UNIVERSITÀ DEGLI STUDI DI FERRARA
DIPARTIMENTO DI ECONOMIA E MANAGEMENT
Via Voltapaletto, 11 – 44121 Ferrara

Quaderno DEM 3/2017

August 2017

Modeling temporal treatment effects with zero inflated semi-parametric regression models: the case of local development policies in France

Hervé Cardot – Antonio Musolesi

Quaderni DEM, volume 6 **ISSN 2281-9673**

Modeling temporal treatment effects with zero inflated semi-parametric regression models: the case of local development policies in France

Hervé CARDOT
Institut de Mathématiques de Bourgogne, UMR CNRS 5584,
Université de Bourgogne Franche-Comté

Antonio MUSOLESI
Department of Economics and Management (DEM),
University of Ferrara and SEEDS

July 19, 2017

Abstract

A semi-parametric approach is considered to estimate the variation along time of the effects of two public policies that were devoted to boost rural development in France. This statistical approach combines the flexibility and modularity of additive models with the ability of panel data to deal with selection bias and to allow for the estimation of dynamic treatment effects. Since we face a kind of zero inflated phenomenon that cannot be dealt with a continuous distribution, we introduce a mixture model with a mass at zero and a continuous density. We find evidence of interesting patterns of temporal treatment effects with relevant nonlinear policy effects. The adopted semi-parametric modeling also offers the possibility of making a counterfactual analysis at an individual level. The methodology is illustrated on a few municipalities for which the evolution of the potential outcomes is estimated and compared under the different possible treatments.

Keywords: Additive Models; Semi-parametric Regression; Panel Data; Policy Evaluation; Temporal Effects; Multiple Treatments; Local Development.

1 Introduction

In response to the deteriorating conditions of distressed areas, many countries, such as USA, UK and France, have established enterprise zone programs (EZ) aimed to increase socio-economic development by means of boosting local employment. At a supranational level, territorial cohesion, convergence and a harmonious development across regions are among the objectives of the European Union which tries to pursue through the structural funds (SF).

Despite their appeal and the high amount of financial resources used, such geographically targeted policies have been criticized with respect to different aspects and doubts have been cast with respect to their effectiveness. Concerning EZ, there exists a number of micro-econometrics works aiming at assessing their economic effects. Most of these studies use standard fixed effects panel data methods, such as difference-in-differences or interactive fixed effects approaches with random growth models (Papke, 1994) or factor models (Gobillon and Magnac, 2016), to account for the non-random assignment of the treatment and provide mixed results (for surveys, see *e.g.* Gobillon et al., 2012; Peters and Fisher, 2004). Looking at the analyses of the effects of regional policies implemented through the European SF, it can be noted that some earlier studies have been carried out by analyzing the convergence process and interpreted the descriptive fact of an increasing divergence across the European regions as an indication that the SF have been ineffective. More recently, some works adopting a causal framework appeared (Becker et al., 2010; Mohl and Hagen, 2010), but also for these policies they provided mixed evidence. In summary, the effectiveness of both EZ and SF is a relevant and contentious issue in the debate regarding local development.

This work proposes a new semi-parametric approach to estimate the variation along time of the treatment effects of these regional policies, combining the flexibility and modularity of additive models with the ability of panel data to deal with selection bias and to allow for the estimation of dynamic policy effects. This represents a new contribution to the literature on regional policy evaluation revealing for the first time some non-linearities as well as interaction effects that have relevant implications for public policy design. The paper also introduces methodological advances which could be useful for future research and, more precisely, the proposed approach is developed along the following directions.

First, we relax the parametric specification to model the regression function, giving a larger flexibility, allowing to unveil possible complex relations and reducing the risk of misspecification. We rely on the rather general framework of additive models and generalized additive models (Hastie and Tibshirani, 1990; Wood, 2017), giving much more flexibility and robustness than usual linear models, but also addressing the curse of dimensionality problem arising in fully nonparametric models. This is extremely relevant for the evaluation of economic public policies devoted to increase economic development or to boost economic growth, for which the number of potential regressors is typically large. Penalized splines are used to represent the non parametric parts of the additive model (Wood, 2004, 2008) and an appealing feature of this approach is its modularity (see, e.g. Ruppert et al., 2003). This means that concepts like main effects, interaction effects, and generalized regression can be viewed as modules that can be put together into an almost endless variety of models. Other prominent features of penalized splines are that: i) they use few knots, thus need less computation than both smoothing splines and fully nonparametric kernel methods; ii) many asymptotic results have also been recently provided (see, *e.g.* Li and Ruppert, 2008, Wood et al., 2016).

Second, we combine spline modeling with panel data. Panel data models have been shown to be very useful for policy evaluation, allowing to account both for selection on observables and selection on unobservables, and permitting to specify the models in terms of potential outcome at different points in time (Heckman and Hotz, 1989; Heckman et al., 1999; Wooldridge, 2005; Hsiao et al., 2011; Lechner, 2015), time being an essential element in the notion of causality (*e.g.* Lechner, 2011). Moreover, despite the fact that there is an increasing availability of relatively long panel data, most of the existing micro-level studies on regional policies focus on static effects. There are some exceptions, suggesting that taking account of dynamic effects is important (see *e.g.* O'Keefe, 2004; Becker et al., 2010). A relevant feature of this work is that the panel structure of the data allows us to estimate in a flexible manner a causal effect that can vary with time.

Third, an original mixture model combining continuous and discrete responses is specified. Specifically, it is often observed that the dependent variable, local employment, does not vary along time, so that when studying its variations along time we face a kind of zero inflated phenomenon that cannot be dealt with a continuous distribution. We thus allow the dependent variable to remain constant in time with a probability that can be strictly larger than zero. The estimation is finally carried out by maximizing the corresponding likelihood function, which is a mixture of a mass at zero and a continuous density.

Fourth, while most of the previous studies focus on one particular policy, either EZ or SF, we will assess the effect of both policies as well as their interaction by adopting a multiple treatments framework (see Frolich, 2004, for a survey). This can be expected to be relevant both in statistical terms, avoiding to mingle the effect of two different, but partially overlapping schemes, and economically, allowing to compare the effects and their evolution over time of two schemes that make use of different instruments to stimulate local development.

Finally, the proposed semi-parametric modeling also permits to estimate what would have been the effects of such policies on particular municipalities by performing a counterfactual estimation at an individual level. The evolutions of the potential outcomes are thus estimated and compared under the different possible treatments for a few municipalities. These municipalities, selected with a clustering k-medoids algorithm (see Kaufman and Rousseeuw, 1990), represent communes with different but typical characteristics within their cluster.

The remainder of the paper is structured as follows. Section two describes the rural policies adopted in France, presents the data and provides some descriptive statistics.

Section three is devoted to the presentation of the econometric framework and of the estimation methodology. Section four provides the presentation and discussion of our main results while section five summarizes and concludes. Additional results are given in supplementary appendices. Appendix A provides detailed information on data, variables and the sample. The results of some placebo tests that discard specifications that are likely to be misspecified are gathered in Appendix B.

2 Description of the policies and data

In France, EZ have been implemented to boost job creation. Such policies are based on fiscal incentives to firms located in deprived areas. Specifically designed to boost employment of rural areas, the ZRR (*Zones de Revitalisation Rurale)* program started the 1st September 1996. A noticeable feature of the program is that the selection of ZRR was clearly not random. A rather complex algorithm was used to determine the eligibility, according to some observable – demographic, economic and institutional – criteria. To be eligible to ZRR, a municipality should be a part of a canton with population density lower than 31 inhabitants per square km (1990 Population Census)[1]. The population or the labor force must also have diminished or the share of the agricultural labor employment must be at least twice the French average. Finally, to be included into the program, the municipality should belong to a pre-existing zoning scheme set up by the European Union, which is called TRDP (*Territoire Rural de Développement Prioritaire*). However, due to political tempering, it is also likely that, beyond such observed criteria, other sources of selection on unobservables could affect the process (Gobillon et al., 2012). A more detailed description of the ZRR program can be found in Behaghel et al. (2015). Figures from census 1999 indicate that about 8% of the French population at that time, resided these zones; such zones cover 39% of the French lands.

Beyond the French experience, EZ have been largely criticized with respect to several aspects, such as the possibility of i) windfall effects to firms who would have hired workers even in absence of the policy; ii) negative spatial spillovers because EZ does not necessarily result in job creation but could cause geographical shifts in jobs from non-EZ to EZ areas; iii) stigmatization of the targeted neighborhood; iv) in absence of tax revenue compensation, EZ could lead to a decrease in the local provision of public services and v) obtaining only a transitory effect on employment and the need for integrated policies against structural unemployment.

At a supranational level, the SF are addressed to help lagging or re-structuring regions, so they are given to regions upon their economic characteristics (such as the per capita GDP or the unemployment level) and then are assigned from the regions

[1]A canton with a population density less than 5 inhabitants per square km is automatically labelled as ZRR without any other requirement.

to firms or to public actors (top-down process) without a clearly expressed assignment mechanism. Then, also for these policies, sources of selection on both observables and unobservables are expected to be relevant. Specifically devoted to boost rural development, the objective 5B programs (1991-93 and 1994-99) allocated financial subsides to firms and public actors located in eligible *"rural areas in decline"*. The eligibility criteria for belonging to an objective 5B area (canton) required that the area has a high share of agricultural employment, a low farming income and a low level of per capita GDP (Gross Domestic Product). The main goal of 5B programs was to improve economic development and local infrastructures, and to support the activities of farms, small and medium sized firms, rural tourism. In 1999, about 16% of French population and 60% of French lands were in these areas.

The resulting French rural policy-zoning scheme has been criticized because the only partial overlap between ZRR and 5B programs can be viewed as a sign of lack of consistency. However, for the estimation of treatment effects, such a partial overlap is a useful source of identification which is exploited in this paper to estimate the specific effect of each policy as well as their interaction effect.

Our sample is obtained by merging different data sets. The municipalities, which correspond to the finest available spatial level, are the statistical units of the analysis and the dependent variable is the number of employees. The data were obtained over a period of ten years, 1993-2002 (for each year data refer to the 1st January), from the INSEE (*Institut National de la Statistique et des Etudes Economiques*) and SIRENE (*Système Informatique pour le Répertoire des Entreprises et de leurs Établissements*) sheet. As explanatory variables, we dispose of ZRR zoning during the period and of the 5B zoning over the period 1994-99. Some other explanatory variables come from the CENSUS. Since the CENSUS data are collected every ten years, and in order to control for the initial conditions, we use data from 1990 CENSUS. Such CENSUS data have been provided by the INSEE in separate sheets, gathering demographic, education and work's qualification information. Finally, we also have at hand information on land use in 1990, obtained thanks to satellite images. After the merging process and some cleanings that are detailed in Appendix A we obtain a sample of 25593 municipalities. Table 1 below provides simple descriptive statistics of some key variables.

Table 1 about here

It can be seen in Table 1 that about 30% of the 25593 municipalities in our sample were under the ZRR scheme. Over the period 1994-99, about 47% of the municipalities were under objective 5B. Examining ZRR and 5B jointly, it appears that 50.9% of the municipalities were under at least one of the two policies. Only 27.4% of the municipalities were, in our sample, under both policies, whereas 20.6% received a support only from 5B program and 2.8% of the municipalities received the incentives only from ZRR. As expected, the treated municipalities present lower socio-economic performances compared to the non-treated ones, with the municipalities under objective 5B alone performing generally better than the other treated municipalities.

3 Model specification and estimation

3.1 Econometric framework and identification hypotheses

We borrow notations from Heckman and Hotz (1989) and Frolich (2004). Let i denote a statistical unit (a municipality in our framework) which is assigned to one of R mutually exclusive development incentives. We denote by Y_{it}^r the potential employment level for municipality i at time t under treatment (incentive) r, for $r \in \{0, 1, \ldots, R-1\}$, with the convention that $r = 0$ corresponds to no treatment. Time t is discrete, taking values in $t_0 < t_1 < \ldots < t_m$. We assume that the incentives are allocated after t_0 and that they may produce an effect from period k, with $t_k > t_0$. All the counterfactuals are assumed to be equal before the treatment begins, that is to say $Y_{it}^r = Y_{it}^0$ for $t_0 \leq t < t_k$ and $r = 1, 2, \ldots, R-1$. As a starting point, we consider the following general model,

$$
\begin{aligned}
Y_{it}^r &= Y_{it}^0 & t_0 \leq t < t_k, \\
&= Y_{it}^0 + \alpha_{it}^r, & t_k \leq t \leq t_m,
\end{aligned}
\tag{1}
$$

where Y_{it}^0 is the employment level for municipality i at time t in the absence of development funds ($r = 0$). For time $t \geq t_k$, α_{it}^r is simply the difference between Y_{it}^r and Y_{it}^0, that is to say the differential effect on the potential outcome, compared to no treatment at all, of treatment r on unit i. With this general model, which will be simplified later, note that α_{it}^r is allowed to vary from one statistical unit to another and also depends on time t.

Consider now a set of characteristics $X_i = (X_{i1}, \ldots, X_{ip})$ observed during the first period of time t_0, which are the *initial conditions*. We suppose that the following model holds,

$$
Y_{it}^0 = g_t^0(X_i) + U_{it}, \qquad t_0 \leq t \leq t_m, \tag{2}
$$

where U_{it} represents unobserved random characteristics and function g_t^0 will be described later in Section 3.2.

Let us denote by D_i, with $D_i \in \{0, 1, \ldots, R-1\}$, the treatment status of municipality i, that is supposed to be a random variable. In order to identify the causal effect, a common practice is to assume the following hypothesis holds (see *e.g.* Imbens and Wooldridge, 2009),

$$
Y_{it}^r \perp\!\!\!\perp D_i \mid X_i,\ U_{it} \qquad \forall r \in \{0, 1, \ldots, R-1\}. \tag{3}
$$

This general condition means that there exist both observable variables (X_i) and unobservable variables (U_{it}) that are related to the potential outcomes (Y_{it}^r) and to the treatment status (D_i), such that given these variables, Y_{it}^r and D_i are independent. This general formulation encompasses the most widely used specifications in the literature. An important particular case of the above condition is generally referred to as conditional independence assumption, unconfoundedness or selection on observables, assuming that

$$Y_{it}^r \perp\!\!\!\perp D_i \mid X_i, \tag{4}$$

so that the information contained in the observed variables X_i makes the potential outcomes unconfounded, that is, conditionally independent of D_i given X_i.

