UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA APLICADA

SUGARCANE, LAND USE AND REGIONAL DEVELOPMENT

André L. S. Chagas
Rudinei Toneto Júnior
Carlos Roberto Azzoni

TD. 009/2011
Programa de Pós-Graduação em Economia Aplicada FE/UFJF

Juiz de Fora

2011

Sugarcane, land use and regional development

André L. S. Chagas[1]
Rudinei Toneto Jr[2]
Carlos Roberto Azzoni[1]

The growing demand for renewable energy sources to replace oil has caused a great expansion of bioenergy, like ethanol of sugarcane. Despite the apparent environmental benefits of using ethanol to replace fossil fuels, to expand production of sugarcane raises other concerns, such as (i) reduction of food supply, due to the substitution of crops, like soybean, (ii) deterioration in potential in social conditions of the producing areas. The expansion of cane sugar occurs in replacement the existing crops, like soybeans, and especially cattle. So, the increase in sugarcane area could press the price of food, reducing the supply of crops and with worrying consequences on the poor. Furthermore, the sugarcane production is characterized by work relationships with limitations when compared with similar relationships in industrial sector, not to mention potential impacts on deforestation and the environment. Despite the strong criticism to the sector's expansion, recent results suggest that its effects are more beneficial than adverse, suggesting that technological advances in food production more than offset the expansion of sugar cane, and that the sugarcane production tends to contribute to the socio-economic development of the producing regions, accounting to comparative advantage for producing regions and to Brazil. This paper aims to discuss the effects that the production of cane sugar has on land use and regional development of regions.

[1] Economics Departament of FEA-USP.
[2] Economics Departament of FEA-USP-RP.

1. Introduction

The growing demand for clean energy sources to replace petroleum has been causing substantial expansion of agro-energy (FAO, 2008) to increase production of fuels, such as ethanol and biodiesel, from agricultural products. These fuels can be obtained from different raw materials. The main ones in the case of ethanol are sugarcane and corn, while for biodiesel the main sources include soybeans, canola (rapeseed), dendê (palm nuts) and castor beans, among others. In Brazil, sugarcane is the principal products to produce ethanol, but to can also be used to produce sugar and other byproducts, as well as for cogeneration of energy.

It is evident and increasing the concern with the destination of sugarcane production to produce energy at the expense of food production. This concern also embraces other forms of production. After all, the land use for energy production potentially reduces the availability of land for the production of food, which presses the price of these, with consequences for the poorest people.

Brazil is highly competitive in the production of grains and sugarcane. The country is the second largest producer soybeans and the major producer of sugarcane of the world. It is responsible for over half the world's ethanol exports and 40% of the sugar exported (F.O. LICHT, 2003).

Despite the environmental benefits from using ethanol from sugarcane instead of fossil fuels, the expansion of sugarcane growing causes some concerns. The first of these is the land use dynamic, since this expansion tends to displace other activities, such as stock breeding, grain and fruit growing (including soybeans), besides pushing the agricultural frontier into native forests. This can cause: (i) reduced supply of foods, the so-called energy versus food conflict; and (ii) displacement of previous activities into frontier areas, leading to deforestation. These points must be considered in the final analysis of the environmental effect of using ethanol (FAO, 2008). Some studies[3] show that Brazil has ample unused tillable lands that can be utilized to grow cane without the need to displace other crops or provoke deforestation, and there are also various areas of degraded pasturage whose use to grow cane would create positive environmental impacts. Even if this situation of land availability is valid for the country as a whole, in the current areas where sugarcane is expanding, it tends to compete for space with existing crops such as oranges and grains, or agro-industrial systems such as stock

[3] For example, see Macedo and Nogueira (2004).

breeding and meat packing. In other words, the expansion of cane growing causes pressures and disputes over land use, especially in regions nearer markets and with better logistics infrastructure. In perspective of the producer regions, would be interesting replace the production of grains for the production of cane sugar?

In this paper, we aim to investigate whether the expansion of sugarcane was favorable to the economic growth of regions. In particular, we seek to determine whether the regions in which there was expansion of sugarcane production grew faster than the other regions. In order to give the comparison between comparable locations, a propensity score matching model space (CHAGAS ET AL., 2011) is employed, as described in the following section.

The next section briefly describes the land use in farm area in Brazil, focuses grains and sugarcane production, the most important temporary agricultural culture of the country. The third section presents the methodology used to identify the possible impacts of growing sugarcane on the social conditions of producing regions. The fourth section presents the results, and the fifth section contains the conclusions.

2. Land use, grains production and sugarcane

In the last two decades, Brazil has advanced in the production of both grains and sugarcane. In early 1990, grain production area occupied more than 35 million hectares, corresponding to about 68% of over 50 million hectares in farming areas. In the same period, the area planted to sugarcane was little more than 4 million hectares, or slightly more than 8% of the total crop area of the country (Table 1). In recent years these numbers have increased to 45 million hectares of planted area of grains, and nearly 9 million hectares in sugarcane. Together these crops account for 80% of the total 65.6 million hectares of crops in areas of the country.

TABLE 1 HERE

Note that the two types of culture showed average growth, when considering the extreme years, despite the decline observed during the nineties. It is also interesting to note that the expansion of both grain and sugarcane was greater than the expansion of agricultural area, which means a reduction in the area of permanent crops (fruits and coffee, for example) in favor of temporary crops (grains and sugarcane).

In spatial terms, the cane is concentrated in the southeast of the country, especially the state of Sao Paulo. Grain production, in turn, is more scattered across the country, with emphasis on the participation of southern and central-west of the country (Figure 1).

