametric vector *d* with a finite dimension. Without random errors and inefficiency, maximum production of output *y* for establishment *j* can be achieved with the use of x_j inputs and by controlling for z_j factors. The production function is thus given by: $y_j = f(x_j, z_j, \theta)$.

The likelihood of inefficiencies creeping up in the production process presupposes the existence of a stochastic component $\eta_j \in (0,1)$ such that production is actually given by

$y_j = f(x_j, z_j, \theta)\eta_j$. When η_j approaches 1 the farm is near the optimal production level defined by the function $f(x, z, \theta)$. When $\eta_j < 1$, the farm is not producing to its full capacity given the technology available to producers and incorporated in the function $f(x, z, \theta)$.

In addition to inefficiencies, production processes are also subject to the effects of random changes in the production environment. These effects can be negligible *per se* but often may lead to changes in the production function. As a result, it is possible to assume the existence of random variables v_j such that the production function becomes $y_j = f(x_j, z_j, \theta)\eta_j \exp(v_j)$.

The above specification is equivalent to the statistical model $\ln y_j = \ln f(x_j, z_j, \theta) + v_j - u_j$, where u_j is a non-negative random variable representing the inefficiency component in the model, i.e., $u_j = -\ln(\eta_j)$.

Production functions can come in many forms, which explore different aspects of the underpinning theory and assumptions about the relationships among variables. A frequently used form, in part due to its generality but also due to the simplicity of use, is the Cobb Douglas specification, which is given by $f(x, z, \theta) = C\prod_{v=1}^{k} x_v^{\beta_v} \exp(z'w)$. As written, $\theta = (\beta, w)$, where $\beta_v > 0$ is the elasticity of input x_v.This form can easily be linearized by applying logs to both sides, yielding $\ln(y_j) = \ln(C) + \sum_{v=1}^{k} \beta_v \ln(x_{v,j}) + \sum_{l=1}^{g} w_l z_{l,j} + v_j - u_j$.

Distinct stochastic specifications for the error component lead to altogether different frontier models. Generally, one assumes that v_j are distributed independently from the inefficiency component u_j. The v_j are assumed to represent a normal random distribution with mean zero and variance σ^2. For u_j, three distinct specifications are possible: exponential with a variance σ^2, half normal with variance σ_u^2 and truncated normal with mean μ and variance δ^2. Expected inefficiencies are given by σ_u^2 for the exponential distribution, $\sqrt{2/\pi}\sigma_u$ for the half-normal, and $\mu + \phi\lambda$ with $\lambda = \phi(\mu/\delta)/\Phi(\mu/\delta)$ for the truncated normal distribution, where $\phi(\cdot)$ and $\Phi(\cdot)$ represent density and distribution functions of the standard normal.

The vector of parameters θ is estimated for n establishments via maximum likelihood, which yields asymtoptic and valid statistical inference. The following likelihood functions are maximized to obtain the parameters θ.

1. Normal-exponential model: $L(\theta^*) = \sum_{j=1}^{n} \left\{ -\ln\sigma_u + \frac{\sigma^2}{2\sigma_u^2} + \ln\left(\frac{-\varepsilon_j - \frac{\sigma^2}{\sigma_u}}{\sigma} \right) + \frac{\varepsilon_j}{\sigma} \right\}$

2. Normal half-normal model: $L(\theta^*) = \sum_{j=1}^{n} \left\{ \frac{1}{2}\ln\left(\frac{2}{\pi}\right) - \ln(\sigma_s) + \ln\left(-\frac{\rho\varepsilon_j}{\sigma_s} \right) - \frac{\varepsilon_j^2}{2\sigma_s^2} \right\}$

3. Normal-truncated normal model:

$$L(\theta^*) = \sum_{j=1}^{n} \left\{ \frac{1}{2}\ln(2\pi) - \ln(\sigma_s) - \ln\Phi\left(\frac{\mu}{\sigma_s\sqrt{\gamma}} \right) + \ln\Phi\left(\frac{(1-\gamma)\mu - \gamma\varepsilon_j}{\sigma_s\sqrt{\gamma(1-\gamma)}} \right) \right\} - \sum_{j=1}^{n} \left\{ \frac{1}{2}\left(\frac{\varepsilon_j + \mu}{\sigma_s} \right)^2 \right\}$$
$$- \sum_{j=1}^{n} \left\{ \frac{1}{2}\left(\frac{\varepsilon_j + \mu}{\sigma_s} \right) \right\}$$

