Sugar Cane Burning and Human Health: An Analysis Using Spatial Difference in Difference

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

20 de June de 2013

Abstract

The production of ethanol and sugar from sugar cane has sharply increased for the last 20 years. If there are overall incentives to substitute the consumption of fossil fuels by biofuels, the increase of production and the expansion of new cultivated areas of sugar cane have eventually an impact on human health and employment mainly at regional levels. To harvest the crop–mostly manually done by low-skill workers–the practice of burning to clean dry grasses and poisonous insects has been executed for years during the dry season, what increase the productivity of workers. However, the burning generates a massive quantity of smoke that spread in the region reaching cities and becoming a potential threat to the human health. The main objective of this paper is to investigate the impact of the burning prohibition of sugar cane on respiratory problems of children, teenagers and elderly people. This analysis proposes a spatial difference-in-difference technique, to control the effect of treatment (the sugarcane culture) over the untreated regions, closest to producing regions. We consider also other control variables such as workforce, income, proportion of woman, elderly and young people in the population, and a time trend variable, to control institutional changes. The data used in this study corresponds to a balanced panel of 644 municipalities from 2002 to 2011.

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 agricultural, representing about 10% of agricultural area in Brazil, and more or less 1% of GDP. The ethanol industry is about 3.5% of Brazilian industrial GDP, and the sector employs more than 6 million of people.

The increase in ethanol production is explained by the incentives to substitute fossil fuels for biofuels. The planted area of sugarcane doubled in the last twenty years. It is the largest increase in production in entire world. In the early nineties, Brazilian production represented 25% percent more than the second producer, India. Now, the difference is more than double.

On the other hand, the increase in sugarcane production and the expansion of new cultivated areas of sugarcane might have an impact on human health and employment, mainly at regional levels. There exist doubts about the impact of sugarcane production, 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 NORONHA (2006).

Some studies have shown that the balance of costs and benefits is positive from the standpoint of the entire country BNDES and CGEE (2008), but the benefits for the cane growing regions may not be as evident, once the producing regions may disproportionately bear the negative impacts of the sector´s presence. The most studied aspect in this respect is the labor market. Many authors encountered negative impacts to manual harvesting ALVES (2006, 2007); BACCARIN et al. (2008).

It is recognized that sugarcane is significantly more valuable by tilled area than many other crops, such as soybeans and corn. As for employment, (Toneto-Jr and Liboni, 2008) observe that sugarcane cultivation generates more jobs than soybean, and only slightly fewer than corn. Thus, because it 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, because of the need for industrial workers and services - transportation, maintenance, etc. - increasing the sector´s indirect effects on the producing region.

(Chagas et al., 2011) evaluates the impact of sugarcane on social regional indicators,

like HDI (Human Development Index) by a spatial propensity score matching, which control the fact that the sugarcane production in one specific region is not aleatory. The results suggest that the presence of sugarcane growing in these areas 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.

The increasing importance of sugarcane production in the country tends to raise more and more arguments about its consequences. This paper provides a contribution to the discussion, by assessing the effects of sugarcane on the respiratory problems of children and adults. We propose a spatial difference-in-difference technique, to control the effect of treatment (the sugarcane culture) on the untreated regions, closest to producing regions. Then, we compare municipalities where burning occurs with others where it does not occur, 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). It takes place in the beginning of harvest, which coincides with the dry season. Many works relates sugarcane burning and increase of fine particulate matter , coarse particulate matter and black carbon concentration, especially in burning hours Lara et al. (2005), and alter positively the air concentration of substances as nitrite, sulfite, oxide of carbon and others Allen et al. (2004). The sugarcane is harvested by unskilled workers, mostly manually. The literature also relate that short- and long-term exposition to classical pollutants (matter, sulfite, nitrite, oxide carbon, etc.) can affect negatively human health capital 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 pre-harvest burning may be as harmful as the pollution from traffic and industries Mazzoli-Rocha et al. (2008), many studies relate the impact of of pre-harvest burning of sugar-cane on health, for specific municipalities or a larger regionArbex et al. (2000, 2004); CANÇADO et al. (2006); Arbex et al. (2007); Ribeiro (2008); Uriarte et al. (2009); Carneseca et al. (2012) . This studies, in general, based in closest effects of burning, considering only the local events of both, burning and respiratory health. They fail, for example, to capture the interactions among the burning events to another place.

