



**HAL**  
open science

## Partially time invariant panel data regression

Hervé Cardot, Antonio Musolesi

► **To cite this version:**

Hervé Cardot, Antonio Musolesi. Partially time invariant panel data regression. 2024. hal-04149063v2

**HAL Id: hal-04149063**

**<https://hal.science/hal-04149063v2>**

Preprint submitted on 15 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Partially time invariant panel data regression

Hervé CARDOT

Institut de Mathématiques de Bourgogne, UMR CNRS 5584,

Université de Bourgogne

Antonio MUSOLESI

Department of Economics and Management (DEM),

University of Ferrara and SEEDS

February 15, 2024

## Abstract

When dealing with panel data, considering the variation over time of the variable of interest allows to get rid of potential individual effects. Even though the outcome variable has a continuous distribution, its variation over time can be equal to zero with a strictly positive probability and thus its distribution is a mixture of a mass at zero and a continuous distribution. We introduce a parametric statistical model based on conditional mixtures, build estimators for the parameters related to the conditional probability of no variation and to the conditional expectation related to the continuous part of the distribution and derive their asymptotic consistency and normality under a specific conditional independence assumption. Consistent confidence intervals are built via an empirical bootstrap approach. In the framework of policy evaluation, we study estimates of treatment effects based on difference-in-differences under the same zero inflation phenomenon and propose estimators of the average treatment effect that are proven to be consistent and asymptotically Gaussian. A small Monte Carlo simulation study assesses the good behavior of the estimators for finite samples and highlights that misspecified models that do not take account of the zero inflation may have a substantial bias. Empirical illustrations based on long time difference for the Mincer wage equation as well as the evaluation of European rural development policies based on the difference-in-differences approach confirm the interest of the proposed statistical modeling, bringing new insights on the size of the bias in commonly used regression models.

*JEL classification:* C21; C23; C25.

*Keywords:* Bootstrap; Heterogeneous Treatment Effects; Mixture of Distributions; Panel Data; Policy Evaluation, Zero Inflation;

# 1 Introduction

In econometric specifications, the dependent variable is often expressed in terms of variation over time. A relevant example includes commonly adopted unobserved effects panel data models, where the typical approach to estimating the parameters of interest consists of adopting a transformation, such as individual differencing over time or within transformation, to eliminate the unobserved component. This is followed by the application of ordinary least squares (OLS) (see, for example, Wooldridge, 2010).

A similar strategy is employed in program evaluation within a difference-in-differences (DID) framework. For identification purposes and to address the issue of selection on unobservables, it is commonly assumed that the conditional independence assumption holds for the difference in the outcome before and after the beginning of the policy. Subsequently, a before-after approach is adopted (Heckman and Hotz, 1989; Lechner, 2011, 2015).

Differencing over time is also employed in many time series models to achieve stationarity. Additionally, it is used in cross-sectional data models when the interest lies in directly modeling outcome variation over time. A typical example is when studying economic growth or employment dynamics as a function of some explanatory variables observed at a given point in time (Sala-i Martin, 1997).

However, while most economic variables such as employment, wages, production, investments, consumption, etc., typically assume non-negative values, modeling the individual deviations of the outcome variable over time has a crucial consequence: these deviations can take either positive or negative values. Importantly, especially at a micro-data level, for a non-negligible fraction of the statistical units under investigation, the variable of interest may not vary over time. Thus, we encounter a *partially time-invariant regression model*.

Relevant examples of such a phenomenon include nominal prices and wages, which exhibit what is generally called nominal rigidity—a resistance to change in response to fluctuations in supply and demand or changes in economic conditions. Another important example is found in employment, especially when focusing on disaggregated units such as municipalities. In these cases, the zero-inflated phenomenon addressed in the paper has been empirically documented (Kahn, 1997; Arulampalam et al., 2004; Álvarez et al., 2006; Cardot and Musolesi, 2020). However, to the best of our knowledge, it has never been fully considered in econometric panel data models.

With this scenario, common zero-inflated approaches, which are based on negative binomial or Poisson distributions and can only deal with non-negative count data, are not appropriate. The data generating process (DGP) under study is also different from the corner solution model,

which arises when the response variable has a continuous distribution over strictly positive values and there is a mass at zero with non-null probability.

This paper aims to provide a mathematical formalization of the zero-inflated empirical phenomenon and present new evidence based on both simulated and real data.

We first consider standard unobserved effects panel data models and propose a statistical parametric model for the long time difference based on a conditional mixture of a continuous linear regression model and a mass at zero. Given a set of covariates, estimators of the parameters modeling the conditional probability of occurrence of the zero variation phenomenon and the continuous linear part are obtained as the minimizers of a contrast function. We prove that under a specific conditional independence assumption the proposed estimators are consistent and asymptotically Gaussian. Additionally, we prove that empirical paired bootstrap approaches can be employed to obtain consistent approximations of the distribution of the unknown parameters and to build confidence intervals for prediction with a given asymptotic confidence level when the conditional probability of observing zero can be expressed as a probit or logit model. It is known, in case of equi-tailed confidence intervals, that bootstrap percentile confidence intervals are superior to the plug in Gaussian asymptotic approximation in the sense that they are second order accurate (see for example Davison and Hinkley, 1997, Chapter 5). Furthermore, we extend the theoretical framework by studying DID estimation under zero inflation and propose an estimator of the average treatment effect (ATE) that is proven to be consistent and asymptotically Gaussian.

A Simulated data example is studied to illustrate the effect of zero inflation on the expected value of the response variable and to check the ability of paired bootstrap procedures to produce reliable confidence intervals. We remark that the zero-inflated phenomenon can produce very different functional relations depending on the underlying parameters and that the linear model provides misleading results. In particular, when the underlying relation is non-monotonic it clearly provides a senseless fit. In contrast, the proposed estimator, which handles the zero-inflation, provides a very faithful description of the underlying DGP. Our simulation also offers evidence of the validity of the non-parametric bootstrap in the proposed zero-inflated framework, even in the case of small samples.

Finally, we illustrate the usefulness of our methodology with two real data examples, providing new insights into the size of the bias in commonly used regression models that assume the variation in the response variable over time has a continuous distribution. By using the data from Baltagi and Khanti-Akom (1990), we initially revisit a classical Mincer wage equation, where the response variable is nominal wage. Additionally, we address the challenge of estimating the ATE of two distinct public policies aimed at boosting local employment in France, based on data from Cardot and Musolesi (2020).

The paper is organized as follows. Section 2 introduces the zero-inflated model within an unobserved effects panel data framework and addresses the problem of estimation. Section 3 extends the previous results by considering DID estimation under zero inflation. Sections 4 and 5 provide a small simulation study and two illustrative examples, respectively. Finally, concluding remarks are given in Section 6 whereas proofs, additional details and information are gathered in an Appendix.

## 2 Partially time invariant panel data model

### 2.1 Model and assumptions

We introduce the following panel data model, allowing the value of the outcome to stay constant at two successive instants. We suppose that we have, for  $i = 1, \dots, n$ , a sample  $(Y_{i,0}, Y_{i,1}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})$  of  $n$  independent realizations of  $(Y_0, \dots, Y_T, \mathbf{x}_0, \dots, \mathbf{x}_T)$ . For each statistical unit  $i$ , we suppose that at time  $t = 0$ , that hereafter will be noted  $t_0$ ,

$$Y_{i,0} = \boldsymbol{\theta}^\top \mathbf{x}_{i,0} + c_i + \epsilon_{i,0} \quad (1)$$

and, at time  $t = 1, \dots, T$ ,

$$Y_{i,t} = \begin{cases} Y_{i,0} & \text{with probability } 1 - \pi_{i,t} \\ \boldsymbol{\theta}^\top \mathbf{x}_{i,t} + c_i + \epsilon_{i,t} & \text{with probability } \pi_{i,t} \end{cases} \quad (2)$$

where  $\epsilon_{i,0}, \dots, \epsilon_{i,T}$  are noise components, satisfying  $\mathbb{E}(\epsilon_{i,t} | \mathbf{x}_{i,t}) = 0$  and  $\mathbb{E}(\epsilon_{i,t}^2 | \mathbf{x}_{i,t}) = \sigma^2$  almost surely. Each individual effect  $c_i$  is supposed to be centered,  $\mathbb{E}(c_i) = 0$  but may be not independent of the regressors, that is to say  $\mathbb{E}(c_i | \mathbf{x}_{i,t}) \neq 0$  in general.

