Department of Economics- FEA/USP

Sugar Cane Burning and Human Health: a Spatial Difference-in-Difference Analysis

ANDRÉ LUIS SQUARIZE CHAGAS
ALEXANDRE N. ALMEIDA
CARLOS ROBERTO AZZONI

WORKING PAPER SERIES Nº 2015-47

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2015-47

Sugar Cane Burning and Human Health: a Spatial Difference-in-Difference Analysis

André Luis Squarize Chagas (achagas@usp.br)

Alexandre N. Almeida (alexnunes@ufscar.br)

Carlos Roberto Azzoni (cazzoni@usp.br)

Abstract:

The production of ethanol and sugar from sugar cane has sharply increased in Brazil in the last 20 years, in a process of substitution of biofuel for fossil fuels. The increase in the production and the expansion of the cultivated area might have impacts on human health and employment, especially at the regional level. So far, the harvest is basically manual, involving low-skill workers. The burning of the cane is meant to increase labor productivity and has been traditionally executed. However, the burning generates a massive quantity of smoke that spread in the region, reaching neighboring cities, thus becoming a potential threat to the human health. The objective of this paper is to measure the impact of burning on respiratory problems of children, teenagers and elderly people. We work with a balanced panel of 644 municipalities, from 2002 to 2011. We use a spatial difference-in-difference technique, to control for the effect of sugarcane burning on non-producing regions in the vicinity of producing regions. We conclude that sugar cane burning significantly increases the incidence of respiratory problems in producing regions. The use of a spatial diff-in-diff model allowed us to find out that the effect on non-producing nearby regions is also significant and quantitatively relevant, at least 66% of the effect on producing regions.

Keywords:: Sugar Cane Burning; Health Condition; Spatial Econometrics

JEL Codes: C14; C21; Q18.

Sugar Cane Burning and Human Health: a Spatial Difference-in-Difference Analysis

André Luis Squarize Chagas

Department of Economics, University of São Paulo, and CNPq (Proc. Proc. 481027/2011-4)

achagas@usp.br

Alexandre N. Almeida

Department of Economics, Federal University of São Carlos

alexnunes@ufscar.br

Carlos Roberto Azzoni

Department of Economics, University of São Paulo

cazzoni@usp.br

22 de julho de 2014

Abstract

The production of ethanol and sugar from sugar cane has sharply increased in Brazil in the last 20 years, in a process of substitution of biofuel for fossil fuels. The increase in the production and the expansion of the cultivated area might have impacts on human health and employment, especially at the regional level. So far, the harvest is basically manual, involving low-skill workers. The burning of the cane is meant to increase labor productivity and has been traditionally executed. However, the burning generates a massive quantity of smoke that spread in the region, reaching neighboring cities, thus becoming a potential threat to the human health. The objective of this paper is to measure the impact of burning on respiratory problems of children, teenagers and elderly people. We work with a balanced panel of 644 municipalities, from 2002 to 2011. We use a spatial difference-in-difference technique, to control for the effect of sugarcane burning on non-producing regions in the vicinity of producing regions. We conclude that sugar cane burning significantly increases the incidence of respiratory problems in producing regions. The use of a spatial diff-in-diff model allowed us to find out that the effect on non-producing nearby regions is also significant and quantitatively relevant, at least 66% of the effect on producing regions.

Key-Words: Sugar Cane Burning, Health Condition, Spatial Econometrics
Jel Classification: C14, C21, Q18

1 Introduction

The production of ethanol and sugar from sugar cane has sharply increased over the last 20 years in Brazil. The culture is one of the most important in Brazilian agriculture, representing about 10% of agricultural area, and close to 1% of GDP. The manufacturing of ethanol accounts for about 3.5% of manufacturing GDP, and the sector employs more than 6 million people. The increase in ethanol production is heavily influenced by governmental incentives to substitute biofuels for fossil fuels. The planted area with sugarcane doubled in the last twenty years, the largest increase in production in the world, and increasing the distance to India, the second most important producers, which now produces only 12.5% of Brazilian production.

This increase in sugarcane production and the expansion of new cultivated areas might have impacts on human health and employment, especially at the regional level. It brings about arguments related to 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 (Noronha, 2006). Some studies have shown that the balance of costs and benefits is positive, from the standpoint of the country as a whole(BNDES and CGEE, 2008), but the benefits for the cane growing regions may not be as evident, for the producing regions may disproportionately bear the negative impacts of the sector's presence. The majority of studies deal with the impacts on the labor market, and many authors stress the negative impacts of manual harvesting (Alves, 2006, 2007; Baccarin et al., 2008). Toneto-Jr and Liboni (2008) observe that sugarcane generates more jobs than soybean, and only slightly fewer than corn. Since the crop generates more value per hectare and more jobs as well, cane growing generates more income per area planted than other staple crops.

Given the importance of transportation costs in relation to the value of sugar cane, processing plants (sugar mills and/or ethanol distilleries) are located near the growing fields. This tends to increase local employment even more, adding workers in the manufacturing process and in transportation, maintenance, etc., adding indirect effects on the producing region. Chagas et al. (2011) assess the impact of sugarcane on regional social indicators, using spatial propensity score matching, which controls for the fact that sugarcane production in a particular region is not random. The results indicate that the presence of sugarcane is not relevant to determine their social conditions, whether for better or worse. It is thus likely that public policies, especially those focused directly on improving education, health and income generation/distribution, have much more noticeable effects on the municipal HDI than sugarcane production. Chagas (2009) considered

the effects of the crop on municipal revenues, and found out also that there is no evidence of any kind of influence.

The increasing importance of sugarcane production in the country keeps raising more and more arguments about its consequences. This paper provides a contribution to the discussion, by assessing its effects on the respiratory problems of children and adults. We employ a spatial difference-in-difference technique, to control for the effect of sugarcane growing on neighboring regions that do not produce it. Thus, we compare municipalities where burning occurs with others where it does not, over the time.