This article is organized in four sections, including this introduction. The next section presents the methodology used to identify the possible impacts of growing sugarcane on the social conditions of producing regions, 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 region interrelated together. Then, we consider a possibility of propagation in the treatment effect, that is, after the treatment, the effect of sugarcane production acts on both region, treated and non-treated. To be clear, consider y_{it} as 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 other, and d_{it} a $n \times 1$ vector in that, such $d_{it} = 1$ if the region 1 is treated in time t and zero, in contrary. Additionally, consider two situation for each region: before and after the treatment. Then, for the before treatment situation, we have

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

where, $y^b_{it,0}$ represents the dependent variable in non-treated region, before the treatment, and $y^b_{it,1}$ the same, for the treated region. Now, after the treatment, we have two impacts - one on treated region e other in non-treated region. But, the impact on non-treated region depends of the proximity of this region to treated one. For clarity, we consider both regions after the treatment in follow

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

In this way, α capture the direct effect of treatment on treated region, and β the indirect effect of treatment over all region, treated and non-treated, conditioned to the neighbor treated, 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 "before" and "after" definitions, we can compute three effects: ATE (Average Treatment Effect), ATET (Average Treatment Effect on the treated), and ATENT (Average Treatment Effect on the non-treated), as follow

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

In matrix notation, as common in Spatial Econometrics literature, 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. μ, α and β are parameters to estimate.

In this perspective, $\mathbf{I}_t \otimes \mathbf{W}D\beta$ represents the indirect effect of treatment on both region, treated and non-treated. This is a additional effect, not estimated in general [1]. This effect, however, is an average effect. It is possible that the indirect effect is different in treated and non-treated regions. For example, if the treatment in treated region is tiny, 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,

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

when estimating β as an average 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.

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

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

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) represent a $nt \times 1$ matrix with D in the principal diagonal and zero, out. Then $\mathbf{W_{ij}}$ represent the neighborhood effects of the j-region on i-region . Substituting in (2), result

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

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 this model does not contains a traditional spatial interaction effect, like 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

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

where, 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 spatial associated, 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

The data used corresponds to an balanced panel of 644 municipalities, from 2002 to 2011. We have information of sugarcane production, year by year, for each locality, provided for IBGE (Statistical and Geography Brazilian Institute), and the Brazilian official department to statistical research. This department makes annually a survey about Brazilian agricultural production, that includes information relates to production, planted area, and harvested area.

In São Paulo state, the planted area with sugarcane represents nearly 50%. The expanding area in the last years given over to cane has prompted a series of questions on the possible conflict between lands used to produce food versus energy. At the national level, of the over 800 million hectares of landmass in Brazil, over 300 million are suitable for farming and ranching activities. Of these, about 60 million are used to grow permanent and temporary crops and some 200 million for animal husbandry. This means there is an ample amount of land available for cane and other crops, and this can benefit from recovery of former pasturage and productivity gains in stock breeding that can release lands. Hence, on closer study this conflict is not borne out on a 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]. It is clear the increase in the production in Northwest region of the state, since the initial year until current year. It is evident too the production sprawl to the west of the state, area until recently occupied by pasture.

[FIGURE 1 HERE]

The production of sugarcane is important to define our treatment variable. We consider as treated al region in that the sugarcane area represents at least 6.7% of total agricultural area. This number represents the median in production area distribution. In the table 1 we report the number of treated area in each year, in the panel. Over time, the treatment increase from 38.4% of the areas until 62.4%, in the end of the period.

