Farm Size and the Determinants of Productive Efficiency in the Brazilian Center-West

Steven M. Helfand

Department of Economics
University of California
Riverside, CA 92521 USA

shelfand@mail.ucr.edu

(909) 827-1572

Contributed paper selected for presentation at the 25th International Conference of Agricultural Economists, August 16-22, 2003, Durban, South Africa

Farm Size and the Determinants of Productive Efficiency in the Brazilian Center-West

Abstract: This paper explores the determinants of technical efficiency, and the relationship between farm size and efficiency, in the Center-West of Brazil. This is the region where agricultural production and total factor productivity have grown the fastest since 1970. It is also a region characterized by unusually large farms. Technical efficiency is studied with Data Envelopment Analysis and county level data disaggregated by farm size and type of land tenure. The efficiency measure is regressed on a set of explanatory variables which includes farm size, type of land tenure, composition of output, access to institutions, and indicators of technology and input usage. The relationship between farm size and efficiency is found to be non-linear, with productivity first falling and then rising with size. Access to institutions, credit, and modern inputs are found to be important determinants of the differences in efficiency across farms. Improved access could strengthen the efficiency advantage of small and medium farms.

Key Words: Productivity, Efficiency, Brazil, Data Envelopment Analysis.

1. Introduction

The majority of studies of agricultural productivity in developing countries support the

view that there is an inverse relationship between productivity and farm size.[1] If correct, land

reform could contribute to improving both equity and efficiency in agriculture. Most of these

studies, however, are based on partial measures of productivity such as yield which are biased in

favor of small producers. It is likely that the inverse relationship would be less pronounced, or

perhaps even reversed, if a measure of total factor productivity (TFP) were used instead. It has

also been suggested that the inverse relationship might weaken in a region characterized by rapid

modernization. This paper explores the relationship between farm size and technical efficiency

in exactly this type of environment. The Center-West (CW) is the region of Brazil where

production and TFP have grown the fastest since 1970.[2] It is also a region characterized by

unusually large farms: the average farm size in the CW is about six times the national average.[3]

---

[1] See Berry and Cline (1979) for one of the early, and most important, contributions to this literature, and Barrett (1996) for a more recent analysis.

[2] Gasques and Conceição (2001) and Vicente et al. (2001a) both estimate annual TFP growth rates for the CW to be in the 4%-5% range for the period 1970-95, or about twice the national average.

[3] Whereas the median farm size class and the mean farm size for Brazil as a whole in 1995 were 10-20 and 73 hectares respectively, in the CW they were 50-100 and 448.

A second objective of this paper is to extend the recent Brazilian literature on total factor productivity growth in agriculture to an analysis of the determinants of productivity. I do this in two ways. First, most recent studies have limited themselves to the measurement of TFP. Only one has sought to analyze the determinants of TFP in an econometric framework (Vicente et al. 2001b). In this paper I regress a measure of technical efficiency on a set of explanatory variables which includes farm size, type of land tenure, composition of output, access to institutions, and indicators of technology and input usage. Second, previous studies have been based on national or state level data. These have provided a valuable first approximation to productivity change in the past several decades. The Brazilian agricultural sector, however, is far too heterogeneous for us to be satisfied with studies conducted at this level. In this paper, I use data from 426 counties, 15 farm size classes, and four types of land tenure. This permits me to avoid many of the issues of aggregation bias present in previous studies, and generates a far richer set of data for studying the determinants of efficiency.

The paper is organized as follows. Section 2 briefly presents the Data Envelopment Analysis methodology that is used to estimate technical efficiency. Section 3 describes the data and the construction of the variables. Section 4 analyzes the empirical results, with an emphasis on farm size, and section 5 provides conclusions.

2. Methodology

Data Envelopment Analysis (DEA) is used in this paper to calculate productive efficiency (Färe et al., 1994).[4] Efficiency is defined in a relative sense, as the distance between observed input-output combinations and a best practice frontier. DEA is one of several techniques that can

be used to calculate a best practice production frontier (Coelli et al. 1997, Kumbhakar and Lovell 2000). Each approach has its advantages and disadvantages. In the present context in which I am working with spatial data derived from 426 counties, DEA seems preferable to a stochastic frontier approach. The econometric theory has yet to be developed for incorporating spatial correlation into a stochastic frontier model.

