

Chagas, Andre

**Conference Paper**

## Comparing social-economic conditions in ethanol production areas in Brazil and United States - a spatial econometric approach

54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Chagas, Andre (2014) : Comparing social-economic conditions in ethanol production areas in Brazil and United States - a spatial econometric approach, 54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124524>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Comparing social-economic conditions in ethanol production areas in Brazil and United States – a spatial econometric approach

Draf version – do not publicized

*André Luis Squarize Chagas*

University of Sao Paulo

Department of Economics

[achagas@usp.br](mailto:achagas@usp.br)

The production of ethanol has increased in United States and Brazil in recent years, because of incentives to reduce gas consumption. In Brazil, there is a big program to use ethanol either as a single fuel or as an additive to gas. In US, the incentive is to use ethanol combined to gas in proportion that can reach to 85 percent of the mix. This situation has prompted the need to assess the economic, social and environmental impacts of this process, both for the country as a whole and for the regions where this has occurred. Doubts that can be raised, for example, concerning the quality of employment, the environmental impacts (soil contamination, atmospheric pollution from burning fields, water use, etc.) and dislocation of other crops to native forests, among others. Even though the balance of costs and benefits is, in general, positive from an overall standpoint, this may not be so in specific producing regions, due to negative externalities. The producing regions may disproportionately bear the negative impacts of the sector's presence. Perhaps the most obvious aspect in this respect is the labor market, since many studies have analyzed the working conditions in the sector, particularly those encountered in manual harvesting, in Brazilian case. It is recognized that sugarcane is significantly more valuable by tilled area than many other crops, such as soybeans, for example. The objective of this paper is to estimate the effect of growing sugarcane on socio-economic indicators (like GDP, HDI or others) in ethanol producing regions, comparing differences in results of Brazilian regions with American ones. In the literature on matching effects, this is interpreted as the effect of the treatment on the treated. Location effects are controlled by spatial econometric techniques, giving rise to the spatial propensity score matching model. This work extends the results from Chagas, Toneto-Jr. and Azzoni (2011), analyzing the producing regions in both countries according its social and economics characteristics, as well the producing characteristics.

Keywords: ethanol production; socioeconomic impacts; spatial propensity score matching

JEL CODE:

## 1. Introduction

The demand for energy sources is one of the main problems to be solved by industrial economies. Since the industrial revolution of the late nineteenth century, many varieties of combinations of energy sources were made in order to provide the supply of the machines, and consumer durables, which is present daily of firms and consumers.

Since the late nineteenth century, oil became viable as the main energy source. Because of its apparent abundance and its relative ease of extraction, he was and still is used intensely, especially with the advent of the automobile industry, which effectively consolidated their need.

However, because this product is not renewable and release a high concentration of carbon dioxide becomes a major determinant of the greenhouse. Thus, the use of this feature is among the main issues discussed in international forums concerned with the reduction of global warming.

Among the alternatives to oil, there are renewable inputs, obtained from sources that allow their reproduction, such as wind, solar, hydro and biofuels from the processing of vegetable matter, which offset the carbon dioxide released by prior absorption of the same gas during growth of the plant. The most widely used biofuel today is ethanol, a product that works in different areas, because of its increasing use as a gasoline additive, present in more countries UDOP (2013).

Brazil, in this case, is on the front line, because its fuel program has a flexible engine, making possible the car drivers use the E100 mixture (100% ethanol). Countries like the United States, and Sweden adopted mixtures of up to 85% ethanol (E85). Thailand, Australia, Canada, China, Colombia, Peru, South Africa, Paraguay, and India have adopted proportions of ethanol in fuel equivalent between 5 to 10%. Besides these, many others are studying this possibility as England, Russia, Cuba, France, Germany, Japan, among others.

Moreover, various social issues arise with respect to ethanol production. According Basaldi and Gomes (2008), the increase in ethanol production has obtained positive points, such as, improving the education of employees, significant reduction of child labor in the sugarcane fields and the significant increase in formal, consequently increasing the contribution to Social Security. However, there are negative factors that are still present in this sector, highlighted: the high workload, exceeding 44 hours per week; the high proportion of employees with remuneration below the minimum wage and increasing the proportion of low aid.