The burning is meant to increase the productivity of workers. It facilitates the harvest, easing access to the plants and reducing work hazards (dry leaves are harmful, there might be poisonous insects and snakes). It takes place in the beginning of harvest, which coincides with the dry season. Many studies relate sugarcane burning to increases in fine particulate matter, coarse particulate matter and black carbon concentration, especially in burning hours (Lara et al., 2005). Allen et al. (2004) observe increases in the concentration of substances as nitrite, sulfite, oxide of carbon, and others. The sugarcane is harvested by unskilled workers, mostly manually. The literature also indicates that exposition to classical pollutants (matter, sulfite, nitrite, oxide carbon, etc.) can affect negatively human health (Sicard et al., 2010), specially for young, elderly and woman people (Braga et al., 1999; Roseiro, 2002; Gonçalves et al., 2005).

The burning of sugarcane generates a massive quantity of smoke that spreads in the region, reaching cities and becoming a potential threat to the human health. Few studies have linked the burning of sugarcane straws with respiratory diseases in the producing regions. Although the pollution from sugarcane burning may be as harmful as the pollution from traffic and manufacturing Mazzoli-Rocha et al. (2008), many studies relate its impact on health, for specific municipalities or for larger regions:Arbex et al. (2000, 2004); Cançado et al. (2006); Arbex et al. (2007); Ribeiro (2008); Uriarte et al. (2009); Carneseca et al. (2012) . These studies consider only the local, or short-distance, effects of burning to respiratory health, ignoring the effects on neighboring municipalities.

This article is organized in four sections, including this introduction. The next section presents the methodology and the data utilized. The fourth section presents the results, and the fifth section contains the final remarks.

2 Methodology and data

2.1 The model

As usual in spatial studies, we consider the interrelations among regions. Therefore, we consider the possibility of the propagation of the treatment e?ect (sugarcane production) on other regions, treated and non-treated. Let y_{it} be a variable of interest, $\mathbf{x}$ a vector of observable characteristic, $\mathbf{w_i}$ a $n \times 1$ vector that associate a region to all others, and d_{it} a $n \times 1$ vector , such that $d_{it} = 1$ if the region 1 is treated in time t and zero, in otherwise. Additionally, consider two situations for each region: before and after the treatment. For the before treatment situation

$$\begin{aligned} y^b_{it,0} &= \mu(\mathbf{x}) + u_{it} \\ y^b_{it,1} &= y^b_{it,0} \end{aligned}$$

where, $y^b_{it,0}$ represents the dependent variable in the non-treated region, before the treatment; $y^b_{it,1}$ is the same variable for the treated region. After the treatment, there are two impacts, one on the treated region e another on the non-treated region. The impact on the non-treated region depends on the proximity of this region to treated one. For clarity, we consider both regions after the treatment such that

$$\begin{aligned} y^a_{it,0} &= \mu(\mathbf{x}) + \mathbf{w_i} d_{it}\beta + u_{it} \\ y^a_{it,1} &= y^a_{it,0} + \alpha \end{aligned}$$

The parameter α captures the direct effect of treatment on treated region; β captures the indirect effect of treatment on all regions, treated and non-treated, conditioned on the neighborhood of the treated region, which is captured by $\mathbf{w_i} d_{it}$. Defining D_{it} a region-i specific indicator of treatment in time t, we can define

$$y_{it} = (1 - D_{it})y_{it,0} + D_{it}y_{it,1} \tag{1}$$

Using the before and after definitions, we can compute three effects: ATE (Average Treatment Effect), ATET (Average Treatment Effect on the Treated), and ATENT (Ave-

rage Treatment Effect on the Non-Treated), as follows

$$\begin{aligned}
\text{ATE} &= E[y^a_{it,1} - y^b_{it,1}] - E[y^a_{it,0} - y^b_{it,0}] \\
&= \alpha \\
\text{ATET} &= E[y^a_{it,1} - y^b_{it,1}] \\
&= \alpha + \mathbf{w_i} d_{it}\beta \\
\text{ATENT} &= E[y^a_{it,0} - y^b_{it,0}] \\
&= \mathbf{w_i} d_{it}\beta
\end{aligned}$$

In matrix notation we have

$$Y = \mu(\mathbf{X}) + (\alpha + \mathbf{I}_t \otimes \mathbf{W}\beta)D + U \tag{2}$$

where Y is a matrix $nt \times 1$ of observations, $\mathbf{X}$ is a $nt \times k$ matrix of covariates, D is a dummy variable that indicate treated regions in each time, $\mathbf{I}_t$ is a square identity matrix of $t \times t$ dimension, $\mathbf{W}$ is a $n \times n$ weight matrix of neighborhood and U is a vector of errors of $nt \times 1$ dimension. The parameters to be estimated are μ, α and β.

$\mathbf{I}_t \otimes \mathbf{W}D\beta$ represents the indirect effect of treatment on both region, treated and non-treated. This is a additional effect, not estimated in general [1]. However, this is an average effect, and does not consider the possibility that the indirect effect could be different in the treated and non-treated regions. For example, if the indirect effect on the treated region is small, because the direct effect is more important, and, in the same time, the indirect effect on non-treated region is profound, because is the only effect to impact this region, then, when estimating β as an average indirect effect to all region, we can underestimate the real effect of the treatment, because β will be estimate as an average of indirect effect on treated and indirect effect on the non-treated regions.