Output distance functions are used to characterize the frontier of a multiple-input multiple-output production technology, and the proportional distance of each observation from the frontier. The following notation is used. The production possibilities set P is the combination of all pairs of inputs x and outputs y that are feasible, where x and y are vectors. Inputs and outputs are assumed to be freely disposable, and P is assumed to be nonempty, closed, and convex. The output distance function D_o is

$$D_o(x, y) = \inf \left\{ \theta : \left(x, \frac{1}{\theta} y \right) \in P \right\} \tag{1}$$

where θ is a non-negative scalar that measures the ratio of the observed vector of outputs to the maximum vector that could be achieved, given the input vector, if all outputs were expanded proportionally. The inverse of the output distance function, the Farrell output-oriented measure of technical efficiency, is used here as a measure of efficiency. The Farrell measure equals one for efficient firms on the frontier, and then increases with inefficiency.

Farrell efficiency measures are found as the solution to a linear programming problem under alternative assumptions about returns to scale. With K farms, N inputs, M outputs, and the assumption of constant returns to scale (CRS), the following linear program must be solved for every observation:

---

[4] Due to page length limitations, the methodology is described only briefly. The interested reader can find a more detailed treatment in the cited references.

$$[D_o(x,y)]^{-1} = \max_{(z,\theta)} \theta$$

$$s.t. \quad \theta_k y_{mk} \leq \sum_{k=1}^{K} z_k y_{mk} \qquad m = 1,...,M \qquad (2)$$

$$\sum_{k=1}^{K} z_k x_{nk} \leq x_{nk} \qquad n = 1,...,N$$

$$z_k \geq 0 \qquad\qquad\quad k = 1,...,K$$

where the z_k are variables which show the intensity with which each farm is used in order to construct the frontier of the production possibilities set. The linear program solves for the maximum value of θ given the constraints that the proportionally expanded vector of outputs and the vector of inputs are in the feasible set, and that the intensity variables are non-negative.

Two caveats are in order. First, the efficiency scores may be very sensitive to measurement error for the farms that define the best practice frontier. Elsewhere, I have explored this issue by removing the 0.5% most efficient observations and re-estimating the frontier. The results across farm sizes remained qualitatively unchanged. Second, it is important to emphasize that the term "inefficiency" is being used here solely to mean the distance between a given farm and the observed best practice frontier. There are many reasons why a farm might not be operating on the best practice frontier, and most of these are unlikely to relate to x-inefficiency. Excluded variables such as land quality, market failures that lead to non-separable household decisions, credit market constraints that lead farms to choose input-output bundles that appear inefficient in relation to unconstrained farms, and different vintages of technology are all possible reasons why rational farmers might not be operating on the frontier. The challenge, it seems to me, is to identify the relative importance of these alternative sources of "inefficiency."

3. The data

The data for this study come from the 1995/96 Agricultural Census in Brazil. For the DEA analysis, I use aggregate output and five inputs. The construction of these variables is briefly explained here. The variables used in the second stage to study the determinants of efficiency come from the same source and are explained jointly with the econometric analysis.

Output (Y): The output variable is defined as the gross value of agricultural output net of three categories of items. First, in an effort to avoid the double counting of animals that takes place when animals are purchased and sold at different stages of the production process, I deducted the value of purchases of cattle, hogs, chickens, and fertilized eggs. Cattle, hogs, and chickens account for over 99% of the value of animal production in the CW. Second, I deducted the value of production of "rural industry," a category which accounts for 2% of the value of agricultural production in the CW. Preliminary estimates revealed that for most products the value added in rural industry was extremely close to zero. Because of data limitations, it was much simpler to exclude rural industry than estimated intermediate inputs. Finally, I excluded the value of "extractive products" and of "forestry products" to be consistent with my decision to exclude forest and woodland areas. These categories of goods only accounted for 1.6% of the value of agricultural output in the region, yet represented 29% of the utilized area.