Figure 1 here

The expansion of sugar cane is more significant. The acreage of this crop has grown, on average, nearly three times the rate of expansion of agricultural area, especially in the last ten years. The expansion of sugarcane has concentrated on South-Central region, highlighting the state of São Paulo, responsible for 60% of the amount produced from cane sugar, and 87% of the country's southeastern region. But, in the most recent period, the expansion occurred in the Midwest, region of the frontier for the sugarcane and the most important country's region for the production of the grains (Figure 2).

Figure 2 here

3. Methodology

3.1. Spatial Propensity Score Matching

We use the spatial propensity score matching (CHAGAS ET AL., 2011) to estimate the effect of growing sugarcane on the GDP per capita of producing regions. Let $D_i = 1$ if the region is on the treatment group, like the group of regions where the sugarcane expansion, for example, and $D_i = 0$ if not. The probability of one region to belonging to one group is affected by factors (X_i), such as the incidence of rain precipitation. From a farmer's point of view, growing sugarcane in a determined place can be interpreted as his/her best response, given the choices available. It is very likely that having other growers nearby can influence his/her decision. This fact introduces a selection bias in comparing regions with different sets of possibilities, and hence different best responses (or, at least, observed responses). The role of the propensity score is to relax these spatial effects. In other words, the spatial dimension to the problem is latent, and the introduction of spatial controls is a necessary precondition for the correct identification of the effects of interest.

The propensity score method was introduced by Rosenbaum and Rubin (1983). Their method controls for the selection bias of different individuals receiving the treatment by estimating the probabilities of receiving treatment, given some observed variables. This probability, $\Pr(D_i = 1 \mid X_i)$, is called the propensity score. Individuals with similar probabilities of receiving the treatment are grouped, so that the result is conditionally independent of whether or not the individual received the treatment, or

$$(Y_0, Y_1) \perp D \mid X \qquad (1),$$

where Y is the result of interest, D is the treatment, and $D \in (0,1)$ and X are covariates. The aim is to estimate the average effect of the treatment on the treated, that is

$$E[(Y_0, Y_1) \mid D = 1, X] = E[(Y_1 \mid D = 1, X) - (Y_0 \mid D = 0, X)] \qquad (2)$$

The value of the counterfactual effect of no treatment on the treated, $E(Y_0 \mid D = 1, X)$, is approximated by the average result of the self-selected group of untreated individuals $E(Y_0 \mid D = 0, X)$ (Heckman; Ichimura; Todd, 1998). Instead of using various conditional covariates, we use the propensity score $P(X) = \Pr(D = 1 \mid X)$, that is, the probability of belonging to the group of cane growing regions, given some determined observed characteristics. This probability is not a random variable, since it is influenced by spatial factors such as climate, quality and land availability, among others. These locational factors can be controlled by the proximity to other producing regions.

Moreover, according to Heckman, Lalonde and Smith (1999), an additional condition for the use of the propensity scoring is the existence of a common support, i.e., that there exist units in both the treatment and control groups for each characteristic X for which comparison is desired. The condition that $0 < P(X) < 1$ assures that for each treated individual there is another matched untreated one, with similar values of X.

The estimation of $P(X) = \Pr(D = 1 \mid X)$ is done by means of a probit or logit model. However, when there are lagged or spatial effects, conventional models calculated by maximum likelihood are not adequate. By construction, the errors of a spatial logit model are heteroskedastic, and estimates based on the hypothesis of homoskedasticity in the presence of heteroskedastic errors are inconsistent (Greene, 2000; Wooldridge, 2001). The general model, considering spatial lags in the dependent variable and in the residuals, called the spatial autocorrelation (SAC) model (Lesage, 1999; Chagas, 2004), can be described in the following form:

$$
\begin{aligned}
y &= \rho \mathbf{W}_1 y + \mathbf{X}'\beta + u \\
u &= \lambda \mathbf{W}_2 u + e \\
e &\sim N(0, \sigma^2 \mathbf{V}) \\
V &= \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix}
\end{aligned}
\tag{3}
$$

where y is a (0,1) dummy variable, **X** are covariates, $\mathbf{W}_1$ and $\mathbf{W}_2$ are neighborhood matrices that control for the effects of the spatial lag; and v_i, $i = 1, ..., n$ are parameters to be estimated, which capture the model's heteroskedasticity[4]. The parameters ρ and λ are, respectively, the effects of the spatial autocorrelation and of the spatial correlation of the residuals. If $\mathbf{W}_1$ and $\mathbf{W}_2$ are the same, it is possible to estimate this general model, but its identification is problematic (LeSage, 1999).

Alternatively, a less general model can be estimated, considering only the spatial autocorrelation, called the spatial autoregression (SAR) model:

$$y = \rho \mathbf{W} y + \mathbf{X}'\beta + e. \tag{4}$$

Another possibility is the spatial error model (SEM), which considers the spatial effect only in the residuals:

$$y = \mathbf{X}'\beta + u \tag{5}$$

[4] Since the dependent variable of a probit model (y) assumes the values 0 or 1, the errors of a spatial autocorrelation model, for example, take on values $-\rho Wy - X\beta$ when $y = 0$, and $\iota - \rho Wy - X\beta$, when $y = 1$. The error term depends on a parameters vector (β) and a constant (ρ), which induces heteroskedasticity (Wooldridge, 2002, p. 470).

$u = \lambda \mathbf{W} u + e$.