The issues related to the environment also directly affect the conditions which workers are subjected. Salles and Itani (2010) emphasize that in places whose fires are still used in the cutting process, both the workers and the local communities are likely to suffer respiratory problems and, in the low rainfall, the effects worsen further. Add up to the problems arising from burning: impacts that reduce the photosynthetic process, which reduces agricultural production and increasing the amount of smoke and soot increases the number of pathogenic micro-organisms in air and water, contributing to the proliferation lavas of the vector mosquito.

Then, the aim of this research is investigate positive and negative aspects of ethanol production, its effects over producer regions, and compare the social indicators in North American regions with those Brazilian ones.

Unfortunately, the work is in the initial phase, which make impossible to advance in conclusions. In the next session, I make a revision of the recent literature about the issues, comparing the advance of the ethanol in USA and Brazil. In the same section, I consider positive and negative aspects of its production and I conclude with some productive considerations about the two technologies. Next section describe the methodology that I intend apply, as soon as data is available.

## **2. Literature revision**

### **2.1. Advancement of ethanol in Brazil and the United States**

According to Balat and Balat (2008), the production of bioethanol increased from 17.25 billion liters in 2000 to over 46 billion in 2007. Despite this amount corresponds to a modest share of world energy consumption - representing 4% of world gasoline produced in the same year - growth projections are optimistic. Countries like the United States, Brazil and some European countries are responsible for the largest biofuels programs in the world. National policies tend to vary according to the raw material and organic farming practices available in the country. It is possible that global production exceeds 125 billion liters in 2020.

The advance in their share of biofuels over other energy matrices was mainly due to the increased production of the ethanol's first generation, which rely on conversion technologies such as fermentation by yeast or varieties of conventional transesterification catalysts. The raw material for this way of production can be grains, vegetable oils and sugar cane, in the U.S. case, is the predominant corn.

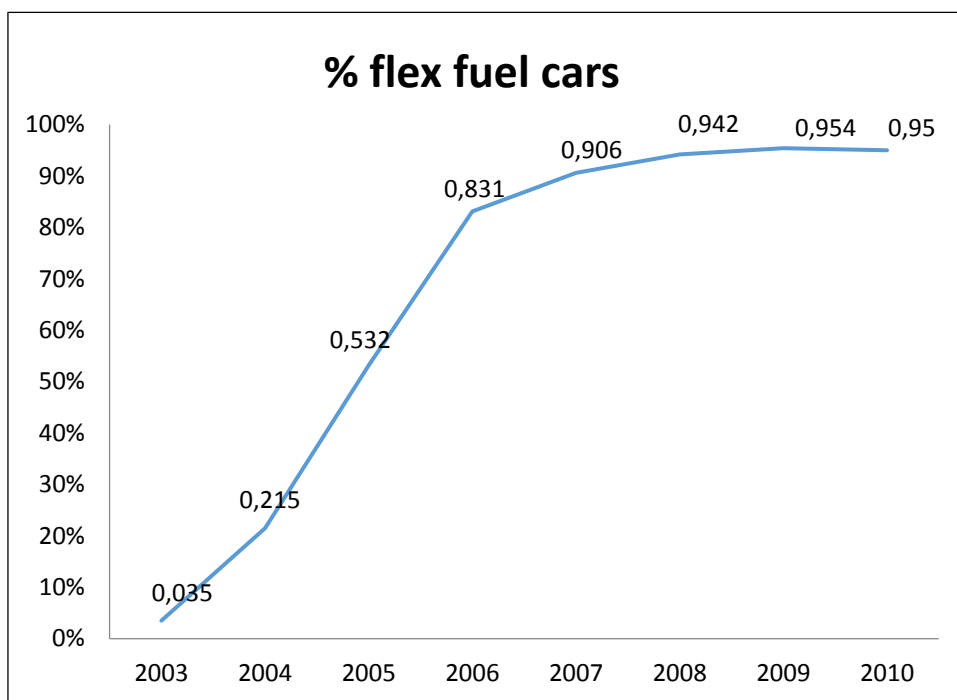
Richards (2013) points out that, although most of the 250 million cars registered in the U.S. still be based to gas, the scenario is changing. Among the reasons cited by the author, the highlights are the 170% increase in the price of a gallon of oil from 2002 to 2012, and the growing environmental awareness in the general population. Ethanol production in the United States increased from less than 2 billion gallons, in 1998, to 14 billion, in 2011. By June 2012, there were 212 ethanol plants in the U.S. in operation or under construction, and at least 150 projected plants to be approved. Ethanol production in the U.S. contributes 400,000 jobs, accounting for US\$42.4 billion to the GDP and US\$29.9 billion in taxes of all kinds ([www.iowacorn.org](http://www.iowacorn.org)).