Since selection bias may not be completely eliminated even after controlling for the observables X_i, it is also important to note that a *before-after* approach may help to address the issue of selection on unobservables. Combining (1) with (2), we have

$$Y_{it}^r - Y_{it_0}^0 = \alpha_{it}^r + \left(g_t^0(X_i) - g_{t_0}^0(X_i)\right) + (U_{it} - U_{it_0}), \quad t_k \le t \le t_m. \tag{5}$$

and we could only require that the conditional independence assumption (4) holds for the difference of the outcome after and before the beginning of the policy. We suppose from now on that

$$Y_{it}^r - Y_{it_0}^0 \perp\!\!\!\perp D_i \mid X_i, \qquad \forall r \in \{0, 1, \ldots, R-1\}. \tag{6}$$

The new conditional independence assumption (6) is more general than (4) and holds for example when the unobservables U_{it} may be described as follows,

$$U_{it} = \phi_{1i} + v_{it} \tag{7}$$

where ϕ_{1i} is a random (individual) time invariant effect, that may be correlated to the treatment variable D_i, and v_{it} is a white noise. Under assumption (6), we have

$$\mathbb{E}\left[U_{it} - U_{it_0} \mid X_i, D_i\right] = \mathbb{E}\left[U_{it} - U_{it_0} \mid X_i\right] = 0 \tag{8}$$

and consequently,

$$\mathbb{E}\left[Y_{it}^r - Y_{it_0}^0 \mid X_i, D_i\right] = \mathbb{E}\left[\alpha_{it}^r \mid X_i\right] + g_t^0(X_i) - g_{t_0}^0(X_i), \quad t_k \le t \le t_m. \tag{9}$$

It is worth mentioning that we could have considered propensity scores (Rosenbaum and Rubin, 1983; Angrist and Hahn, 2004; Imai and Van Dyk, 2004) in place of X, in the conditioning variables appearing in (6). This would have also ensured that D is conditionally independent of the potential outcomes while achieving dimensional reduction. One drawback of this approach, which can be effective for estimating mean effects on the treated or on the whole population, is interpretation (see e.g. Imbens and Wooldridge, 2009) as well as the fact that the propensity scores may not be highly relevant variables to estimate accurately the variations of the conditional potential outcomes, given the vector of covariates X. Indeed, we can split the vector of all the available covariates X into four parts,

$$X = \left(X_{Y\cap D}, X_{\bar{Y}\cap D}, X_{Y\cap\bar{D}}, X_{\bar{Y}\cap\bar{D}}\right),$$

where $X_{Y\cap D}$ is the set of covariates that are related both to $Y_t^r - Y_{t_0}^0$ and D and $X_{\bar{Y}\cap D}$ is the set of covariates that are independent of $Y_t^r - Y_{t_0}^0$ but are related to D. Note that these two sets, $X_{Y\cap D}$ and $X_{\bar{Y}\cap D}$, represent the variables entering the propensity score function. The set $X_{Y\cap \bar{D}}$ is the set of covariates that are related to $Y_t^r - Y_{t_0}^0$ but are independent of D and $X_{\bar{Y}\cap \bar{D}}$ is the set of covariates that are independent of $Y_t^r - Y_{t_0}^0$ and D (see Figure 3.1).

$$\begin{array}{cccccc} & & D & \longrightarrow & Y_t^r - Y_{t_0} & \\ & \nearrow & \uparrow & \nearrow & \uparrow & \\ X_{\bar{Y}\cap D} & & X_{Y\cap D} & & X_{Y\cap \bar{D}} & \end{array}$$

Figure 1: The expected causal relation between $Y_t^r - Y_{t_0}$, X and D

The smallest set of conditioning variables required to satisfy condition (6) is $X_{Y\cap D}$. However, one of the aims in this work is to estimate, at an individual level, the variations over time of the potential effects of the different policies. This is why we also take account of the set of variables $X_{Y\cap \bar{D}}$ in a way that is as flexible as possible to have a better prediction of the potential outcomes. As a result, our statistical approach is built by modeling in a non parametric way the relation between $Y_t^r - Y_{t_0}^0$ and X and by selecting, among all the available variables, the variables that belong to one of the two sets $X_{Y\cap D}$ and $X_{Y\cap \bar{D}}$. Note that if we were interested in the best possible estimation of the propensity scores, i.e. the scores giving the probability of receiving policy r, for $r = 0, \dots, R-1$, our statistical models would have focused on the sets of variables $X_{\bar{Y}\cap D}$ and $X_{Y\cap D}$.

In the following Sections it is assumed that the set of covariates X is restricted to $X_{Y\cap D}$ and $X_{Y\cap \bar{D}}$. Other observed variables that could be considered are those that influence selection into the program even if they do not affect directly the outcome, *i.e.* $X_{\bar{Y}\cap D}$. Introducing these variables in the regression function may help to solve the problem of selection on observables, provided there is no misspecification error, using the terminology by Heckman and Hotz (1989). Appendix A provides further comments on this issue. The variable selection procedure is described in Section 4 and in Appendix A.

3.2 A flexible semi-parametric additive modeling approach

Suppose we have now a sample $(Y_{it}^{D_i}, X_i, D_i)_{i=1,\dots,n}$, for $t \in \{t_0, \dots, t_m\}$. We can write

$$Y_{it}^{D_i} = \sum_{r=0}^{R-1} Y_{it}^r \mathbb{1}_{\{D_i = r\}} \tag{10}$$

where the indicator function satisfies $\mathbb{1}_{\{D_i=r\}} = 1$ if $D_i = r$ and zero else. Consequently, we can express, with model (5), the variation along time of the employment level of municipality i, $Y_{it}^{D_i} - Y_{it_0}^0$, as follows,

$$Y_{it}^{D_i} - Y_{it_0}^0 = \sum_{r=1}^{R-1} \mathbb{1}_{\{D_i=r\}}\alpha_{it}^r + \left(g_t^0(X_i) - g_{t_0}^0(X_i)\right) + \left(U_{it} - U_{it_0}\right), \quad t_k \leq t \leq t_m. \tag{11}$$

The term α_{it}^r which reflects in equation (11) the impact of treatment r should be equal to zero when $t_0 \leq t < t_k$.

In the econometric literature, α_{it}^r is nearly often expanded as a linear function of the covariates (see *e.g.* Heckman and Hotz, 1989, eq. 3.9). The linearity assumption is strong and a miss-specification of the relation between $Y_{it}^r - Y_{it_0}$ and the regressors may lead to wrong results and interpretation of the policy effect. We thus prefer to consider a more general model that can take account of non linear effects nonparametrically via an additive form (Hastie and Tibshirani, 1990; Wood, 2017).

The expected value that would be obtained at time t for a municipality with characteristics X_i under no treatment, is supposed to be additively modeled as follows,

$$g_t^0(X_i) - g_{t_0}^0(X_i) = \mu_t^0 + \sum_{j=1}^{p} g_{jt}^0(X_{ij}), \tag{12}$$

where $g_{jt}^0(.),\ j = 1, \ldots, p$, are unknown smooth univariate functions. The identifiability constraints

$$\sum_{i=1}^{n} g_{jt}^0(X_{ij}) = 0, \quad j = 1, \ldots, p,$$

ensure that μ_t^0 represents the mean value of the variation of the potential outcome between t and t_0 if all the units would have received no incentives at all ($D_i = 0$ for all $i = 1, \ldots, n$).

We also suppose that the differential policy effect α_{it}^r can be expressed, given the vector of covariates X_i, with the following additive model,

$$\alpha_{it}^r = \alpha_t^r + \sum_{j=1}^{p} g_{jt}^r(X_{ij}), \tag{13}$$

where $g_{jt}^r(.),\ j = 1, \ldots, p$ are unknown smooth functions satisfying the identifiability constraints

$$\sum_{i=1}^{n} g_{jt}^r(X_{ij}) = 0, \quad j = 1, \ldots, p.$$

Consequently, α_t^r represents the mean effect, over the whole sample, at period t of treatment r and the function g_{jt}^r reveals how the mean impact of the policy r is modulated by the characteristics X_i of each considered statistical unit.

Note that a simple extension of (13) consists in considering interactions between covariates instead of additive effects. For $2 \leq d \leq p$, the additive effects of d covariates, $g_{1t}^r(X_{i1}) + g_{2t}^r(X_{i2}) + \cdots + g_{dt}(X_{id})$ can be replaced by a more general multivariate function $g_{1,2,\ldots,d,t}^r(X_{i1}, X_{i2}, \ldots, X_{id})$ that could allow a more flexible fit to the data, at the expense of a more difficult interpretability and, because of the curse of dimensionality, less precise estimates.

The behavior of functions g^r_{jt} is of central interest and our general model encompasses the following particular cases,

a) No effect of the policy r compared to no treatment at all, when $\alpha^r_t = 0$ and $g^r_{jt} = 0$ for all $t \geq t_k$.

b) Linear trends in time when $\alpha^r_t = \alpha^r_0 + \alpha^r_1 t$ and linear effects of the covariates when $g^r_{jt}(X_{ij}) = \beta^r_{jt} X_{ij}$.

c) Polynomial trends in time and polynomial effects of the covariates, as well as smooth threshold effects.

3.3 A mixture model combining continuous responses and a mass at zero

A relevant feature of this study is that the statistical units are generally demographically small and we observe no variation at all of the dependent variable along time for a non negligible fraction of the municipalities, *i.e.* $Y^{D_i}_{it} = Y_{it_0}$. Table 2 below shows that the modal value of $Y^{D_i}_{it} - Y_{it_0}$ is indeed 0 for all the values of t, with t varying between 1994 to 2002 and t_0 corresponding to the year 1993. We can remark that the fraction of zeros decreases with t and varies with the treatment status. Figure 2 depicts the estimated density function of the dependent variable, $Y^{D_i}_{it} - Y_{it_0}$, for $t = 1994$.

Table 2 about here
Figure 2 about here

This empirical fact leads us to modify the model introduced in (11) in order to take account of this important feature of the data. There is a kind of *zero inflated effect* that can not be dealt with a continuous distribution. We thus allow $Y^{D_i}_{it} - Y^0_{it_0}$ to be equal to zero with a probability that may be strictly larger than 0. Let us denote by $\Delta^{D_i}_{it} = \sum_{r=0}^{R-1} \mathbb{1}_{\{D_i = r\}} \Delta^r_{it}$, with $\Delta^r_{it} = Y^r_{it} - Y^0_{it_0}$. We propose to describe the distribution of the counterfactual variation of the level of employment Δ^r_{it} as a mixture of a mass at 0 and a continuous distribution. This means that, for $a < 0 < b$,

$$\mathbb{P}\left[a < \Delta^r_{it} < b \mid X_i\right] = \mathbb{P}\left[\Delta^r_{it} = 0 \mid X_i\right] + \left(1 - \mathbb{P}\left[\Delta^r_{it} = 0 \mid X_i\right]\right) \int_a^b f^r_t\left(u; X_i\right) du, \quad (14)$$

where $f^r_t(u; X_i)$ denotes the conditional density of Δ^r_{it} at point $u \in mathbbR$ given the covariates X_i and the fact that $\Delta^r_{it} \neq 0$. Combined with (11), the zero inflated model (14) gives us the conditional average potential outcome of policy r given X_i,

$$\mathbb{E}\left[Y^r_{it} - Y_{it_0} \mid X_i\right] = \left(1 - \mathbb{P}\left[\Delta^r_{it} = 0 \mid X_i\right]\right) \times \left[\alpha^r_{it} + \left(g^0_t(X_i) - g^0_{t_0}(X_i)\right)\right]$$

and the conditional average potential outcome with no policy

$$\mathbb{E}\left[Y_{it}^0 - Y_{it_0} \mid X_i\right] = \left(1 - \mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right]\right) \times \left[g_t^0(X_i) - g_{t_0}^0(X_i)\right],$$

so that the expected difference between the potential outcomes at time t with treatment r and without treatment is now given by

$$\begin{aligned}\mathbb{E}\left[Y_{it}^r - Y_{it}^0 \mid X_i\right] &= \left(1 - \mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right]\right) \times \alpha_{it}^r \\ &\quad - \left(\mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right]\right) \times \left[g_t^0(X_i) - g_{t_0}^0(X_i)\right]. \quad (15)\end{aligned}$$

Previous expression (15), which explicitly takes account of the zero inflation feature of the counterfactual outcome variations, is of central importance in this paper. However, it is more difficult to interpret than (1) since it is composed of two main terms which may act in opposite directions.

Suppose for example that, for $t > t_0$, there is a general positive trend for employment, meaning that $g_t^0(X_i) - g_{t_0}^0(X_i) \geq 0$ and $\alpha_{it}^r \geq 0$. Then, while the first term, $\left(1 - \mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right]\right) \alpha_{it}^r$, will always be positive, the sign of the second one, $\left(\mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right]\right) \times \left[g_t^0(X_i) - g_{t_0}^0(X_i)\right]$, will depend on the relative strength of the two probabilities that are involved. If policy r makes it increase the probability of variation compared to no policy at all, *i.e.* $\mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right] \leq 0$, then the mean variation $\mathbb{E}\left[Y_{it}^r - Y_{it}^0 \mid X_i\right]$ will be positive. Otherwise, if $\mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right] > 0$, then the two terms in (15) will have opposite signs and the sign of $\mathbb{E}\left[Y_{it}^r - Y_{it}^0 \mid X_i\right]$ will depend on the strength of each term.

We suppose that the probability that $Y_{it}^{D_i} - Y_{it_0} = 0$ given the covariates can be expressed with a generalized additive model and a logit link function. Using a similar decomposition as in (10), we consider the following logistic regression models, for $t = t_1, \ldots, t_m$,

$$\operatorname{logit}\left(\mathbb{P}\left[Y_{it}^{D_i} - Y_{it_0}^0 = 0 \mid X_i\right]\right) = \beta_{0t}^0 + \sum_{r=1}^{R-1} \mathbb{1}_{\{D_i = r\}} \delta\beta_{0t}^r + \sum_{j=1}^{p} \beta_{jt}(X_{ij}), \quad (16)$$

where $\beta_{jt}(.)$ are unknown smooth univariate functions. For our purpose, the most important parameters are the differential effects $\delta\beta_{0t}^r$, $r = 1, \ldots, R-1$. For example, if $\delta\beta_{0t}^r > 0$, then the probability no variation is larger under policy r compared to no policy at all ($r = 0$) given the covariates $X_i = (X_{i1}, \ldots, X_{ip})$. Recall that the unknown functions $\beta_{jt}(X_{ij})$ are not necessarily linear and that it would be possible to consider a more sophisticated model that could take interaction effects into account, replacing $\beta_{jt}(X_{ij})$ by $\beta_{jt}^r(X_{ij})$, for $r = 1, \ldots, R-1$.