A strategy to choose among these models is first to estimate the most general one (SAC). If the coefficients of the two spatial effects are accepted, this is the best model among the three. If not, the model is estimated associated with the significant knowledge from the previous step.

In the form specified, the models have many more parameters to be estimated than degrees of freedom, preventing the use of the usual techniques. LeSage (1999, 2000) introduced Bayesian estimates, employing techniques based on Monte Carlo Markov chains (MCMC) by means of Gibbs and Metropolis-Hastings sampling. The basic idea of the Monte Carlo method is to characterize the joint (posterior) distribution of the quantities of interest (parameters), and using modern computational techniques, simply to generate a sample of the distribution (taking selections randomly) and calculate the statistics from this sample. With a sufficiently large number of draws, the statistics can approximate the population parameters. Since the initial draws are performed based on an initial (prior) estimate, Franzese Jr. and Hays (2007) suggested that 5,000 to 10,000 draws be taken, and to discard the first 1,000 (called burn-in)[5]. Another model selection criterion arises from this procedure. At each step of the simulation, the cases are recorded when ρ and λ lie within the acceptance interval (-1 to 1). If this rate is very low, the model might be misspecified.

3.2. Kernel Matching

The effect of the treatment on the treated is calculated by comparing the performance of the treated group (denoted by Y_1, indexed by I_1) with that of the untreated group (denoted by Y_0, indexed by I_0), through the following equation (Heckman; Ichimura; Todd, 1998):

$$E[(Y_1 - Y_0) \mid D = 1, P(X)] = \frac{1}{N_1} \sum_{i \in I_1} [Y_{1i} - \sum_{i \in I_0} W_{N_0 N_1}(i, j) Y_{0j}] \qquad (6)$$

[5] In our spatial propensity score estimates, we used 10,000 drawings and discarded the first 1,000.

where $W_{N_0 N_1}(i, j)$ is usually a matrix of positive weights, defined so that for each $i \in I_1$, $\sum_{i \in I_0} W_{N_0 N_1}(i, j) = 1$, and N_0 and N_1 are the numbers of observations in I_0 and I_1, respectively.

A kernel estimator is used to choose the weights in such a way that observations that are nearer in terms of their distances measured by $|P(X_i) - P(X_j)|$ receive greater weight. This weighting is given by a kernel function, which must integrate to one and be continuous and symmetric about the origin (Härdle; Linton, 1994).

$$K(u) = K(-u); \int_{-1}^{1} K(u) du = 1 \qquad (7)$$

A frequently used functional form is the "biweight" (or quartic), expressed by

$$\begin{aligned} K(u) &= \frac{15}{16}(u^2 - 1)^2 \text{ for} |u| < 1 \\ &= 0 \text{ otherwise} \end{aligned} \qquad (8)$$

where $u = \dfrac{P(X_i) - P(X_j)}{h}$.

Implementing the estimation via a kernel function requires choosing a suitable bandwidth (h). The smaller h is, the less weight is given to larger distances and the greater the weight given to more proximate observations. The consistency of nonparametric estimators requires the bandwidth to approach zero as the sample size increases, but not necessarily at the same speed (Todd, 1999).

The approximation of the score distribution, by means of the kernel function, is

$$\hat{f}_h(P(X)) = \frac{1}{n} \sum_{i=1}^{n} K(P(X) - P(X_i)) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K_h\left(\frac{P(X) - P(X_i)}{h} \right) \qquad (9)$$

Following Cameron and Trivedi (2005), we choose the bandwidth according to Silverman's plug-in estimate

$h^* = 2.778N^{-0,2}\min(s, iqr/1.349)$

where N is the sample size, s is the sample standard deviation and iqr is the sample interquartile range, given by the difference between the 3rd and 2nd quartiles. Sensitivity tests are performed to check the sensitivity of the results to the choice of the bandwidth h.

4. Results

4.1. Data base

We use data on grains and sugarcane cultivation, from 1991 to 2009, available from the Municipal Agricultural Survey (*Pesquisa Agrícola Municipal* - PAM), conducted by the Brazilian Institute of Geography and Statistics (IBGE, the official census bureau). Since during the past decade new municipalities have been created by splitting off from existing ones, we grouped them into minimum comparable areas (MCA). There are a total of 4,248 MCAs in the database.

We made three different tests that are reported here. They were excluded in all areas of the northern region. Also excluded were areas where the average share of agricultural GDP in total GDP was lower than the lowest relative share of agricultural GDP in the total GDP (which occurred in 1999, the year that the agricultural GDP accounted for 4.7% of GDP).

The first test in the treatment group included all municipalities in the period 2000 to 2009, expanded the area planted with sugar cane farming, compared with the period 1991 to 1999. The control group of this test was composed of all municipalities in which there was maintenance or reduction in cane area.

The second test, in the treatment group included the municipalities in which there was a reduction in grain acreage and that expansion of sugarcane has represented at least 50% of this reduction. In the control group were included in the municipalities where the grain areas have been maintained or increased.

Finally, the last test, use it as treatment group the regions in which the cultivated areas of grain and sugarcane have increased, and, as control, has become the regions where production of grains and sugar cane have been maintained or reduced.

To calculate the propensity score were also used data of average temperature and precipitation of AMCs. These data are available in IPEADATA and were estimates from the climatic data base of CRU CL 2.0 10 'of the Climate Research Unit at the

University of East Anglia (UEA-CRU) in England[6] for the period 1961-1991. Climate and temperature are variables that distinguish the regions and producing cane sugar is best suited to certain climatic conditions than others.