In the above equations $\varepsilon_j = v_j - u_j$ represents the difference $\ln y_j - \ln f(x_j, z_j, \theta)$ between the response variable and deterministic part of the model, $\sigma_s^2 = \sigma^2 + \sigma_u^2$, $\rho = \sigma_u/\sigma$ and $\gamma = \sigma_u^2/\sigma_s^2$. The parameter θ^* includes θ and the additional parameterization used in the inefficiency component. Effects associated with contextual variables affecting technical efficiencies are modeled using the parameters involved in the specification of the distributions associated with inefficiency (e.g. half normal, exponential and truncated normal). The exponential and half normal distributions postulate that $\sigma_u^2 = \exp(m'b)$, where m is a vector of covariates and b is the corresponding vector effects. For the truncated normal distribution, the conditional mean of technical inefficiency can be explained as well and is given by $\mu = m'b$. The expected value of inefficiency in any case is a monotonic function of the linear construct. Heteroscedasticity in the inefficiency component can be modeled imposing a similar type of specification. This option is typically used in the exponential half normal specifications, and is not available in Stata (2011) for the normal truncated. However, the truncated normal allows for a specification that explains the variation in the conditional mean of the efficiencies. Under the three distributions and regardless of whether the conditional mean or variance of the inefficiency term is explained, the estimation of the error term component as a function of contextual variables is done simultaneously with the estimation of the production function, via maximum likelihood.

In this work we have used a normal-half normal distribution and resorted to contextual variables to explain both the error components and the production function. Thus, the measure of technical efficiency (te_j) is estimated by (Stata, 2011):

$$te_j = \left(\frac{1-\Phi\left(\sigma_* - \mu_{*j}/\sigma_*\right)}{1-\Phi\left(-\mu_{*j}/\sigma_*\right)} \right) \exp\left(-\mu_{*j} + \frac{1}{2}\sigma_*^2\right), \text{ where, } \mu_{*j} = -\varepsilon_j \frac{\sigma_u^2}{\sigma_s^2},\ \sigma_* = \frac{\sigma_u \sigma}{\sigma_s}.$$

3. Data

3.1. General aspects

The data components involved in this work drew from two agricultural censuses: 1995-1996 and 2006. There were three key types of variables necessary to conduct the stochastic frontier analysis: inputs, outputs and selected explanatory variables for the inefficiency and random components of the production function.

For the inputs and outputs, data were collected drawing from value of/expenditures in outputs and inputs. The choice of values as opposed to quantities arose from two main factors: first, using value of output allows for aggregation of all outputs and simplifies things considerably econometrically, as it eliminates the need of a distance function approach[1]. Second, as the goal of the analysis is to consider the poor and the impact of Embrapa on them, using value of outputs enabled us to analogize the total output of a given farm to its income. More specifically, we constructed output brackets to proxy income brackets and to classify farmers as poor or otherwise (see next section for details).

The list below provides the complete set of inputs and outputs used to construct the variables used in the analysis. Most of the variables used are straight-forward and do not require further explaining. The labor variable, however, does. We have used as proxy for labor the combined costs of family and hired labor (in either salary or other forms of payment) per farm, as provided by the census. It would have been better to have used a measure which more directly captured labor efforts put into production, particularly as family labor tends to be under-reported. Rada and Buccola (2012) have found that the intensity of family labor is in fact three times higher than hired labor for the four agricultural censuses they analyzed in Brazil. In their work, they have used number of days of family laborers as a measure of labor. This suggests that our labor elasticities are probably underestimated, though this is unlikely to generate a serious problem because the elasticities we found are in the same order (technology is the highest, followed by labor and land) as the ones found by Rada and Buccola (2012). However, we do acknowledge this issue as a potential limitation to this work.

[1] This becomes particularly important since all the analysis had to be performed the IBGE's headquarters in Rio de Janeiro, due to data confidentiality issues.