Model (2), which is central in this work, indicates that, at each instant, two regimes are possible. With probability  $\pi_{i,t}$ , there is a non null variation of the outcome between  $t$  and  $t_0 = 0$  which can be described by the values of some regressors and a noise component. In the second regime, which occurs with probability  $1 - \pi_{i,t}$ , there is no variation of the outcome  $Y$  between  $t$  and  $t_0$ . We introduce the sequence of Bernoulli variables  $Z_{i,t}$ , taking values in  $\{0, 1\}$ , and defined by  $Z_{i,t} = 0$  if  $Y_{i,t} = Y_{i,0}$  and  $Z_{i,t} = 1$  else, for  $t = 1, \dots, T$ . Taking the difference to eliminate the unobserved individual effect  $c_i$ , we get with (1) and (2),

$$Y_{i,t} - Y_{i,0} = Z_{i,t} \times \left[ \boldsymbol{\theta}^\top (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) + \epsilon_{i,t} - \epsilon_{i,0} \right] + (1 - Z_{i,t}) \times 0. \quad (3)$$

The distribution of  $\Delta Y_{i,t} = Y_{i,t} - Y_{i,0}$  is thus a mixture of a continuous distribution and a Dirac at zero.

We denote by

$$\Delta_c Y_{i,t} = \boldsymbol{\theta}^\top (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) + \epsilon_{i,t} - \epsilon_{i,0}, \quad (4)$$

the potential continuous variation of  $Y_i$  between  $t$  and  $t_0$ . We suppose furthermore that the probability of variation can be expressed, given  $\mathbf{x}_{i,t}$ , via a parametric model,

$$\pi_{i,t} = \pi(\mathbf{x}_{i,t}, \boldsymbol{\beta}_t). \quad (5)$$

for some known link function  $\pi(\cdot, \cdot)$  but unknown parameter  $\boldsymbol{\beta}_t$  which is allowed to vary with  $t$ . This includes logistic and probit regression. For example,  $\log(\pi_{i,t}/(1-\pi_{i,t})) = \boldsymbol{\beta}_t^\top \mathbf{x}_{i,t}$  corresponds to logistic regression and  $\pi_{i,t} = \Phi(\boldsymbol{\beta}_t^\top \mathbf{x}_{i,t})$  corresponds to probit regression when  $\Phi(w) = \mathbb{P}(W \leq w)$ ,  $W$  being a centered Gaussian random variable with unit variance. The parameters to be estimated are  $\boldsymbol{\beta}_t, t = 1, \dots, T$  and  $\boldsymbol{\theta}$ .

We assume that the following conditional independence assumptions hold for  $t = 1, \dots, T$ ,

$$(\mathbf{H}_{1,t}) \quad \Delta_c Y_t \perp\!\!\!\perp Z_t \mid \mathbf{x}_t, \mathbf{x}_0$$

Assumption  $(\mathbf{H}_{1,t})$  ensures that we have at hand a sufficient rich set of variables  $\mathbf{x}_t$  and  $\mathbf{x}_0$  such that  $\Delta_c Y_t$  and  $Z_t$  can be supposed to be conditionally independent. It is similar to assumption (17.38) in Wooldridge (2010) for the Hurdle model in which  $Y$  only takes positive values. Note that with (4), assumption  $(\mathbf{H}_{1,t})$ , can be rewritten

$$(\mathbf{H}_{1,t}) \quad \epsilon_t - \epsilon_0 \perp\!\!\!\perp Z_t \mid \mathbf{x}_t, \mathbf{x}_0 \quad (6)$$

We directly get, with (3), (5) and assumption  $(\mathbf{H}_{1,t})$  that

$$\mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] = \pi(\mathbf{x}_t, \boldsymbol{\beta}_t) \boldsymbol{\theta}^\top (\mathbf{x}_t - \mathbf{x}_0), \quad t = 1, \dots, T. \quad (7)$$

If, furthermore,  $\pi(\mathbf{x}, \boldsymbol{\beta})$  is differentiable with respect to  $\mathbf{x}$ ,

$$\frac{\partial \mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} = \pi(\mathbf{x}_t, \boldsymbol{\beta}_t) \boldsymbol{\theta} + \boldsymbol{\theta}^\top (\mathbf{x}_t - \mathbf{x}_0) \frac{\partial \pi(\mathbf{x}_t, \boldsymbol{\beta}_t)}{\partial \mathbf{x}_t}, \quad (8)$$

meaning that the sign and the amplitude of the effects of a variation of  $\mathbf{x}_t$  on  $Y_t - Y_0$  depend on  $\boldsymbol{\theta}$  but also on  $\frac{\partial \pi}{\partial \mathbf{x}_t}$ , the variation of the probability of observing no change in time.

**Remark 1.** Note that if we do not take account of the zero inflation phenomenon the best linear approximation, in the mean squared error sense, to the conditional expectation of  $\mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]$  given in (7), is equal to  $\tilde{\boldsymbol{\theta}}^\top (\mathbf{x}_t - \mathbf{x}_0)$ , with

$$\tilde{\boldsymbol{\theta}} = \left( \mathbb{E} \left[ (X_t - X_0)(X_t - X_0)^\top \right] \right)^{-1} \mathbb{E} [(X_t - X_0) \Delta Y_t] \quad (9)$$

where  $\Delta Y_t = Z_t (\boldsymbol{\theta}^\top (X_t - X_0) + \epsilon_t - \epsilon_0)$ . Thus, considering that the noise components are i.i.d, we have

$$\mathbb{E} [(X_t - X_0) \Delta Y_t] = \mathbb{E} \left[ Z_t (X_t - X_0)(X_t - X_0)^\top \right] \boldsymbol{\theta}$$

Unless  $Z_t = 1$  almost surely, meaning that there is no zero inflation phenomenon, we have that  $\tilde{\boldsymbol{\theta}}^\top(\mathbf{x}_t - \mathbf{x}_0) \neq \boldsymbol{\theta}^\top(\mathbf{x}_t - \mathbf{x}_0)$ . Note that in the particular case in which the random binary variable  $Z_t$  is independent of  $X_t$ , meaning that  $\pi(\mathbf{x}_t, \boldsymbol{\beta}) = \pi_t$ , we have  $\mathbb{E}[Z_t(X_t - X_0)(X_t - X_0)^\top] = \mathbb{E}[Z_t] \mathbb{E}[(X_t - X_0)(X_t - X_0)^\top]$  and  $\tilde{\boldsymbol{\theta}}^\top(\mathbf{x}_t - \mathbf{x}_0) = \pi_t \boldsymbol{\theta}^\top(\mathbf{x}_t - \mathbf{x}_0) = \mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]$ . In the general case in which  $Z_t$  does depend on the covariates, the estimation of conditional expectation given by  $\tilde{\boldsymbol{\theta}}^\top(\mathbf{x}_t - \mathbf{x}_0)$  will be biased.

**Remark 2.** Note that the within transform, which is widely used in panel data to remove the individual effects  $c_i$ ,  $i = 1, \dots, n$  would also provide estimators which are biased and not consistent. Indeed, we get with (3)

$$\bar{Y}_i = Y_{i,0} + \boldsymbol{\theta}^\top (\bar{Z}_i \bar{\mathbf{x}}_i - \bar{Z}_i \mathbf{x}_{i,0}) + \bar{Z}_i \bar{\epsilon}_i - \bar{Z}_i \epsilon_{i,0} \quad (10)$$

where  $\bar{Y}_i = (T+1)^{-1} \sum_{t=0}^T Y_{i,t}$ ,  $\bar{Z}_i = (T+1)^{-1} \sum_{t=0}^T Z_{i,t}$  and  $\bar{Z}_i \bar{\mathbf{x}}_i = (T+1)^{-1} \sum_{t=0}^T Z_{i,t} \mathbf{x}_{i,t}$  and  $\bar{Z}_i \bar{\epsilon}_i = (T+1)^{-1} \sum_{t=0}^T Z_{i,t} \epsilon_{i,t}$  and the convention  $Z_{i,0} = 1$ . Thus, the within transform can be expanded as follows,

$$\begin{aligned} Y_{i,t} - \bar{Y}_i &= Z_{i,t} \left[ \boldsymbol{\theta}^\top (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) + \epsilon_{i,t} - \epsilon_{i,0} \right] - \boldsymbol{\theta}^\top (\bar{Z}_i \bar{\mathbf{x}}_i - \bar{Z}_i \mathbf{x}_{i,0}) - (\bar{Z}_i \bar{\epsilon}_i - \bar{Z}_i \epsilon_{i,0}) \\ &= \boldsymbol{\theta}^\top [Z_{i,t} (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) - (\bar{Z}_i \bar{\mathbf{x}}_i - \bar{Z}_i \mathbf{x}_{i,0})] + Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) - (\bar{Z}_i \bar{\epsilon}_i - \bar{Z}_i \epsilon_{i,0}). \end{aligned} \quad (11)$$