We also use data on the biome classification belongs to the municipality. Officially, there are six biomes in Brazil, namely, savanna, grasslands, wetlands, plains, Amazonian and Atlantic forest (IBGE). For each region was calculated the proportion of territory belonging to each of the biomes.

Regional development was measured as the average annual growth of per capita GDP hall. GDP and per capita GDP is calculated by IBGE municipal and published annually.

The inverse distance is used for the neighborhood matrix (Anselin, 1998; Chagas, 2004). The distances were obtained from the geographic coordinates of the center of each municipality (latitude and longitude). For the MCAs (aggregations of municipalities), this was the average latitude and longitude, weighted by the average population of each member municipality between 1991 and 2000. Unlike the usual practice, we used geodesic rather than Euclidian distance. By this criterion, more distant neighboring places receive a lower weight than in the case of Euclidian distance (but greater than in the case of neighborhood matrices that consider only places that share borders). This criterion is more suitable, since sugarcane production tends to be influenced by the proximity and not necessarily by contiguity. We have calculated the neighborhood matrix considering a neighborhood radius of 150 kilometers.

4.2. Spatial Propensity Scoring

To calculate the spatial propensity score, we use climate variables, like temperature and precipitation, that control regional specification related to farm production. Additionally, we use the biome incidence to control soil specification in the MCA specific. Finally, we considered the neighborhood spatial effects that capture both the fact that an MCA with neighboring MCAs that are in treatment group is more likely to belongs to the treatment group (dependence or spatial autocorrelation) and institutional similarities, for example, controlled by the spatial dependence specification in the error term.

[6] http://www.cru.uea.ac.uk.

The use of this limited set of covariates is justified by parsimonious point of view. Observable factors that are important in explaining the per capita GDP results of a given place may not be so significant to explain the production of sugarcane. Therefore, our approach for the estimation of the propensity score is more parsimonious from the outset, by restricting as much as possible the number of independent variables.

Erro! Fonte de referência não encontrada. presents the result of estimating the spatial propensity score by means of spatial logit estimation, as suggested by LeSage (1999). We follow a strategy specific to the general, as in Florax, Folmer, and Rey (2003)[7]. The general (SAC model) and the SEM models produced weaker results than the SAR model. The model's fit (pseudo-R^2) is lower both SAC and SAR, for all tests. The efficiency is also lower – the variance (σ^2) is greater in the SAC and in the SEM cases than in the SAR case. Besides this, the acceptance rate, i.e., the frequency with which, in iterations, the parameter for spatial effects was in the range [-1, 1], (particularly the coefficient associated with the spatial errors component) is very low (just over half, both for SAC, as in the case SEM). Finally, there is the counterintuitive result, in the SAC case, that the spatial errors are negative. For these reasons, the SAR model is preferred before than the SAC and the SEM.

TABLE 2 HERE

With respect to climatic variables, higher temperatures in the months March-May and December-February increase the likelihood of production of sugarcane. During the first period occurs planting the crop in the central-south, the main producing region. In the second period, in general, occurs the harvest. For the test two, the higher temperature in the months March-May increase the likelihood of belonging to the treatment group – the group of regions that increase sugarcane production in detriment of grains production. On the other hand, the higher temperature in the months September-November reduces this likelihood. About the test 3, only the coefficient of the period September-November is significant at 5%, and the higher temperature in these months increases the likelihood of belonging to the treatment group – in this case, the regions at the increase at the production both sugarcane and grains.

[7] The authors suggest that for the case of linear models, this strategy is preferable to the reverse strategy (general-to-specific). We take these results as valid for the case of nonlinear models too.

For the rain's case, higher precipitation in the period June-August, and the lower precipitation in March-May period increases the likelihood of belong in the group of production sugarcane's region. Already for test two, all coefficient of precipitation are significant, being the coefficients of the periods June-August and September-November are positives, and the periods March-May and December-November are negatives. The contrary occur in the test 3, in that the higher precipitation on periods June-August and September-November reduces the likelihood belong in the treatment group.

It is interesting the result about the coefficients of the biomes. At all tests, the Swampland coefficients are not significant, reflecting the absence of both sugarcane and grains in this biome. And, for the others, the coefficients are negative and significant (the Atlantic forest biome is no significant for the test 2). This result is consequence, probably, of the higher presence of regions in control group that in the treatment group, for all biomes.

Finally, in all tests, the coefficient of spatial autocorrelation is positive and significant, i.e., the presence of neighbors in the treatment group increases the likelihood belongs in this group.

4.3. Matching

The table 3 presents the average growth of per capita GDP for the regions of interest of the study, without the control of spatial propensity score matching. Note that without this control, the regions that increased the sugarcane production grew, in average, 12.2% per year, rate 0.16 p.p. higher than the regions in that the sugarcane not increased. This result, however, is not statistically significant.

TABLE 3 HERE

Still in this table, without controls, apparently, the regions in that the sugarcane grew in substitution of the grains, the average growth of per capita GDP was minor that the growth of regions in which the grains area was not reduced. In this case, the average growth of the regions that increased sugarcane was 11.8% per year, while the control regions increased by 12.3% per year, and this result is statistically significant.

Finally, when one considers the group of regions in which the area of sugarcane and grain area were expanded, the average growth of per capita GDP was 12.3% per year, against the growth of 11.8% in the regions in which the sugarcane and grains areas sugarcane and grains areas were maintained or reduced. Again, this result is statistically significant.