- Y(*output*): Value of production of cattle, swine, goats, equines, buffaloes, donkeys, asinine, mules, sheep, other birds, rabbits, apiculture, sericulture, ran culture, aquaculture, horticulture, flowers, forestry, agro industry, permanent crops, temporary crops, extractive activities; BRL (R$).
- Land (*xterra*): 4 percent of land expenses, the rent paid for the land; BRL (R$).
- Labor (*xtrab*): Salaries or other forms of compensation paid to family and hired laborers; BRL (R$).
- Capital (*xtec*): Machinery, improvements in the farm, equipment rental, value of permanent crops, value of animals, value of forests in the establishment, value of seeds, value of salt and fodder, value of medication, fertilizers, manure, pesticides, expenses with fuel, electricity, storage, services provided, raw materials, incubation of eggs and other expenses (value of permanent crops, forests, machinery, improvements on the farm, animals and equipment rental were depreciated at a rate of 6 percent over a number of years; varying according to the category); BRL (R$).

Variables explaining efficiency:

- Embrapa score: Expert perceptions of Embrapa's influence in meso-regions; Number – scaled from 1 to 3.
- Access to technical assistance: Whether farmer had access to technical assistance; Dummy-variable.
- Regions: Controls for the five different regions; Dummy-variables.
- Output strata: Controls for the 16 different strata; Dummy-variables.
- Indicator of whether farm had a net output<0: Given endogeneity of indicator variable, predicted probabilities were used; Probability 0-1.

3.2. Sampling

The sampling strategy adopted followed a stratified random sampling approach applied to two different agricultural censuses: 2006 and 1995-1996. For 2006, the sample contained 258,684 establishments out of a population of 4,614,030 farms. A slightly bigger sample of 284,923 was obtained for 1995-1996 out of a population of 4,722,101 establishments. Given that our interest lies mainly in understanding poor farmer's behavior and that most farmers are indeed poor, we adopted a proportional sampling technique as described in Cochran (1977). The sampling was based on both regional and gross output strata (which here we proxied to income to be able to classify farmers) (Alves et al., 2001, 2006, 2012). The allocation criteria took into account a standard deviation of R$ 50 to estimate the average gross output in the census of 1995-1996 and R$ 150 in the 2006 census, with a 95% probability. The reference for defining output brackets was the statutory minimum wage in force in Brazil. Given the rise in income that took place in the country and factors such as inflation, a minimum wage R$ 300.00 was adopted for 2006 and R$ 100.00 for 1995-1996.

The combined stratification of (5) regions and (3) output brackets led to a total of 16 strata, the last of which included rich (i.e. with high output values) farmers (which were grouped together regardless of the region). Across the five regions (North, Northeast, South, Southeast, Midwest), output brackets for 2006 were defined as follows: A - annual gross output in the range (0, 7,200.00], B - annual gross output in the range (7,200.00, 36,000.00], C - (36,000.00, 720,000.00]. Farmers were considered rich if their output exceeded R$ 720,000.00. Output brackets for the 1995-1996 census can be obtained by dividing the 2006 values by 3, since the minimum wage in 1995-1996 was 1/3 of that in 2006.

The sampling and stratification approach described above allow for the use of stochastic frontiers to analyze patters of efficiency within regions and across output brackets. However, stochastic frontier methods are fairly demanding on data, requiring that information on output as well as other inputs (labor, land and capital inputs) be not only valid but extensive. With this in mind the addition of the 16th (the rich) stratum becomes particular relevant, as this group accounted for nearly 28,000 establishments in 2006.

Combining the two censuses also required ensuring that a similar suite of variables were common in the two years. As a result, the samples for both census years were significantly reduced to 74,149 establishments in 2006 and only 15,477 farms, resulting in a total sample of 89,626.

An additional classification looked at farmer's specialization (whether crops, livestock or mixed) and was regarded and treated as a contextual variable in the econometric approach (Cochran, 1977). The specialization criteria were defined as follows:

1. crop area / total area > 0.5 – crop specialization;
2. pasture area / total area > 0.5 – livestock specialization;
3. both less than or equal to 0.5 – mixed specialization.