We first remark that if  $Z_{i,t} = 1, t = 1, \dots, T$ , we get back to the classical Within transform regression equation. If  $Z_{i,t}$  is supposed to be independent of  $\mathbf{x}_i = (\mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})$ , with  $\mathbb{E}[Z_{i,t}] = \pi_t$ , then taking the expectation at both sides of (11), we get

$$\mathbb{E}[Y_{i,t} - \bar{Y}_i \mid \mathbf{x}_i] = \boldsymbol{\theta}^\top [\pi_t (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) - \bar{\pi} (\bar{\mathbf{x}}_i - \mathbf{x}_{i,0})], \quad (12)$$

with  $\bar{\pi} = (T+1)^{-1} \sum_{t=0}^T \pi_t$  and  $\bar{\mathbf{x}}_i = (T+1)^{-1} \sum_{t=0}^T \mathbf{x}_{i,t}$ . The first moment equality (12) confirms that, even when  $Z_{i,t}$  is independent of the covariates in the model, a linear regression model for  $Y_{i,t} - \bar{Y}_i$  based on  $\mathbf{x}_{i,t} - \bar{\mathbf{x}}_i$  will generally lead to biased estimators of  $\boldsymbol{\theta}$ .

## 2.2 Empirical example: Mincer wage equation

Nominal wage holds significant importance in economic theory (Friedman, 1971; Brenner, 1980) and has been extensively employed as a dependent variable in various research streams. Importantly, it is well-established that nominal wage is partially time-invariant. The phenomenon of wage stickiness, which refers to the fact that workers' earnings do not adjust quickly to changes in labor market conditions, is a notable economic characteristic. This aspect is not attributable to measurement errors or missing data but represents a salient feature in economic dynamics. The literature on this subject is extensive, covering both theoretical and empirical perspectives.

Haley (1990) conducted a survey of different theoretical explanations for sticky wages, including efficiency wage models (Stiglitz, 1984), insider-outsider theories (Lindbeck et al., 1989), implicit contracts, unions' behavior (Calmfors and Driffill, 1988), and social and psychological factors (Akerlof and Yellen, 1990). Moreover, empirical studies have already documented that the zero-inflated phenomenon addressed in the paper applies for nominal wages (Kahn, 1997; Arulampalam et al., 2004).

A significant area of research where the nominal wage typically serves as a response variable is the classical Mincer wage equation (Mincer, 1974). For illustrative purposes, we employ the dataset described in Baltagi and Khanti-Akom (1990), which consists of a panel of 595 individuals observed from 1976 to 1982 and is drawn from the Panel Study of Income Dynamics. In Baltagi and Khanti-Akom (1990) among others, the Mincer wage equation is fitted, with the logarithm of earnings in nominal terms, denoted as  $Y_{i,t} = \log(WAGE_{i,t})$ , which is modeled as the sum of a linear function of years of education ( $edu_{i,t}$ ) and a quadratic function of full-time work experience ( $exp_{i,t}$ ). The model is also extended by incorporating additional socio-economic variables  $\mathbf{x}^*_{i,t}$ , thus  $\mathbf{x}_{i,t} = \left[ edu_{i,t}, exp_{i,t}, exp_{i,t}^2, \mathbf{x}^{*\top}_{i,t} \right]^\top$ .

These data are consistent with the DGP that is described in the previous section. First note that the response variable in levels,  $\log(WAGE_{i,t})$ , can be supposed to be continuous. However, due to wage stickiness or nominal rigidities, the long-differenced variable  $\log(WAGE_{i,t}) - \log(WAGE_{i,0})$  can no longer be treated as a continuous variable. When looking for instance at the difference  $\log(WAGE_{i,1}) - \log(WAGE_{i,0})$  between time  $t = 1$  (corresponding to year 1977) and time  $t_0$  (year 1976), we observe that for around 18.5% of the observations the variation in time of the wages is equal to zero. The fraction of zeros varies between 18.5% for  $t = 1$  and 0% for  $t = T$ , and it is equals to 3.5% when considering all observations, for  $t = 1, \dots, T$  (detailed results are available upon request). As displayed in Figure 1, taking the difference over time induces a zero-inflated phenomenon that cannot be dealt properly by a standard continuous distribution model, while a mixture distribution combining a mass at zero and a continuous distribution, as in (3), seems to be more relevant.

===== Figure 1 =====

Second, consistently with (5), the probability  $\pi_{i,t}$  of observing a non-null variation in  $\log(WAGE_{i,t})$  between  $t$  and  $t_0$ , is significantly affected by (some of) the explanatory variables  $\mathbf{x}_{i,t}$ . This is consistent with the literature focusing on the key factors that contribute to wage rigidity, a subject extensively explored by seminal works (Blanchard and Katz, 1991; Blanchard and Gali, 1992; Pissarides, 2009; Druant et al., 2012). These determinants span a spectrum of dimensions: i)



industry-specific factors, encompassing labor demand, supply dynamics, and cyclical sensitivity; ii) occupational factors, with a particular focus on unionization; iii) employee characteristics, including education, experience, and skill levels; iv) firm-specific considerations, especially the size of the organization; and v) geographical factors.

Finally, as far assumption  $(\mathbf{H}_{1,t})$  is concerned, note that despite this assumption is not directly testable from data, it is a rather weak assumption that is likely to be fulfilled in many empirical applications. In the Mincer wage equation, while it seems rather unlikely that the probability of the event  $WAGE_{i,t} = WAGE_{i,0}$  does not depend on any characteristic of individual  $i$ , i.e.  $\Delta_c Y_t \perp\!\!\!\perp Z_t$ , assuming that exists some contemporaneous and lagged variables  $\mathbf{x}_t$  and  $\mathbf{x}_0$  such that  $\Delta_c Y_t$  and  $Z_t$  are conditionally independent is a much more credible situation, as lagged and contemporaneous levels of education and experience (among others) could explain  $Z_t$  and this could make  $Z_t$  conditionally independent to  $\Delta_c Y_t$ .

### 2.3 Definition of the estimators

We define, for  $i = 1, \dots, n$  and  $t = 1, \dots, T$ ,

$$\Delta Y_{i,t} = Y_{i,t} - Y_{i,0} \quad (13)$$

The estimation of  $\boldsymbol{\theta}$  and  $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T$  can be performed by minimizing the functional

$$\Psi_n(\boldsymbol{\theta}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T) = \sum_{t=1}^T \Psi_{1n,t}(\boldsymbol{\beta}_t) + \Psi_{2n}(\boldsymbol{\theta}), \quad (14)$$

with

$$\Psi_{1n,t}(\boldsymbol{\beta}_t) = -\frac{1}{n} \sum_{i=1}^n \left( Z_{i,t} \ln \left( \frac{\pi(\mathbf{x}_{i,t}, \boldsymbol{\beta}_t)}{1 - \pi(\mathbf{x}_{i,t}, \boldsymbol{\beta}_t)} \right) + \ln(1 - \pi(\mathbf{x}_{i,t}, \boldsymbol{\beta}_t)) \right) \quad (15)$$

and

$$\Psi_{2n}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \left( \Delta Y_{i,t} - \boldsymbol{\theta}^\top (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) \right)^2. \quad (16)$$

Note that  $\Psi_{1n,t}(\boldsymbol{\beta}_t)$  is simply the opposite of the likelihood criterion for  $\boldsymbol{\beta}_t$  and  $\Psi_{2n}(\boldsymbol{\theta})$  is a least squares criterion defined over the subsample of varying outcomes. We define the estimators  $\widehat{\boldsymbol{\theta}}$  and  $\widehat{\boldsymbol{\beta}}_t$ ,  $t = 1, \dots, T$  as follows

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \Psi_{2n}(\boldsymbol{\theta}) \quad (17)$$

$$\widehat{\boldsymbol{\beta}}_t = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \Psi_{1n,t}(\boldsymbol{\beta}_t) \quad (18)$$

Identification of parameter  $\boldsymbol{\theta}$  is ensured with the following assumption,

$$(\mathbf{H}_2) \quad \mathbf{Q}_\pi = \mathbb{E} \left[ \sum_{t=1}^T Z_t (\mathbf{x}_t - \mathbf{x}_0) (\mathbf{x}_t - \mathbf{x}_0)^\top \right] \text{ exists and is a full rank matrix.}$$

Assumption  $\mathbf{H}_2$  is a classical assumption required to get the identifiability of the regression parameter  $\boldsymbol{\theta}$ , in the specific subpopulation in which the variation in time of  $Y$  is not equal to zero. Assumption  $(\mathbf{H}_2)$  is similar to assumption FD.2 in Wooldridge (2010) (Chapter 10) but also takes into account the zero-inflation phenomenon.