Consider, for clarity, the follow decomposition in $\mathbf{W}$ matrix,

[1] The exception are (Angelucci and De Giorgi, 2009; Kaboski and Townsend., 2012; Berniell et al., 2013). The difference of our work to these is that with the spatial econometrics it is possible to control for different structures of neighborhood, not captured by these studies

$$\mathbf{I}_t \otimes \mathbf{W} = \mathbf{W_{11}} + \mathbf{W_{12}} + \mathbf{W_{21}} + \mathbf{W_{22}}$$

where

$$\begin{aligned}
\mathbf{W_{11}} &= \text{diag}(D) \times (\mathbf{I}_t \otimes \mathbf{W}) \times \text{diag}D) \\
\mathbf{W_{12}} &= \text{diag}(D) \times (\mathbf{I}_t \otimes \mathbf{W}) \times \text{diag}(1-D) \\
\mathbf{W_{21}} &= \text{diag}(1-D) \times (\mathbf{I}_t \otimes \mathbf{W}) \times \text{diag}(D), \text{ and} \\
\mathbf{W_{22}} &= \text{diag}(1-D) \times (\mathbf{I}_t \otimes \mathbf{W}) \times \text{diag}(1-D)
\end{aligned}$$

As D is a dummy variable associated to treatment information, and diag(D) is a $nt \times 1$ matrix with D in the principal diagonal and zero elsewhere. Then, $\mathbf{W_{ij}}$ represent the neighborhood effects of the j-region on i-region . Substituting in (2), results

$$Y = \mu(\mathbf{X}) + [\alpha + (\mathbf{W_{11}} + \mathbf{W_{12}} + \mathbf{W_{21}} + \mathbf{W_{22}})\beta]D + U$$

Then, it is clear that β represents an average effect, as we mentioned above. A more realistic model consider different effect for dissimilar $\mathbf{W}$ matrix. As, by construction, $\mathbf{W_{12}}D$ and $\mathbf{W_{22}}D$ are a $\mathbf{0}$-vector, the unrestricted model is

$$Y = \mu(\mathbf{X}) + [\alpha + (\mathbf{W_{11}}\beta_1 + \mathbf{W_{21}}\beta_2)]D + U \tag{3}$$

The models in (2) and (3) represent a spatial dif-in-dif model (SDID model). It is important to register that they do not contain a traditional spatial interaction effect, like in the SAR and SEM models Anselin (1988); LeSage and Pace (2009). But, we can model the control effects, $\mu(\mathbf{X})$, including an auto-regressive spatial term, or the error as a spatial error model,

$$\mu(\mathbf{X}) = \rho(\mathbf{I_t} \otimes \mathbf{W})Y + \mathbf{X}\gamma'$$

and/or

$$E = \lambda(\mathbf{I_t} \otimes \mathbf{W})U$$

In the first equation $\mathbf{X}$ is a $n \times k$ vector of observable characteristics, $\mathbf{W}$ is a spatial weight matrix of $n \times n$ dimension, γ is a $1 \times k$ parameter vector to be estimated, and ρ is the spatial auto-regressive parameter. And, in the second one, E is an error vector, does not spatially correlated, λ is the spatial error parameter to be estimated.

In this way, a complete version of 2 and 3 models is

$$Y = [\mathbf{I_{nt}} - \rho(\mathbf{I_t} \otimes \mathbf{W})]^{-1}\{\mathbf{X}\gamma' + [\alpha + (\mathbf{W_{11}}\beta_1 + \mathbf{W_{21}}\beta_2)]D + [\mathbf{I_{nt}} - \lambda(\mathbf{I_t} \otimes \mathbf{W})]^{-1}U\} \quad (4)$$

2.2 Data

We use a balanced panel of 644 municipalities, from 2002 to 2011. Information of annual sugarcane production, planted area, and harvested area for each municipality is provided by IBGE, the Brazilian statistical agency. The planted area with sugarcane in São Paulo state represents nearly 50% of the total planted area in the country. The total area of Brazil is over 800 million hectares of landmass, of which over 300 million are suitable for farming and ranching activities. At present, only about 60 million are used to grow permanent and temporary crops and some 200 million are used for animal ranching. This means that there is an ample amount of land available for cane and other crops. There is another possible source of available land, provided by the recovery of pasture areas and from productivity gains in general. Hence, the idea of sugar cane production growth generating a conflict with other crops and even forests does not seem to be present at the national scale Chagas et al. (2008).

In the figure 1 we mapping the evolution of sugarcane production in São Paulo state, by municipalities, in the period 2002-2011[2]. The increase in production in northwest part of the state is quite clear, starting from the first year. The production sprawl to the west of the state over areas previously occupied by cattle ranching is also evident.

[FIGURE 1 HERE]

[2]We select some years in this periods, but the evolution is evident.

We consider as treated, all region in which the sugarcane area represents at least 6.7% of total agricultural area, the median production area. Table 1 1 reports the number of treated region in each year, and the proportion of total agricultural area of the state. In period from 2002 to 2011, the sugarcane area grew from 38.4% to 62.4%.

[TABLE 1 HERE]

Our variable of interest is the number of people hospitalized because of respiratory problems (per thousand), at the municipality level. The data was provided by DATA-SUS, from the Department of Public Health[3] and includes information on both private and public institutions. Figure 2 portrays the evolution of hospitalization in São Paulo state, by region, from 2002 until 2011, showing a reduction in the number of cases of hospitalization over time.

[FIGURE 2 HERE]

This result can be due to the reduction of sugarcane burning, since recent legislation prohibits burning in certain areas and times. The controlled burning of sugarcane is regulated by the federal act (Act 2661/98) and in the State of São Paulo has a specific law more restrictive (State Law 11.241/02). Besides the regulatory pressures to reduce the emission of pollutants and their harmful effects, there are economic benefits in the use of cane by-products, such as straws, leaves and bagasse. There are also issues related to the labor market, as the increasing formalization of labor and the enhancement of the workforce. An Environmental Protocol signed between mill owners, sugarcane producers and the government of the state of São Paulo establishes the end of the burning in the mechanized areas until 2014, and in all other areas, in 2017. It is possible that the trend in hospitalization numbers will follow the change in the burning practice. Table 2 emphasizes this point.