4. Econometric Results

We now move to the analysis of stochastic production frontier conducted on the Brazilian agricultural census data. The results we present here are derived from a total of five models estimated. The first model encompasses all producers for the year 2006, irrespective of their specialization, whether crop, livestock or mixed. Three subsequent models look at each individual specialty separately. And a final model adds the census data of 1995-1996 to the 2006 census data. The underlying goal of these estimations was to identify and understand the behavior of those most affected by Embrapa's technologies, in addition to providing insights into the different specialization, as well as the progress between the two census years.

For the global model, the production function takes the following form, as indicated previously:

$$\ln(y_j) = \beta_0 + \beta_1 \ln(xtrab_j) + \beta_2 \ln(xterra_j) + \beta_3 \ln(xtec_j) + \beta_4 D_{1j} + \beta_5 D_{2j} + \beta_6 D_3 + \beta_7 D_{4j} + v_j - u_j$$

where ln is the natural log, *y* represents gross output, *xtrab* are expenditures with labor, *xterra* are expenditures with land and *xtec* are expenditures with technological inputs for the *j*th producer. D1 to D4 are dichotomous dummy variables for four regions, which are compared against the Midwest. v_j and u_j represent the random error and inefficiency components of the model, respectively. The normal half-normal distribution was chosen as it best fits the data.

The random error component (v_j) is defined as a function of output classes and thus takes into account the heteroscedasticity present in the sample. This, therefore, controls the regional variability in the production function and in the output classes by explaining the variance of the error term.

The variance of the inefficiency component (u_j) can also be explained by contextual variables. Here, we chose the following variables: probability of negative net output (*p*), agricultural research efforts (*score*), access to technical assistance (*assitec*) and regional dummies. The probability of negative net output was estimated in a first-stage via a probit regression (results are not reported here, but are available upon request). This was done to control for endogeneity in the output equation. We, therefore, resorted to an instrumental variable approach, which regressed the dichotomous dummy variable indicating whether the farm had a negative net output against a set of exogenous variables. These variables included the inputs described above and the following variables, which in addition to serving as instruments are also relevant in explaining changes in technical inefficiency: output strata; farmer's experience; type of production (whether crops, livestock or mixed); education of head of household; age; family size; access to credit; access to cooperatives; technical assistance; whether the farmer was in urban or rural setting; whether the farmer rented or owned his farm. Then, fitted values of the estimated probabilities were added to the regression outlined above. As specified, the predicted probabilities therefore account for some of the usual explanatory variables that explain changes in technical efficiency. These variables were not added in the efficiency component due to lack of convergence that arose from correlation with the regional dummies. The analysis is conditional on *p*. The standard deviations of the estimators were calculated via the bootstrap technique available in Stata 11 (Stata, 2011), based on 1,000 replications.

A total of 74,296 observations were used in the estimation. As indicated in table 1, panel a, the global results were largely significant both for the deterministic part of the model and for the stochastic components. An indication of the good fit of the model can be obtained by correlating and squaring the observed and predicted values, which is by definition the R^2. We obtained an R^2 of 93%. All relevant effects for explaining the variation in technical inefficiency are significant and negative, indicating that technical inefficiency decreases as a given contextual variable increases. A logical implication of this is that technical efficiencies are increasing as a function of the defined variables.

The output strata are indicated by the variables est1-est15. These represent the three output brackets defined for each region. The regions are in the following order: North, Northeast, Southeast, South and Midwest. As these variables are defined as dummies, their results are relative to the "rich" farmers, i.e. those who obtained a gross output greater than 200 minimum wages per month. The regional dummy variables are reg_1-reg_4 and represent the North, Northeast, South and Southeast, respectively. The coefficients therefore represent differences between these regions and the Midwest region. All continuous variables are presented in natural logs.

Land elasticities presented in table 2 are considerably lower than elasticities of labor and especially technological inputs. This result has strong implications for technology diffusion and suggests that farmers that do not adopt technological inputs face a dire predicament in terms of output.