Utilized area (X1): 90% of the value of output in the CW comes from temporary crops and cattle. With this in mind, I constructed the area variable to include natural and permanent pastures, land utilized for crops, and productive land that was not being used. These categories accounted respectively for 58%, 7%, and 2% of the utilized land. The excluded categories were forest and woodland (29%), and unusable land (4%).

Labor (X2): First, I constructed a variable for family labor by treating males and females as equals, and by only counting half of the family labor under 14 years of age. Second, I constructed a variable for hired labor along the lines of Guanziroli et al. (2001). These authors devised an approach to correct for the fact that a) temporary workers should not be counted as full-time equivalent workers, and b) many farms, especially the large ones, hire contracting firms in an effort to avoid paying social security and other labor taxes. Finally, I added family and hired labor together in order to construct the labor variable in full-time equivalent units.

Tractors (X3): Tractors are measured in the equivalent of a 75 horsepower tractor, which was the midpoint of the modal horsepower class (50-100hp).

Animals (X4): Many studies of TFP have not used animals as an input, even when they were considered part of output. I aggregated animals based on their relative prices in the CW of Brazil. The stock of animals in cattle equivalents was then used as a proxy for the stock of capital in animal production.

Purchased inputs (X5): An input variable was created based on the expenditures for a) fertilizer, b) chemicals (such as pesticides and herbicides), c) seeds, d) fuel, and e) feed and medicine for animals.

A final step in the construction of the variables involved creating "representative farms" for each farm size, of each type of land tenure, in each county. This was necessary because I did not have access to farm level data. I used data from all 426 counties in the CW (excluding the Federal District) that was aggregated into 15 size classes and 4 types of tenure. After removing unusable observations, the final dataset covered 237,595 establishments aggregated into 9,304 representative farms, implying an average of 25 establishments per representative farm.

4. Empirical results

The median value of output per hectare (Y/X1) and the median value of output per unit of labor (Y/X2) were calculated for each farm size class. The data were consistent with a broad body of international evidence on the relationship between farm size and productivity. On the one hand, there is a strong inverse relationship between value per hectare and farm size, reflecting the intensive use by small farms of their scarce factor of production--land. Labor productivity, in contrast, is much higher on large farms where the opportunity cost of labor is greater. The relationship between total factor productivity (or technical efficiency under CRS) and farm size is influenced by the partial productivities of land, labor, and all other inputs.

The technical efficiency scores by farm size were calculated according to the methodology in Section 2 and the assumption of constant returns to scale. The efficiency scores were then regressed on different combinations of explanatory variables as reported in Table 1. All six regressions were estimated with county level fixed effects to take account of spatial heterogeneity, such as differences in soil quality, that was not captured in the regressors. The equations were also estimated with a GLS procedure that allowed for heteroscedasticity across the 426 counties. Regression (6) went one step further and allowed for spatial correlation in the errors across counties using a SUR framework.[5] In practice, the estimation techniques mattered little and the estimated coefficients from all of the regressions were quite robust to alternative specifications and to the inclusion/exclusion of different sets of variables.

---

[5] The estimation was conducted as if the data set contained a pooling of 426 counties and 60 "time periods." The time periods referred to the 60 combinations of farm sizes and types. Thus, in this context "contemporaneous" observations were those that were of the same size and type in different counties. While this approach is more restrictive than the standard spatial econometric approach because it limits the estimation of spatial relationships to observations of the same size and type, it is also more general than the standard approach because it does not impose an explicit notion of neighborhood (i.e., zeros and ones in a W matrix) and allows for non-zero spatial correlation across all pairs of counties.