Brazil, in turn, began to invest in ethanol to meet the demand for sustainable energy, from the partial replacement of an imported product, oil, to a nationally produced, ethanol by sugarcane. This option was a strategic option of government. In 1975, it implanted the Proálcool Program. Since then, the production of ethanol and its primary feedstock in Brazil, sugar cane, grown significantly. In 2007, the country produced around 40% of ethanol in the world (Balat, Balat, 2008). In the period 1980 to 2012, the annual growth rates of ethanol production, and sugarcane were, on average, 8.7% and 7.4%, respectively. In the Central-South region, biggest responsible for the production, sugarcane grown on average 8.8% per year, and ethanol production, 9.2%. ([www.unicadata.com.br](http://www.unicadata.com.br)).

Public policies in Brazil contributed decisively to the growth of this production. They made possible the development either institutionally as technologically enable the ethanol production on a large scale. In 1984, 50% of cars produced in Brazil already were powered by bioethanol, and jumped to 98% until 1988. However, government

intervention began to decline in 1993, despite the add obligation of ethanol to gas. In 1998, only 3% of producing national car were powered by ethanol.

The design of flex fuel engine, around 2003, makes possible the car drivers choose between gas and ethanol, in any proportion. Initially, consumers had tax incentive to buy new flex fuel cars, which stimulated either the automobile industries, as ethanol production sector. In 2010, over 95% of cars that were sold in Brazil were flex-fuel (UNICA, 2013). Compared to 21.5%, in 2004, it is possible note the significant advance in this sector. This is clear in the figure 1.



Source: Anfavea Yearbook

Figure 1. Proportion of cars that run on ethanol

The participation of sugarcane and ethanol in Brazil's GDP amounted to US\$48 billion, in 2012. By 2020, US\$90 billion is expected (UNICA, 2013). In terms of exports of ethanol, the scenery is also optimistic, with US\$15 billion the balance of exports, in 2012, and US\$26 billion expected for 2020. Such growth can generate over 1 million new jobs, both direct and indirect, besides contributing to US\$54 billion in new industries and US\$23 billion in agriculture.

## 2.2. Potential positive and negative impacts of production at national and local level

The impacts of this growth in production, from 2000 until today, were significant. However, the interpretations of these results are differ in some points analyzed in the literature. In this section, I exposed the positive and negative views, showing different perspectives on the context of the expansion of sugarcane production its influence in social indicators. A more extension view, see Chagas et al. (2012) and Chagas (2014).

The positive outlook indicates a significant improvement in the balance of payments, after all. Brazil has importing a smaller amount of energy source and, in addition, develops a supply chain with skilled manpower, including: biochemical, physical, chemical and many other for the processing of sugar cane and its technological field of production, also impacting on income in this new local industry.

The most controversial aspect of sugarcane growing is related to working conditions (Alves, 2006, 2007; Mendonça, 2006a; 2006b; Baccarin; Alves; Gomes, 2008). Alves (2006) calls attention to the extreme physical exertion required of workers in the sector, especially those engaged in manual harvesting. Although this aspect is still a problem at present, legislation in the most relevant producing areas has changed to make mechanical harvesting mandatory in the next few years. On the positive side, there is some evidence that wages are higher than in other agricultural sectors (Toneto-Jr and Liboni, 2008; Hoffman and Oliveira, 2008). However, Piketty, Menezes and Duarte (2008) have shown that the crop has not played a significant role in reducing poverty and inequality in the country. Indeed, for the state of São Paulo (Brazil's main cane producing state), the authors concluded that the sector contributed to the concentration of income.