[TABLE 2 HERE]

Therefore, we introduce a trend variable to take it into account. Additionally, we consider control variables at the municipality level, such as the proportion of workers in the population, the mean income, proportion of elderly and young people, and the proportion of woman in the population. The ?rst two variables control for socio-economic

[3]http://www2.datasus.gov.br/DATASUS/index.php.

conditions, since they may influence the family?s ability to prevent health problems. The remaining variables control for the susceptibility of being affected by respiratory health problems in different age groups, as suggested by the literatureBraga et al. (1999); Roseiro (2002); Gonçalves et al. (2005). Table 3 reports descriptive statistics about the variables considered, and the table 4 reports the linear correlation matrix.

[TABLE 3 HERE]

[TABLE 4 HERE]

3 Preliminary results

This section presents preliminaries results of our estimations of two models, based on expressions (2) and (3). In the first one we consider only the average indirect treatment effect, and test for the pooled specification, the spatial fixed effect and the pooled spatial error model. We use the Elhorst's routine for panel data models (Elhorst, 2010a,b, 2011). Elhorst consider a ML estimation, mainly because the number of studies considering IV/GMM estimators of spatial panel data models is still relatively sparse [4].

Elhorst (2010b) provides a Matlab routines to estimate spatial panel data models, including the bias correction procedure proposed by (Lee and Yu, 2010) if the spatial panel data model contains spatial and/or time-period fixed effects, the direct and indirect effects estimates of the explanatory variables proposed by (LeSage and Pace, 2009), and a selection framework to determine which spatial panel data model best describes the data.

We consider a k-nearest distance matrix of neighbor. To choose the order of the k-nearest neighbor we estimate a pooled model, without spatial effects, and compute the Moran I on the residual of this regression. We replicate this procedure to k between 3 and 20, and choose the order that maximize the Moran I. Thus, we choose the matrix for which there are the greatest spatial auto correlation.

Then, we compute three models, as follow

[4]One exception is (Kelejian et al., 2006), who considered IV estimation of a spatial lag model with time-period fixed effects. They point out that this model cannot be combined with a spatial weights matrix whose non-diagonal elements are all equal to $1/(N?1)$. In this situation, the spatially lagged dependent is asymptotically proportional and thus collinear with the time-period fixed effects as N goes to infinity.

$$
\begin{aligned}
&\text{pooled model} \\
y_{it} &= \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta \mathbf{w_i'} d_t + u_{it} \\
&\text{sar model} \\
y_{it} &= \rho \mathbf{w_i'} y_t + \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta \mathbf{w_i'} d_t + u_{it} \\
&\text{sem model} \\
y_{it} &= \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta \mathbf{w_i'} d_t + u_{it} + \lambda \mathbf{w_i'} u_t
\end{aligned}
$$

where y_{it} is a hospitalizations per thousand in inhabitants, in region i in time t; $\mathbf{X_{it}}$ represents a vector of control variables, including a constant term; $\mathbf{w_i}$ is a vector of neighbor weights, u_{it} is a error term; D_{it} is a indicator of treatment and d_t is vector of all indicators in t; y_t and u_t are vectors for all region, i time t. The parameters ρ, α, γ, and β are estimated. This set of model estimate a indirect effect of the treatment as an average effect between all region, treated and non-treated.

A less restrictive set of model is

$$
\begin{aligned}
&\text{pooled model} \\
y_{it} &= \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta_1 \mathbf{w'_{NT}} d_t + \beta_2 \mathbf{w'_{TT}} d_t + u_{it} \\
&\text{sar model} \\
y_{it} &= \rho \mathbf{w_i'} y_t + \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta_1 \mathbf{w'_{NT}} d_t + \beta_2 \mathbf{w'_{TT}} d_t + u_{it} \\
&\text{sem model} \\
y_{it} &= \mathbf{x_{it}}\gamma' + \alpha D_{it} + \beta_1 \mathbf{w'_{NT}} d_t + \beta_2 \mathbf{w'_{TT}} d_t + u_{it} + \lambda \mathbf{w_i'} u_t
\end{aligned}
$$

where $\mathbf{w_{NT}}$ is a vector with only weight of neighbor treated related to neighbor non-treated, and $\mathbf{w_{TT}}$ is a vector with weight of neighbor treated related to neighbors treated too. To choose between the models, we use a Lagrange Multiplier (LM) test for a spatially lagged dependent variable, for spatial error correlation, and their counterparts robustified against the alternative of the other form [5] (Anselin et al., 1996). These tests have become very popular in empirical research. Recently, Anselin L (2006) also specified the first two LM tests for a spatial panel

[5]Software programs have built-in routines that automatically report the results of these tests. Matlab routines have been made available by Donald Lacombe at ?http://oak.cats.ohiou.edu/ lacombe/research.html?. Elhorst provided the routines for the panel data case.

$$LM_\rho = \frac{[\mathbf{e}'(\mathbf{I_t} \otimes \mathbf{W})\mathbf{Y}/\hat{\sigma}^2]^2}{J} \quad \text{and} \quad LM_\lambda = \frac{[\mathbf{e}'(\mathbf{I_t} \otimes \mathbf{W})\mathbf{e}/\hat{\sigma}^2]^2}{T \times T_W}$$

where $\mathbf{e}$ denotes the residual vector of a pooled regression model without any spatial or time-specific effects, or of a panel data model with spatial and/or time period fixed effects. Finally, J and T_W defined by

$$\begin{aligned} J &= \frac{1}{\hat{\sigma}^2}[(\mathbf{I_T} \otimes \mathbf{W})\mathbf{X}\hat{\beta})'(\mathbf{I_{NT}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I_T} \otimes \mathbf{W})\mathbf{X}\hat{\beta} + TT_W\hat{\sigma}^2], \\ T_W &= tr(\mathbf{WW} + \mathbf{W}'\mathbf{W}) \end{aligned}$$