3.4 Estimation procedure

We observe, for a statistical unit i, the realized outcomes $Y_{it}^{D_i}$ at instants $t = t_0, \ldots, t_m$, whereas the counterfactuals Y_{it}^r, for $r \neq D_i$, cannot be observed. The estimation of the parameters and functions defined in (12), (13) and (16), relies on the sample $(Y_{it}^{D_i}, X_i, D_i)_{i=1,\ldots,n}$, for $t \in \{t_0, \ldots, t_m\}$. We assume that there are no spatial interactions between the statistical units so that $(Y_{it}^{D_i}, X_i, D_i)$ and $(Y_{\ell t}^{D_\ell}, X_\ell, D_\ell)$ can be supposed to be independent if $i \neq \ell$. See however Section 4.6, in which the possibility of spatial spillover effects is introduced via the definition of additional covariates that take account of the treatments received by the neighboring municipalities. The $t_m - t_0$ samples $(Y_{it}^{D_i} - Y_{it_0}^0, X_i, D_i)_{i=1,\ldots,n}$, with $t = t_1, \ldots, t_m$ are used separately to estimate the parameters of interest and the regression functions.

The fact that the considered mixture in (14) is a mixture of a continuous variable and a discrete variable makes the computation of the likelihood simple. Indeed, as far as the continuous part is concerned, the probability of no variation is equal to zero and we can proceed as if the two underlying distributions were adjusted separately. For each instant t, the likelihood, given the sample $(Y_{it}^{D_i} - Y_{it_0}^0, X_i, D_i)_{i=1,\ldots,n}$ is equal to

$$\mathcal{L}_t = \prod_{i=1}^{n} p_{it}^{T_{it}} (1 - p_{it})^{1-T_{it}} f_t^{D_i} (Y_{it}^{D_i} - Y_{it_0}^0; X_i, D_i)^{1-T_{it}}$$

where $T_{it} = \mathbb{1}_{\{\Delta_{it}^{D_i}=0\}}$ is the indicator function of no variation between t and t_0 and $p_{it} = \mathbb{P}\left[\Delta_{it}^{D_i} = 0 \mid X_i, D_i\right]$. Taking account now of the different policies, the log-likelihood can be expressed as follows,

$$\ln \mathcal{L}_t = \sum_{i:T_{it}=1} \ln p_{it} + \sum_{i:T_{it}=0} \ln(1 - p_{it}) \tag{17}$$

$$+ \sum_{i:T_{it}=0} \sum_{r=0}^{R-1} \mathbb{1}_{\{D_i=r\}} \ln f_t^r (Y_{it}^{D_i} - Y_{it_0}^0; X_i, D_i), \tag{18}$$

so that the probability of no variation can be estimated separately by maximizing the terms at the right-hand side of (17), whereas the additive models related to the continuous variation of $Y_t^D - Y_{t_0}^0$ are estimated by maximizing the function at the right-hand side of (18). This means that in practice, the subsample $\{i \mid T_{it} = 0\}$ is used for the adjustment of the additive models related to the continuous part. The estimation of the unknown functional parameters introduced in (12), (13) and (16), which are supposed to be smooth functions, is performed thanks to the `mgcv` library in the R language (see Wood, 2017, for a general presentation). The regression functions to be estimated are expanded in spline basis and a penalized likelihood criterion is maximized. Penalties, tuned by smoothing parameters, are added to the log-likelihood in order to control the trade off between smoothness of the estimated functions and fidelity to the data. To select the values of the smoothing parameters, restricted maximum likelihood (REML) estimation was preferred over alternative approaches such as Generalized Cross Validation (GCV) or Akaike's Information Criterion (AIC), since such approaches may undersmooth and are more likely to develop multiple minima than REML. Pointwise confidence intervals that take account of the smoothing parameter uncertainty (Wood et al., 2016) and variable selection is performed following Marra and Wood (2011).

4 Results

4.1 Parameters of interest

We focus on the assessment of ZRR and 5B as well as their joint mean effect. The partial overlap of these two schemes makes possible the identification of the interaction effect of ZRR and 5B. We thus adopt a framework with $R = 4$ multiple potential outcomes, and consider the generalized treatment variable, $\mathtt{D}_i \in \{0, ZRR, 5B, ZRR\&5B\}$ indicating the programme in which municipality i actually participated. The modality 0 indicates that the municipality i did not receive any policy, ZRR (respectively $5B$) indicates that the municipality i received incentives only from ZRR (respectively only from 5B) and $ZRR\&5B$ indicates that the municipality i received incentives both from ZRR and 5B.

The mean differential effect, $\mathbb{E}\left[Y_{it}^r - Y_{it}^0 \mid X_i\right]$, of policy r compared to no policy, for a unit with characteristics X_i, is expressed in (15) and depends on different ingredients. Estimations of α_{it}^r and $\left[g_t^0(X_i) - g_{t_0}^0(X_i)\right]$ are related to the continuous part of the model while the discrete one provides us information about $\mathbb{P}\left[\Delta_{it}^r = 0 \mid X_i\right]$ and $\mathbb{P}\left[\Delta_{it}^0 = 0 \mid X_i\right]$.

As far as the continuous part of the model is concerned, parameter α_t^{5B} measures the mean differential effect, over the whole sample, of policy 5B compared to no policy at all ($r = 0$) whereas the joint effect of ZRR and 5B is given by $\alpha_t^{ZRR\&5B}$. Finally, concerning the effect of ZRR, it is worth to note that only few municipalities (precisely 722) are treated in this case. Consequently, we prefer to focus our attention on the 7014 municipalities that receive incentives both from 5B and ZRR and we calculate the following differential effect $\alpha_t^{ZRR} = \alpha_t^{ZRR\&5B} - \alpha_t^{5B}$. This differential effect simply represents the difference between the outcome when receiving incentives both from ZRR and 5B and the outcome when only 5B applies. The same reasoning applies to the interpretation of the parameter $\delta\beta_{0t}^r$ when dealing with the estimation of the conditional probability of a null employment variation in (16).

4.2 Homogeneous temporal treatment effects

In a first step, additive models are fitted on the subsamples $\{i \mid Y_{it}^{D_i} - Y_{it_0} \neq 0\}$. For easier comparison with previous studies, we first focus on the mean temporal effects of the different policies over the whole population. We consequently fit a simple model for α_{it}^r as in (13), assuming that $\alpha_{it}^r = \alpha_t^r$, for $t_k \leq t \leq t_m$ and $r \in \{1, \ldots, R-1\}$. This specification is also considered when performing the placebo tests (see Appendix B).

A backward variable selection procedure has been employed to select the variables to be introduced in the regression functions defined in (12) such that the conditional independence assumption (6) holds. This procedure leaded us to retain 11 variables among the 16 initial variables (the selected variables are those reported in Table 1).

Note that we consider pre-treatment covariates, say X_{pre}, in the set of observable variables X to ensure that D causes X and Y causes X do not occur.[2] This is likely to be relevant in our economic context where it could be expected that the covariates prior the introduction of the policy, such as for example the share of qualified workers or the existing stock of infrastructure, cause both the inclusion in the program D, and the potential local employment Y ($X_{pre} \to D$ and $X_{pre} \to Y$). After the introduction of the policy, the level of such covariates, say X_{post}, is likely to be affected by its past values X_{pre}, by the treatment D and finally also by the response variable Y. Indeed, in the example dealing with the variables mentioned above, the share of qualified workers and the stock of infrastructure may be directly affected by the policy ($D \to X_{post}$) and since the introduction of the policy could have also increased local employment ($D \to Y$), this may in turn stimulate the creation of new infrastructure/qualified hires ($Y \to X_{post}$). In such a causal framework, X_{pre} should be controlled for whereas X_{post} should not (see Lee, 2005).

Second, note that the vector X_i may also contain the initial level of employment. Including the initial outcome as a regressor is particularly relevant if the average outcomes of the treated and the control groups differ substantially at the first period, as in this case (see *e.g.* Imbens and Wooldridge, 2009; Lechner, 2015)

As expected, the initial outcome was found highly significant and has been included. In almost all cases the linearity was clearly rejected in favor of nonlinear regression functions. We also remark that not imposing a linear relation in (12) leads to retain a larger number of significant variables compared to simpler linear regression models since there are only 6 significant variables when imposing linear relations. Note finally that almost the same results would have been obtained if we would have employed the double penalty variable selection approach proposed by Marra and Wood (2011) (detailed results are available upon request).

A first relevant fact is that, according to our estimations (Table 3), ZRR has produced a very short-run (*abrupt but transitory*) and quite low mean effect on local employment. The estimated value of α_t^{ZRR} for the pre-program years is close to zero and is clearly not significant (years 1994 and 1995 are useful to perform the placebo tests). Then, the effect grows and rises up to 1.820 (significant at level 5%) in 1999. Afterwards, it sharply decreases and becomes close to zero again. The effect is not significant anymore at the end of the period, reaching similar values to those obtained for the pre-intervention time period.

This result of a short-run and rather low impact of ZRR on local employment has to be compared with previous studies. As far as the French experience is concerned, our result can be seen as a refinement of Behaghel et al. (2015) who did not find any significant average effect of ZRR at a canton level over the period. Our results are globally consistent with this latter work, indicating that for nearly all the years under study ZRR had no significant effect. However, this also shows that allowing in the

[2] Lee (2005) labels *collider* the situation when both D and Y cause X.

model the effect of the policy to vary in time can be useful to show the existence of temporal effects over short periods of time, which otherwise could be missed when looking at average effects over time.

It can be noted that beyond the French experience, the literature generally provides mixed evidence. In some papers a significant effect on employment (Papke, 1994; Ham et al., 2011) is noted for such policies whereas some other works indicate that EZ have been ineffective (Bondonio and Engberg, 2000; Neumark and Kolko, 2010). Similarly to our study, O'Keefe (2004) finds a transitory effect of the California's program on employment.

Next, we examine the temporal mean effect of 5B, which is measured by the parameter α_t^{5B}. The time pattern is clearly different from ZRR, with a *gradual start, long-term duration* effect on local employment since $\widehat{\alpha}_t^{5B}$ grows overtime, becomes significant at 5% when $t = 1999$, reaching the value of 1.64, and remains quite stable until the end of the period (*i.e* $t = 2002$), with significance levels around 10%. This result appears to be consistent not only with some previous studies that described a positive effect of SF on regional growth (Cappelen et al., 2003; Ederveen et al., 2006) but also with the intrinsic nature of such a program which tries to pursue long-run growth using structural instruments such as the development of new infrastructure. Moreover, it also confirms the results of Becker et al. (2010) who focus attention on the effect of Objective 1 on regional growth for NUTS2 and NUTS3 regions and find that on average the effect takes four years to become significant and increases afterwards up to the sixth and last available year after its introduction.

Finally, the mean joint impact of 5B and ZRR, estimated by $\widehat{\alpha}_t^{ZRR\&5B}$, increases quite constantly until 1999, becoming significant for $t = 1997$, just after the introduction of ZRR and reaching a peak of 3.47 (more than twice than the impact of 5B) for $t = 1999$. It then decreases sharply for $t = 2000$ where it stabilizes at about 2.5 and remains significant until the end of the period.

In summary, we provide some new results to the empirical literature on regional development policies as well as a clear evidence supporting the idea that modeling temporal treatment effects is a key for properly assessing the effect of such policies. This general result is likely to apply for many other public policies.

4.3 Heterogeneous treatment effects via non-parametric interactions

The analysis performed in Section 4.2 can be extended in several directions. In this paper, since a major interest lies in assessing possible heterogeneous treatment effects, we examine how the effect of a policy may vary with some economic or demographic characteristics of the municipalities. For that purpose, we consider a generalization of model (13) in which interactions between variables are allowed. Such a model can also be useful to estimate counterfactual outcomes at an individual level taking account of the characteristics of the particular municipality under study, which was not possible with the model estimated in the previous Section.

For computational reasons and because of the curse of dimensionality, model selection has been performed. Only two significant variables are retained to fit α_{it}^r, i.e. the individual differential effects of policy r compared to no policy: the initial level of employment (SIZE) of the municipality and its population density (DENSITY). Using an approximate ANOVA test procedure (see Wood, 2017), an additive structure for α_{it}^r, *i.e* $\alpha_{it}^r = \alpha_t^r + g_{1t}^r(\texttt{SIZE}_i) + g_{2t}^r(\texttt{DENSITY}_i)$ is strongly rejected for all years t in favor of a more general model based on bivariate regression functions,

$$\alpha_{it}^r = \alpha_t^r + g_t^r(\texttt{SIZE}_i, \texttt{DENSITY}_i), \quad r = 1, \ldots, R-1.$$

This means that a fully nonparametric interaction between these two variables must be taken into account to evaluate the effect of a policy.

Results are presented in Table 4 and Figures 3, 4 and 5. A first result that emerges is that the estimates of the parametric part of (13), representing the mean effect of the policies, i.e. α_t^r, are very similar to those obtained adopting the model employed in the previous section and which does not allow to expand α_{it}^r as a function of the covariates. This is particularly true for $\widehat{\alpha}_t^{ZRR}$ and $\widehat{\alpha}_t^{ZRR\&5B}$ having a very similar pattern in terms of magnitude and `p-values` with respect to the above presented estimates.

Next, the examination of the nonparametric part $g_t^r(.)$ of (13), reveals how the mean impact of the policy r is modulated by the characteristics in terms of density and size of each considered statistical unit. In almost all cases, the smooth functions appear to be highly significant, using a Bayesian approach to variance estimation (Marra and Wood, 2012), with generally quite high effective degrees of freedom, thus indicating rather complex functions (see Wood, 2017, for a detailed discussion). Concerning the specific nonparametric effect of each policy, the results are as follows. For all the treatments, we first note that both the magnitude and the shape of the nonparametric effect vary with time. Looking at ZRR, the estimated smooth function $\widehat{g}_t^{ZRR}(\texttt{SIZE}, \texttt{DENSITY})$ is very flat and close to zero at the beginning and at the end of the period whereas it becomes clearly nonlinear with a bell-shaped pattern for a period of a few years after the introduction of the policy. The maximum of these functions is generally reached for levels of DENSITY slightly above 50 and for levels of SIZE at about 150, even if the location of these maxima slightly change over time. For the last two years, the maximum is reached for slightly smaller and denser municipalities. Note that in the plots, the domain of SIZE and DENSITY has been appropriately reduced to focus on municipalities not having a too large sizes or very high levels of density.