When this assumption is fulfilled, we have that for large  $n$ , the estimator of parameter  $\boldsymbol{\theta}$  is uniquely defined as follows,

$$\widehat{\boldsymbol{\theta}} = \left( \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top \right)^{-1} \left( \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \Delta Y_{i,t} \Delta \mathbf{x}_{i,t} \right), \quad (19)$$

where  $\Delta \mathbf{x}_{i,t} = \mathbf{x}_{i,t} - \mathbf{x}_{i,0}$ .

Then, using (7), estimates of the expected variation of the outcome can be derived as follows,

$$\widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] = \pi(\mathbf{x}_t, \widehat{\boldsymbol{\beta}}_t) \widehat{\boldsymbol{\theta}}^\top (\mathbf{x}_t - \mathbf{x}_0), \quad t = 1, \dots, T. \quad (20)$$

If, furthermore,  $\pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}_t)$  is differentiable with respect to  $\mathbf{x}$ , we can define with (8), an estimate of the effect of a variation of  $\mathbf{x}_t$  on the variation of the outcome,

$$\frac{\partial \widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} = \pi(\mathbf{x}_t, \widehat{\boldsymbol{\beta}}_t) \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\theta}}^\top (\mathbf{x}_t - \mathbf{x}_0) \frac{\partial \pi(\mathbf{x}_t, \widehat{\boldsymbol{\beta}}_t)}{\partial \mathbf{x}_t}. \quad (21)$$

## 2.4 Some asymptotic properties

Our notations are borrowed from van der Vaart (1998), and we denote by  $U_n = o_p(1)$  the fact that the sequence  $(U_n)_{n \geq 1}$  of random variables (vectors or matrices) converges to zero in probability when  $n$  tends to infinity, whereas the convergence in distribution of the sequence towards a Gaussian random vector with expectation  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Gamma}$  is denoted by  $U_n \rightsquigarrow \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ .

It can be proven under hypotheses  $(\mathbf{H}_{1,t})$  and  $(\mathbf{H}_2)$  that  $\widehat{\boldsymbol{\theta}}$  is a consistent estimator of  $\boldsymbol{\theta}$  that is asymptotically Gaussian as  $n$  tends to infinity, as shown in the following proposition.

**Proposition 2.1.** *Suppose that models (1) and (2) hold and assume that hypotheses  $(\mathbf{H}_{1,t}), t = 1, \dots, T$  and  $(\mathbf{H}_2)$  are fulfilled. Then as  $n$  tends to infinity,*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = o_p(1)$$

and

$$\sqrt{n} \left( \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \rightsquigarrow \mathcal{N} \left( 0, \mathbf{Q}_\pi^{-1} \mathbf{Q}_{Z,\epsilon} \mathbf{Q}_\pi^{-1} \right),$$

where  $\mathbf{Q}_{Z,\epsilon}$  is the covariance matrix of  $\sum_{t=1}^T Z_t (\epsilon_t - \epsilon_0) \Delta \mathbf{x}_t$ .

**Remark 3.** *If we suppose furthermore that the increments of the residuals  $(\epsilon_{i,t} - \epsilon_{i,0})$  are independent of  $Z_t$  and  $\mathbf{x}_t$ , and are i.i.d, with common variance  $\sigma^2$ , then the covariance matrix  $\mathbf{Q}_{Z,\epsilon}$  satisfies  $\mathbf{Q}_{Z,\epsilon} = \sigma^2 \mathbf{Q}_\pi$ , and under the the assumptions of Proposition 2.1,*

$$\sqrt{n} \left( \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \rightsquigarrow \mathcal{N} \left( 0, \sigma^2 \mathbf{Q}_\pi^{-1} \right).$$

If  $\pi(\mathbf{x}, \boldsymbol{\beta})$  is of a logit or probit shape and if the set of assumptions

$$(\mathbf{H}_{3,t}) \quad \mathbb{E} \left[ \mathbf{x}_t \mathbf{x}_t^\top \right] \quad \text{is a full rank matrix}$$

hold for  $t = 1, \dots, T$ , the parameters  $\boldsymbol{\beta}_t$  can be estimated efficiently with maximum likelihood approaches (see Newey and McFadden (1994) for probit regression and Hjort and Pollard (2011) for logistic regression) and that maximum likelihood estimators  $\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_T$  are consistent and asymptotically Gaussian as  $n$  tends to infinity, with limiting covariance matrix denoted by  $\boldsymbol{\Gamma}_\beta$ . Note that there is no need to impose that  $\boldsymbol{\beta}$  belongs to some compact space, thanks to the concavity in the parameters of the log likelihood for probit (see Newey and McFadden, 1994) and logistic (see Hjort and Pollard, 2011) regression models.

**Proposition 2.2.** *Suppose that models (1) and (2) hold and assume that hypotheses  $(\mathbf{H}_{1,t}), t = 1, \dots, T$  and  $(\mathbf{H}_2)$  and  $(\mathbf{H}_{3,t}), t = 1, \dots, T$  are fulfilled. Suppose also that  $\pi(\boldsymbol{\beta}_t, \cdot)$  is a logit or probit link function for  $t = 1, \dots, T$ . Then as  $n$  tends to infinity,*

$$\sqrt{n} \left( \begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\beta} \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{Q}_\pi^{-1} \mathbf{Q}_{Z,\epsilon} \mathbf{Q}_\pi^{-1} & 0 \\ 0 & \boldsymbol{\Gamma}_\beta \end{pmatrix} \right),$$

where  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T)$  and  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_T)$ .

## 2.5 Bootstrap confidence intervals

We are now interested in computing confidence intervals for the expected effect  $\mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0] = \pi(\mathbf{x}_t, \boldsymbol{\beta}_t) \boldsymbol{\theta}^\top (\mathbf{x}_t - \mathbf{x}_0)$  as well as the partial effects  $\frac{\partial \mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t}$ . Note first that under previous hypotheses, we directly get with the help of the continuous mapping theorem (see van der Vaart (1998), Theorem 2.3), that, given  $\mathbf{x}_t$  and  $\mathbf{x}_0$ ,  $\widehat{\mathbb{E}}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]$  defined in (20) converges in probability to  $\mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]$  and  $\frac{\partial \widehat{\mathbb{E}}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t}$  defined in (21) converges in probability to  $\frac{\partial \mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t}$  as  $n$  tends to infinity.

With Proposition 2.2 and the use of the Delta method, it is possible to build confidence intervals for  $\mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]$  and the partial effects  $\frac{\partial \mathbb{E}[Y_t - Y_0 | \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t}$ . However, this approach relies on the estimation of the asymptotic variance which has a complicated expression and is not so simple to implement in statistical softwares.

The percentile method based on paired bootstrap which is simple to implement and reasonably time-consuming in our parametric framework, may be preferred for at least two reasons. On the one hand, it is theoretically superior to the normal approximation limits for equi-tailed confidence intervals in the sense that it is second order accurate whereas confidence intervals based on asymptotic normal approximations are only first order accurate (see Chapter 5 in Davison and

Hinkley, 1997). On the other hand it does not require to compute estimates of the asymptotic variance.