and "tr" denotes the trace of a matrix. In view of these formulas, the robust counterparts of these LM tests for a spatial panel will take the form

$$\begin{aligned} \text{robust}LM_\rho &= \frac{[\mathbf{e}'(\mathbf{I_t} \otimes \mathbf{W})\mathbf{Y}/\hat{\sigma}^2]^2 - [\mathbf{e}'(\mathbf{I_t} \otimes \mathbf{W})\mathbf{e}/\hat{\sigma}^2]^2}{J - TT_W}, \\ \text{robust}LM_\lambda &= \frac{[\mathbf{e}'(\mathbf{I_t} \otimes \mathbf{W})\mathbf{e}/\hat{\sigma}^2 - TT_W/J \times \mathbf{e}'(\mathbf{I_T} \otimes \mathbf{W})\mathbf{Y}/\hat{\sigma}^2]^2}{T \times T_W} \end{aligned}$$

Elhorst (2010a) indicates that these tests in panel data models, when they include spatially lagged independent variables, must be investigated further.

The result are reported in table 5. The first three models consider the average indirect effect of the treatment; the last one considers the less restricted model, with different impacts for different groups of region.

For pooled models (model 1 and 4), the results are intuitive, in general. The proportion of children in the population presents counterintuitive negative, and significant, coefficients in both models. All control variables are significant, except the proportion of women in the population, which is not significant in all models. In both models, the treatment presents a strong impact, much larger than in the other models, and the indirect effect is negative and non-significant (Model 1).

This is the kind of result one would have applying conventional techniques. These fail to take into account properly the interrelations between the regions. But, if we compare the difference between Model 1 and Model 4, we can see an interesting result.

The separation of the indirect effects on the treated and non-treated regions increases the direct effect on the treated regions. It also gives a counterintuitive result related to the indirect effects on the treated region, which is negative. This result, in particular, might be due to relevant omitted variables, despite its highly significant level. In fact, when we include spatial controls, this result does not remain. For the non-treated region, the effect is positive and very significant. The indirect effect on the non-treated region is about half of the effect on the treated regions.

As for the indirect effect, the difference in results remains, when we compare the average indirect effect to the separated effect. The impact on the non-treated regions seems very important and very different than the indirect impact on the treat regions. In the SAR model (model 5), the indirect effect on the non-treated (0.78) is about 80% of the direct effect of the treatment (0.97). If we estimate to the mean, the indirect effect in Model 2 (0.44) is only 66% of the direct effect (0.65), and both are smaller.

These are preliminary results only and other tests are needed before se have stronger conclusions. A counterintuitive result relates to the fact that the SAR model prevails over the SEM model. It is not reasonable assume that hospitalization due respiratory diseases cause more hospitalization on neighbors. It would be more intuitive that the error in the region was impacted to the neighbors error. It is possible that some omitted relevant variable remain in this case.

4 Final remarks

The increase in ethanol production in recent years provoked many debates. In this paper we try to address on of these, more specifically, the impact on human health due to the burning of sugarcane. We proposed a new methodology to evaluate the impacts, using aggregate data at municipality level and information about the neighbors. More specifically, we consider the indirect effect of the treatment on the treated and on non-treated regions.

Our results suggest that our method makes it possible to better identify the impact, not only in the treated regions, but on the non-treated too. The research is in the beginning, but the results look promising.

Referências

Allen, A., Cardoso, A. and Rocha, G. da (2004). Influence of sugar cane burning on aerosol soluble ion composition in southeastern brazil. *Atmospheric Environment* 38: 5025 – 5038.

Alves, F. J. C. (2006). Por que morrem os cortadores de cana? *Saúde e Sociedade* 15: 90–98.

Alves, F. J. C. (2007). *Migrantes: trabalho e trabalhadores no complexo agroindustrial canavieiro - os heróis do agronegócio brasileiro.*. São Carlos: EDUFSCar, chap. 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? 21–54.

Angelucci, M. and De Giorgi, G. (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review* 99: 486–508.

Anselin, L. (1988). *Spatial Econometrics: Methods and models*. Dordecht: Kluwer Academic Publishers.

Anselin, L., Bera, A. K., Florax, R. and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26: 77–104.

Anselin L, J. H., Le Gallo J. (2006). *The econometrics of panel data, fundamentals and recent developments in theory and practice*. Dordrecht: Kluwer Academic Publishers, chap. Spatial panel econometrics. 3rd ed., 901–969.

Arbex, M. A., Bohm, G. M., Saldiva, P. H. and Conceição, G. (2000). Assessment of the effects of sugar cane plantation burning on daily counts of inhalation therapy. *J Air Waste Manag Assoc* 50: 1745–9.

Arbex, M. A., Cançado, J. E. D., Pereira, L. A. M., Braga, A. L. F. and Saldiva, P. H. N. (2004). Queima de biomassa e efeitos sobre a saúde. *Bras Pneumol* 30: 158–175.

Arbex, M. A., Martins, L. C., Oliveira, R. C., Pereira, L. A. A., Arbex, F. F., Cançado, S. P. H. N., J. E. D. and Braga, A. L. F. (2007). Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. *Journal of Epidemiology and Community Health* 61: 395–400.

Baccarin, J. G., Alves, F. J. C. and Gomes, L. F. C. (2008). Emprego e condições de trabalho dos canavieiros no centro-sul do Brasil, entre 1995 e 2007. In *Anais do XLVI Congresso da Sober*. Rio Branco: Sociedade Brasileira de Economia e Sociologia Rural.

Berniell, L., Mata, D. de la and Valdés, N. (2013). Spillovers of health education at school on parents' physical activity. *Health Economics* n/a: n/a–n/a.