In terms of the variation in technical inefficiencies, the effects of technical assistance, probability of negative net output and the importance of Embrapa have the right and expected signs and, thus, reduce technical inefficiency. In addition to purely examining the coefficients, it is often helpful to look at the selected post-estimation results. Key among these is the average technical efficiencies, which are given by the expression in equation 5. Table 3 reports technical efficiencies by region and by output bracket along with two other important indicators: average scores of Embrapa's importance as defined in the previous section and the average probability of positive net output. Some important aspects emerge from table 3: higher output brackets generally observe a slightly higher score of the importance of Embrapa; the highest scores are found in the Northeast and in the Midwest, suggesting a fairly significant difference among regions in terms of Embrapa's relevance. As for technical efficiencies, the "rich" are by far the most efficient and efficiency for other classes grows in conjunction with output brackets (i.e. higher output bracket, higher efficiency). From a regional perspective, technical efficiencies are larger in the Northeast and the South and lower in the Midwest.

The next set of results presented in 8 panels b, c and d look at the different impacts of Embrapa for the three different types of specialization: crops, livestock and mixed. Marginally significant results for livestock and mixed are contrasted with a fairly significant one for crops. For livestock, technical assistance becomes non-significant. In terms of input elasticities, they tend to follow the general trend observed in the general model.

The inclusion of agricultural census data for 1995-1996 in the analysis leads to the results shown in table 4. Given the differences in the availability of variables between the two censuses, the joint model was slightly changed. A time dummy variable was added along with interactions of the time dummy (y2006) with inputs, as well as the perception of Embrapa's importance. The variable *score* was taken to be constant in the two periods. The intercept of the production function is negative, indicating softening of the technical component in the period. The elasticity of technological inputs is significantly higher in 2006, suggesting higher importance of the use of technological inputs to increase production. The score variable (*score*), technical assistance (*assitec*) and probability of negative net output (p) are statistically significant and act to reduce technical inefficiency. Positive interaction between the year dummy and the score variable, however, indicates an increase in technical inefficiency in the period between censuses.

Table 1: Production function estimation results for global model, crops, livestock and mixed (StdDv stands for Standard Deviation).

	Coefficient	StdDv	Coefficient	StdDv	Coefficient	StdDv	Coefficient	StdDv
	Global (a)		Crops (b)		Livestock (c)		Mixed (d)	
ly								
lxtrab	0.2102	0.0037	0.2345	0.0065	0.1332	0.0059	0.2035	0.0067
lxterra	0.0901	0.0029	0.1380	0.0055	0.0642	0.0037	0.0742	0.0048
lxtec	0.6399	0.0055	0.5856	0.0080	0.7126	0.0083	0.6727	0.0092
reg_1	0.0959	0.0792	0.5585	0.1226	-0.0942	0.0334	0.2966	0.0756
reg_2	-0.1064	0.0442	0.0577	0.0855	-0.1145	0.0289	-0.1461	0.0700
reg_3	0.0335	0.0253	0.1304	0.0541	0.0114	0.0271	0.0974	0.0694
reg_4	-0.0581	0.0317	-0.2527	0.0441	-0.0440	0.0295	0.0766	0.0657
_cons	2.2497	0.0466	2.4985	0.0798	2.0541	0.0643	2.0404	0.1046
lnsig2v								
est1	-1.1407	0.1186	-0.8359	0.5777	-1.1922	0.1803	-1.6584	0.1603
est2	-1.6203	0.0958	-1.0483	0.1886	-2.2108	0.1576	-1.8217	0.1050
est3	-0.7374	0.1061	0.1085	0.2294	-1.4519	0.1618	-0.6512	0.1395
est4	-0.8819	0.0458	-0.7738	0.0695	-0.9928	0.0672	-1.2989	0.0715
est5	-1.3773	0.0481	-0.8926	0.0703	-1.8395	0.0780	-1.7447	0.0688
est6	0.0951	0.0434	0.6007	0.0552	-0.3851	0.0761	-0.2400	0.0798
est7	-1.2383	0.1752	-0.6557	0.3165	-1.5812	0.2351	-1.8300	0.2137
est8	-2.4837	0.1033	-2.1078	0.2007	-2.7770	0.1100	-2.4073	0.1365
est9	-1.6268	0.0570	-1.3806	0.1000	-1.7948	0.0791	-1.5810	0.1189
est10	-1.9556	0.2095	-1.6446	0.2268	-2.7099	0.3100	-3.2403	0.4069
est11	-2.8076	0.0975	-2.6631	0.1096	-2.6766	0.0976	-3.1327	0.1015
est12	-1.7319	0.0486	-1.6003	0.0697	-1.7839	0.0829	-1.8571	0.0857
est13	-1.6244	12.0834	-0.5879	16.4775	-1.7399	3.0225	-3.4488	7.3275
est14	-2.4917	0.4411	-1.3088	14.2345	-3.0890	0.1748	-3.0790	0.3218
est15	-1.9668	0.0937	-1.5402	0.2166	-1.9391	0.1083	-2.3471	0.1833
_cons	1.0639	0.0217	0.8779	0.0314	1.0395	0.0385	1.3509	0.0373
lnsig2u								
p	7.5244	0.1891	5.8072	0.2516	10.9839	0.2413	7.2470	0.1781
score	-0.0594	0.0851	-0.1218	0.1147	-0.0441	0.0301	-0.0700	0.0800
assitec	-0.1165	0.0330	-0.2427	0.0502	0.0070	0.0454	-0.1339	0.0658
reg_1	0.2133	61.5350	1.3244	4.4339	-0.0034	0.0766	0.2032	0.1803
reg_2	0.1156	0.2064	1.3377	0.4411	-0.0732	0.0727	-0.3268	0.2846
reg_3	-0.0317	0.2346	0.7388	0.8607	-0.1362	0.0696	-0.1644	0.3027
reg_4	0.0364	0.2847	0.3510	0.1752	-0.0622	0.0728	-0.0779	0.2225
_cons	-5.2017	0.3230	-3.8314	0.3492	-8.8069	0.2505	-4.6294	0.2634