Recall that a sample is made of  $(Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})$ ,  $i = 1, \dots, n$ , which are supposed to be  $n$  independent realizations of  $(Y_0, \dots, Y_T, \mathbf{x}_0, \dots, \mathbf{x}_T)$ . Consider a paired bootstrap sample  $(Y_{i,0}^*, \dots, Y_{i,T}^*, \mathbf{x}_{i,0}^*, \dots, \mathbf{x}_{i,T}^*)$ ,  $i = 1, \dots, n$  drawn independently and with equal probability from the empirical distribution of  $(Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})$ ,  $i = 1, \dots, n$ . We denote by  $(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\beta}}^*)$  the bootstrap estimate of  $(\boldsymbol{\theta}, \boldsymbol{\beta})$  defined as the minimizer of  $\Psi_n(\boldsymbol{\theta}, \boldsymbol{\beta})$  evaluated over the bootstrap sample. For  $\mathbf{u}_1 \in \mathbb{R}^p$  and  $\mathbf{u}_2 \in \mathbb{R}^{Tp}$ , we denote by  $F_{n,B}(\mathbf{u}_1, \mathbf{u}_2)$  the conditional joint cumulative distribution function of the bootstrap estimator, given the data:

$$F_{n,B}(\mathbf{u}_1, \mathbf{u}_2) = \mathbb{P} \left[ \sqrt{n} \left( (\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\beta}}^*) - (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) \right) \leq (\mathbf{u}_1, \mathbf{u}_2) \mid (Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})_{i=1}^n \right]$$

where the inequality should be understood component-wise.

We denote by  $F_n(\mathbf{u}_1, \mathbf{u}_2) = \mathbb{P} \left[ \sqrt{n} \left( (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) - (\boldsymbol{\theta}, \boldsymbol{\beta}) \right) \leq (\mathbf{u}_1, \mathbf{u}_2) \right]$  the joint cumulative distribution function corresponding to the multivariate Gaussian distribution given in Proposition 2.2. We can state the following proposition, which ensures that the bootstrap procedures provide consistent approximation to the distribution of  $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$  and can be useful to build consistent confidence intervals for  $\mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]$  and  $\frac{\partial \mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t}$  with a given asymptotic confidence level.

**Proposition 2.3.** *Suppose that models (1) and (2) hold and assume that hypotheses  $(\mathbf{H}_{1,t})$ ,  $t = 1, \dots, T$  and  $(\mathbf{H}_2)$  and  $(\mathbf{H}_{3,t})$ ,  $t = 1, \dots, T$  are fulfilled. Suppose also that  $\pi(\boldsymbol{\beta}_t, \cdot)$  is a logit or probit link function for  $t = 1, \dots, T$ . Then as  $n$  tends to infinity,*

$$\sup_{\mathbf{u}_1 \in \mathbb{R}^p, \mathbf{u}_2 \in \mathbb{R}^{Tp}} |F_{n,B}(\mathbf{u}_1, \mathbf{u}_2) - F_n(\mathbf{u}_1, \mathbf{u}_2)| = o_p(1).$$

Given  $\mathbf{x}_t$  and  $\mathbf{x}_0$ , we also have

$$\begin{aligned} & \sup_{u \in \mathbb{R}} \left| \mathbb{P} \left[ \sqrt{n} \left( \widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] - \mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] \right) \leq u \right] \right. \\ & \left. - \mathbb{P} \left[ \sqrt{n} \left( \widehat{\mathbb{E}}^*[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] - \widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] \right) \leq u \mid (Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})_{i=1}^n \right] \right| = o_p(1). \end{aligned}$$

and

$$\begin{aligned} & \sup_{u \in \mathbb{R}^p} \left| \mathbb{P} \left[ \sqrt{n} \left( \frac{\partial \widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} - \frac{\partial \mathbb{E}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} \right) \leq u \right] \right. \\ & \left. - \mathbb{P} \left[ \sqrt{n} \left( \frac{\partial \widehat{\mathbb{E}}^*[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} - \frac{\partial \widehat{\mathbb{E}}[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} \right) \leq u \mid (Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})_{i=1}^n \right] \right| = o_p(1). \end{aligned}$$

where  $\widehat{\mathbb{E}}^*[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0] = \pi(\mathbf{x}_t, \hat{\boldsymbol{\beta}}_t^*) \hat{\boldsymbol{\theta}}^{*\top} (\mathbf{x}_t - \mathbf{x}_0)$  and

$$\frac{\partial \widehat{\mathbb{E}}^*[Y_t - Y_0 \mid \mathbf{x}_t, \mathbf{x}_0]}{\partial \mathbf{x}_t} = \pi(\mathbf{x}_t, \hat{\boldsymbol{\beta}}_t^*) \hat{\boldsymbol{\theta}}^* + \hat{\boldsymbol{\theta}}^{*\top} (\mathbf{x}_t - \mathbf{x}_0) \frac{\partial \pi(\mathbf{x}_t, \hat{\boldsymbol{\beta}}_t^*)}{\partial \mathbf{x}_t}.$$

The first part of this Proposition is a consequence of Theorem 2.4 in Bose and Chatterjee (2003), which heavily relies on the fact that the estimators are defined as the minimizers of a convex objective function  $\Psi_n(\boldsymbol{\theta}, \boldsymbol{\beta})$  that is a twice differentiable in  $\boldsymbol{\theta}$  and  $\boldsymbol{\beta}$ . We could consider other sufficiently smooth link functions to model the Bernoulli variable  $Z$  and Proposition 2.3 would remain true provided that the criterion  $\Psi_{1n}(\cdot)$  is a twice differentiable convex function of  $\boldsymbol{\beta}$ . The second part of Proposition 2.3 is based on an application of the Delta method for bootstrapped estimates (see Theorem 23.9 in van der Vaart, 1998).

A direct consequence of Proposition 2.3 is that percentile bootstrap confidence intervals for  $\pi(\mathbf{x}_t, \boldsymbol{\beta}_t)\boldsymbol{\theta}^\top(\mathbf{x}_t - \mathbf{x}_0)$  and  $\frac{\partial}{\partial \mathbf{x}_t}(\pi(\mathbf{x}_t, \boldsymbol{\beta}_t)\boldsymbol{\theta}^\top(\mathbf{x}_t - \mathbf{x}_0))$  are consistent, with second order accuracy when considering equi-tail intervals (for instance with confidence 90%, based on the 5% and 95% quantiles).

### 3 Program evaluation with difference-in-differences

In the last decades, there has been a huge amount of literature on DID estimation and on its relation with standard unobserved effects panel data models (Heckman and Hotz, 1989; Wooldridge, 2005; Abadie, 2005; Lechner, 2015; Lee and Kang, 2006; Heckman et al., 1997, 1998). Recent works have provided further insights. Some of them have investigated the assumptions that are needed to yield estimated coefficients having a causal interpretation and, in particular, have considered various settings such as allowing for heterogeneous treatment effects, variation in treatment timing, and dynamic treatment effects (De Chaisemartin and d’Haultfoeuille, 2020; Goodman-Bacon, 2021; Han, 2021; Sun and Abraham, 2021).

We explore another direction and show that the presence of a zero inflation phenomenon when considering the difference-in-differences approach gives rise to heterogeneous nonlinear treatment effects.

#### 3.1 Model, assumptions and estimand

We suppose now that we aim at evaluating a treatment effect, among  $R - 1$  possible exclusive treatments, set up at time  $t_\tau$ , on an outcome  $Y_t$  at time  $t > t_\tau$ . The value of  $Y_t$  is made at discrete instants in time,  $t = 0, \dots, t_\tau, \dots, T$ . We suppose that at time  $t_0$ , we have for  $i = 1, \dots, n$ ,

$$Y_{i,0} = \boldsymbol{\theta}_{0,0}^\top \mathbf{x}_i + c_i + \epsilon_{i,0}$$

where  $\boldsymbol{\theta}_{0,0}$  is an unknown vector of regression coefficients,  $c_i$  is an unobserved individual effect and  $\epsilon_{i,0}$  is a noise component satisfying  $\mathbb{E}(\epsilon_{i,0}|\mathbf{x}_i) = 0$  and  $\mathbb{E}(\epsilon_{i,0}^2|\mathbf{x}_i) = \sigma_0^2$  almost surely. We denote by  $D_i^r$ , for  $r \in \{0, 1, \dots, R - 1\}$ , the binary treatment indicator variable that takes value