The joint nonparametric effect of ZRR and 5B, $\widehat{g}_t^{ZRR\&5B}(\texttt{SIZE}, \texttt{DENSITY})$ behaves similarly in terms of shape and time pattern but with a stronger effect for the years 1999 and 2000. Finally, the estimated nonparametric surface measuring the effect of 5B, $\widehat{g}_t^{5B}(\texttt{SIZE}, \texttt{DENSITY})$, is generally quite flat, even if some positive effects appeared for rather low levels of DENSITY and for $t \geq 1999$.

These results confirm our previous findings of a gradual start, long-term duration mean effect of 5B SF on local employment and of a short-term mean impact of ZRR. They also reveal the existence of nonlinear and non-additive effects of such policies with respect to the initial level of employment (`SIZE`) and to the population density (`DENSITY`).

4.4 Estimating the conditional probability of a null employment variation along time

Generalized additive models based on binomial regression with logit link function are fitted to estimate the probability that a variation of the response between t and t_0 does not occur, given the treatment status and the initial conditions. This conditional probability is expressed in (16).

We focus on the parameters $\delta\beta_{0t}^{r}$, see Section 3.3, for $t_k \leq t \leq t_m$ and $r \in \{1, \ldots, R-1\}$, while the results about the effect of the initial conditions, which enter nonparametrically via the smooth functions $\beta_{jt}(X_{ij})$, are not discussed here but are available upon request.

Again, a backward variable selection procedure has been employed to select the variables to be introduced in the model. The estimation results are presented in Table 4 and indicate that the 5B program has a negative effect on the probability that employment does not vary along time. The estimated parameter $\widehat{\delta\beta}_{0t}^{5B}$ is always negative, in a significant way for nearly all instants t. Referring to (15), this means that $\mathbb{P}\left[\Delta_{it}^{5B} = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^{0} = 0 \mid X_i\right] < 0$. Looking at the effects of ZRR, it can be noted that the estimated parameter $\widehat{\delta\beta}_{0t}^{ZRR}$ is always positive, but is not significant in most of the cases, so that $\mathbb{P}\left[\Delta_{it}^{ZRR} = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^{0} = 0 \mid X_i\right]$ is not significantly different from zero. Finally, the estimated joint policy effect $\widehat{\delta\beta}_{0t}^{ZRR\&5B}$ is always very close to zero and is never significant.

Looking at the total policy effect in (15), these results suggest that, for both ZRR and the joint policy ZRR&5B, this effect is mainly driven by the first part of the expression, i.e. by $(1 - \mathbb{P}\left[\Delta_{it}^{r} = 0 \mid X_i\right]) \times \alpha_{it}^{r}$. Conversely, for 5B, there is an additional effect coming from the second part of the expression, since, as noted before, $\mathbb{P}\left[\Delta_{it}^{5B} = 0 \mid X_i\right] - \mathbb{P}\left[\Delta_{it}^{0} = 0 \mid X_i\right] < 0$.

Note as well that while descriptive statistics (Table 2) indicate that treated municipalities, compared to non-treated ones, more often experienced no variation along time of the employment level, when using the proposed causal framework, the estimation results suggest that *ceteris paribus*, the mean differential policy effect on the probability to experience a null employment variation along time is negative or null, thus providing a completely different result with respect to the descriptive pattern of the data.

4.5 Counterfactual analysis: dynamic policy effects at an individual level

Our statistical model, which is estimated nonparametrically on a large sample (see Table 4), can take account of non-linear and local effects and thus make it possible to conduct a counterfactual analysis at an individual level. This relevant feature is illustrated on a few representative municipalities for which the evolutions of the potential outcomes are estimated and compared under the different possible treatments. These municipalities have been chosen with a clustering partition around medoids procedure (see Kaufman and Rousseeuw, 1990) with four clusters so that they represent four different homogeneous groups. Descriptive statistics are given in Table 5.

Using (15) and (13) we can estimate what would have been the evolution of the expected global effect of each municipality under each policy, taking account of the zero inflation effect. We are also interested in building confidence intervals. Due to the complexity of our statistical estimations at an individual level, which are products of predictions obtained with generalized additive models, the standard delta method cannot be used easily. We consider instead the more flexible bootstrap approach (see *e.g.* Efron and Tibshirani, 1993) to approximate the distribution of the conditional counterfactual outcome of each selected municipality i having characteristics X_i.

We draw $B = 1000$ bootstrap samples and for each bootstrap sample b, with $b = 1, \dots B$, we make the following estimation of the expected counterfactual evolution (see (15)),

$$\begin{aligned}\widehat{\mathbb{E}}^b\left[Y_{it}^r - Y_{it}^0 \mid X_i\right] = &\left(1 - \widehat{\mathbb{P}}^b\left[\Delta_{it}^r = 0 \mid X_i\right]\right) \times \widehat{\alpha}_{it}^{r,b} \\ &- \left(\widehat{\mathbb{P}}^b\left[\Delta_{it}^r = 0 \mid X_i\right] - \widehat{\mathbb{P}}^b\left[\Delta_{it}^0 = 0 \mid X_i\right]\right) \times \left[\widehat{g}_t^{0,b}(X_i) - \widehat{g}_{t_0}^{0,b}(X_i)\right],\end{aligned}$$

where $\widehat{\mathbb{P}}^b\left[\Delta_{it}^r = 0 \mid X_i\right]$ is the estimated probability, with sample b, of no employment variation and $\widehat{\alpha}_{it}^{r,b}$, $\widehat{g}_{t_0}^{0,b}$ and $\widehat{g}_t^{0,b}(X_i)$ are the fitted values. Then, we can deduce, using the percentile method, bootstrap confidence intervals for the conditional expectation $\mathbb{E}\left[Y_{it}^r - Y_{it}^0 \mid X_i\right]$, *i.e* the mean effect at time t on municipality with characteristics X_i of treatment r compared to no treatment.

Bootstrap results are drawn in Figures 6, 7, 8 and 9 for the four municipalities under study. The first selected municipality, which is named DSI1, is an extremely dense and urbanized municipality, with values of `DENSITY` and `URB` greater than the 95th percentile. It is also very rich in terms of `INCOME` and big in terms of `SIZE`, with values of these variables about the 80th percentile. For this municipality, we estimate a positive evolution of employment in the absence of any policy. We can also note that, according to our model, ZRR, 5B and the joint policies ZRR&5B would have no significant effect on the evolution of employment for the considered period. The second municipality, named DS3, is rather dense, urbanized and big, with values of `DENSITY`, `URB and SIZE` about the 75th percentile of our sample. The value of `INCOME` is close to the median. We note that policy 5B has an inverted U effect over time, that is only significant for $t = 1997$ and $t = 1998$, whereas both ZRR and the joint policy ZRR&5B have a strongly significant positive impact on employment, over the whole period. Such an impact increases over the years and reaches a peak for $t = 1999$ and then it seems to be quite stable for the following years. The third municipality, DSI3, is quite close to the median values in terms of `DENSITY` and `SIZE`. For this municipality, considering simultaneously ZRR and 5B permits to improve the positive effects of 5B, particularly between 1998 and 2000. All the policies produce an effect with an inverted U time pattern, even if the effect of ZRR is not significant for most of the years. Finally, for the last municipality DSI7, which is a small and poor municipality, there is a clear positive effect of 5B over all the period (except the last year), with again an inverted U pattern over time. For this municipality, ZRR has instead no significant effect over the whole period.

These results first point out that considering a flexible nonparametric model that allows for heterogeneous effects is crucial in order to have a better understanding of the potential effect of the policies. Second, they specifically highlight that ZRR and 5B are likely to produce effects that vary according to the typology of the municipalities: while 5B produces an effect for very small and rural municipalities, ZRR seems to be more effective for bigger and more dense/urbanized areas. We also note that these two policies do not seem to have significant impact for very dense and large municipalities.

4.6 Extension: spatial spillovers

The proposed model is flexible and modular enough so that it can be extended in various directions. As an illustrative example, we address the relevant issue of the possible existence of policy effects on neighboring municipalities, *i.e.* spatial spillover effects (see *e.g.* Behaghel et al., 2015). To save space, the analysis is restricted to the continuous part of the model. One standard way to deal with this issue consists in introducing, in the model, explanatory variables accounting for the absence or the presence of the policies in the neighboring municipalities. Ex ante, for both ZRR and 5B, the spillovers may be either positive arising directly through a higher labor demand and/or indirectly from agglomeration economies or negative if some substitution effects occur. In practice, the identification of spillovers is an intricate empirical matter, requiring the definition of the neighborhood and the choice of an adequate channel of transmission. We focus here on purely geographic spillovers and adopt a very restrictive notion of neighborhood by considering the spillovers arising from the municipalities sharing a common border. Among the 25593 municipalities under study, 10523 municipalities have all their neighboring municipalities that do not receive any funds, 2496 municipalities have all their neighboring municipalities that are under 5B but not under ZRR while for 239 municipalities, the entire neighborhood is under ZRR but not under 5B. There is also a group of 7888 municipalities that have some neighboring municipalities under 5B and some other neighboring municipalities which are under ZRR. Finally, there is a group of 4447 municipalities with all the neighboring municipalities under both 5B and ZRR.

With this classification in mind, we build a new categorical variable, denoted by $\mathtt{WD}_i \in \{0, 5\_ALL, Z\_ALL, 5\&Z\_SOME, 5\&Z\_ALL\}$, with modalities corresponding to the above mentioned categories and the corresponding parameters are noted $\omega_t^{5\_ALL}, \omega_t^{Z\_ALL}, \omega_t^{5\&Z\_SOME}$ and $\omega_t^{5\&Z\_ALL}$. These parameters capture the spillover effects by measuring the mean differential effect, over the whole sample, with respect to the reference category which is chosen to be 0, *i.e.* the category of municipalities having neighboring municipalities that do not receive any funds. The new variable $\mathtt{WD}_i$ is then added as an additional explanatory variable in the regression functions given in (12) and (13). The estimation results indicate no significant spillover effects, meaning that both ZRR and 5B produced an effect that remains spatially localized. Geographic spillovers are never statistically significant with `p-values` being always very far from standard significance levels. Note finally that the absence of significant spillover effects still holds when considering many alternative definitions of `WD` based on different considerations about geographic proximity (detailed results are available upon request). This result is consistent with a recent literature on regional policy evaluation suggesting that policy spillovers do not occur or at best, they are modest in magnitude (see e.g. Becker et al., 2010; Behaghel et al., 2015; Gobillon et al., 2012).

Interestingly, it appears that if we consider a more flexible model that allows non-parametric interactions effects we get a different picture. In particular, some interactive spillovers appear now highly significant. Note also that after a model selection procedure, the same variables that have been employed in Section 4.3 are retained in the model, that is `SIZE` and `DENSITY`, to interact with `WD`. We also get again that an additive structure is rejected in favor of a bivariate smooth function. This result provides additional empirical support to the importance of considering flexible models in order to let the data a chance to speak.

We finally provide a brief comment to the results presented in Figure 10.[3] For $\mathtt{WD}_i \in \{5\_ALL, 5\&Z\_ALL\}$, we find evidence of significant interactive spillover effects. A relevant result is that, for both modalities, spillovers are very low or even negative for low levels of both `SIZE` and `DENSITY`, while they become positive and reach their maximum level for municipalities characterized by high levels of both variables.

5 Concluding remarks

In this paper, we introduce a semi-parametric approach to estimate the variation along time of regional treatment effects in France. We rely on additive models and generalized

[3] As in previous figures, the domain of the continuous variables has been appropriately reduced to the regions where the effects are significant. To save space we focus only on $t = 1999$.

additive models, giving more flexibility than linear models to fit the data, and we exploit the longitudinal structure of the data to account for selection bias and to estimate dynamic policy effects. Since we face a kind of a zero inflated phenomenon that cannot be dealt with a continuous distribution, we introduce a mixture model that combines a Dirac mass at zero and a continuous density.

We find that while the enterprise zone program yields an abrupt but transitory average effect, we note a gradual start, long-term duration average effect of the structural funds on local employment. The results also reveal that the effect of such policies varies nonlinearly with respect to some covariates. The proposed model also allows to perform a counterfactual analysis at an individual level. This relevant feature is illustrated on a few municipalities for which the evolution over time of the potential outcomes are estimated and compared under the different possible treatments.

This work provides new results about the pattern of temporal treatment effects and nonlinear interactions, as well as some guidance for future research. It first suggests, within a flexible semi-parametric regression framework, a way to deal with an excess of zeros by considering a mixture of a continuous and a discrete distribution. This may be relevant for other policy evaluations when the dependent variable does not vary along time for a non-negligible fraction of the units. Second, the consideration of a model in which the effect of the policy is expanded as a nonlinear function of the covariates provides a richer framework that allows for refined analysis and permits to perform a counterfactual estimation at individual levels. This could be relevant in many cases in which heterogeneous policy effects are likely to be present or when there is an interest in units having some peculiar characteristics.

Acknowledgements: The authors would like to thank Virginie Piguet (INRA Dijon, CESAER) for her valuable help during the preparation of the data and Dr Bertrand Schmitt (INRA Paris, DEPE) for fruitful discussions about regional development policies in France. Calculations were performed using HPC resources from PSIUN CCUB (Centre de Calcul de l'Université de Bourgogne, France).

References

Angrist, J. and Hahn, J. (2004). When to control for covariates? panel asymptotics for estimates of treatment effects. *Review of Economics and statistics*, 86(1):58–72.

Becker, S. O., Egger, P. H., and Von Ehrlich, M. (2010). Going NUTS: The effect of EU structural funds on regional performance. *Journal of Public Economics*, 94(9):578–590.

Behaghel, L., Lorenceau, A., and Quantin, S. (2015). Replacing churches and mason lodges? Tax exemptions and rural development. *Journal of Public Economics*, 125:1–15.

Bondonio, D. and Engberg, J. (2000). Enterprise zones and local employment: evidence from the states' programs. *Regional Science and Urban Economics*, 30(5):519–549.