Table 2: Input elasticities.

Inputs	Elasticities	Confidence interval	Share *
Labor	0.210	(0.204; 0.216)	22.34
Land	0.090	(0.085; 0.095)	09.57
Capital	0.640	(0.631; 0.649)	68.09
Sum of coefficients	0.940	(0.920; 0.960)	100.00

* relative to the sum of elasticities

Table 3. Average technical efficiencies, Embrapa scores and predicted probabilities of having a positive net output by region and output bracket.

Region	Output	*te*	*score*	*1-p*
North	(0, 2]	0.326	1.51	0.192
	(2, 10]	0.571	1.586	0.441
	(10, 200]	0.671	1.608	0.534
Northeast	(0, 2]	0.519	2.770	0.340
	(2, 10]	0.734	2.776	0.608
	(10, 200]	0.846	2.768	0.780
Southeast	(0, 2]	0.236	1.877	0.084
	(2, 10]	0.438	1.747	0.259
	(10, 200]	0.699	1.821	0.526
South	(0, 2]	0.282	1.931	0.108
	(2, 10]	0.617	1.923	0.464
	(10, 200]	0.770	1.926	0.620
Midwest	(0, 2]	0.157	2.244	0.037
	(2, 10]	0.377	2.258	0.204
	(10, 200]	0.593	2.220	0.378
-	>200	0.864	2.137	0.786

Table 4. Estimation results for global model using both censuses (StdDv stands for Standard Deviation).

	Coefficient	StdDv		Coefficient	StdDv		Coefficient	StdDv
ly			lnsig2v			lnsig2u		
l*xtrab*	0.2627	0.0057	est1	-0.9847	0.0861	*p*	70.524	0.0709
l*xterra*	0.1550	0.0054	est2	-14.275	0.0728	*score*	-0.2733	0.0691
l*xtec*	0.4229	0.0059	est3	-0.5940	0.0890	*scoreint*	0.2252	0.0702
A	-0.0447	0.0068	est4	-0.8870	0.0341	*assitec*	-0.1617	0.0250
B	-0.0624	0.0058	est5	-12.048	0.0383	reg_1	0.2517	0.0561
C	0.2164	0.0072	est6	0.1790	0.0391	reg_2	0.2311	0.0505
reg_1	0.1181	0.0287	est7	-11.598	0.1558	reg_3	-0.0179	0.0460
reg_2	-0.0735	0.0229	est8	-22.724	0.0729	reg_4	0.0739	0.0442
reg_3	0.0189	0.0207	est9	-15.183	0.0517	y2006	-24.390	0.1739
reg_4	-0.0604	0.0185	est10	-18.140	0.1182	_cons	-23.154	0.1694
y2006	-14.494	0.0333	est11	-25.359	0.0484			
_cons	36.817	0.0334	est12	-17.360	0.0421			
			est13	-14.228	0.2374			
			est14	-23.128	0.3376			
			est15	-18.397	0.0801			
			_cons	0.8732	0.0204			

Table 5 provides a comparison of the two census in terms of inputs. Note that the column identified by "%" indicates the relative impact of each input in the total variation of gross output resulting from one percent increases in each input.