Camargo-Jr. and Toneto-Jr. (2008) have found a positive association between sugarcane growing and sugar and alcohol production and socioeconomic indicators. In general, municipalities with strong involvement in the sugar-alcohol sector perform better on socioeconomic indicators, and in some cases even outperform the greater São Paulo Metropolitan Region (SPMR), the state's main region in economic terms. Silva (2008) also found the same positive impact when no cross-effects on other variables are considered. However, when consideration is made for the fact that the sector's presence can affect local human development through its impact on other variables, he found that the situation is reversed and the sector's presence has net negative impacts.

Other issues concerning about the environment, such as deforestation of new areas for planting sugarcane in most regions of Brazil and, added to this, the advance of fires, which shows positive correlation with respiratory problems of the local population in the State of São Paulo (Chagas, Almeida, Azzoni 2013). With respect to the effects of sugarcane burning, Salles and Itani (2008) point out that this practice is exercised to facilitate their subsequent cutting, it exposes workers and neighboring populations to sugarcane areas to air pollutants that cause serious respiratory health effects and to the environment. So when fires occur, the lower rainfall and the effects are compounded.

Arbex (2004) showed that the burning generates an increase in mortality, hospitalizations, use of drugs for respiratory and cardiovascular problems, and impaired pulmonary function. The same work also shows indirect impacts related to decreased photosynthetic production of plants, impacting negatively on agricultural production, since the ultraviolet rays A and B are blocked by the amount of smoke, which increases the number of pathogenic microorganisms in air and water.

Noronha and Ortiz (2006) criticized the way that occurs sugarcane production in Brazil, claiming that monoculture has resulted in major damage to society and the environment, among them, the concentration of land ownership, wealth and income, the destruction of forests, contamination of air, soil and water, and the expulsion of rural populations.

Piketty (2008) examined the Brazilian case, more specifically two areas: the state of Sao Paulo and Pernambuco. The result showed that the planting of sugarcane decreased the concentration of income in the period from 1993 to 1999, when inflation in Brazil was relatively high, and the government had adopted the fixed exchange rate policy. In the posterior period (2000-2006), with floating exchange rates, the authors conclude that the

effect was the opposite. There has been a concentration of income in the sector and subsector, this is, planting of this product has not promoted social equity.

Chagas, Azzoni, and Toneto-Jr (2011) show that the balance of costs and benefits of planting sugarcane in Brazil is positive when viewed on a national scale, however, I cannot say the same for a regional analysis in which the influences of externalities negative impact of greatest relevance. Through spatial propensity score matching methodology, to make regional analysis comparable, and to verify the impact of sugarcane production in HDI indicators, the authors conclude that there are not effects associated to sugarcane production in regions, either positive or negative.

### 2.3. Comparison of corn production in the U.S. with the cane in Brazil

The U.S. production is a good basis for comparison to Brazil, because there are the higher ethanol producers and both have a regional concentration in production: Brazilian Center-South region, and North American North-Central region. Perhaps, the concentration of this production may have comparative advantages with respect to a production scattered evenly throughout the country.

The goal of this section is to show the main differences between the sugarcane production in Brazil and the USA. By ABAG survey, I can see the advantages and disadvantages of producing sugarcane-based ethanol. Besides its yield is higher, your cost is lower.

Table 1 shows some comparative advantages of Brazilian sugarcane production through maize production in U.S.A. Some parameters, such as income, cost of production and selling price, show that ethanol from sugarcane is clearly better in terms of comparative advantage than ethanol from corn. The only point at which the corn has advantage is its ability to store.

Table 1: Some comparable variables between sugarcane and maize productions, 2008

Parameter	Unity	Sugarcane	Maize
Productivity	t/ha	90,0	8,1
Cost (include production and transport costs)	BRL/liter	0,9	1,1
Stock	Period	36 hours after cutting	1 year or more
Energy requirement	kcal x1000	10,509	8,115
Energy input/output	kcal	1 / 4.60	1 / 3.84
Ethanol production	liter/ha	8,100	3,000
	liter/ton	90	371
Conversion rate	kg/ 1000L	11,110	2,690
Total Energy Expenditure	kcal/ 1000L	1,518,000	6,597,000
Total Annual Production	Billion liters	15.8	17.2

Energy Balance	kcal input/ output	1 / 3.24	1 / 1.29
Production cost	US\$/L	0.28	0.45
Selling Price	US\$/L	0.42	0.92