Brown, J. D., Earle, J. S., and Telegdy, A. (2006). The productivity effects of privatization: Longitudinal estimates from Hungary, Romania, Russia, and Ukraine. *Journal of political economy*, 114(1):61–99.

Cappelen, A., Castellacci, F., Fagerberg, J., and Verspagen, B. (2003). The impact of EU regional support on growth and convergence in the European Union. *JCMS: Journal of Common Market Studies*, 41(4):621–644.

Ederveen, S., Groot, H. L., and Nahuis, R. (2006). Fertile soil for structural funds? a panel data analysis of the conditional effectiveness of European cohesion policy. *Kyklos*, 59(1):17–42.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, New York.

Friedlander, D. and Robins, P. K. (1995). Evaluating program evaluations: New evidence on commonly used nonexperimental methods. *The American Economic Review*, pages 923–937.

Frolich, M. (2004). Programme evaluation with multiple treatments. *Journal of Economic Surveys*, 18(2):181–224.

Gobillon, L. and Magnac, T. (2016). Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls. *Review of Economics and Statistics*, 98(3):535–551.

Gobillon, L., Magnac, T., and Selod, H. (2012). Do unemployed workers benefit from enterprise zones? The French experience. *Journal of Public Economics*, 96(9-10):881–892.

Ham, J. C., Swenson, C., İmrohoroğlu, A., and Song, H. (2011). Government programs can improve local labor markets: Evidence from state enterprise zones, federal empowerment zones and federal enterprise community. *Journal of Public Economics*, 95(7):779–797.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, Ltd., London.

Heckman, J. and Hotz, V. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Amer. Statist. Assoc.*, 84:862–874.

Heckman, J., Lalonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, chapter 31, pages 1865–2097. Elsevier.

Hsiao, C., Ching, H. S., and Wan, S. (2011). A panel data approach for program evaluation-measuring the impact of political and economic interaction of Hong-Kong with Mainland China. *Journal of Applied Econometrics*, 27:705–740.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, 52(12):5186–5201.

Imai, K. and Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding groups in data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. An introduction to cluster analysis, A Wiley-Interscience Publication.

Lechner, M. (2011). The relation of different concepts of causality used in time series and microeconometrics. *Econometric Reviews*, 30:109–127.

Lechner, M. (2015). Treatment effects and panel data. In Baltagi, B., editor, *The Oxford Handbook of Panel Data*. Oxford University Press.

Lee, M.-J. (2005). *Micro-econometrics for policy, program, and treatment effects*. Oxford University Press on Demand.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalised splines. *Biometrika*, pages 415–436.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. *Comput. Statist. Data Anal.*, 55(7):2372–2387.

Marra, G. and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74.

Mohl, P. and Hagen, T. (2010). Do EU structural funds promote regional growth? new evidence from various panel data approaches. *Regional Science and Urban Economics*, 40(5):353–365.

Neumark, D. and Kolko, J. (2010). Do enterprise zones create jobs? evidence from californian enterprise zone program. *Journal of Urban Economics*, 68(1):1–19.

O'Keefe, S. (2004). Job creation in Californians enterprise zones: A comparison using a propensity score matching model. *Journal of Urban Economics*, 55:131–150.

Papke, L. (1994). Tax policy and urban development. evidence from the indiana enterprise zone program. *Journal of Public Economics*, 54:37–49.

Peters, A. and Fisher, P. (2004). The failures of economic development incentives. *Journal of the American Planning Association*, 70:27–37.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rosenbaum, P. R. (2002). Observational studies. In *Observational Studies*, pages 1–17. Springer.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). Semiparametric regression. In *Cambridge series in statistical and probabilistic mathematics*. Cambridge university press.

Sheather, S. J. (2004). Density estimation. *Statistical Science*, 19(4):588–597.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Amer. Statist. Assoc.*, 99(467):673–686.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):495–518.

Wood, S. N. (2017). *Generalized additive models: an introduction with R, second edition*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL.

Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *J. Amer. Statist. Assoc.*, 111:1548–1563.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random coefficient and treatment-effect panel data models. *The Review of Economics and Statistics*, 87:395–390.

A Data, variables and sample

After the merge of the different sheets provided by the INSEE containing information on local employment, the demographic structure, education and land use, we get a data set containing 36000 municipalities, that is the 98,5% of the French municipalities. While the paper focuses specifically on rural development policies, it is worth recalling that a relevant fraction of the municipalities received structural funds (1994-99) not specifically devoted to rural development. These are the Objective 1 and the Objective 2 funds. Objective 1 has the explicit aim of fostering per capita GDP growth in regions that are lagging behind the EU average - defined as those areas with a per capita GDP of less than 75 per cent of the EU average - and of promoting aggregate growth in the EU. Objective 2 covers regions struggling with structural difficulties and aims to reduce the gap in socio-economic development by financing productive investment in infrastructures, local development initiatives and business activities. Table A1 describes the distribution of the municipalities according to the ZRR and the structural funds schemes (1994-99).

Structural Funds/ZRR	0	1	total
0	10831	401	11232
1	378	268	646
2	6815	590	7405
5B	6641	10076	16717
total	24665	11335	36000

Table A1 Distribution of the municipalities according to ZRR and Structural Funds schemes

Among the 646 municipalities under Objective 1, 350 are located in Corsica. All the Corsica's municipalities available in our dataset are under the Objective 1. Among them, 268 were also under ZRR scheme. The remaining 296 municipalities under Objective 1 are located in the region Nord-Pas de Calais and were not under ZRR. Given the small number of municipalities under the Objective 1 and their specific characteristics, we decided to remove them from the analysis. This simplifies greatly the framework of the analysis without losing a relevant amount of information, getting a dataset containing 35354 municipalities.

Concerning the Objective 2, we initially estimated the proposed model by including a treatment variable defined as $D_i \in \{0, EU2, 5B, ZRR, ZRR\&EU2, ZRR\&5B\}$, which also accounts for the Objective 2, $EU2$ ($ZRR\&EU2$) indicating that the municipality i receive incentives only from Objective 2 (from both Objective 2 and ZRR). However, the estimated parameters $\widehat{\alpha}_t^{EU2}$ and $\widehat{\alpha}_t^{ZRR\&EU2}$ ($\widehat{\delta\beta}_{0t}^{EU2}$ and $\widehat{\delta\beta}_{0t}^{ZRR\&EU2}$) were always very close to zero and never significant with `p-values` very far from standard significance levels. This result along with the fact that the interest of this paper is on rural development, motivated the use of the treatment variable defined in Section 4.1 where the Objective 2 municipalities are considered as if they had not received any treatment. The use of such a variable simplifies the analysis and the presentation of the results without losing relevant information, also provided that the our parameters of interest α_t^{5B}, $\alpha_t^{ZRR\&5B}$ and α_t^{ZRR} ($\delta\beta_{0t}^{5B}$, $\delta\beta_{0t}^{ZRR\&5B}$ and $\delta\beta_{0t}^{ZRR}$) are fondamentally not affected by such a choice.

Concerning the definition of the variables, the dependent variable Y_{it} (i indicating the municipality; and t the time $t = 1993, ..., 2003$) measures the number of employees and it has been calculated from the SIRENE data sheet covering manufacture, trade and services, while the initial full set of regressors (measured at time $t = 1990$) is composed of the following 16 variables:

Initial outcome

$\texttt{SIZE}_i \equiv Y_{it_0}$ is the initial outcome, i.e the level of employment at t_0, with t_0 equals to 1993;

Socio-economic and demographic variables

$\texttt{DENSITY}_i \equiv (population)_i \,/\, \left(surface\ in\ terms\ of\ \ km^2\right)_i$;

$\texttt{OLD}_i \equiv (population\ over\ 65)_i \,/\, (total\ population)_i$;

$\texttt{INC}_i \equiv (net\ taxable\ income)_i \,/\, (total\ population)_i$;

$\texttt{FACT}_i \equiv (number\ of\ factory\ workers)_i \,/\, (total\ population)_i$;

$\texttt{EXE}_i \equiv (number\ of\ executive\ workers)_i \,/\, (total\ population)_i$;

$\texttt{FARM}_i \equiv (number\ of\ \ farmers)_i \,/\, (total\ population)_i$;

$\texttt{UNIV}_i \equiv \frac{(number\ of\ people\ with\ a\ master\ level\ degree\ called\ \text{“Maîtrise universitaire”})_i}{(total\ population)_i}$;

$\texttt{BTS}_i \equiv \frac{(number\ of\ people\ with\ a\ \ technical\ degree\ called\ \text{“Brevet de Technicien Supérieur”})_i}{(total\ population)_i}$;

$\texttt{NOEDU}_i \equiv (number\ of\ people\ without\ a\ degree)_i \,/\, (total\ population)_i$;

Land use

$\texttt{AGRI}_i \equiv (farmland\ surface)_i \,/\, (total\ surface)_i$;

$\texttt{CULT}_i \equiv (cultivated\ land\ surface)_i \,/\, (total\ surface)_i$;

$\texttt{URB}_i \equiv (urban\ surface)_i \,/\, (total\ surface)_i$;

$\texttt{IND}_i \equiv (industrial\ surface)_i \,/\, (total\ surface)_i$;

$\texttt{ARA}_i \equiv (arable\ surface)_i \,/\, (total\ surface)_i$;

$\texttt{GRA}_i \equiv (grassland\ surface)_i \,/\, (total\ surface)_i$;

The socio-economic and demographic variables come from standard INSEE sources while the variables measuring land use have been obtained from the “Corine Land Cover” base (providing remote sensing images which have been merged with the French map at a municipality level).

The retained models, those results are presented in Section 4, have been obtained using a backward selection procedure starting from the above set of potential explanatory variables. Backward selection provided almost the same results as the double penalty approach proposed by Marra and Wood (2011), those detailed results are available upon request. More precisely, we selected the variables equation-by-equation for $t = 1994, ..., 2002$, by setting the threshold level for the `p-values` to 0.01 and in the end, to use the same explanatory variables for all t, we choosed the variables that were 1% significant at least for one time period, t. According to the notation used in Section 3, these variables are noted as $X_{Y \cap D}$ and $X_{Y \cap \bar{D}}$.

Concerning the estimation of the the conditional probability of a null employment variation along time which is expressed in eq. (16), we retained the following variables:

$$X_i^{\{logit\}} = (SIZE_i, DENSITY_i, UNIV, INC, FACT_i, EXE, FARM, BTS_i, NOEDU, ARA_i, URB_i, IND_i, GRA)\,,$$

while for the continuous part of the model referring to the subsample $\{i\,|Y_{it}^{D_i} - Y_{it_0} \neq 0\}$, the variables that we selected are:

$$X_i^{\{continuous\}} = (SIZE_i, DENSITY_i, OLD_i, INC, FACT_i, BTS_i, CULT_i, AGRI_i, ARA_i, URB_i, IND_i)\,.$$

Also note that according to Heckman and Hotz (1989, pg. 865), selection bias may also arise from the presence of variables that may influence selection into the program even if they do not affect directly the outcome and introducing these variables into the regression solves this additional source of selection bias. Using the notation employed in Section 3, these variables are noted as $X_{\bar{Y}\cap D}$. We determine these variables by exploiting recent advances in generalized additive models permitting the estimation of multinomial logistic regression (Wood et al., 2016). This allows a flexible estimation of a generalized propensity score $\mathbb{P}[D_i \mid X_i]$ as a function of additive smooth components. Again we used the backward selection and finally we added 3 more variables that appeared to affect selection into the programs and that were not selected directly from the outcome equation. These variables are $FARM_i$, $NOEDU_i$ and GRA_i. However, adding these variables does not produce relevant changes to the estimates of the mean effects and detailed estimation results are available upon request.

Finally, let broadly recall the trimming procedure we used to determine the sample for the estimation. We dropped outlier observations which have been identified using a variety of methods such as the visual inspection of the distribution via kernel density estimation, standard boxplot, adjusted boxplot for skewed distributions (Hubert and Vandervieren, 2008), bivariate inspection and bivariate boxplot (Rousseeuw et al., 1999). The variables we collected generally present an asymmetric distribution and in some cases are characterized by an extremely long right tail. This is the case of $SIZE_i$ (skewness=151) and $DENSITY_i$ (skewness=15.69), which have a crucial role in the model with interactions. For these two variables we ended as follows. For $SIZE_i$, we keep municipalities for which $SIZE_i < 500$, 500 representing the $92th$ percentile while for $DENSITY_i$ we select municipalities having $DENSITY_i < 1000$, 1000 being about the $97th$ percentile. In both cases, the range of the variable has been greatly reduced, from 1128000 to 499 in the first case and from 21940 to 999 in the second one. After the cleaning, the sample used for the estimation contains 25593 municipalities. For such a sample, we globally do not observe problems in terms of lack of overlap. This feature makes the average treatment effect relevant for policy purposes.

B Placebo tests

As underlined by Heckman et al. (1999), when different methods produce different inference would suggest that selection bias is important and that some of the adopted estimators are likely to be misspecified. In order to detect misspecified models, we implement both 'pre-program' and 'post program' tests along the lines depicted by Heckman and Hotz (1989) and implemented empirically in some previous papers (see e.g. Brown et al., 2006; Friedlander and Robins, 1995). These tests are based on the idea that a valid estimator would correctly adjust for differences in pre-program (resp. post-program) outcomes between future (resp. past) participants and non-participants, otherwise the estimator is rejected.