In explaining production increases, labor becomes less important as indicated by the significant decrease in its elasticity, shown in table 4. This suggests that agriculture became more mechanized in 2006. In a similar fashion, land, which was barely significant in 1995-1996, completely lost significance in 2006. The coefficient C reveals the substantial importance of technological inputs. The coefficients A, B and C, meaning l*xtrab*, l*xterra* and l*xtec* were added in table 4 and used to construct table 5.

Table 5. Comparison of input coefficients between 1995-1996 and 2006.

Variables	1995-1996		2006	
	Coefficients	%	Coefficients	%
Labor	0.263	31.3	0.210	22.3
Land	0.155	18.4	0.090	9.6
Capital	0.423	50.3	0.640	68.1
Total	0.841	100.0	0.95	100.0

5. Conclusion and Recommendations

This paper looked at the impacts of Embrapa on output of farmers by using a stochastic frontier approach. It did so by taking into account the important role technical efficiencies play in moving farmers ahead towards more efficient production, and in doing so being able to produce more with less or the same. This, in turn, makes them better off and able to obtain higher outputs from adequate input use.

Several key results are worth emphasizing. To start, the relative size of the estimated elasticities indicates the dominance of technological inputs over other inputs, which is also supported by the findings of Rada and Buccola (2012). This effect is exacerbated when the two censuses are combined. Technical assistance along with Embrapa contributes to the reduction of technical inefficiencies, thus increasing technical efficiencies, with the effect persisting even when 1995 is added.

However, the results by specialization – crops, livestock and mixed – vary slightly. For one, technical assistance becomes non-significant for the livestock group, while Embrapa remains marginally significant for livestock and mixed but highly significant for crops. Overall inefficiency has increased between the two censuses indicating that technical change has taken place in Brazil during that period and, therefore, the frontier is further out, as we would expect.

The predicted probabilities obtained from the probit regression, which instrumented the dummy variable indicating whether or not farmers had a negative net output, was also significant and of extreme importance in reducing inefficiencies. While this finding is likely to be obvious, it highlights the importance of controlling for farmers that are not capable of generating a sufficient or at least non-negative net output.

Across output groups, wealthier farmers tend to be more efficient than poorer farmers, which is consistent with the low use of inputs described in the descriptive section of the paper. Regional differences were also notable, with high efficiency levels for the Northeast and South and dismal levels for the Midwest. The other two regions fell in between.

What does all of this mean for Embrapa?

1. The censuses data of 1995-1996 and 2006 show that traditional inputs are no longer able to explain much of growth as previously, in large part due to the dominance of the technology effect (modern inputs). At the same time, production has become largely concentrated with few farms generating 51% of the gross output in 2006. As agriculture has grown tremendously in no small part due to technology advancements, it is clear that Embrapa has lagged in its effort to diffuse technology and probably extending technical assistance to more farmers across the country will be a significant challenge.
2. Technology is knowledge created by research and applied by producers through production systems. Thus, it seems evident that only a selected few larger farmers were able to fruitfully develop production systems that benefit from technology. Small scale

agriculture needs to be reassessed and refocused to be able access technology and become profitable.

3. Technical inefficiency increased between periods, as we indicated above. This evidence of technical change suggests that extension services have an even greater role to play in providing access to technology to millions of farmers. Embrapa's research and extension need to go hand in hand.
4. Technical assistance had a positive effect on technical efficiency. At the same time, negative net outputs are strongly associated with inefficiency. This suggests that the technical assistance being provided is outweighed by the lack of financial and managerial skills of farmers. These skills have thus far not been provided by technology through technical assistance.