[TABLE 1 HERE]

We consider as our interested variable of interest the hospitalization for respiratory problems per one thousand habitants. This data is provided by the computer department of the public health system (DATASUS)[3]. This department collects and systematizes administrative information about the health system, not only of the public health system, but of the private system too. This information is very disaggregated, in spatial terms. For our interest, we take the data in county level. In the figure 2 we can see the evolution of the hospitalization in São Paulo state, by region, from 2002 until 2011. It is evident the reduction in the cases of hospitalization over the time.

[FIGURE 2 HERE]

This result can be due to reduce on burning sugarcane. Nowadays, the legislation prohibits certain types of fires 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). The actual trend is that this practice be ended in a few years, both by regulatory pressures to reduce the emission of pollutants and their harmful effects as by the economic stimulus resulting from full use of the cane (juice, straw, leaves and bagasse), or even issues related to the labor market, formalization of the labor and enhancement of workforce. In São Paulo state, the Environmental

[2]We select some years in this periods, but the evolution is evident.
[3]http://www2.datasus.gov.br/DATASUS/index.php.

Protocol signed between the plants, sugarcane producers and the government establishes the end of the burning in the mechanized areas until 2014, and in all areas, in 2017. For this, it is possible that the trend in hospitalization accompanies the end of the burning. The table 2 makes this fact more evident.

[TABLE 2 HERE]

Then, we consider a trend variable as a control, to adjust for this empirical evidence. Additionally, we can consider other control variables, such proportion of workers in the population, mean income, proportion of elderly and young people, and proportion of woman in the population. The first two variables control to socio-economic conditions, which makes possible that people prevent health problems. The remain variables control the groups more susceptible to respiratory health problems, as pointed in literature Braga et al. (1999); Roseiro (2002); GONÇALVES et al. (2005). Table 3 reports statistical descriptive 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. We test two set of model, based on 2 and 3, as follow. In the first one, we consider only the average indirect treatment effect, and test the pooled specification, spatial fixed effect and pooled spatial error model. We based on Elhorst routine for panel data models (**?**Elhorst, 2010; **?**). Elhorst consider a ML estimation. The main, because the number of studies considering IV/GMM estimators of spatial panel data models is still relatively sparse [4].

(Elhorst, 2010) provides a Matlab routines to estimate spatial panel data models, including the bias correction procedure proposed by (Lee and 2010, 2010) if the spatial panel data model contain spatial and/or time-period fixed effects, the direct and indirect

[4]One exception is (**?**), 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.

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 de order that maximize the Moran I, so we choose the matrix for which there are the greatest spatial auto correlation.

Then, we compute three models, as follow

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

where y_{it} is a hospitalizations per thousand in habitants, 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 will be estimate. 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 neighbour 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] (L 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

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

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

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

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

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

(**?**) call attention for these tests in panel data model, when a model includes spatially lagged independent variables must still be investigated.

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

[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.

For pooled models (model 1 and 4), the results are intuitive, in general. The only one is the result to proportion of children in the population, for the both pooled models, which are negative and significant values. In the pooled model still, all control variables are very significant, unless the proportion of woman in the population - which is not significant in all models. In both models, the treatment seems a strong impact, much higher than in the other models, and the indirect effect is negative and non-significant (in the model 1).

Hold up in this result a little more. It is the result that a research will have, applying the conventional techniques. And this can justified the fact that the researches not explore more the interrelation between regions, because a result not significant is not a incentive result. But, if we compare the difference between the model 1 and the model 4, we can see a provocative result. The separation in the indirect effects on treated and non-treated regions increase the direct effect on treated regions. And give us a counterintuitive result relate to indirect effect on treated region, which is negative. This result, in particularly, can be due a omitted relevant variables, despite a high significant level. In fact, when we include spatial controls, this result does not remain.

But, for the non-treated region, the effect is very interesting. The signal is very significant and positive. Then, the indirect effect on non-treated is about half of the effect on treated regions.

We improve this investigation, including spatial controls. The LM tests, computed over the residuals of pooled model, are very conclusive to inclusion of this control, and the robust LM test can suggest that model with lag in dependent variable is more significant than in the error. Nevertheless, we report both results.