Regression (1) estimated the unconditional relationship between the log of technical inefficiency and the log of farm size. Quadratic and cubic terms were added as well to capture non-linearities that were observed in a prior graphical analysis of the results. All three coefficients were statistically significant at least at the 1% level. In Figure 1 I have plotted the curve "size" based on the coefficients from regression (1). The curve depicts a non-linear relationship between inefficiency and farm size, with inefficiency first rising and then falling as farm size grows. The modal farm size class of 20-50 ha might be considered as a target size for land reform in this region, and for this reason has its level of inefficiency set equal to one in the Figure. Relative to this group, which accounts for 21% of the farms in the region, farms in the 200-1000 range are estimated to be about 23% less efficient. While farms smaller than 20 ha are even more efficient, the analysis here will focus on farms above 20 hectares which represent 78% of the farms in the region and more then 95% of all variables other than labor.[6]

In addition to farm size, the next five regressions controlled for differences in efficiency due to land tenure status. Dummies were included for renters, sharecroppers, and occupants to differentiate them from landowners. The results indicate that renters were slightly more efficient than owners, while occupants and sharecroppers (in all cases but one) were less efficient. It is likely that renters were more efficient in this region because they were a more homogenous group of large, market oriented farmers relative to owners who were the majority. Although statistically significant, the impact of differences in land tenure on efficiency was generally only in the 2-4% range. The coefficients on size were essentially unchanged.

In addition to land tenure, regression (3) controls for differences in the composition of output. Relative to the excluded category, cattle, producers that specialized in the higher value

---

[6] Some authors, such as Rezende (2002), argue that small family farms are not viable in the CW due to the long drought period and the relative scarcity of off farm employment.

products (temporary crops, horticulture, and permanent crops) had lower levels of inefficiency. A coefficient equal to –1 implied that a one percentage point increase in the share of production coming from temporary crops, for example, translated into a 1% increase in technical efficiency. Producers of hogs and chickens were little different than those of cattle, and producers of "other" products were less efficient. Based on the coefficients from this regression, the conditional relationship between size and inefficiency is graphed in Figure 1 as "composition of output." Because cattle production accounted for nearly half of the value of output for farms over 50 ha, but less than 25% for farms under 10 ha, controlling for these differences reduces the efficiency disadvantage of large producers. Thus, the "composition" curve is higher than the "size" curve below the 20-50 ha class and lower than it above this size. The advantage of farms in the 20-50 ha class over farms in the 200-1000 ha class drops to a maximum of 18%.

In addition to the above mentioned variables, regression (4) incorporates variables intended to capture access to institutions and public goods. Thus, the variables credit, electricity, technical assistance, and cooperatives all measure the share of establishments in each representative farm with access to these items. These variables range between zero and one. Market access, in contrast, is intended to capture the degree to which market oriented farms are likely to be more efficient than farms producing for their own consumption. This variable is measured as the value of sales divided by the value of production. All five variables had a statistically significant impact on inefficiency. Access to credit, electricity, and technical assistance had the largest impact, while market orientation had the smallest. Once again, the conditional relationship between size and inefficiency was plotted in Figure 1 based on the coefficients from this regression. The curve "institutions" shifts up substantially relative to the previous scenario, reflecting the fact that large farms tended to have preferential access to these

institutions, and controlling for this advantage would increase the relative inefficiency of these farms. Thus, farms in the 1000-2000 range were now 45% less efficient than farms in the 20-50 ha class, or more than twice what was observed in the base case. This is powerful evidence of the importance of providing small farms with access to these institutions.

Regressions (5) and (6) control, in addition, for a variety of factors related to the level of technology and the use of inputs. The only difference between the two is that (6) uses the spatial SUR approach that was described above. The coefficient estimates are almost identical, and in what follows I focus on (5). All variables are measured as the share of establishments that report using the designated item. Thus, an increase in the share of farms using machines in production (rather than just human labor or animals), irrigation, fertilizers, pest and disease control, or soil conservation, all contribute to reducing inefficiency. The use of mechanical milking has the largest effect of all of these variables. However, the size of this effect is likely due to the correlation between the use of milking machines and specialization in the production of milk--a relatively high value activity--rather than a pure effect attributable to the use of this technology. The curve "technology/inputs" in Figure 1 shows the relationship between size and inefficiency based on the coefficients from regression (5). The curve shows that if we could also control for the differences in the use of technology and inputs, the relative inefficiency of large farms would be even greater than in the previous scenarios. The relative disadvantage of some farm sizes rises to nearly 60% in this case. However, due to the indivisibilities that exist with certain technologies, such as tractors, providing small and medium farmers with greater access to these inputs might require institutional innovations such as the development of rental markets for the services provided by these inputs.