6. References

Aigner, D., C.A.K. Lovell, and P. Schimidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, vol. 6: 21-37.

Alston, J. M., T.J. Wyatt, P.G. Pardey, M. C. Marra, and C. Chan-Kang. 2000. *A Meta-Analysis of Rates of Return to Agricultural R&D: Ex Pede Herculem?*. IFPRI Research Report 113. Washington, D.C.: International Food Policy Research Institute.

Alves, E., G.S. Souza, and A.S.P Brandão. 2001. A Situação do Produtor com Menos de 100 Hectares. *Revista de Política Agrícola*. vol. (1): 27-36.

Alves, E., G.S. Souza, and C.A.V. Oliveira. 2006. Desempenho dos Estabelecimentos do PRONAF. *Revista de Política Agrícola*. vol. 15(4): 5-23.

Alves,E., G.S. Souza, and D.P. Rocha. 2012. Lucratividade na Agricultura. *Revista de Política Agrícola*. (12).

Belloumi, M. and M.S; Matoussi. 2006. A Stochastic Frontier Approach for Measuring Technical Efficiencies of Date Farms in Southern Tunisia. *Agricultural and Resource Economics Review* vol. 35(2): 285-298.

Christiaensen, L. and L. Demery. 2007. *Down to Earth: Agriculture and Poverty Reduction in Africa*. Washington, D.C.: World Bank, 2007.

Cochran, W.G. *Sampling Techniques 3rd ed*. 1977. New York: John Willey. 428p.

Coelli, T.J., D.S.P. Rao, and C.J. O'Donnel. 2005. *An Introduction to Efficiency and Productivity Analysis*. 2nd ed. New York: Springer. 367p.

Almeida, O. A. 2009. *Exports, Energy, Food: The Multiple Functions of Brazilian Agriculture*. CeSPI Working Papers 61/2009. Rome: Centro Studi di Politica Internazionale.

Fan, S. and X. Zhang. 2008. Public Expenditure, Growth and Poverty Reduction in Rural Uganda. *African Development Review/Revue Africaine de Développement* vol. 20 (3): 466-496.

Fan, S.; B. Yu, and S. Jitsuchon. 2008. Does Allocation of Public Spending Matter in Poverty Reduction? Evidence from Thailand. *Asian Economic Journal* vol. 22(4): 411-430.

Greene, W.H. 2011. *Econometric Analysis*. New York: Prentice Hall. 1232p.

Kumbhakar, S.C., R. Ortega-Argiles, L. Potters, M. Vivarelli and P. Voigt. 2012. Corporate R&D and Firm Efficiency: Evidence from Europe's Top R&D Investors. *Journal of Productivity Analysis*. vol. 37(2): 125-140.

Khumbhakar, S.C., and C.A.K. Lovell. 2000. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press. 344p.

Nin-Pratt, A. and E. C. Magalhães, E. (in preparation). The impact of seed programs on Ethiopia's farmers efficiency.

Pardey, P.G. , J. M. Alston, C. Chan-Kang, E. C. Magalhães, and S.A. Vosti. 2006. International and Institutional R&D Spillovers: Attribution of Benefits Among Sources for Brazil's New Crop Varieties. *American Journal of Agricultural Economics*. vol. 88 (1): 104-123.

Rada, N.E. and S.T. Buccola. 2012. Agricultural policy and productivity: evidence from Brazilian censuses. *Agricultural Economics*. vol.43(4): 355-367.

Stata. 2011. *Stata 12 Base Reference Manual Set - Frontier - Stochastic Frontier Models*. College Station, TX: Stata Press. 586p.

Thirtle, C., L. Lin, and J. Piesse. 2003. The Impact of Research-Led Agricultural Productivity Growth on Poverty Reduction in Africa, Asia and Latin America. *World Development.* vol. 31(12): 1959-1975.
Vicente J. R. 1999. Eficiência na Produção Agrícola Paulista e Seus Determinantes. *Economia Aplicada/Brazilian Journal of Applied Economics*. vol. 3 (2): 263-287.
Zhang, Y., X. Wang, T. Glauben, and B. Brummer. 2011. The Impact of Land Reallocation on Technical Efficiency: Evidence from China. *Agricultural Economics*. vol. 42 (4): 495-507.