Source: ABAG

### 3. Methodology

#### 3.1. Spatial Propensity Score Matching

Our objective is to estimate the effect of growing sugarcane on the HDI-M of producing regions. Let  $D_i = 1$  if sugarcane is produced in region  $i$  and  $D_i = 0$  if not. The probability of one region to belonging to one group is affected by factors ( $X_i$ ), such as proximity to a mill/distillery, for example. From a farmer's point of view, growing cane 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,  $\Pr(D_i = 1 | X_i)$ , 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  $(Y_0, Y_1) \perp D | X$  (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

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

The value of the counterfactual effect of no treatment on the treated,  $E(Y_0 | D = 1, X)$ , is approximated by the average result of the self-selected group of untreated individuals  $E(Y_0 | D = 0, X)$  (Heckman; Ichimura; Todd, 1998). Instead of using various conditional covariates, I use the propensity score  $P(X) = \Pr(D = 1 | X)$ , 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  $P(X) = \Pr(D = 1 | X)$  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:

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

where  $y$  is a (0,1) dummy variable,  $\mathbf{X}$  are covariates,  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are neighborhood matrices that control for the effects of the spatial lag; and  $v_i, i = 1, \dots, n$  are parameters to be estimated, which capture the model's heteroskedasticity<sup>1</sup>. The parameters  $\rho$  and  $\lambda$  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:

$$\begin{aligned}
 y &= \mathbf{X}' \beta + u \\
 u &= \lambda \mathbf{W}u + e
 \end{aligned} \tag{5}$$

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

---

<sup>1</sup> 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  $-\rho \mathbf{W}y - \mathbf{X}' \beta$  when  $y = 0$ , and  $1 - \rho \mathbf{W}y - \mathbf{X}' \beta$ , when  $y = 1$ . The error term depends on a parameters vector ( $\beta$ ) and a constant ( $\rho$ ), which induces heteroskedasticity (Wooldridge, 2002, p. 470).

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)<sup>2</sup>. Another model selection criterion arises from this procedure. At each step of the simulation, the cases are recorded when  $\rho$  and  $\lambda$  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):

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

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

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

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

$$K(u) = \frac{15}{16} (u^2 - 1)^2 \text{ for } |u| < 1 \quad (8)$$

$$= 0 \text{ otherwise}$$

---

<sup>2</sup> In our spatial propensity score estimates, I used 10,000 drawings and discarded the first 1,000.

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

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

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

Following Cameron and Trivedi (2005), I 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  $h$ .

### 3.3. Considerations on Bias

Let  $V$  be the bias in the estimate of the treatment effect on the treated, defined as

$$V = E[(Y_0 | D = 1, P(X)) - (Y_0 | D = 0, P(X))] \quad (10)$$

This is the bias that comes from utilizing the average results of the comparison group as a proxy for the average results of the participants in the program if they had not participated. According to Heckman, Ichimura and Todd (1997), the bias can be divided into three basic components: lack of a common support; unobservable errors; and differences in the results that remain even after taking into consideration the observable characteristics and performing comparisons in a common support region. This last component is due to differences in the unobservables, known as selection bias. This bias arises when, for given values of  $X$ , there is a systematic relation between participation in the program and the results, i.e., there are unobserved variables that jointly influence the results and participation in the program, conditional on the observable variables. To deal with this bias, the best way would be to allocate subjects to the program at random, because in this way one can guarantee that participants and nonparticipants would have the same expected outcome without the program.

I should remark that matching methods (as is the case of the propensity score) only eliminate two of the three sources of bias. The first type is eliminated by the matching within a region with common support. The careful matching of the comparison group, based on observable characteristics, eliminates the second bias component. However, matching methods only deal with observable characteristics, leaving the problem of latent heterogeneity, which causes a possible bias in estimating the program's impact. Thus, the

propensity score method permits reducing, but not eliminating, the bias caused by unobservable factors. How much the bias is reduced crucially depends on the richness and quality of the control variables used to compute the propensity score and carry out the matching (Becker; Ichino, 2002).