BNDES and CGEE (2008). *Bioetanol de cana-de-açúcar: energia para o desenvolvimento*. Rio de Janeiro: BNDES.

Braga, A. L., CONCEIÇÃO, G. M., Pereira, L. A., Kishi, H. S., Pereira, J. C., Andrade, M. F., GONÇALVES, F. L., Saldiva, P. H. and Latorre, M. R. (1999). Air pollution and pediatric respiratory hospital admissions in sÃ£o paulo, brazil. *Journal of Environmental Medicine* 1: 95–102.

Cançado, J. E. D., Saldiva, P. H. N., Pereira, L. A. A., Lara, L. B. L. S., Artaxo, P., Martinelli, L. A., Arbex, M. A., Zanobetti, A. and Braga, A. L. F. (2006). The impact of sugar cane burning emissions on the respiratory system of children and the elderly. *Environ Health Perspect.* 114: 725–729.

Carneseca, E. C., Achcar, J. A. and Martinez, E. Z. (2012). Association between particulate matter air pollution and monthly inhalation and nebulization procedures in Ribeirão Preto, São Paulo State, Brazil. *Cadernos de Saúde Pública* 28: 1591 – 1598.

Chagas, A. L. S. (2009). Três ensaios sobre o setor produtor de cana-de-açúcar no Brasil. Ph. d. dissertation,, University of São Paulo, São Paulo.

Chagas, A. L. S., Toneto-Jr, R. and Azzoni, C. R. (2008). Teremos que trocar energia por comida? análise do impacto da expansão da produção de cana-de-açúcar sobre o preço da terra e dos alimentos. *Economia (Brasília)* 39-61: 39–61.

Chagas, A. L. S., Toneto-Jr, R. and Azzoni, C. R. (2011). A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in brazil. *International Regional Science Review* 35: 48 – 69.

Elhorst, J. P. (2010a). *Handbook of applied spatial analysis*. Springer, chap. Spatial panel data models. 377–407.

Elhorst, J. P. (2010b). Matlab software for spatial panels. In *Paper presented at 4th World Conference of the Spatial Econometric Association, Chicago*.

Elhorst, J. P. (2011). Spatial panel models. York, UK: The University of York.

Gonçalves, F., Carvalho, L., Conde, F., Latorre, M., Saldiva, P. and Braga, A. (2005). The effects of air pollution and meteorological parameters on respiratory morbidity during the summer in São Paulo City. *Environment International* 31: 343 – 349.

Kaboski, J. P. and Townsend., R. M. (2012). The impact of credit on village economies. *American Economic Journal: Applied Economics* 4: 98–133.

Kelejian, H. H., Prucha, I. R. and Yuzefovich, Y. (2006). Estimation problems in models with spatial weighting matrices which have blocks of equal elements. *Journal of Regional Science* 46: 507–515.

Lara, L., Artaxo, P., Martinelli, L., Camargo, P., Victoria, R. and Ferraz, E. (2005). Properties of aerosols from sugar-cane burning emissions in southeastern brazil. *Atmospheric Environment* 39: 4627 – 4637.

Lee, L. F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154: 165–185.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press.

Mazzoli-Rocha, F., MAGALHÃES, C. B., Malm, O., Saldiva, P. H., Zin, W. A. and Faffe, D. S. (2008). Comparative respiratory toxicity of particles produced by traffic and sugar cane burning. *Environ Res.* 108: 35–41.

Noronha, S. e. a. (2006). *Agronegócio e biocombustíveis: uma mistura explosiva - Impactos da expansão das monoculturas para a produção de Bioenergia*. Rio de Janeiro: Núcleo Amigos da Terra.

Ribeiro, H. (2008). Queimadas de cana-de-açúcar no Brasil: efeitos à saúde respiratória. *Revista de Saúde Pública* 42: 370–6.

Roseiro, M. N. V. (2002). Morbidade por problemas respiratórios em Ribeirão Preto-SP, de 1995 a 2001, segundo indicadores ambientais, sociais e econômicos. Tese de doutorado, Universidade de São Paulo, escola de Enfermagem de Ribeirão Preto.

Sicard, P., Mangin, A., Hebel, P. and Malléa, P. (2010). Detection and estimation trends linked to air quality and mortality on french riviera over the 1990-2005 period. *Sci Total Environ* 408: 1943–50.

Toneto-Jr, R. and Liboni, L. (2008). Mercado de Trabalho da cana-de-açúcar. In *Anais do I Workshop do Observatório do Setor Sucroalcooleiro*. Ribeirão Preto.

Uriarte, M., Yackulic, C. B., Cooper, T., Flynn, D., Cortes, M., Crk, T., Cullman, G., McGinty, M. and Sircely, J. (2009). Expansion of sugarcane production in sÃ£o paulo, brazil: Implications for fire occurrence and respiratory health. *Agriculture, Ecosystems & Environment* 132: 48 – 56.