These placebo tests are performed here looking at the effect of ZRR, because the availability of some years prior the introduction of the ZRR incentives, occurred in September 1996, allows us to conduct 'pre-program' tests, while for the program 5B, introduced in 1994, there is not enough statistical information before its introduction. More precisely, we focus attention on the continuous part on the model, and precisely on the parameter α_t^{ZRR}, and compare the *before-after* to the *random growth* model. The former approach as in eq. (5) is motivated by a model in which U_{it} is assumed having the form of eq. (7) while the random growth specification (Heckman and Hotz, 1989; Wooldridge, 2005) assumes the following specification for U_{it} :

$$U_{it} = \phi_{1i} + \phi_{2i}t + v_{it} \tag{19}$$

allowing individual parameters (ϕ_{1i}, ϕ_{2i}) to be correlated with the treatment indicator variable D_i.To estimate the model, we adopt the same tranformation as in Heckman and Hotz (1989), that is $\left[Y_{it}^r - Y_{it_0}^0 - (t - t_0)\left(Y_{it_0}^0 - Y_{it_0-1}^0\right)\right]$ and the underlying conditional independence assumption on a transformed equation can be written as

$$\left[Y_{it}^r - Y_{it_0}^0 - (t - t_0)\left(Y_{it_0}^0 - Y_{it_0-1}^0\right)\right] \perp\!\!\!\perp D_i \mid X_i, \qquad \forall r \in \{0, 1, \ldots, R-1\}. \tag{20}$$

The 'pre-program' test is generally implemented by setting $t < k$ and by testing the significance of the treatment effect α_t^r. If α_t^r is significantly different from 0 then the underlying model fails to pass the test. However, even if the logic is compelling, if a shock or an anticipation effect close to the time of the treatment affects only one group but not the other, the results from such a test are potentially misleading. This problem has also been summarized under the heading *"fallacy of alignment"* (Heckman et al., 1999). In our case, treated firms could (shortly) postpone hiring in order to obtain the public incentives, so that using quite longer lags can be useful in order to obtain an effective test and avoiding to overestimate the treatment effect (Brown et al., 2006; Friedlander and Robins, 1995).

Accordingly, we first use all the available information in the data and use the most distant data before the introduction of the policy to set t_0 and propose, for the *before-after* specification, three tests by setting ($t_0 = 1993$, $t = 1994$), ($t_0 = 1993$, $t = 1995$) and ($t_0 = 1993$, $t = 1996$), respectively. Next we set $t_0 = 1994$. This allows both to make the *before-after* and the *random growth* estimators directly comparable and to verify the robustness of the previous tests to a change in the starting point t_0.

A post program test has an identical structure to the pre-program test except that for such a test $t > k$, when neither group receives the treatment. As for the pre-program test, we alternatively set t_0 to 1993 and 1994, whereas for t we use the last two years in the sample, that is 2001 and 2002. The interpretation of this kind of test could be however more problematic than that of the pre-program test since it could be that a policy has a permanent or a long-term impact on the outcome. However, the fact that some previous studies pointed out that various EZ have only a short run impact on employment makes the post-program test of a certain empirical relevance here. Moreover, even if it cannot be excluded à priori that a rural policy produces an effect only for some few years, it is difficult to imagine a situation in which its effect become negative after some years from its adoption. So a negative and significant estimate of α_t for $t > k$ would suggest that the model is misspecified.

Table B1 about here

The results from such tests (Table B1) provide interesting insights which are summarized below. A first relevant result is that, when analyzing the *before-after* model, setting t_0 alternatively to 1993 and 1994 has no effect on the results of the tests. Secondly, it seems ex-post that the results of the *before-after* specification which does not include the initial conditions are quite unsatisfactory, specifically looking at the post-program tests since the effect of the policy decreases overtime becoming not only negative but also statistically significant at the end of the period for $t = 2001$ and $t = 2002$, with `p-values` very close to zero. Such a negative and decreasing overtime estimates for the post treatment periods could indicate that the assumptions underlying the identification of the causal effect are still too restrictive to obtain a credible result. This could arise because i) the treated municipalities are expected to have a different (i.e. lower) time trend than non treated ones even in absence of the policy; ii) some observable factors can be related to the policy placement (also affecting the outcome variable), those omission from the model causes the so called *overt bias*, to adopt the terminology from Lee (2005) and Rosenbaum (2002). A third relevant result is that adding (nonparametrically) the initial conditions greatly improves the results of the tests (this specification passes both pre and post program tests) and provides much more credible results. Moreover, non reported results indicate that using an additive model instead of a linear specification improves greatly the alignment.

A central issue concerns the comparison of the *before-after* with the *random growth.* If the initial conditions are not included the *random growth*, similarly to the *before-after*, does not pass the post-program tests and provide negative and decreasing overtime estimates of the treatment effect with with `p-values` below standard levels. When the initial conditions are included, the results are as follows. While the *before-after* clearly passes the tests with estimates close to zero and not significant (`p-values` are equal to 0.771 and 0.847), *the random growth* still provides estimates of α_t for the post-program period which are highly negative (-3.681 and -4.787) and show a decreasing trend overtime with associated `p-values` equal to 0.300 and 0.232, which are much lower than those obtained with the *before-after* specification. Looking at the estimates for all available t may provide further insights. The *random growth* provides estimates of the effect of ZRR that are negative for all t, are relatively high in magnitude and are increasing in absolute value with t.

These tests suggest the use of a *before-after* specification added with the initial conditions and allowing for nonparametric effects of such initial conditions. For such a model, a very good alignment is obtained pre and post treatment. We do not intend to claim that we have found the 'true' model but a purpose of this paper has been to reduce the risk of misspecification by relying on semi-parametric modeling and variable selection and by discarding specifications that fail to provide a good alignment.

	FULL SAMPLE		**ZONING SCHEME**							
			0		*5B*		*ZRR&5B*		*ZRR*	
n	25593		12580		5277		7014		722	
%			0.491		0.206		0.274		0.028	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
SIZE	56.37	92.281	68.201	101.316	57.22	93.453	36.665	70.317	35.558	67.252
DENSITY	56.02	73.590	78.084	91.599	49.603	49.364	24.215	27.592	27.300	33.782
OLD	0.174	0.064	0.143	0.052	0.179	0.056	0.222	0.060	0.190	0.057
INC	4838	1160.315	5302.966	1166	4577.307	951.730	4207.766	917.3673	4773.478	1044.728
FACT	0.141	0.053	0.152	0.050	0.142	0.052	0.119	0.054	0.140	0.058
BTS	0.028	0.022	0.031	0.021	0.027	0.020	0.026	0.022	0.024	0.022
CULT	0.029	0.119	0.039	0.140	0.0314	0.120	0.012	0.069	0.013	0.067
AGRI	0.204	0.181	0.180	0.180	0.264	0.188	0.206	0.169	0.176	0.144
ARA	0.325	0.306	0.421	0.319	0.216	0.252	0.232	0.263	0.342	0.283
URB	0.022	0.039	0.032	0.046	0.018	0.033	0.007	0.019	0.009	0.214
IND	0.001	0.005	0.001	0.007	0.0008	0.005	0.0003	0.003	0.0004	0.003

Table 1: Descriptive statistics. The precise definition of the variables can be found in Appendix A.

	FULL SAMPLE			ZONING SCHEME												
				0			*5B*			*ZRR&5B*			*ZRR*			
t	Mean	M	Mode	Mean	M	Mode(%)	Mean	M	Mode(%)	Mean	M	Mode(%)	Mean	M	Mode(%)
1994	-0.06	0	0(26)	0.21	0	0(22)	-0.44	0	0(24)	-0.20	0	0(33)	-0.66	0	0(36)
1995	-0.27	0	0(20)	0.06	0	0(16)	-0.56	0	0(18)	-0.59	0	0(25)	-1.10	0	0(27)
1996	1.08	0	0(16)	1.69	0	0(12)	0.78	0	0(14)	0.35	0	0(21)	-0.17	0	0(23)
1997	-1.01	0	0(13)	-0.99	0	0(11)	-1.16	0	0(11)	-0.88	0	0(18)	-1.58	0	0(19)
1998	0.76	0	0(12)	1.27	0	0(10)	0.65	0	0(11)	0.12	0	0(17)	-0.76	0	0(18)
1999	4.58	1	0(11)	5.30	1	0(9)	4.67	1	0(10)	3.67	1	0(15)	0.06	0	0(15)
2000	5.28	1	0(11)	6.49	1	0(9)	5.39	1	0(9)	3.52	1	0(15)	0.29	1	0(14)
2001	8.72	2	0(10)	10.75	2	0(8)	9.01	2	0(9)	5.49	1	0(13)	2.45	1	0(14)
2002	12.88	2	0(9)	16.32	2	0(7)	13.09	3	0(8)	7.36	1	0(13)	4.66	1	0(13)

Table 2: The evolution of employment overtime. Values refer to Δ Employment calculated as Employment(t)-Employment(t_0). Time t is allowed to vary between 1994 to 2002 and t_0 is equal to 1993. M indicates the median. The values between brackets indicate the relative frequency in terms of percentage of the modal value.

t	α_t^{ZRR}	α_t^{5B}	$\alpha_t^{ZRR\&5B}$
1996	0.537	0.161	0.698
	(0.475)	(0.812)	(0.353)
1997	0.784	0.835	1.619
	(0.303)	(0.229)	(0.034)
1998	0.963	0.7721	1.735
	(0.242)	(0.303)	(0.035)
1999	1.824	1.646	3.470
	(0.048)	(0.050)	(0.001)
2000	1.298	1.230	2.529
	(0.183)	(0.162)	(0.009)
2001	1.003	1.630	2.633
	(0.358)	(0.098)	(0.015)
2002	0.4612	1.837	2.2985
	(0.719)	(0.114)	(0.073)

Table 3: Preliminary results: homogeneous temporal treatment effects. `p-values` are in brackets.

	CONTINUOUS PART			DISCRETE PART		
t						
	α_t^{ZRR}	α_t^{5B}	$\alpha_t^{ZRR\&5B}$	$\delta\beta_{0t}^{ZRR}$	$\delta\beta_{0t}^{5B}$	$\delta\beta_{0t}^{ZRR\&5B}$
1996	-0.049	0.114	-0.010	0.108	-0.055	0.053
	(0.958)	(0.868)	(0.990)	(0.061)	(0.348)	(0.334)
1997	-0.1418	0.894	0.764	0.147	-0.147	-0.001
	(0.882)	(0.201)	(0.415)	(0.017)	(0.018)	(0.993)
1998	0.2119	0.864	1.087	0.0874	-0.124	-0.037
	(0.8367)	(0.251)	(0.277)	(0.168)	(0.050)	(0.535)
1999	2.159	1.378	3.537	0.047	-0.131	-0.083
	(0.063)	(0.100)	(0.001)	(0.463)	(0.044)	(0.178)
2000	1.372	0.721	2.381	0.054	-0.098	-0.043
	(0.258)	(0.438)	(0.044)	(0.419)	(0.142)	(0.491)
2001	1.0862	1.376	2.454	0.051	-0.131	-0.079
	(0.418)	(0.173)	(0.059)	(0.460)	(0.058)	(0.225)
2002	-0.174	1.017	1.279	0.124	-0.089	0.034
	(0.912)	(0.408)	(0.406)	(0.071)	(0.212)	(0.594)
	g_t^{ZRR}	g_t^{5B}	$g_t^{ZRR\&5B}$			
1996	10.666	7.766	10.574			
	(3.33e-08)	(4.72e-04)	(1.67e-09)			
1997	11.019	5.725	11.034			
	(3.91e-08)	(0.144)	(2.83e-10)			
1998	10.911	3.495	10.960			
	(3.26e-10)	(0.033)	(2e-16)			
1999	12.703	3.029	12.703			
	(1.25e-15)	(7.88e-04)	(5.15e-15)			
2000	13.195	7.750	13.232			
	(3.24e-16)	(6.66e-07)	(1.36e-14)			
2001	10.144	5.088	8.695			
	(2e-16)	(2.15e-04)	(2e-16)			
2002	7.977	7.842	8.285			
	(5.43e-13)	(4.73e-10)	(2.47e-13)			

Table 4: Main results. For the continuous part, $\alpha_{it}^r = \alpha_t^r + g_t^r(SIZE, DENSITY)$ and non-isotropic tensor product splines (Wood, 2006) are used for the bivariate functions $g_t^r(SIZE, DENSITY)$. For such nonparametric components: we report the effective degrees of freedom with `p-values` in brackets. For the parametric components of both continuous and discrete parts, α_t^r and $\delta\beta_{0t}^{ZRR\_b}$, we report the estimated coefficient with `p-values` in brackets.

Municip.	DENSITY	SIZE	INCOME	OLD	FACT	BTS	AGRIH	URB
DSI1	218.85	105	5772	0.11	0.19	0.016	0.08	0.23
DS3	61.26	48	4324	0.30	0.06	0.037	0.19	0.032
DSI3	41.87	25	6300	0.20	0.13	0.038	0.03	0.028
DSI7	22.74	10	3724	0.14	0.16	0.007	0.22	0.015

Table 5: Descriptive statistics of the municipalities selected for counterfactual analysis.

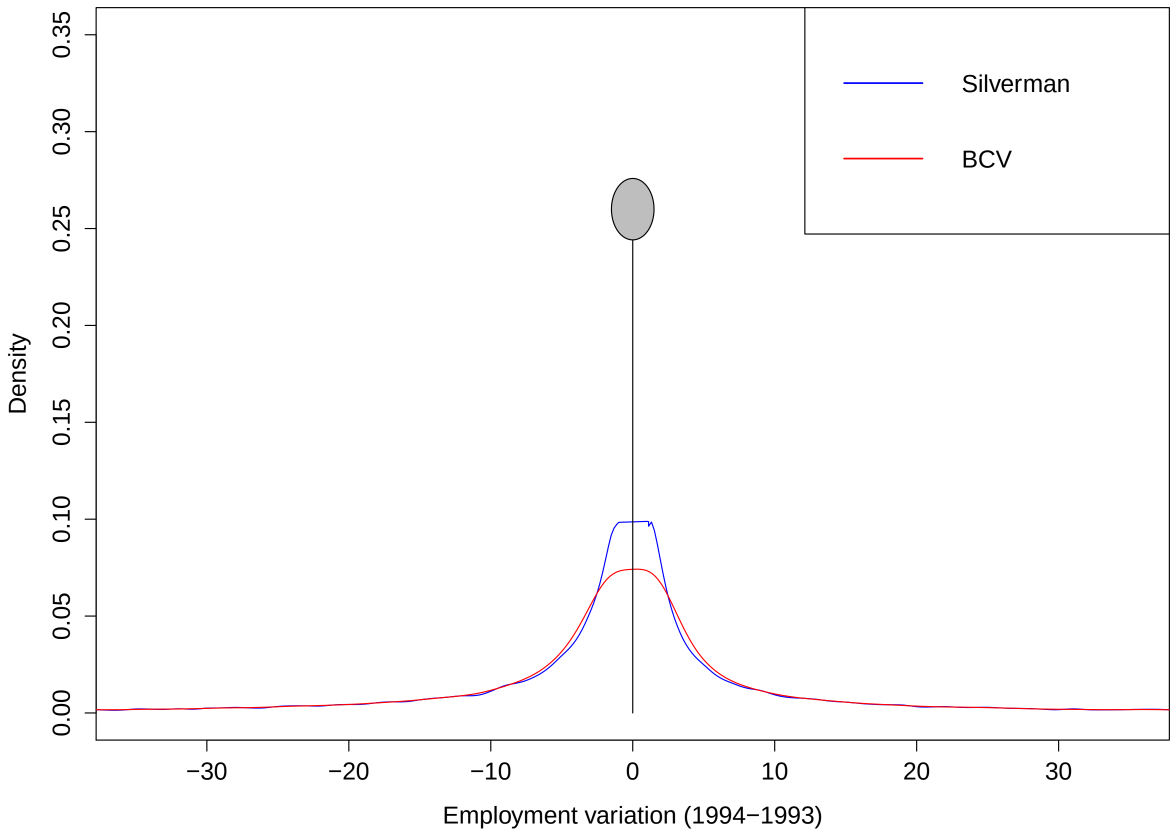

Figure 2: The estimated distribution of $Y_{it}^{D_i} - Y_{it_0}$ for $t = 1994$ and $t_0 = 1993$. The probability of observing no variation is estimated by the proportion of observations such that $Y_{it}^{D_i} - Y_{it_0} = 0$ whereas the continuous density of $Y_{it}^{D_i} - Y_{it_0} \neq 0$ is estimated thanks to kernel density estimators, with two different standard ways of selecting the bandwidth value. Silverman: Silverman's rule of thumb; BCV: Biased Cross Validation (see Silverman, 1986; Sheather, 2004).