5. Conclusions

In this paper I used Data Envelopment Analysis (DEA) to estimate the technical efficiency of farms in the Center-West of Brazil, and then studied the determinants of efficiency with regression techniques. Future research should explore the robustness of the results presented here by comparing them with estimates from a stochastic frontier production function.

There are important policy implications that can be derived from the analysis in this paper. The results indicated that access to credit institutions and goods that are often provided by the public sector, such as electricity and technical assistance, were among the most important determinants of differences in efficiency. Other important determinants included the use of inputs such as irrigation and fertilizers, and differences in the composition of output. These results identify the types of policies and production practices that would contribute to increased technical efficiency in this region.

The relationship between farm size and technical efficiency was also studied and it was shown to be more complex than what is normally believed. Rather than an inverse relationship, where productivity falls as farm size rises, a U-shaped relationship was found. For farms up to about 200 hectares, efficiency did fall as farm size rose, but beyond this size it started to rise again. The most important reasons why the inverse relationship broke down relate to preferential access by large farms to institutions and services that help lower inefficiency (such as credit, technical assistance, and rural electricity) as well more intensive use of the technologies and inputs that raise productivity. If one could create an environment in which small to medium size farms (20-200 ha, for example) had equal access to productivity enhancing institutions, and improved access to modern technologies and inputs, then these farm could still produce more efficiently than farms in the 2,000-20,000 ha range. Thus, even in the Center-West of Brazil, a

region characterized by extremely large farms and relatively high levels of technology, land reform continues to provide the possibility of simultaneously improving equity and efficiency. Its success, however, is strongly conditioned by the complementary institutions, investments, and services that permit small and medium size farms to compete on a level playing field.

References

Barrett, C. B. 1996. "On Price Risk and the Inverse Farm Size-Productivity Relationship," *Journal of Development Economics* 51(2):193-216.

Berry, R Albert, and William Cline. 1979. *Agrarian Structure and Productivity in Developing Countries*, Baltimore: Johns Hopkins University Press.

Coelli, Tim, D.S. Prasada Rao, and George E. Battese, *An Introduction to Efficiency and Productivity Analysis*, Norwell: Kluwer Academic Publishers, 1999.

Färe, Rolf, Shawna Grosskopf and C.A. Knox Lovell, *Production Frontiers*, Cambridge: Cambridge University Press, 1994.

Fulginiti, Lilyan E., and Richard K. Perrin, "LDC Agriculture: Nonparametric Malmquist Productivity Indexes," *Journal of Development Economics*, Vol. 53, 1997, pp. 373-90.

Gasques, J.G. and J.C.P. Conceição, "Transformações Estruturais da Agricultura e Produtividade Total Dos Fatores," in *Transformações da Agricultura e Políticas Públicas,* Brasília: IPEA, 2001.

Guanziroli, Carlos, Ademar Romeiro, Antônio M. Buainain, Alberto Di Sabbato, and Gilson Bittencourt, *Agricultura Familiar e Reforma Agrária no Século XXI*, Rio de Janeiro: Editora Garamond Ltda., 2001.

Helfand, Steven M., and Luis P. Brunstein, "The Changing Structure of the Brazilian Agricultural Sector and the Limitations of the 1995/96 Agricultural Census," *Revista de Economia e Sociologia Rural,* Vol. 39, No. 3, July/Sept. 2001, pp. 179-203

Instituto Brasileiro de Geografía e Estatistica (IBGE), *Censo Agropecuário 1995-96,* Rio de Janeiro: IBGE, 1998.

Kumbhakar, Subal C., and C.A. Knox Lovell, *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press, 2000.

Rezende, Gervásio Castro de, "Agricultural Growth and Agrarian Structure in the *Cerrado* Regions of Brazil: The Role of Land Prices, Natural Resources and Technology," mimeo, 2002.