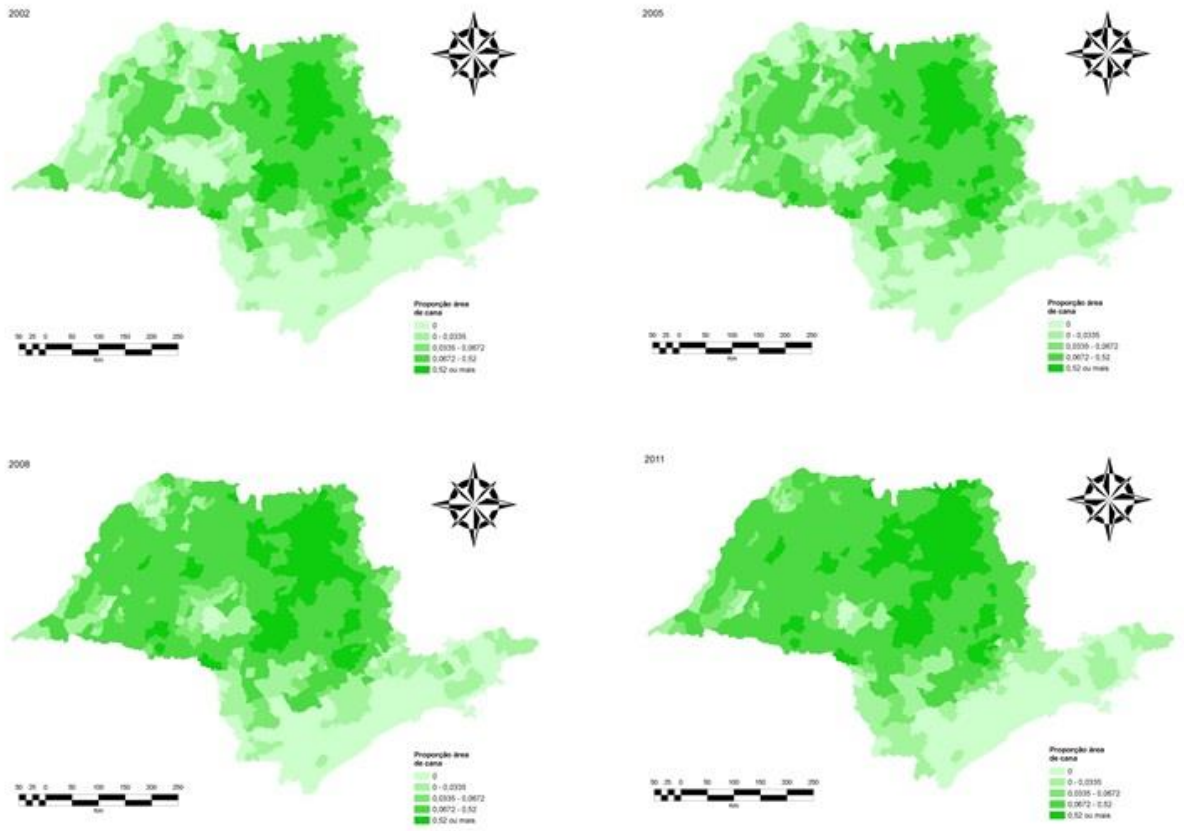

Figura 1: Sugarcane production in São Paulo state by municipality, 2002, 2005, 2008, and 2011
Source: IBGE, Municipal Agricultural Research.

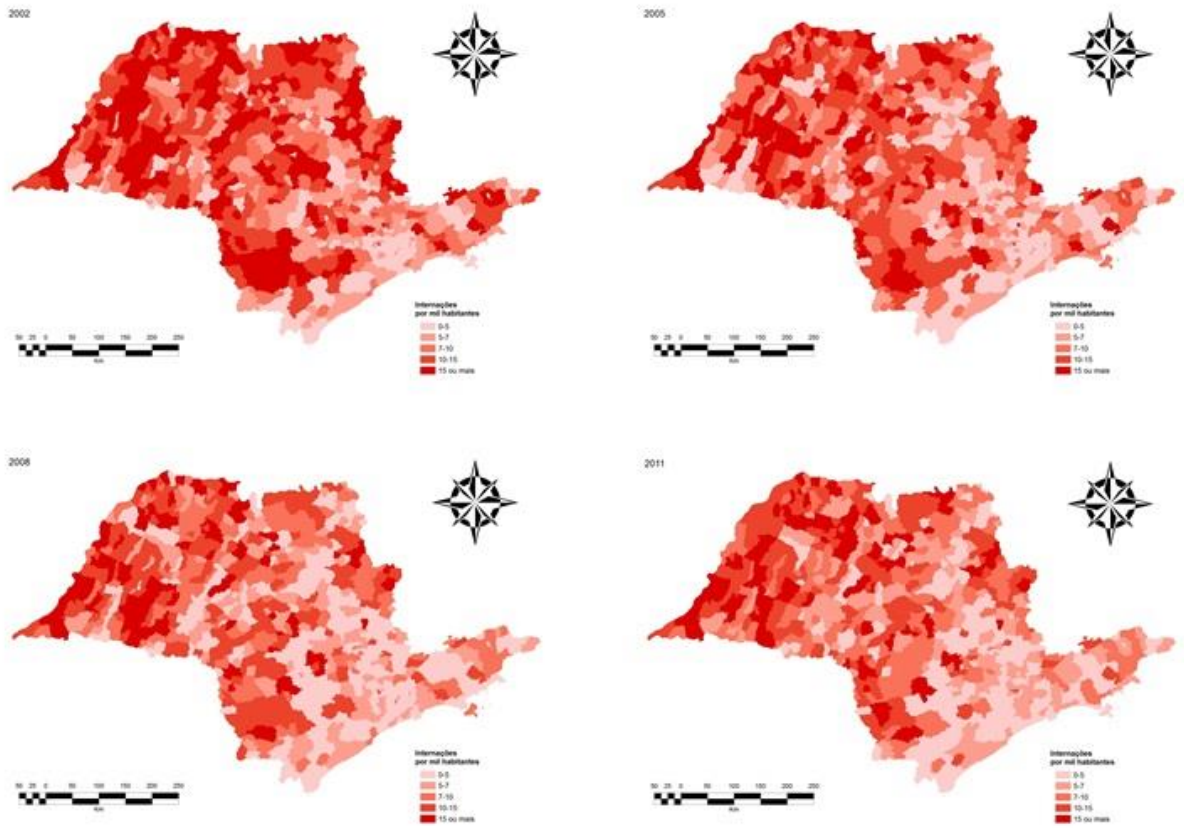

Figura 2: Hospitalization due to respiratory problems in São Paulo state by municipality, 2002, 2005, 2008, and 2011
Source: Datasus, Health Ministry.