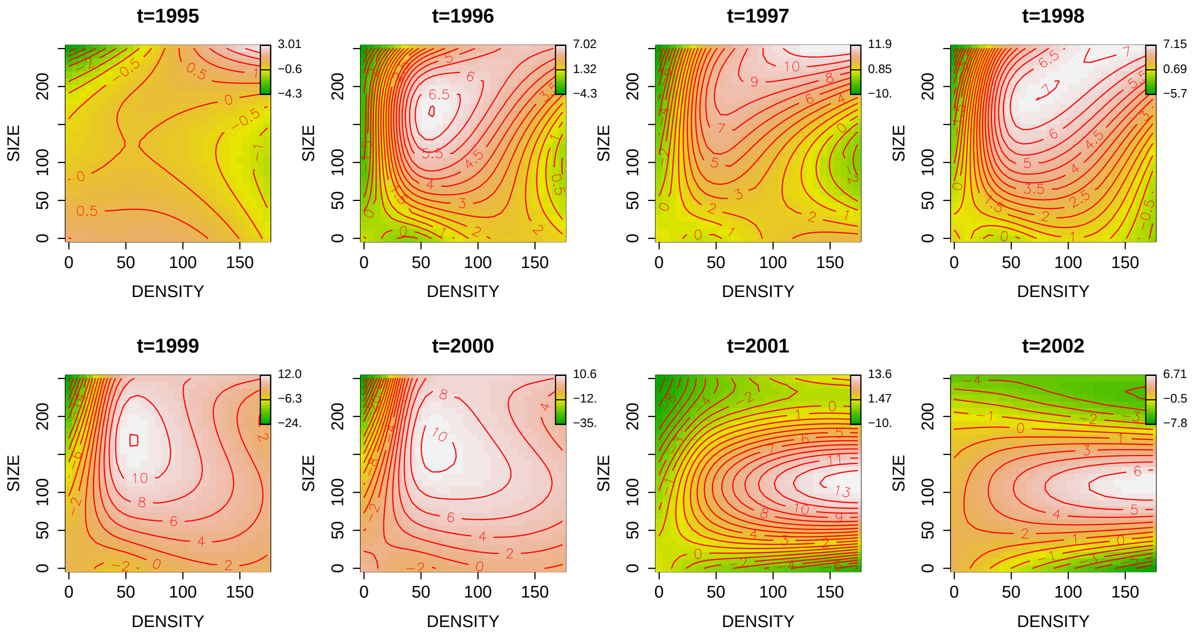

Figure 3: Contour plots of $\widehat{g}_t^{ZRR}(\texttt{SIZE}, \texttt{DENSITY})$.

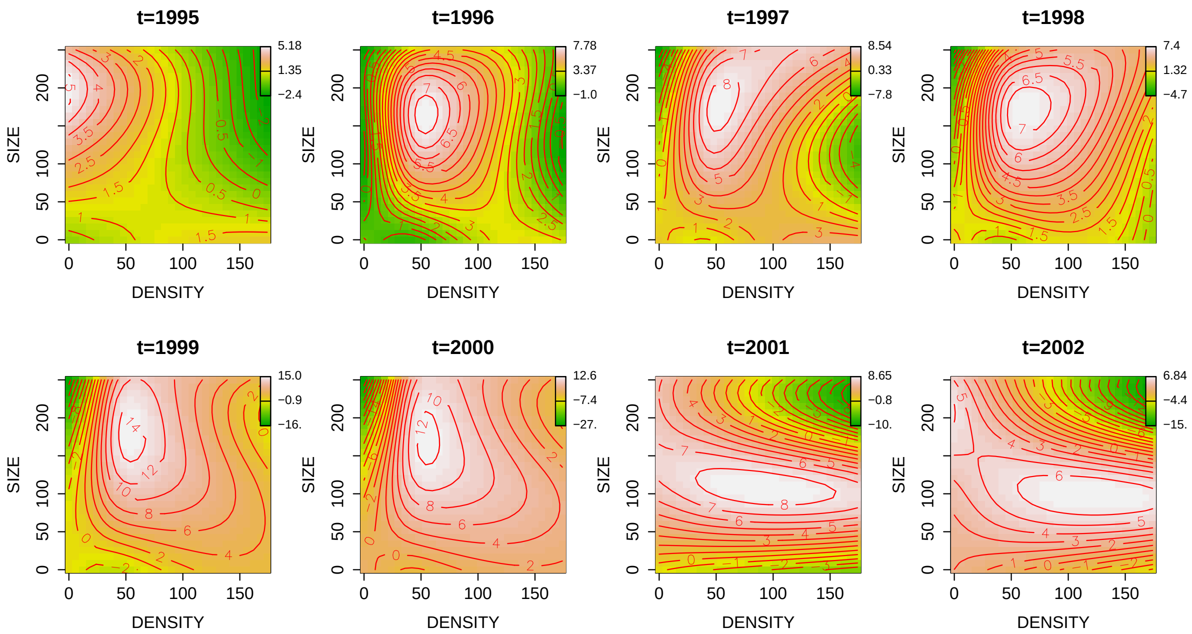

Figure 4: Contour plots of $\widehat{g}_t^{ZRR\&5B}(\texttt{SIZE}, \texttt{DENSITY})$.

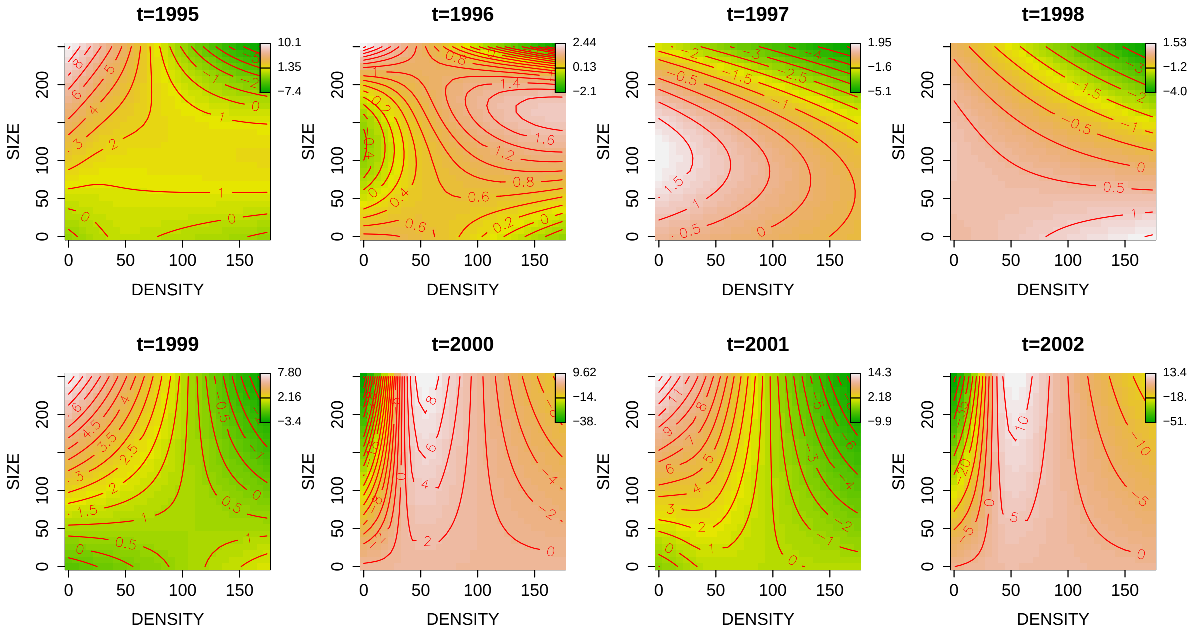

Figure 5: Contour plots of $\widehat{g}_t^{5B}(\texttt{SIZE}, \texttt{DENSITY})$.

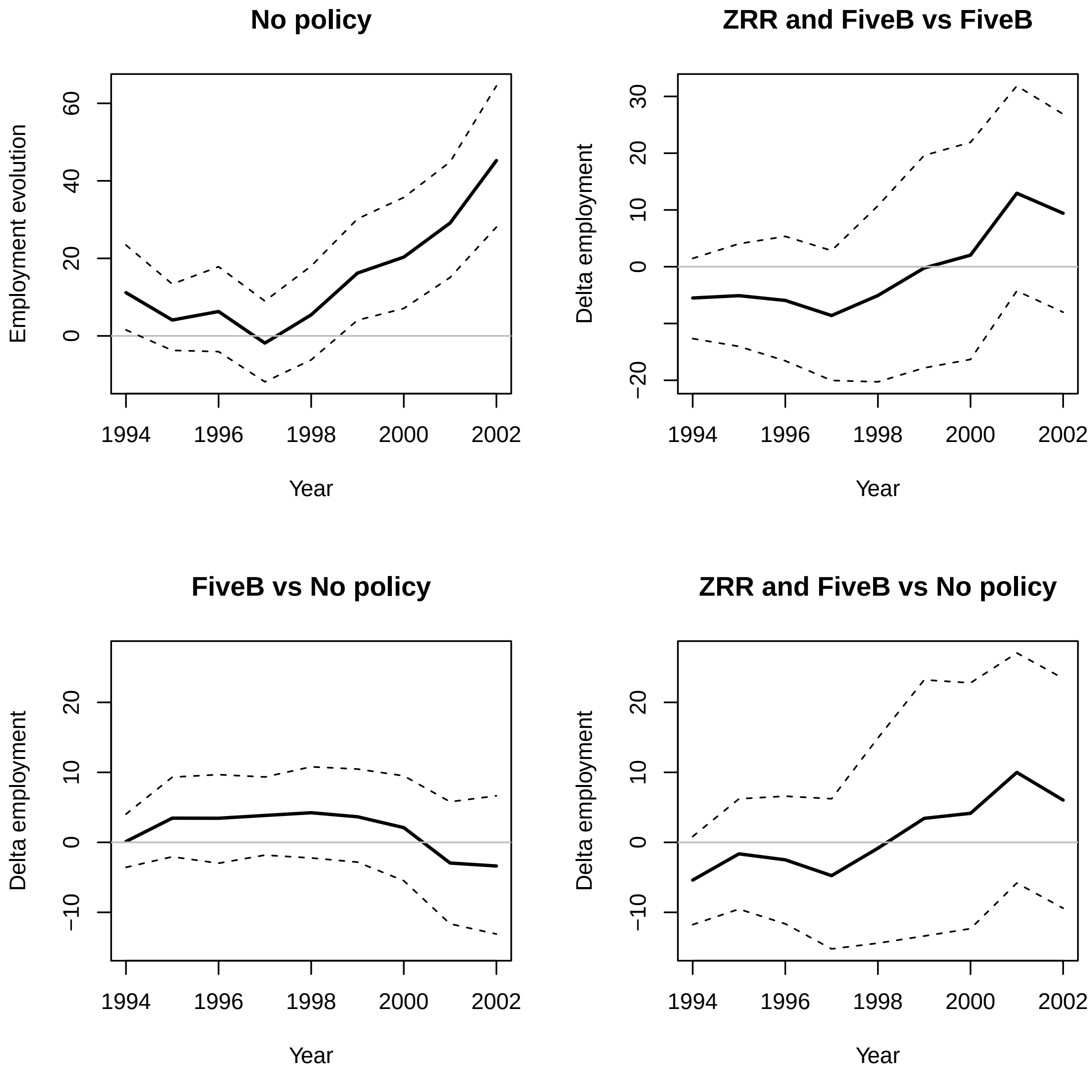

Figure 6: Counterfactual estimation of the evolution of the employment level for municipality DSI 1. The first plot (top left) represents the estimated evolution of employment when no funds are given to the municipality. The others plots represent the difference of evolution between the joint policies ZRR and Five B compared to only Five B (top right), between Five B and no policy (bottom left) and between the joint policies ZRR and Five B compared to no policy at all. Mean values are drawn in plain line and 95% bootstrap confidence intervals in dotted line.