1 if treatment  $r$  has been applied to statistical unit  $i$  and 0 otherwise, with the convention that  $r = 0$  corresponds to no treatment. The  $R - 1$  possible treatments are supposed to be mutually exclusive, so that by definition  $\sum_{r=0}^{R-1} D_i^r = 1$ . We suppose that the potential outcome at time  $t \geq t_\tau$ , under treatment  $r$ , can be expressed as follows

$$Y_{i,t}^r = \begin{cases} Y_{i,0} & \text{with probability } 1 - \pi_{i,t}^r \\ \boldsymbol{\theta}_{r,t}^\top \mathbf{x}_i + c_i + \epsilon_{i,t}^r & \text{with probability } \pi_{i,t}^r \end{cases} \quad (22)$$

Note that it is only possible to observe one value of  $Y_{i,t}^r$ , which is equal to  $Y_{i,t} = \sum_{r=0}^{R-1} D_i^r Y_{i,t}^r$ , among the  $R$  potential outcomes  $Y_{i,t}^0, \dots, Y_{i,t}^{R-1}$ . The potential outcome difference between time  $t = 0$  and time  $t \geq t_\tau$ , under treatment  $r$ , is thus equal to

$$Y_{i,t}^r - Y_{i,0} = Z_{i,t}^r \left( (\boldsymbol{\theta}_{r,t} - \boldsymbol{\theta}_{0,0})^\top \mathbf{x}_i + \epsilon_{i,t}^r - \epsilon_{i,0} \right) + (1 - Z_{i,t}^r) 0, \quad (23)$$

where  $Z_{i,t}^r$  is a (counterfactual) binary variable indicating which regime governs the evolution of the outcome between  $t \geq t_\tau$  and time 0, with  $Z_{i,t}^r = 0$  if there is no variation of the outcome and  $Z_{i,t}^r = 1$  otherwise.

Our aim is to estimate, given  $\mathbf{x}_i$ , the average treatment effect at time  $t$  under treatment  $r$  compared to no treatment,

$$\text{ATE}^r(t, \mathbf{x}) = \mathbb{E} (Y_t^r - Y_t^0 \mid \mathbf{x}). \quad (24)$$

We assume that, for  $r = 0, \dots, R - 1$  and  $t \geq t_\tau$ , the set of confounding variables  $\mathbf{x}$  ensures that

$$(\mathbf{H}_{4,t}^r) \quad \epsilon_t^r - \epsilon_0 \perp\!\!\!\perp Z_t^r \mid \mathbf{x}.$$

Assumption  $(\mathbf{H}_{4,t}^r)$  is similar to assumption  $(\mathbf{H}_{1,t})$  discussed in Section 2.

We also assume that, for some known parametric model and unknown parameter  $\boldsymbol{\beta}_{r,t}$ , we have

$$\mathbb{P} [Z_t^r = 1 \mid \mathbf{x}_i] = \pi(\mathbf{x}_i, \boldsymbol{\beta}_{r,t}). \quad (25)$$

Then, hypotheses  $(\mathbf{H}_{4,t}^r)$  and  $(\mathbf{H}_{4,t}^0)$  allow to get the following decomposition for the conditional average treatment effect, for treatment  $r \neq 0$ .

**Proposition 3.1.** *If assumption  $(\mathbf{H}_{4,t}^r)$  and  $(\mathbf{H}_{4,t}^0)$  are in force for  $r \neq 0$ , and models (23) and (25) are true, then*

$$\text{ATE}^r(t, \mathbf{x}) = \left( \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t}) \tilde{\boldsymbol{\theta}}_{r,t} - \pi(\mathbf{x}, \boldsymbol{\beta}_{0,t}) \tilde{\boldsymbol{\theta}}_{0,t} \right)^\top \mathbf{x}.$$

where  $\tilde{\boldsymbol{\theta}}_{r,t} = \boldsymbol{\theta}_{r,t} - \boldsymbol{\theta}_{0,0}$ .

Note that if there is no zero-inflation phenomenon, we get the classical result,  $\text{ATE}^r(t, \mathbf{x}) = (\boldsymbol{\theta}_{r,t} - \boldsymbol{\theta}_{0,t})^\top \mathbf{x}$ . The proof of Proposition 3.1 is direct and thus omitted.

We also deduce directly from Proposition 3.1 that the marginal effect of a variation of  $\mathbf{x}$  is equal to

$$\frac{\partial \text{ATE}^r(t, \mathbf{x})}{\partial \mathbf{x}} = \left( \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t}) \tilde{\boldsymbol{\theta}}_{r,t} - \pi(\mathbf{x}, \boldsymbol{\beta}_{0,t}) \tilde{\boldsymbol{\theta}}_{0,t} \right) + \frac{\partial \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t})}{\partial \mathbf{x}} \tilde{\boldsymbol{\theta}}_{r,t}^\top \mathbf{x} - \frac{\partial \pi(\mathbf{x}, \boldsymbol{\beta}_{0,t})}{\partial \mathbf{x}} \tilde{\boldsymbol{\theta}}_{0,t}^\top \mathbf{x}. \quad (26)$$

The presence of the terms  $\pi(\mathbf{x}, \boldsymbol{\beta}_{r,t})$  and  $\pi(\mathbf{x}, \boldsymbol{\beta}_{0,t})$  in the expression of  $\text{ATE}^r(t, \mathbf{x})$  induces a non linear effect of the set of covariates  $\mathbf{x}_t$  on the average effect of treatment  $r$ .

### 3.2 Empirical example: local employment evolution and rural policies

We exploit the French data set used by Cardot and Musolesi (2020), which covered 25,593 municipalities over the period 1993-2002. Employment variation over time was modeled as a function of local development policies and of some confounding (pre-treatment) covariates that are indicated as relevant by the related literature on local employment growth, such as demographics, education, work qualifications, land use and the initial level of employment.

Figure 2 depicts the estimated distribution of the variation of employment in time  $\text{EMP}_{i,1} - \text{EMP}_{i,0}$ ,  $\text{EMP}_{i,t}$  being the employment level in French municipalities with time  $t = 1$  corresponding to year 1994 and time  $t_0$  to year 1993.

===== Figure 2 =====

Also in that case, the distribution of the variation in time of the dependent variable can be approximated by a mixture of a mass at 0 (0 representing more than 25 % of the municipalities) and a continuous density function, whose support is defined over both positive and negative values.

Moreover, consistently with (25), the probability of observing a non-null variation of  $\text{EMP}_{i,t}$  overtime is significantly affected by (some of) the explanatory variables  $\mathbf{x}$  (detailed results are available upon request). Since  $Z_{i,t}^r$  is largely explained by the size of the municipality, and by other socio-economic characteristics, so that introducing these variables in the regression function could make  $\epsilon_t^r - \epsilon_0$  conditionally independent to  $Z_t^r$ , as stated in the identification condition ( $\mathbf{H}_{4,t}$ ).

### 3.3 Estimation of the conditional average treatment effect

Suppose we have a sample  $(Y_{i,t}, Y_{i,0}, D_i^0, \dots, D_i^{R-1}, \mathbf{x}_i)$  for  $i = 1, \dots, n$ . The observed value of the outcome  $Y_{i,t}$  can be written as follows

$$Y_{i,t} = \sum_{r=0}^{R-1} D_i^r Y_{i,t}^r. \quad (27)$$

We define the binary variable  $Z_{i,t}$  as  $Z_{i,t} = 1$  if  $(Y_{i,t} - Y_{i,0}) \neq 0$  and  $Z_{i,t} = 0$  otherwise.

For treatment  $r$  and time  $t$ , the vectors of parameters  $\boldsymbol{\beta}_{r,t}$  and  $\tilde{\boldsymbol{\theta}}_{r,t} = \boldsymbol{\theta}_{r,t} - \boldsymbol{\theta}_{0,0}$  can be estimated by minimizing the function  $\Psi_{n,t}^r(\boldsymbol{\theta}, \boldsymbol{\beta}) = \Psi_{1,n,t}^r(\boldsymbol{\beta}) + \Psi_{2,n,t}^r(\boldsymbol{\theta})$ , where

$$\Psi_{1,n,t}^r(\boldsymbol{\beta}) = -\frac{1}{n} \sum_{i=1}^n D_i^r \left[ Z_{i,t} \ln \left( \frac{\pi(\mathbf{x}_i, \boldsymbol{\beta})}{1 - \pi(\mathbf{x}_i, \boldsymbol{\beta})} \right) + \ln(1 - \pi(\mathbf{x}_i, \boldsymbol{\beta})) \right] \quad (28)$$

is the opposite of the log likelihood and where, as in Section 2.5, the conditional probability  $\pi(\boldsymbol{\beta}_{r,t}, \mathbf{x}) = \mathbb{P}[Z_t^r = 1 | \mathbf{x}]$  is supposed to be of a probit or logit shape.