Tabela 1: Number of treated region and proportion on total, 2002-2011

Year	Number of treated region	Prop. Total
2002	247	0.384
2003	255	0.396
2004	259	0.402
2005	278	0.432
2006	306	0.475
2007	333	0.517
2008	360	0.559
2009	383	0.595
2010	397	0.616
2011	402	0.624

Source: IBGE, authors calculations.

Tabela 2: Hospitalization due respiratory health problem, by region, 2002-2011

Year	Hospitalization due to respiratory problems			
	Mean	Std. Dev	Maximum	Minimun
2002	11.678	7.216	45.366	0.810
2003	11.337	7.332	41.895	0.272
2004	10.742	6.831	45.326	0.000
2005	9.996	6.291	42.415	0.000
2006	10.632	6.860	54.111	0.262
2007	9.780	6.271	54.756	1.175
2008	8.745	5.690	37.122	0.949
2009	9.798	6.367	39.896	1.291
2010	9.355	6.204	46.392	1.166
2011	9.264	5.967	45.065	0.458

Tabela 3: Summary statistics for the variables

Variable	**Mean**	**Sdt. Dev.**	**N**
Treatment	0.5000	0.5000	6440
$\mathbf{W_{11}}D$	0.4199	0.4625	6440
$\mathbf{W_{11}}D$	0.0752	0.2073	6440
Workers	0.1957	0.1396	6440
Income	6.4751	0.2857	6440
Old	0.1220	0.0295	6440
Child	0.2345	0.0331	6440
Female	0.9801	0.0650	6440

Tabela 4: Linear correlation between variable of the model

Variabels	Treatment	$\mathbf{W_{11}}D$	$\mathbf{W_{12}}D$	Workers	Income	Old	Child	Female
Treatment	1.0000							
$\mathbf{W_{11}}D$	0.9080	1.0000						
$\mathbf{W_{12}}D$	-0.3626	-0.3292	1.0000					
Workers	0.1395	0.1499	-0.0331	1.0000				
Income	0.1464	0.1742	-0.1312	0.3482	1.0000			
Old	0.0605	0.0513	0.1541	-0.0992	-0.2296	1.0000		
Child	-0.2404	-0.2399	-0.1126	-0.1218	-0.1602	-0.7256	1.0000	
Female	-0.0567	-0.0453	-0.0107	0.1581	0.2548	0.1098	-0.0098	1.0000

Tabela 5: Preliminary results

	Variable independent: hospitalization											
variable	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
trend	-2.8888		-1.8905		-2.4929		-2.8289		-1.7804		-2.3914	
	(-8.6207)	***	(-2.8006)	***	(-3.5428)	***	(-8.4481)	***	(-2.6239)	***	(-3.3866)	***
treatment (D)	1.2647		0.6536		0.6875		2.4427		0.9730		1.0738	
	(5.0803)	***	(3.5379)	***	(3.8217)	***	(6.5633)	***	(3.5426)	***	(3.7341)	***
WD	-0.2284		0.4350		0.6009							
	(-.801)	ns	(1.6539)	*	(2.1909)	**						
W11D							-1.4001		0.1380		0.2431	
							(-3.5009)	***	(.4263)	ns	(.7067)	ns
W12D							1.0108		0.7797		1.0131	
							(2.4935)	**	(2.2761)	**	(2.7798)	***
Controls												
workers	-1.9435		-1.6324		-1.5049		-1.9775		-1.5886		-1.4700	
	(-3.2286)	***	(-2.7553)	***	(-2.5586)	**	(-3.2893)	***	(-2.6788)	***	(-2.4991)	**
income	-4.7911		-0.3993		-0.5210		-4.6447		-0.3858		-0.5118	
	(-14.3312)	***	(-1.0103)	ns	(-1.2949)	ns	(-13.9099)	***	(-.976)	ns	(-1.2723)	ns
old	23.0847		16.0961		17.5064		22.8372		15.2617		16.7494	
	(6.913)	***	(1.4071)	ns	(1.4544)	ns	(6.8377)	***	(1.333)	ns	(1.3905)	ns
child	-25.8023		15.1254		15.4772		-24.3103		16.8650		17.0221	
	(-10.1153)	***	(1.3289)	ns	(1.3113)	ns	(-9.6967)	***	(1.4751)	ns	(1.4382)	ns
female	1.8961		2.6305		3.0166		1.9315		2.5046		2.8885	
	(1.5257)	ns	(.8646)	ns	(.9619)	ns	(1.5567)	ns	(.823)	ns	(.9209)	ns
rho			0.2060						0.205991			
			(14.3192)	***					(14.3227)	***		
lambda					0.2070						0.2090	
					(.)	ns					(.)	ns
Constant	43.9791		3.4448		5.8086		42.4605		3.05174772		5.4684	
	42.8369	***	(.9656)	ns	(1.6288)	ns	(43.3934)	***	(.8533)	nd	(1.5301)	ns
R-square	0.1179		0.8136		0.8046		0.1203		0.8136		0.8047	
corr-rsqa	0.1179		0.074		0.0744		0.1203		0.0743		0.0748	
s2	38.1749		8.0654		8.0584		38.0703		8.0659		8.0523	
log-likel	-20865.78		-15907.099		-15905.25		-20856.945		-15905.971		-15903.782	
Nobs	6440		6440		6440		6440		6440		6440	
Nvar	9		8		9		10		9		10	
# Fixed effects			644		644				644			
# interactions			1		14				1		16	
rho min			-1		-0.99				-1		-0.99	
rho max			1		0.99				1		0.99	
LM lag, prob				771.8073	0.0000					760.877	0	
LM lag rob, prob				689.6016	0					681.424	0	
LM error, prob				83.9612	0					80.9182	0	
LM error rob, prob				1.7555	0.1852					1.4656	0.226	

Notes: ns = not significant.
*** = significant to 1%.
** = significant to 5%.
* = significant to 10%.