### 3.4. Sensitivity Analysis Utilizing the Rosenbaum Bounds Method

As previously mentioned, unobserved factors can bias the estimates of the treatment effect on the treated when using the propensity score method. Since it is impossible to estimate the magnitude of the selection bias in non-experimental data, an important tool can be employed to evaluate it in estimating the causal effects. This method is called bound analysis. It evaluates the potential impact of the selection bias that arises due to unobserved variables. Here I use the method known as Rosenbaum bounds (Rosenbaum, 2002; Diprete; Gangl, 2004; Resende and Oliveira, 2008), the idea of which is to estimate the influence of a possible omitted variable on the selection bias existing in the probability of participation in the treatment, which can possibly impair the conclusions on the causal effects<sup>3</sup>.

Sensitivity analysis can be used to test the robustness of the results to the presence of bias due to an omitted covariate. Here this analysis aims to evaluate the effect of a possibly omitted variable on the decision to grow sugarcane, as well as the results on the social indicators that can change our conclusions.

To understand the Rosenbaum bounds, let  $\pi_i$  represent the odds of student  $i$  receiving the treatment (growing cane);  $\pi_i/(1 - \pi_i)$  is the odds ratio. Assume the log odds ratio can be expressed as a generalized function of observables,  $X_i$ , and a binary, unobserved term,  $u_i$  that taking values in  $[0, 1]$ . Formally,

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \kappa(X_i) + \gamma u_i \quad (11)$$

Where  $\gamma$  represents the effect of  $u_i$  on the decision to grow sugarcane, and  $\kappa(\cdot)$  is a unknown function. Thus, the relative odds ratio of two observationally identical students is given by

$$\frac{\frac{\pi_i}{1 - \pi_i}}{\frac{\pi_j}{1 - \pi_j}} = \frac{e^{\kappa(X_i) + \gamma u_i}}{e^{\kappa(X_j) + \gamma u_j}} = e^{\gamma(u_i - u_j)} \quad (12)$$

where in the last passage, I use the fact that  $\kappa(X_i) = \kappa(X_j)$  for two individuals with the same observed characteristics.

Therefore, if there are no differences in the unobserved variables ( $u_i = u_j$ ) and these variables do not influence the probability of participation ( $\gamma = 0$ ), the odds ratio will be equal to 1, implying there is no selection bias. It follows, then, that if their odds of participation differ – that is, if the odds ratio is different from 1 – any selection bias can

---

<sup>3</sup> For a more complete discussion, see Rosenbaum (2002)

only be due to the presence of unobservable factors. The sensitivity analysis evaluates how much the program's effect is altered by the change in the values of  $\gamma$  and  $u_i - u_j$ .

In practice, this means examining the bounds of the odds ratio of participation. Rosenbaum (2002) showed that (12) is bounded as follows:

$$\frac{1}{\Gamma} \leq \frac{\pi_i(1-\pi_j)}{\pi_j(1-\pi_i)} \leq \Gamma \quad (13)$$

where  $\Gamma = e^\gamma$ .

The matched regions have the same probability of participation only if  $e^\gamma = 1$ . However, if  $e^\gamma = 2$ , regions who appear to be similar (in terms of  $X$ ) could differ in their odds of receiving the treatment by as much as a factor of 2. Hence, according to Rosenbaum (2002),  $e^\gamma$  is a measure of the degree of departure from a study free of selection bias.

#### 4. Remarks

Unfortunately, the work is in the initial phase, which make impossible to advance in conclusions.

#### 5. References

ABAG – Associação Brasileira do Agronegócio. Disponível em < <http://www.abag.com.br> > acesso: 9/12/2013.

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.

Alves, F. J. C. 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.

Arbex, M. A. et al. Queima de biomassa e efeitos sobre a saúde. *J Bras Pneumol* 2004; 30(2) 158-175. São Paulo – FMUSP. Acesso em 25 de fevereiro de 2014

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.

Balat, M., Balat, H. Recent Trends in global production and utilization of bio-ethanol fuel. *Applied Energy*. 2008, p. 2773-2282. Acesso em 3 de fevereiro de 2014.

Basaldi, O. V., Gomes, E. G. Mercado assalariado da cana de açúcar: Uma análise das regiões tradicionais no período de 1992 a 2006. *EMBRAPA*. Brasil, 20 de Julho de 2008. p. 1-18. Acesso em 8 de janeiro de 2014.