The conclusion does not change in relation to the pooled models. The controls appearance very weak controls now. Only proportion of workers in the population remains significant and negative, as expected. For the treatment effect, the magnitude of impact decreasing in all situation, compared to the pooled model. Is this can due to the fact that a part of this respiratory problem is consequence of the other regional factors than the sugarcane culture, like climatic conditions and seasonal effects. But, the impact of sugarcane culture appearance to be positive and significant in treated regions.

About 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 indirect impact on the treat one. In the SAR model (model 5), the indirect effect on non-treated (0.78) is circa 80% of the direct effect of treatment (0.97). If we estimate to the mean, the indirect effect in the model 2 (0.44) is only 66% of the direct effect (0.65), and both are smaller.

These are preliminary results only and other test need to make until the final paper. A counterintuitive result relate to the SAR model to overcome the SEM model. It is not reasonable assume that hospitalization due respiratory diseases cause more hospitalization on neighbour. It would be more intuitive that the error in the region was impacted to the neighbours error. It is possible that some omitted relevant variable remain in this case.

4 Final remarks

The increase in ethanol in the last years provoked many debates. In this paper we try to address on of this, more specifically, the impact on human health due to burning sugarcane.

We proposed a new methodology to evaluate impact, using aggregate data in regional level and information about the neighbors. More specifically, we consider the indirect eff ect of treatment on treated and non-treated region.

Our results suggest the correction on our assumption. In fact, the results suggest that our method makes it possible to identify better the impact, not only in the treated regions, but on non treated too.

The research is in beginning, but the results look promising.

Referências

A.G. Allen, A.A. Cardoso, and G.O. da Rocha. Influence of sugar cane burning on aerosol soluble ion composition in southeastern brazil. *Atmospheric Environment*, 38(30):5025 – 5038, 2004. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2004.06.019. URL `http://www.sciencedirect.com/science/article/pii/S1352231004006144`.

F. J. C. ALVES. Por que morrem os cortadores de cana? *Saúde e Sociedade*, 15(3):90–98, Spet/Dec 2006.

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

Manuela Angelucci and Giacomo De Giorgi. Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, 99(1): 486–508, September 2009. doi: 10.1257/aer.99.1.486. URL `http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.486`.

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

Jayet H Anselin L, Le Gallo J. *The econometrics of panel data, fundamentals and recent developments in theory and practice*, chapter Spatial panel econometrics., pages 901–969. Kluwer Academic Publishers, Dordrecht, 3rd edition, 2006.

M. A. Arbex, G. M. Bohm, P. H. Saldiva, and G. CONCEIÇÃO. Assessment of the effects of sugar cane plantation burning on daily counts of inhalation therapy. *J Air Waste Manag Assoc*, 50(10):1745–9, 2000.

M. A. Arbex, J. E. D. CANÇADO, L. A. M. PEREIRA, A. L. F. Braga, and P. H. N. Saldiva. Queima de biomassa e efeitos sobre a saúde. *Bras Pneumol*, 30(2):158–175, 2004.

M. A. Arbex, L. C. Martins, R. C. Oliveira, L. A. A. Pereira, F. F. Arbex, Saldiva P. H. N. CANÇADO, J. E. D., and A. L. F. Braga. Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. *Journal of Epidemiology and Community Health*, 61(5):395–400, 2007.

J. G. BACCARIN, F. J. C. ALVES, and L. F. C. GOMES. 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, 2008. Sociedade Brasileira de Economia e Sociologia Rural.

Lucila Berniell, Dolores de la Mata, and Nieves Valdés. Spillovers of health education at school on parents' physical activity. *Health Economics*, n/a:n/a–n/a, 2013. ISSN 1099-1050. doi: 10.1002/hec.2958. URL `http://dx.doi.org/10.1002/hec.2958`.