Vicente, José R., Lilian C. Anefalos, and Denise V. Caser, "Produtividade Agrícola no Brasil, 1970-1995," *Agricultura em São Paulo, SP*, Vol. 39, No. 2, 2001a, pp. 1-23.

-----, "Produtividade Total de Fatores na Agricultura Brasileira: Influência de Capital Humano, Insumos Modernos e Recursos Naturais," mimeo, 2001b.

Table 1
Regression Results for the Determinants of Technical Efficiency

Dependent Variable: ln(inefficiency)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Farm size						
Size	0.49	0.50	0.43	0.43	0.45	0.45
Size^2	-0.05	-0.05	-0.04	-0.03	-0.03	-0.03
Size^3	0.00	0.00	0.00	0.00$^\#$	0.00$^\#$	0.00$^+$
Land tenure (relative to owners)						
Renter		-0.33	-0.06	-0.12	-0.10	-0.08
Sharecropper		-0.13	0.16	0.10	0.13	0.14
Occupant		0.17	0.21	0.15	0.16	0.19
Composition of output (relative to cattle)						
Temporary crops			-1.12	-1.00	-0.79	-0.76
Horticulture			-1.34	-1.22	-0.76	-0.59
Permanent crops			-1.15	-1.12	-1.04	-0.97
Hogs and chickens			-0.11$^\#$	-0.06$^\#$	-0.04$^\#$	-0.05
Other			0.97	0.85	0.86	0.88
Access to institutions/public goods						
Credit				-0.51	-0.43	-0.43
Electricity				-0.37	-0.29	-0.31
Technical assistance				-0.33	-0.22	-0.23
Cooperatives				-0.15	-0.10	-0.08
Market access				-0.06	-0.05	-0.02
Technology and inputs						
Machines					-0.12	-0.11
Irrigation					-0.30	-0.37
Fertilizer					-0.26	-0.29
Pest and disease control					-0.17	-0.14
Soil conservation					-0.23	-0.29
Mechanical milking					-0.83	-0.81
Artificial insemination					0.02$^\#$	0.08
Adjusted R^2	0.87	0.87	0.89	0.90	0.90	---

Notes:
1) All regressions have county fixed effects (not reported) and allow for heteroscedasticity across counties using a GLS procedure. Regression (6) also uses a spatial SUR. See text for details.
2) All coefficients are significant at least at the 1% level unless designated with the following symbols:
   + = statistically significant at the 5% level.
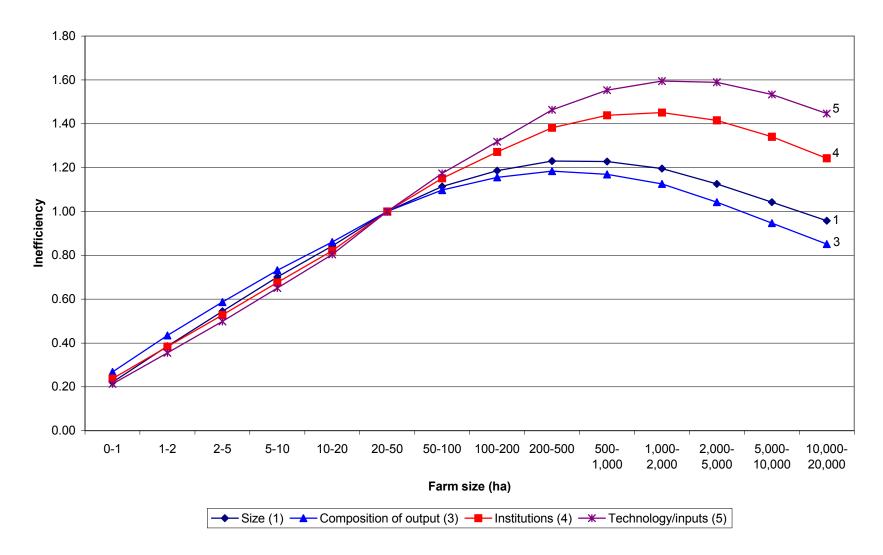   # = not statistically significant at the 5% level.

Figure 1: Conditional Effect of Farm Size on Inefficiency