Boundy, B., Diegel, S. W., Lynn, W. Davis, S. C. Biomass Data Book: Edition 4. *US department of Energy*. p. 9-12, publicado em setembro de 2011. Acesso em 25 de outubro de 2013.

Camargo, A. S. and Toneto-Jr, R. Indicadores socioeconômicos e a cana de açúcar no Estado de São Paulo. *In Anais do IWorkshop do Observatório do Setor Sucroalcooleiro*. Ribeirão Preto, pg 1-11, 2008.

Cavalcanti, C. Concepções da economia ecológica: suas relações com a economia dominante e a economia ambiental. *Estudos. avançados* [online]. 2010, vol.24, n.68, p. 65-70. ISSN 0103-4014.

Chagas, A. L. S, Azzoni, C. R. Toneto-Jr, R. A Spatial Propensity Score Matching Evaluation of the Social Impacts of Sugarcane Growing on Municipalities in Brazil. *International Regional Science Review*. Brasil, vol. 35, n. 1, p. 48-69, 2012.

Chagas, A. L. S.; Almeida, A. N.; Azzoni, C. R. Sugarcane burning and human health: an analysis using spatial difference in difference. In: VII World Conference of the Spatial Econometrics Association, 2013, Washington. *Annals of VII World Conference of the Spatial Econometrics Association*, 2013.

Chagas, A. L. S. Socio-Economic and Ambient Impacts of Sugarcane Expansion in Brazil: Effects of the Second Generation Ethanol Production. In: Silvio Silverio da Silva, Anuj Kumar Chandel. (Org.). *Biofuels in Brazil - Fundamental Aspects, Recent Developments, and Future Perspectives*. 1 ed., 2014.

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.

IEA – Instituto de Economia Agrícola/SP. Previsões e Estimativas das Safras Agrícolas no Estado de São Paulo, Safras 2005/06. IEA/4º Levantamento, 2006. Disponível em [www.iea.sp.gov.br](http://www.iea.sp.gov.br). Acesso em 19 de março de 2013.

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., Ortiz, L. Agronegócio e biocombustíveis: uma mistura explosiva: impactos na expansão das monoculturas para a produção de biocombustíveis. *Fundação Heirich Böll*, Brasil, p.2, 2006. Acesso em 18 de março de 2013.

Nyko, D., Garcia, Jorge L. F., Milanez, A. Y., Dunham, F. B., A corrida tecnológica pelos biocombustíveis de segunda geração: uma perspectiva comparada. *BNDES*, setorial 32, Brasil, p.5-48, 2010. Acesso em 19 de março de 2013.

Piketty, M. G.; Menezes, T. M, and Duarte, J. B. N. A. Sugar Cane in Brazil, poverty and equity: evidences for the 1992-2006 period. In *Anais do XXXIV Congresso da ANPEC*, Salvador, 2008. ANPEC.

Raízen – Soluções de energia limpa e renovável. Produção de etanol. Disponível em <http://pt.raizen.com.br/pt-br/segmentos-de-atuacao/producao-de-etanol>. Acesso em 20 de Março de 2013.

Richard, E. Green Jobs: Biofuels: Careers in Biofuels. *Bureau Labour of Statistic*. Estados Unidos, Janeiro de 2013. Report 10. Disponível no link: [www.bls.gov/green/biofuels/biofuels.pdf](http://www.bls.gov/green/biofuels/biofuels.pdf) . Acesso em 5 de fevereiro de 2013

Salles, K., Itani, A. Condições de trabalho de cortadores de cana-de-açúcar. Anais do IV Congresso de Iniciação Científica. *Senac*. p.65-70. Acesso em 10 de fevereiro de 2014.

Toneto-JR, R.; Liboni, L. B. Evolução Recente do Mercado de Trabalho da Cana-de-Açúcar no Brasil (1995-2006). Organizações Rurais e Agroindustriais (UFLA), v. 10, p. 1-20, 2008.

UDOP – União dos produtores de bioenergia. Disponível em < <http://www.udop.br> >  
Data de acesso: 21/08/2013.

World Bank, 2008. World Development Report 2008. Agriculture Development. Washington D.C., World bank, 365 p. Acesso em 23 de Março de 2013.