In the table 4, we present the average growth of per capita GDP for the regions of interest of the study, with the control of spatial propensity score matching, as above described. In this case, the regions that increased the sugarcane production grew, in average, 12.2% per year, rate 0.27 p.p. higher than the comparable regions in that the sugarcane not increased. In contrary of the without-control case, this result is statistically significant at 10%.

TABLE 4 HERE

Too in contrary to the above, the result for the test two is not statistically significant, despite is negative. In those regions in that the sugarcane grew in substitution of the grains, the average growth of per capita GDP was 11.8% per year, while the control regions increased by 12.2% per year.

Finally, when one considers the group of regions in which the area of sugarcane and grain area were expanded, the average growth of per capita GDP was 12.3% per year, against the growth of 11.7% in the more comparable regions in which the sugarcane and grains areas sugarcane and grains areas were maintained or reduced. This result is statistically significant at 10%.

5. Conclusions

The expanding production of sugarcane in Brazil in the recent has aroused a intensive debate about the land use, moreover the potential conflict to the energy use and food use. Of a regional point of view, this debate is reflected in the form of allocate the land – for sugarcane culture or the grains culture.

We implemented a spatial propensity score matching test, as suggested in Chagas et al. 2011, for investigate this question. This methodology is useful because it deals with the fact that one cannot immediately compare average indicators of cane producing regions with those of non-producing ones, since the probability of production is not a random variable. Thus, spatial factors need to be considered to control for the probability of producing or not.

The results suggest that, in the regions in that the sugarcane culture grew, the per capita GDP grew more than in the more comparable regions in that the sugarcane not grew. About the potential conflict in land use for cane or grains cultures, despite the regions in that the cane substitute the grains has grew minor that the control regions, this result is not statistically significant. For this reason, our result is not conclusive.

But, if there is the option of growth the two cultures, this appears the more suitable decision. The regions that increased both cane and grains cultures grew more than the more comparable regions.

6. References

ALVES, F. J. C. Por que morrem os cortadores de cana? **Saúde e Sociedade**, vol.15 no. 3. São Paulo, pp. 90-98, Sept./Dec. 2006.

_______. Migração de trabalhadores rurais do Maranhão e Piauí para o corte de cana em São Paulo - será este um fenômeno casual ou recorrente da estratégia empresarial do Complexo Agroindustrial Canavieiro? In NOVAES, R.; ALVES, F. J. C. (Orgs.). **Migrantes: trabalho e trabalhadores no complexo agroindustrial canavieiro** - os heróis do agronegócio brasileiro. São Carlos: EDUFSCar, 2007, pp. 21-54.

ANSELIN, L. **Spatial Econometrics**: Methods and models. Dordecht: Kluwer, 1988.

BACCARIN, J. G.; ALVES, F. J. C.; GOMES, L. F. C. Emprego e condições de trabalho dos canavieiros no centro-sul do Brasil, entre 1995 e 2007. **Anais do XLVI Congresso da Sober**. Rio Branco: Sociedade Brasileira de Economia e Sociologia Rural, 2008.

BECKER, S. O.; ICHINO, A. Estimation of average treatment effects based on propensity score. **Stata Journal**, v. 2, n. 4, pp. 358-377, 2002. Available at: <http://www.sobecker.de/pscore.html.>. Accessed on April 12, 2009.

BNDES; CGEE. **Bioetanol de cana-de-açúcar**: energia para o desenvolvimento. Rio de Janeiro: BNDES, 2008.

CAMARGO-JR., A. S.; TONETO-JR, R. Indicadores sócio-econômicos e a cana-de-açúcar no estado de São Paulo. **Anais do I Workshop do Observatório do Setor Sucroalcooleiro**. Ribeirão Preto, 2008.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics**: methods and applications. Cambridge: Cambridge University Press, 2005.

CHAGAS, A. L. S. **Externalidades da aglomeração**: microfundamentação e evidências empíricas. University of São Paulo, master’s dissertation, 2004.

CHAGAS, A. L. S. ; TONETO-JR, R. ; Azzoni, C. R. A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in Brazil. **International Regional Science Review**, 2011.

DIPRETE, T.; GANGL, M. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. **Sociological Methodology**, v. 34, n. 1, pp. 271-310, Dec. 2004.

FAO. **The State of Food and Agriculture 2008**: Biofuels: Prospects, risks and opportunities. Rome, 2008.

FLORAX, R.J.G.M., FOLMER, H., REY, S.J., 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. Regional Science and Urban Economics 33, 557 579.

FRANZESE-JR, R. J.; HAYS, J. C. The spatial probit model of interdependent binary outcomes: Estimation, interpretation, and presentation. **24th Annual Summer Meeting**