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

A. L.F. Braga, G. M.S. CONCEIÇÃO, L. A.A. Pereira, H. S. Kishi, J. C.R. Pereira, M. F. Andrade, F. L.T. GONÇALVES, P. H.N. Saldiva, and M. R.D.O. Latorre. Air pollution and pediatric respiratory hospital admissions in sÃ£o paulo, brazil. *Journal of Environmental Medicine*, 1(2):95–102, 1999. ISSN 1099-1301. doi: 10.1002/

(SICI)1099-1301(199904/06)1:2<95::AID-JEM16>3.0.CO;2-S. URL http://dx.doi.org/10.1002/(SICI)1099-1301(199904/06)1:2<95::AID-JEM16>3.0.CO;2-S.

J. E. D. CANÇADO, P. H. N. Saldiva, L. A. A. Pereira, L. B. L. S. Lara, P. Artaxo, L. A. Martinelli, M. A. Arbex, A. Zanobetti, and A. L. F. Braga. The impact of sugar caneburning emissions on the respiratory system of children and the elderly. *Environ Health Perspect.*, 114(5):725–729, 2006. doi: 10.1289/ehp.8485.

E. C. Carneseca, J. A. Achcar, and E. Z. Martinez. Association between particulate matter air pollution and monthly inhalation and nebulization procedures in Ribeirão Preto, São Paulo State, Brazil. *Cadernos de Saúde Pública*, 28(8):1591 – 1598, 2012. ISSN 0102-311X. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2012000800017&nrm=iso.

A. L. S Chagas, R. Toneto-Jr, and C. R. Azzoni. 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, 2008.

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

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

F.L.T. GONÇALVES, L.M.V. Carvalho, F.C. Conde, M.R.D.O. Latorre, P.H.N. Saldiva, and A.L.F Braga. The effects of air pollution and meteorological parameters on respiratory morbidity during the summer in São Paulo City. *Environment International*, 31(3): 343 – 349, 2005. ISSN 0160-4120. doi: http://dx.doi.org/10.1016/j.envint.2004.08.004. URL http://www.sciencedirect.com/science/article/pii/S0160412004001424.

J. P. Kaboski and R. M. Townsend. The impact of credit on village economies. *American Economic Journal: Applied Economics*, 4(2):98–133, 2012.

Anselin L, Bera A. K., Florax R, and Yoon M. J. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104, 1996.

L.L. Lara, P. Artaxo, L.A. Martinelli, P.B. Camargo, R.L. Victoria, and E.S.B. Ferraz. Properties of aerosols from sugar-cane burning emissions in southeastern brazil. *Atmospheric Environment*, 39(26):4627 – 4637, 2005. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2005.04.026. URL http://www.sciencedirect.com/science/article/pii/S135223100500395X.

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

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

F. Mazzoli-Rocha, C. B. MAGALHÃES, O. Malm, P. H. Saldiva, W. A. Zin, and D. S. Faffe. Comparative respiratory toxicity of particles produced by traffic and sugar cane burning. *Environ Res.*, 108(1):35–41, Sep 2008.

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

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

M. N. V. Roseiro. *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, 2002.

P. Sicard, A. Mangin, P. Hebel, and P Malléa. Detection and estimation trends linked to air quality and mortality on french riviera over the 1990-2005 period. *Sci Total Environ*, 408(8):1943–50, Mar 2010.

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

M. Uriarte, C. B. Yackulic, T Cooper, D. Flynn, M. Cortes, T. Crk, G. Cullman, M. McGinty, and J. Sircely. Expansion of sugarcane production in sÃ£o paulo, brazil: Implications for fire occurrence and respiratory health. *Agriculture, Ecosystems & Environment*, 132(1â2):48 – 56, 2009. ISSN 0167-8809. doi: http://dx.doi.org/10.1016/j.agee.2009.02.018. URL `http://www.sciencedirect.com/science/article/pii/S0167880909000760`.