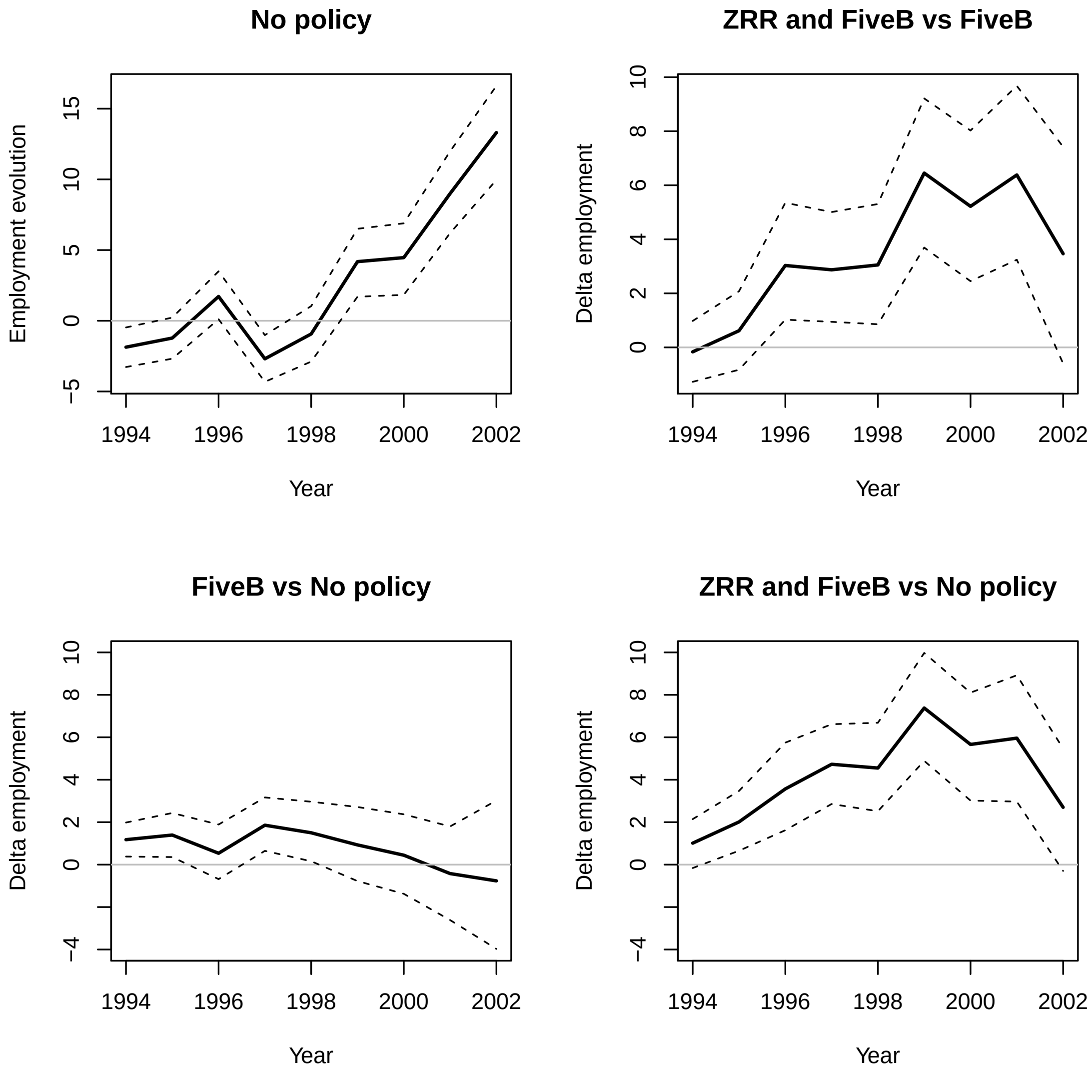

Figure 7: Counterfactual estimation of the evolution of the employment level for municipality DS3. The first plot (top left) represents the estimated evolution of employment when no funds are given to the municipality. The others plots represent the difference of evolution between the joint policies ZRR and Five B compared to only Five B (top right), between Five B and no policy (bottom left) and between the joint policies ZRR and Five B compared to no policy at all. Mean values are drawn in plain line and 95% bootstrap confidence intervals in dotted line.

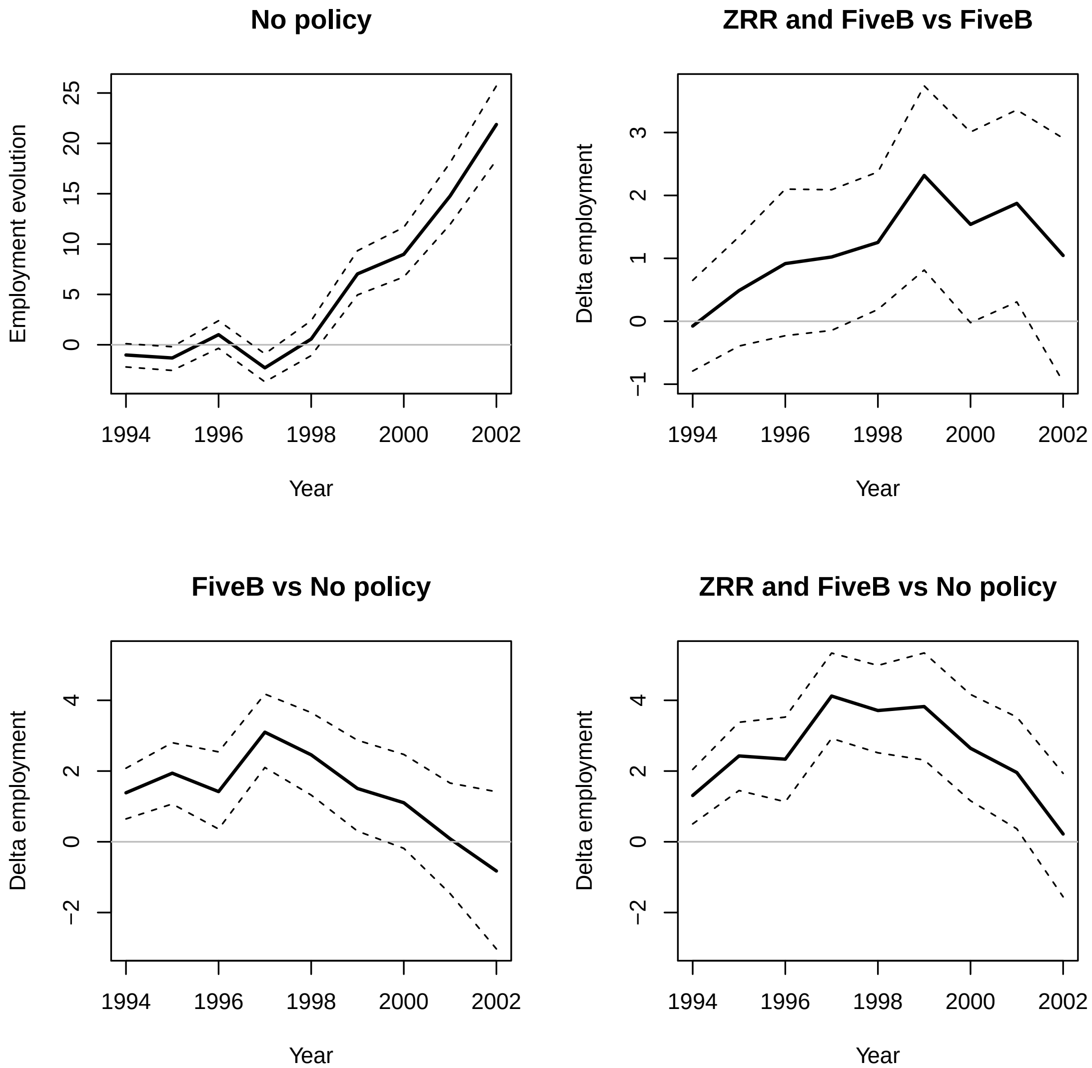

Figure 8: Counterfactual estimation of the evolution of the employment level for municipality DSI 3. The first plot (top left) represents the estimated evolution of employment when no funds are given to the municipality. The others plots represent the difference of evolution between the joint policies ZRR and Five B compared to only Five B (top right), between Five B and no policy (bottom left) and between the joint policies ZRR and Five B compared to no policy at all. Mean values are drawn in plain line and 95% bootstrap confidence intervals in dotted line.

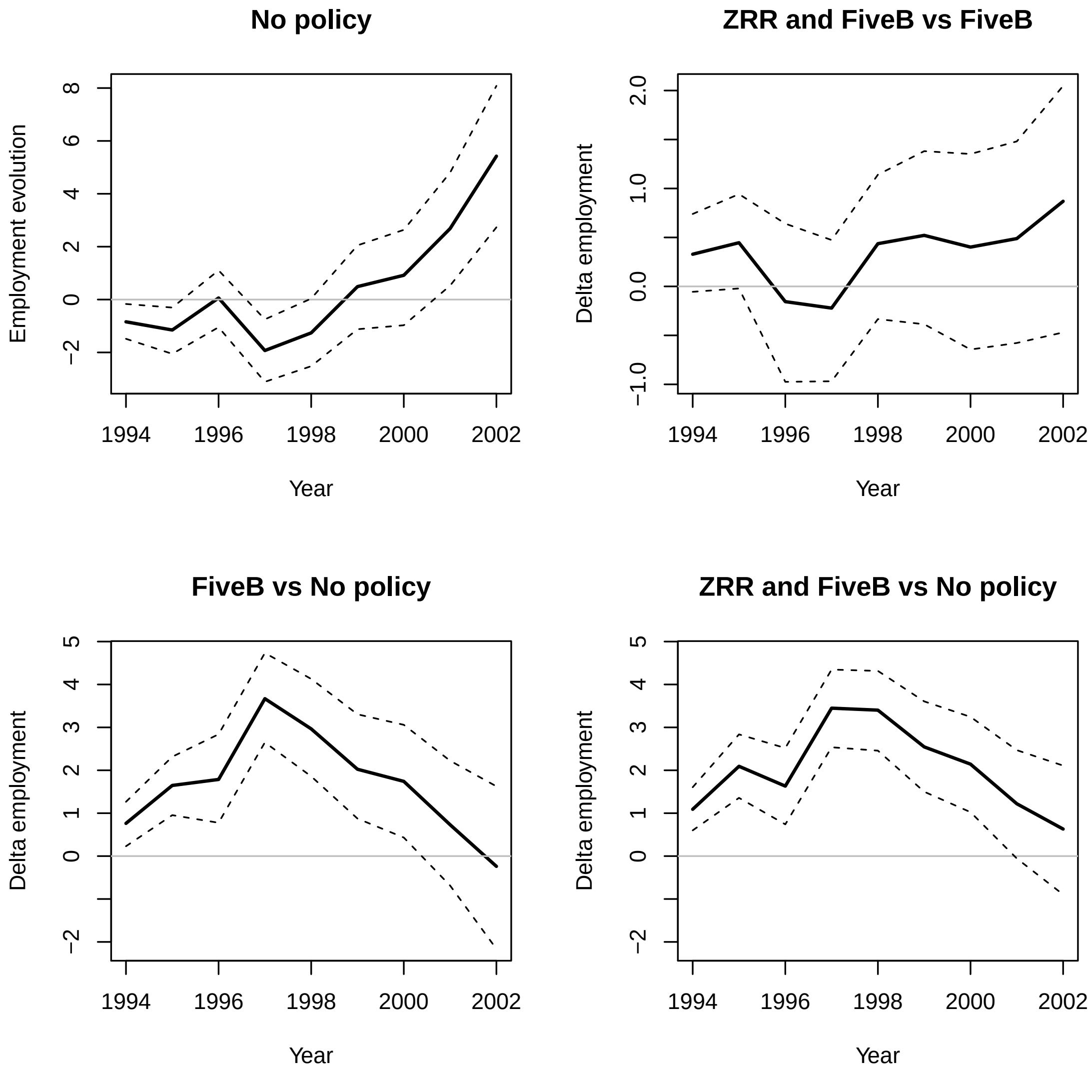

Figure 9: Counterfactual estimation of the evolution of the employment level for municipality DSI 7. The first plot (top left) represents the estimated evolution of employment when no funds are given to the municipality. The others plots represent the difference of evolution between the joint policies ZRR and Five B compared to only Five B (top right), between Five B and no policy (bottom left) and between the joint policies ZRR and Five B compared to no policy at all. Mean values are drawn in plain line and 95% bootstrap confidence intervals in dotted line.

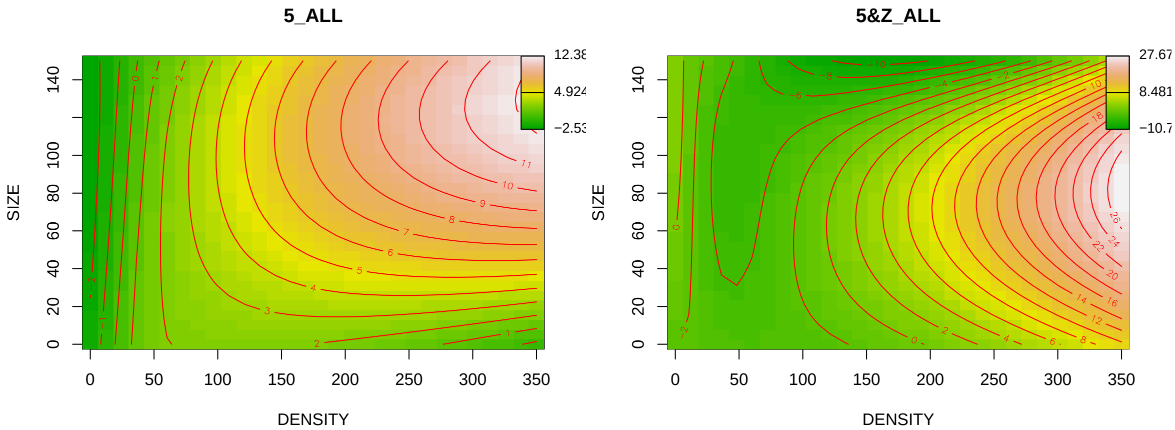

Figure 10: Spillover effects. Contour plots.

	Pre-program test						Post-program test			
	α_t^{ZRR}	p-value	α_t^{ZRR}	p-value	α_t^{ZRR}	p-value	α_t^{ZRR}	p-value	α_t^{ZRR}	p-value
	t=1994 t_0=1993		t=1995 t_0=1993		t=1996 t_0=1993		t=2001 t_0=1993		t=2002 t_0=1993	
Before-After										
No control variables	0.271	0.551	-0.113	0.814	-0.459	0.445	-3.523	3.245e-05	-5.775	3.855e-08
Control variables (X_i)	0.777	0.207	0.262	0.703	0.537	0.475	1.003	0.358	0.461	0.719
	t=1995 t_0=1994		t=1996 t_0=1994				t=2001 t_0=1994		t=2002 t_0=1994	
Before-After										
No control variables	-0.462	0.296	-0.728	0.190			-3.707	3.672e-06	-5.984	2.625e-09
Control variables (X_i)	-0.541	0.386	-0.200	0.768			0.301	0.771	-0.237	0.847
Random growth										
No control variables	-0.627	0.311	-1.193	0.233			-5.376	0.0444	-7.819	0.010
Control variables (X_i)	-1.114	0.167	-1.338	0.299			-3.681	0.300	-4.787	0.232

Table B1: Placebo tests