of the Society for Political Methodology, July 20, 2007. Available at: <http://polmeth.wustl.edu/retrieve.php?id=715>. Accessed on May 31, 2009.
GREENE, W. H. **Econometric Analysis**. 4th ed. New Jersey: Prentice Hall, 2000.
HÄRDLE, W; LINTON, O. Applied nonparametric methods. In ENGLE, R. F.; MACFADDEN, D. L. (eds.) **Handbook of Econometrics**, vol. 4, Amsterdam: Elsevier Science, 1994, pp. 2295-2339.
HECKMAN, J.J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. In ASHENFELTER, O.; CARD, D. (eds.) **Handbook of Labor Economics**, vol. 3. Amsterdam: Elsevier Science, 1999, pp. 1865-2097.
HECKMAN, J.J; ICHIMURA, H; TODD, P. Matching as an econometric evaluation estimator. **Review of Economic Studies**, vol. 65, No. 2, April 1998, pp. 261-294.
_______. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. **Review of Economic Studies**, vol. 64, no. 4, Oct. 1997, pp. 605-654.
HOFFMAN, R.; OLIVEIRA, F. C. R. Remuneração e características das pessoas ocupadas na agro-indústria canavieira no Brasil, de 2002 a 2006. **Anais do XLVI Congresso da Sober**. Rio Branco: Sociedade Brasileira de Economia e Sociologia Rural, 2008.
LESAGE, J.P. **Spatial Econometrics**. Mimeo, 1999. Available at: <http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf>. Accessed on March 1, 2009.
______. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. **Geographical Analysis,** 32(1):19-35, 2000.
LEUVEN, E.; SIANESI, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Available at: <http://ideas.repec.org/c/boc/bocode/s432001.html>. Accessed on June 1, 2009.
MENDONÇA, M. L. The WTO and the destructive effects of the sugarcane industry in Brazil. **Land Research Action Network**. 13 Fev, 2006a. Disponível em < http://www.landaction.org>. Acessado em 01/06/2008.
MENDONÇA, M. L. The WTO and the devastating impacts of the sugarcane industry in Brazil. **Land Research Action Network**, 2006b. Disponível em < http://www.landaction.org>. Acessado em 01/06/2008.
NORONHA, S. et al. **Agronegócio e biocombustíveis**: uma mistura explosiva - Impactos da expansão das monoculturas para a produção de Bioenergia. Rio de Janeiro: Núcleo Amigos da Terra, 2006.
PIKETTY, M. G.; MENEZES, T. M; DUARTE, J. B. N. A. Sugar cane in Brazil, poverty and equity: Evidence for the 1992-2006 period. **Anais do XXXIV Congresso da ANPEC**. Salvador: ANPEC, 2008.
PNUD; IPEA; FJP. **Atlas de desenvolvimento humano no Brasil**. Brasília, 2003.
RESENDE, A. C. C.; OLIVEIRA, A. M. H. C. Avaliando resultados de um programa de transferência de renda: o impacto do Bolsa-Escola sobre os gastos das famílias brasileiras. **Estudos Econômicos**, vol. 38, n.2, 2008, pp. 235-265.

ROSENBAUM, P.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, 70, pp. 41-55, 1983.

ROSENBAUM, P. **Observational Studies**, Springer, New York. 2002.

SILVA, R. Setor sucroalcooleiro no estado de São Paulo: mensurando impactos sócio-econômicos. **Anais do I Workshop do Observatório do Setor Sucroalcooleiro**. Ribeirão Preto, 2008.

TODD, P. **A practical guide to implementing matching estimators**. Mimeo. 1999. Available at: <http://athena.sas.upenn.edu/~petra/papers/prac.pdf>. Accessed on May 31, 2009.

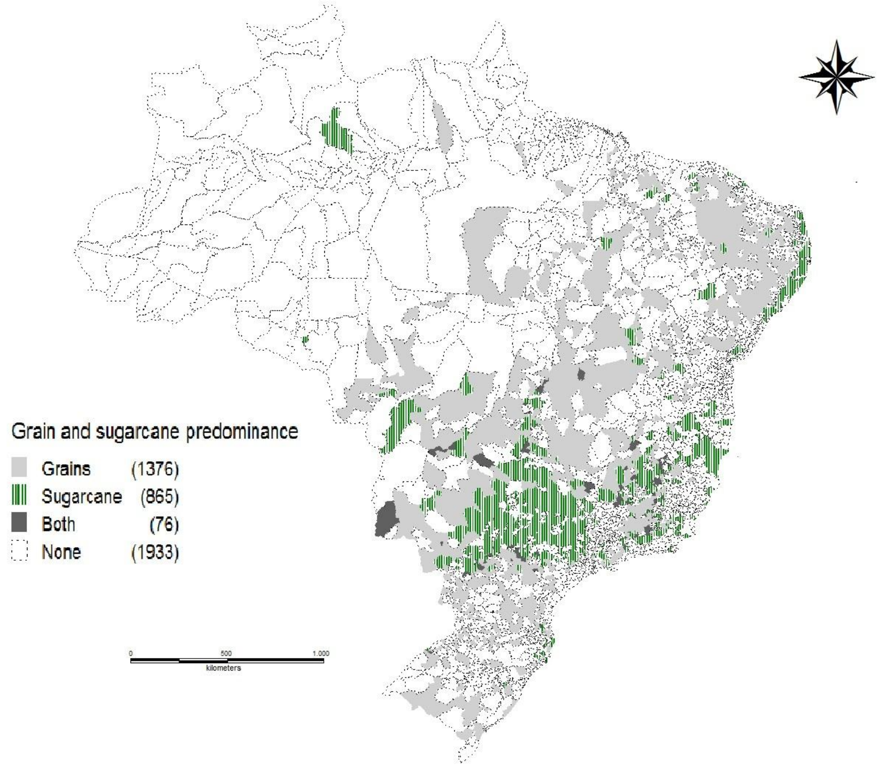
TONETO-JR., R.; LIBONI, L.B. Mercado de Trabalho da cana-de-açúcar. **Anais do I Workshop do Observatório do Setor Sucroalcooleiro**. Ribeirão Preto, 2008.

WOOLDRIDGE, J. M. **Econometric Analysis of Cross-Section and Panel Data**. Cambridge: MIT Press, 2001.