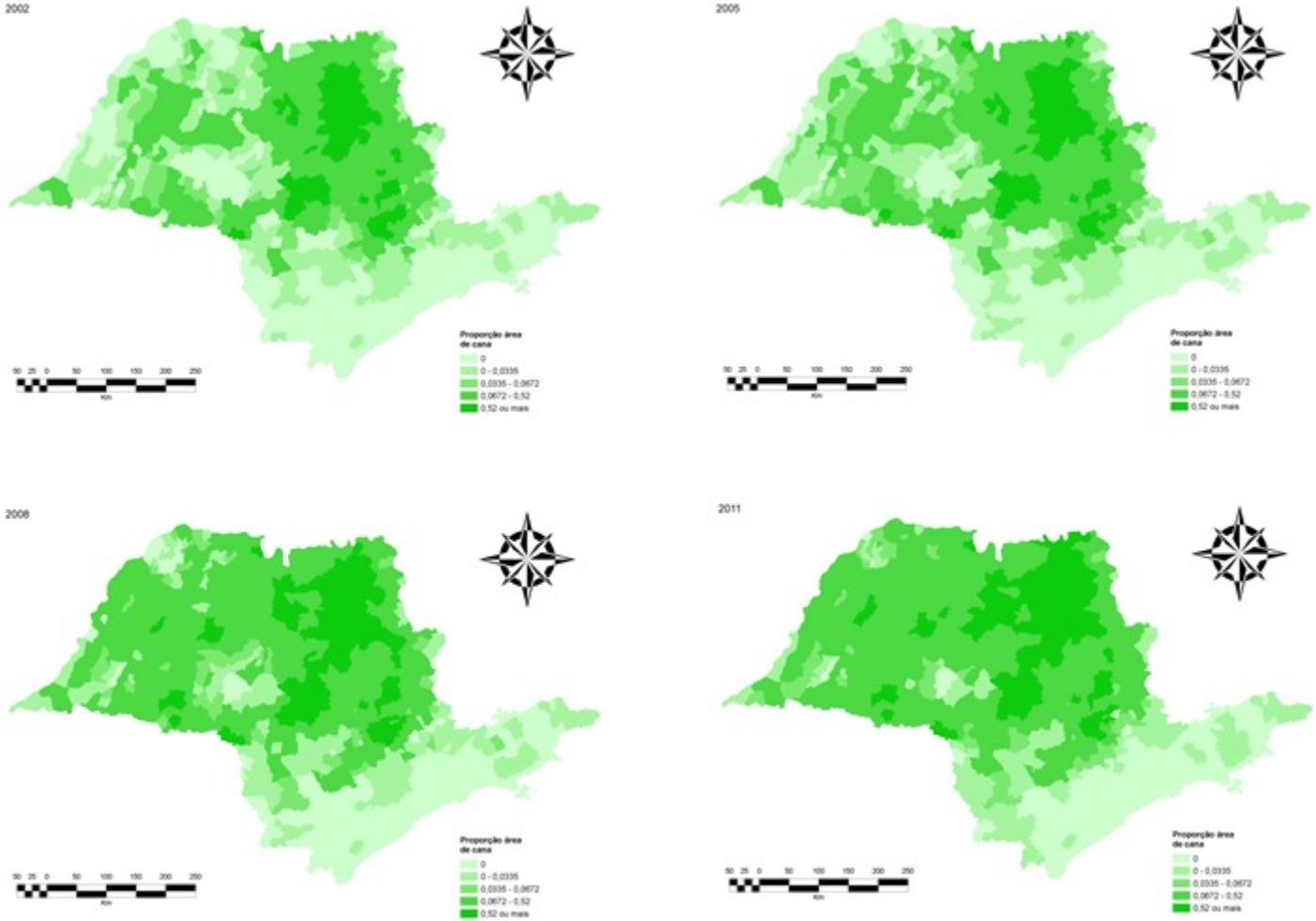

Figura 1: Sugarcane production in São Paulo state by municipality, 2002, 2005, 2008, and 2011
Source: IBGE, Municipal Agricultural Research.