Function  $\Psi_{2,n,t}^r(\boldsymbol{\theta})$  to be minimized is a least squares criterion

$$\Psi_{2,n,t}^r(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n Z_{i,t} D_i^r \left( (Y_{i,t} - Y_{i,0}) - \mathbf{x}_i^\top \boldsymbol{\theta} \right)^2. \quad (29)$$

Assuming that  $\sum_{i=1}^n D_i^r Z_{i,t} \mathbf{x}_i \mathbf{x}_i^\top$  is a full rank matrix (which is true with high probability, as seen in Section 3.4 under hypothesis  $(\mathbf{H}_{6,t}^r)$ ), function  $\Psi_{2,n,t}^r$  has a unique minimizer,

$$\widehat{\boldsymbol{\theta}}_{r,t} = \left( \sum_{i=1}^n D_i^r Z_{i,t} \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \left( \sum_{i=1}^n D_i^r Z_{i,t} \mathbf{x}_i (Y_{i,t} - Y_{i,0}) \right). \quad (30)$$

Replacing the unknown parameters in the expression of  $\text{ATE}^r(t, \mathbf{x})$  given in Proposition 3.1 by their estimators, we get the estimate

$$\widehat{\text{ATE}}^r(t, \mathbf{x}) = \left( \pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}_{r,t}) \widehat{\boldsymbol{\theta}}_{r,t} - \pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}_{0,t}) \widehat{\boldsymbol{\theta}}_{0,t} \right)^\top \mathbf{x} \quad (31)$$

for the conditional average treatment effect at time  $t$  for treatment  $r$ .

### 3.4 Consistency of the estimated conditional treatment effect

We assume in the following that, for  $t \geq t_\tau$ ,

$$(\mathbf{H}_{5,t}) \quad \left( Y_t^0 - Y_0, \dots, Y_t^{R-1} - Y_0 \right) \perp\!\!\!\perp (D^0, \dots, D^{R-1}) \mid \mathbf{x}.$$

Condition  $(\mathbf{H}_{5,t})$  is a classical conditional independence assumption in the econometric literature on policy evaluation and multiple treatment effects. With (23) and  $(\mathbf{H}_{4,t}^r)$ , it implies that

$$(\epsilon_{0,t} - \epsilon_0, \dots, \epsilon_{R-1,t} - \epsilon_0) \perp\!\!\!\perp (D^0, \dots, D^{R-1}) \mid \mathbf{x} \quad (32)$$

and

$$(Z_t^0, \dots, Z_t^{R-1}) \perp\!\!\!\perp (D^0, \dots, D^{R-1}) \mid \mathbf{x}. \quad (33)$$

We denote by  $\pi_t^r(\mathbf{x}) = \mathbb{P}[D^r Z_t^r = 1 | \mathbf{x}]$  the probability of receiving treatment  $r$  and that  $Y_t^r - Y_0$  is different from zero, given  $\mathbf{x}$ . Note that if  $D_i^r = 1$ , we only observe  $Z_{i,t}^r$  and  $Y_{i,t}^r - Y_{i,0}$  for the unit  $i$  in the sample. Hypothesis  $(\mathbf{H}_{5,t})$  implies that the distribution of  $Z_t^r$  given  $\mathbf{x}$  is independent



on  $D^j$ , for  $j \in \{0, \dots, R-1\}$ . Note that when  $(\mathbf{H}_{5,t})$  holds and model (23) is true, we have  $\pi_t^r(\mathbf{x}) = \pi(\mathbf{x}, \boldsymbol{\beta}_{rt})\mathbb{P}[D^r = 1|\mathbf{x}]$ .

We also assume that the following assumption holds,

$$(\mathbf{H}_{6,t}) \quad \mathbb{E}\left(\pi_t^r(\mathbf{x})\mathbf{x}\mathbf{x}^\top\right) = \mathbf{Q}_t^r \text{ where } \mathbf{Q}_t^r \text{ is a non-singular matrix, } r = 0, \dots, R-1.$$

Condition  $(\mathbf{H}_{6,t})$  is fulfilled under the classical assumption that  $\mathbb{E}(\mathbf{x}\mathbf{x}^\top)$  is a full rank matrix and  $\pi_t^r(\mathbf{x}) > 0$  almost surely. This identifiability condition ensures the existence of a unique estimator  $\widehat{\boldsymbol{\theta}}_{r,t}$  when the sample size  $n$  is large enough. It also implies that  $\mathbb{P}[D^r = 1] > 0$ , for  $r = 0, 1, \dots, R-1$ .

We can now state the consistency and asymptotic normality of the estimators of the parameters defined in models (23) and (25).

**Proposition 3.2.** *Suppose that model (22) holds. Assume also that hypotheses  $(\mathbf{H}_{4,t})$ ,  $(\mathbf{H}_{5,t})$ , and  $(\mathbf{H}_{6,t})$  are fulfilled. Then as  $n$  tends to infinity,*

$$\widehat{\boldsymbol{\theta}}_{r,t} - \widetilde{\boldsymbol{\theta}}_{r,t} = o_p(1).$$

If  $\mathbb{P}[Z_t^r = 1|\mathbf{x}] = \pi(\mathbf{x}, \boldsymbol{\beta}_{rt})$  is of a probit or logit shape, we have, as  $n$  tends to infinity,

$$\widehat{\boldsymbol{\beta}}_{r,t} - \boldsymbol{\beta}_{r,t} = o_p(1).$$

and

$$\sqrt{n} \left( \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{0,t} \\ \widehat{\boldsymbol{\theta}}_{r,t} \\ \widehat{\boldsymbol{\beta}}_{0,t} \\ \widehat{\boldsymbol{\beta}}_{r,t} \end{pmatrix} - \begin{pmatrix} \widetilde{\boldsymbol{\theta}}_{0,t} \\ \widetilde{\boldsymbol{\theta}}_{r,t} \\ \boldsymbol{\beta}_{0,t} \\ \boldsymbol{\beta}_{r,t} \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left( 0, \begin{pmatrix} \boldsymbol{\Gamma}_{\theta,t}^r & 0 \\ 0 & \boldsymbol{\Gamma}_{\beta,t}^r \end{pmatrix} \right),$$

where  $\boldsymbol{\Gamma}_{\theta,t}^r$  is the block diagonal asymptotic covariance matrix of  $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{0,t} - \widetilde{\boldsymbol{\theta}}_{0,t}, \widehat{\boldsymbol{\theta}}_{r,t} - \widetilde{\boldsymbol{\theta}}_{r,t})$  and  $\boldsymbol{\Gamma}_{\beta,t}^r$  is the asymptotic covariance matrix of  $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{0,t} - \boldsymbol{\beta}_{0,t}, \widehat{\boldsymbol{\beta}}_{r,t} - \boldsymbol{\beta}_{r,t})$ .

We deduce the following corollary from previous Proposition.

**Corollary 3.3.** *Under the assumptions of Proposition 3.2 as  $n$  tends to infinity and for all  $\mathbf{x} \in \mathbb{R}^p$  we have*

$$\widehat{ATE}^r(t, \mathbf{x}) - ATE^r(t, \mathbf{x}) = o_p(1).$$

Furthermore, if  $(\boldsymbol{\beta}_{r,t}, \widetilde{\boldsymbol{\theta}}_{r,t}) \neq (\boldsymbol{\beta}_{0,t}, \widetilde{\boldsymbol{\theta}}_{0,t})$ ,

$$\sqrt{n} \left( \widehat{ATE}^r(t, \mathbf{x}) - ATE^r(t, \mathbf{x}) \right) \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Delta}_t^r(\mathbf{x}))$$

for some covariance matrix  $\boldsymbol{\Delta}_t^r(\mathbf{x})$ .

The proof of Corollary 3.3 is a direct consequence of the continuous mapping theorem and the Delta method. It is thus omitted.