Table 1: Area planted annually – grains, sugarcane and total farming (in millions of hectare)

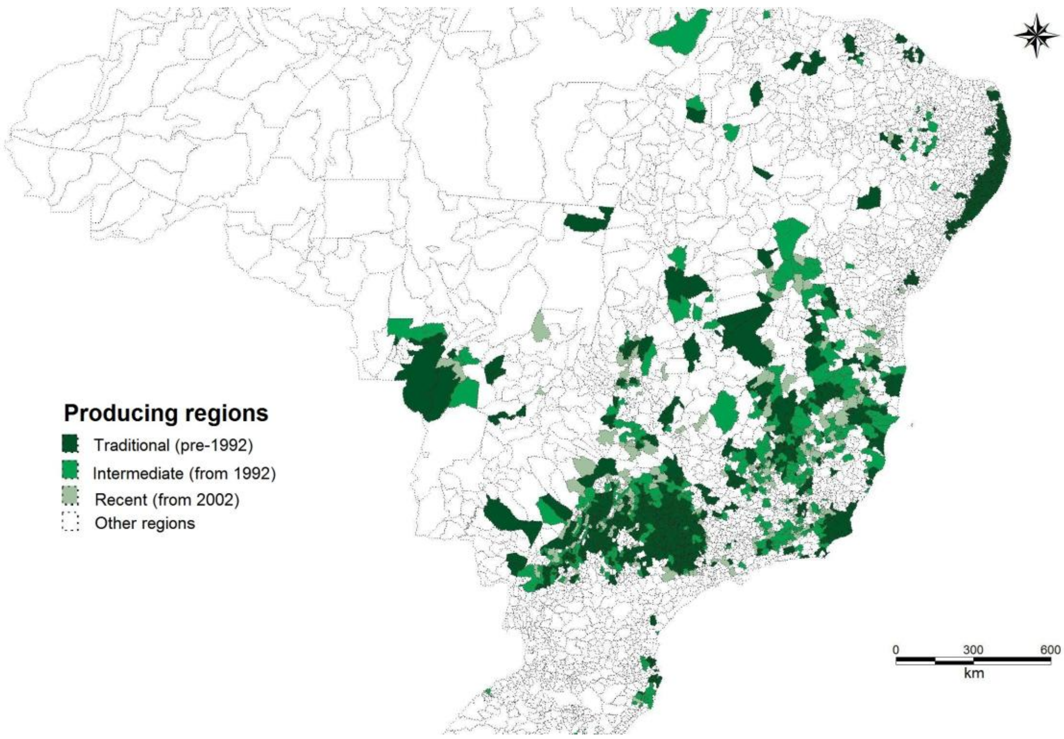
Year	Grains	Sugarcane	Farming
	hectare millions		
1990	36.42	4.32	53.15
1991	35.22	4.24	51.75
1992	35.73	4.22	52.27
1993	34.41	3.95	49.23
1994	37.74	4.36	52.82
1995	36.71	4.64	51.85
1996	32.46	4.83	46.82
1997	33.49	4.88	48.30
1998	33.04	5.05	48.54
1999	35.26	4.98	50.70
2000	36.02	4.88	51.82
2001	35.68	5.02	51.64
2002	38.33	5.21	54.51
2003	42.01	5.38	58.46
2004	45.38	5.63	63.04
2005	46.00	5.82	64.32
2006	44.11	6.39	62.57
2007	43.33	7.09	62.35
2008	45.22	8.21	65.53
2009	45.53	8.76	65.63
Annual average growth	1.18%	3.79%	1.12%