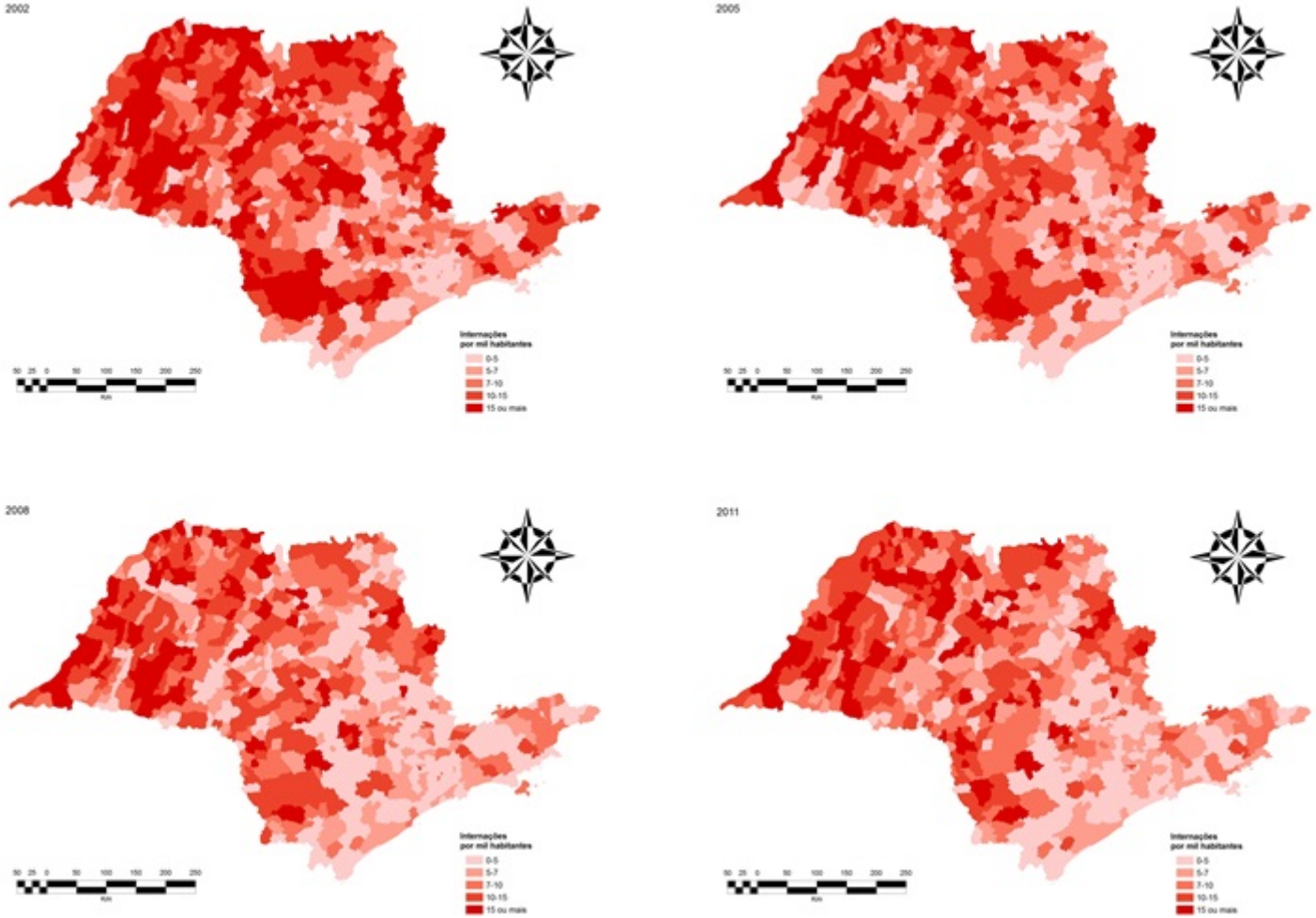

Figura 2: Hospitalization due to respiratory problems in São Paulo state by municipality, 2002, 2005, 2008, and 2011
Source: Datasus, Health Ministry.

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%.