Source: Municipal Agricultural Research, IBGE



Source: Municipal Agricultural Research, IBGE

Figure 1: Grains and sugarcane predominance



Source: Municipal Agricultural Research, IBGE

Figure 2: Expansions regions of the sugarcane

Table 2: Logit model to estimate the spatial propensity score matching

		SAR Mod.			SEM Mod.			SAC Mod.		
		Test 1	Test 2	Test 3	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
Pseudo-R^2	=	0.2022	0.4578	0.1640	0.1800	0.4158	0.0854	-0.0160	0.3007	0.1494
σ^2	=	1.1069	1.2111	1.0991	1.0266	1.0096	1.0040	2.0214	3.0329	2.6968
no. of obs., no. of var.	=	3510 , 14	1735 , 14	1274 , 14	3510 , 14	1735 , 14	1274 , 14	3510 , 14	1735 , 14	1274 , 14
no. 0, 1 y-values	=	1773 , 1737	1065 , 670	742 , 532	1773 , 1737	1065 , 670	742 , 532	1773 , 1737	1065 , 670	742 , 532
accept rate ρ	=	0.9998	0.9998	0.9994				0.6360	0.6174	0.7866
accept. rate. λ	=				0.4942	0.4882	0.4982	0.4501	0.5538	0.9603
Posterior Estimates										
Variable		Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Preciptation										
jun to ago		0.0026 **	0.0141 *	-0.0058 **	0.0022 ns	0.0171 *	-0.0061 ***	0.0008 ns	0.0204 *	-0.0099 ns
		(2.023)	(6.307)	(-2.368)	(.002)	(.017)	(-.006)	(.001)	(.020)	(-.010)
mar to may		-0.0053 *	-0.0166 *	0.0018 ns	-0.0033 ns	-0.0196 *	0.0044 ns	-0.0040 ns	-0.0207 *	0.0062 ns
		(-3.640)	(-6.170)	(.744)	(-.003)	(-.020)	(.004)	(-.004)	(-.021)	(.006)
sep to nov		0.0023 ns	0.0126 *	-0.0051 ***	0.0017 ns	0.0174 *	-0.0070 ***	0.0015 ns	0.0175 **	-0.0109 ns
		(1.543)	(4.649)	(-1.950)	(.002)	(.017)	(-.007)	(.002)	(.017)	(-.011)
dec to feb		0.0020 ns	-0.0161 *	0.0071 **	0.0039 ns	-0.0161 *	0.0089 **	0.0086 ns	-0.0201 **	0.0142 ***
		(1.298)	(-6.018)	(2.497)	(.004)	(-.016)	(.009)	(.009)	(-.020)	(.014)
Temperature										
jun to ago		-0.3288 *	0.0497 ns	-0.2443 ns	-0.2998 ***	0.0667 ns	-0.2763 ns	-0.2439 ns	-0.0691 ns	-0.3209 ns
		(-3.209)	(.271)	(-1.397)	(-1.655)	(.240)	(-1.010)	(-.895)	(-.139)	(-.678)
mar to may		0.1849 ***	0.4122 **	-0.1281 ns	0.1805 ns	0.6404 *	-0.0245 ns	0.1288 ns	1.0163 **	-0.0095 ns
		(1.915)	(2.383)	(-.780)	(1.157)	(2.634)	(-.098)	(.538)	(2.124)	(-.022)
sep to nov		0.0403 ns	-0.4117 *	0.3665 **	-0.0013 ns	-0.6808 *	0.1846 ns	0.0960 ns	-1.0518 **	0.1688 ns
		(.477)	(-2.642)	(2.480)	(-.009)	(-2.982)	(.804)	(.413)	(-2.188)	(.426)
dec to feb		0.1805 **	0.0713 ns	0.0882 ns	0.1656 ns	0.0630 ns	0.2236 ns	0.0792 ns	0.2391 ns	0.3046 ns
		(2.270)	(.491)	(.633)	(1.198)	(.304)	(1.067)	(.375)	(.601)	(.810)
Biome										
Caatinga		-2.4460 *	-3.1588 *	-2.7248 *	-2.1930 *	-2.3388 *	-3.7524 *	-2.8875 *	-3.1372 *	-4.9484 *
		(-5.915)	(-5.646)	(-2.730)	(-5.014)	(-4.078)	(-2.780)	(-4.494)	(-3.275)	(-2.629)
Savanna		-1.6586 *	-1.2528 **	-2.5377 **	-1.0612 **	-0.1295 ns	-3.4568 **	-1.4800 **	-0.4177 ns	-4.4728 **
		(-3.863)	(-2.181)	(-2.393)	(-2.350)	(-.240)	(-2.404)	(-2.353)	(-.448)	(-2.30)
Swampland		-1.2653 ns	-1.0147 ns	-4.6028 ns	-0.2971 ns	0.1056 ns	-5.2735 ***	0.7953 ns	4.5159 ns	-5.4829 ns
		(-1.266)	(-.729)	(-.896)	(-.256)	(.075)	(-1.848)	(.372)	(1.021)	(-1.050)
Pampa		-2.1605 *	-2.1757 *	-2.0743 ***	-1.8390 *	-1.9304 ***	-2.8634 ***	-2.7454 *	-2.0825 ns	-3.8038 ***
		(-4.230)	(-2.658)	(-1.718)	(-3.121)	(-1.886)	(-1.771)	(-3.033)	(-1.460)	(-1.750)
Amazon		-2.1662 *	-2.6042 *	-2.4881 **	-1.8579 *	-1.9626 **	-3.3005 **	-2.8345 **	-4.5049 **	-3.9849 ***
		(-4.371)	(-3.288)	(-2.108)	(-3.057)	(-2.180)	(-2.066)	(-2.557)	(-1.982)	(-1.746)
Atlantic forest		-1.4123 *	-0.8706 ns	-2.0329 ***	-0.9457 **	-0.0311 ns	-2.9481 **	-1.3363 **	-0.0810 ns	-3.8034 **
		(-3.349)	(-1.536)	(-1.941)	(-2.153)	(-.060)	(-2.076)	(-2.219)	(-.095)	(-1.998)
ρ		0.3468 *	0.2935 *	0.4920 *				0.9273 *	0.9239 *	0.8112 *
		(6.915)	(5.992)	(11.573)				(48.913)	(33.644)	(3.992)
λ					0.6865 *	0.6472 *	0.5730 *	-0.9675 *	-0.8320 *	-0.7235 **
					(15.743)	(12.631)	(13.154)	(-15.751)	(-7.651)	(-2.152)

* Significant at 1%; ** Significant at 5%; *** Significant at 10%; ns: no significant.

T-statistic between parentheses.

Source: Prepared by the authors.

Table 3: Average growth for the regions – without control models

	GDP per capita (mean)		
	Test 1	**Test 2**	**Test 3**
Group			
Treatment	12.159	11.778	12.298
Control	11.995	12.349	11.826
Difference	0.164	-0.571	0.472
t-estat	1.363	-3.265	2.458
Prob	0.173	0.001	0.014
	Observations		
Group			
Treatment	1737	670	532
Control	1773	1065	742

Source: Prepared by the authors.

Table 4: Average growth for teh regions – spatial propensity score matching models

	GDP per capita (mean)		
	Test 1	**Test 2**	**Test 3**
Group			
Treatment	12.159	11.778	12.298
Control	11.888	12.153	11.692
Difference	0.271	-0.375	0.606
t-estat	1.848	-1.345	1.807
Prob	0.065	0.179	0.071
	Observations		
Group			
Treatment	1737	670	532
Control	1773	1065	742

Source: Prepared by the authors.