The expression for the asymptotic variance  $\mathbf{\Delta}_t^r(\mathbf{x})$  of  $\widehat{\text{ATE}}^r(t, \mathbf{x})$  can be derived with the delta method. It is complicated and not given here. As in Section 2.5, paired bootstrap approaches are not difficult to employ and give reliable (and consistent) confidence intervals since the estimators are obtained as minimizers of  $\Psi_{n,t}^r(\boldsymbol{\beta}, \boldsymbol{\theta})$ , which is a twice differentiable convex functional.

## 4 A simulation study

To illustrate with a very simple example the effect of zero inflation on the expected value of the response variable and to check the ability of paired bootstrap procedures to produce reliable confidence intervals, we consider the following toy model. A time  $t_0$ , we suppose that the outcome variable satisfies, for  $i = 1, \dots, n$ ,

$$Y_{i,0} = \theta_0 + \theta x_{i,0} + c_i + \epsilon_{i,0},$$

whereas at time  $t_1 > t_0$ , we observe

$$Y_{i,1} = \begin{cases} Y_{i,0} & \text{with probability } 1 - \pi(x_{i,1} - x_{i,0}, \beta_0, \beta) \\ \theta_1 + \theta x_{i,1} + c_i + \epsilon_{i,1} & \text{with probability } \pi(x_{i,1} - x_{i,0}, \beta_0, \beta) \end{cases}$$

We thus have that

$$\mathbb{E}[Y_{i,1} - Y_{i,0} | x_{i,0}, x_{i,1}] = \pi(x_{i,1} - x_{i,0}, \beta_0, \beta) \times (\theta_1 - \theta_0 + \theta(x_{i,1} - x_{i,0})) \quad (34)$$

We generate artificial data as follows. The variation  $x_{i,1} - x_{i,0}$  are independent and uniformly distributed in the interval  $[-2, 2]$ . The error terms  $\epsilon_{i,1} - \epsilon_{i,0}$  are independent normally distributed random variables with mean 0 and variance  $\sigma_\epsilon^2 = 0.5$ . The probability of variation is described by the following probit model :

$$\pi(x_{i,1} - x_{i,0}, \beta_0, \beta) = \mathbb{P}[\beta_0 + \beta(x_{i,1} - x_{i,0}) + \nu > 0]$$

where the distribution of  $\nu$  is a standard Gaussian, independent of  $\epsilon$ . The constant terms  $\tilde{\theta}_0 = \theta_1 - \theta_0$  and  $\beta_0$  are both equal to 1, while the slope parameters  $\theta$  and  $\beta$  take different values corresponding to different scenarios. Hypothesis  $(\mathbf{H}_{1t})$ , for  $t = 1$ , as well as hypothesis  $(\mathbf{H}_2)$  are satisfied, noting that

$$\mathbf{Q}_\pi = \mathbb{E} \left( \pi(x_1 - x_0, \beta_0, \beta) \begin{pmatrix} 1 & 0 \\ 0 & (x_1 - x_0)^2 \end{pmatrix} \right)$$

is a definite positive matrix.

We generate data considering different values for  $\theta$ , with  $\theta \in \{2, 1, 0.6, 0.2, -0.2, -0.6, -1, -2\}$ , while  $\beta = 2$  and draw, in Figure 3, the expected variation of the outcome given in (34), non linear estimates obtained with (20), as well as linear estimates based on the empirical version of (9)<sup>1</sup>. Pointwise confidence intervals, with 95% confidence, built via the bootstrap procedure described below (see also Section 2.5) are also drawn in Figure 3.

The algorithm is the following, based on  $B = 1000$  bootstrap replications:

- Repeat for  $b = 1$  to  $b = B$ 
  - Draw from the initial sample a paired bootstrap sample,  $(Y_{1,0}^*, Y_{1,1}^*, x_{1,1}^* - x_{1,0}^*), \dots, (Y_{n,0}^*, Y_{n,1}^*, x_{n,1}^* - x_{n,0}^*)$  with equal probability sampling with replacement.
  - Estimate, with probit and OLS, the conditional probability of not observing zero and the continuous part of the zero-inflated model, respectively, and then compute, as in (20), the estimated expected value  $\widehat{\Delta Y}^b = \pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}^b)(\widehat{\boldsymbol{\theta}}^b)^\top \mathbf{x}$ .

Then, non-parametric bootstrap confidence intervals with confidence  $\alpha$  are built by considering the quantiles of order  $\alpha/2$  and  $1 - \alpha/2$  for the estimated expected value.

===== Figure 3 =====

It clearly appears from the plots in Figure 3 that the zero-inflated phenomenon can produce very different functional relations depending on the parameters  $\beta$  and  $\theta$ . When  $\beta$  and  $\theta$  have the same sign the relation is monotonic, otherwise, when they have opposite signs, the resulting relation can also be non-monotonic. Clearly, as  $\beta$  (resp.  $\theta$ ) gets closer to 0 the resulting relation approaches linearity (resp. a probit shape).

As far as the estimation is concerned, the proposed estimator, which handles the zero inflation, provides a very faithful description of the underlying DGP. Additionally, the true underlying relation is always within the bootstrapped bands, which closely follow the DGP. In contrast, the linear model always provides misleading results, and in particular, when the underlying relation is non-monotonic it clearly provides a senseless fit.

By considering samples with moderate sizes,  $n = 200$ , we also evaluate the ability of the bootstrap procedure to build reliable confidence intervals. Results are plotted in Figure 4, for a nominal level of  $1 - \alpha = 0.95$ . We note that irrespective of the values of the parameters  $\beta$  and  $\theta$ ,

---

<sup>1</sup>Additional results obtained by considering other values for  $\beta$ , are available upon request.

the empirical coverages are most often very close to the nominal ones. The only exception is when  $X$  is in the range between  $-2$  and  $-1$ , where the empirical coverage is about 0.9. Overall, these results offer evidence of the validity of the non-parametric bootstrap in the proposed zero-inflated framework even in the case of a small sample size.

===== Figure 4 =====

## 5 Empirical illustrations

### 5.1 Mincer wage equation

We reexamine the classical problem of estimating a wage equation using panel data and present evidence that standard approaches, which rely on OLS estimation techniques after consideration of individual differencing over time, such as long difference (LD) and first difference (FD) estimators, are prone to bias in the presence of the zero-inflated phenomenon described earlier. We find that this bias is sizeable, even when the fraction of observations equal to zero is relatively small.

#### 5.1.1 Overview of the model and data

We employ the dataset detailed in Baltagi and Khanti-Akom (1990), as discussed in Section 2, where the long-differenced dependent variable,  $\log(WAGE_{i,t}) - \log(WAGE_{i,0})$ , exhibits a distribution that can be approximated by a mixture of a mass at 0 and a continuous density function, with a support defined over both positive and negative values. As far as the explanatory variables are concerned, in addition to years of education (*edu*) and full-time work experience (*exp*), the dataset also contains the number of weeks worked (*wks*) and some dummy variables: occupation (*occ* = 1 if the individual is a blue-collar worker), industry (*ind* = 1 if the individual works in manufacturing), geographical location (*south* = 1 and *smsa* = 1 if the individual resides in the south and in a metropolitan area, respectively), marital status (*ms* = 1 if the individual is married), union coverage (*union* = 1), sex (*fem* = 1 if the individual is female), and race (*blk* = 1 if the individual is black).

The Mincer wage equation, derived from a theoretical model of schooling choice and post-schooling training decisions, serves as the cornerstone of a vast literature in empirical economics. Its simplicity and ability to accurately depict reality make it a prominent topic of study (Card, 1999). Numerous studies have delved into the empirical validity of this specification and its implications (Heckman et al., 2006). A notable debate revolves around the functional form, particularly the adoption of a quadratic form for experience. According to Murphy and Welch (1990), the

quadratic specification tends to provide a poor approximation of the underlying concave function. It tends to overstate initial earnings, overstate earnings at mid-career, and understate earnings at retirement. Subsequent proposals suggested the use of higher-order polynomial functions (Lemieux, 2006). More recently, studies employing non-parametric regression models have offered additional insights. For example, Henderson and Souto (2018) presented evidence of a concave but monotonic relation, aligning with the primary findings of Murphy and Welch (1990).

In this paper, we opt for a log-log specification, which offers several advantages. Firstly, it allows for the identification of the parameter of work experience when time dummies are incorporated into the model, a feature that the log-level specification lacks.<sup>2</sup> Second, the log-log specification also encompasses a variety of non-linear relations between *WAGE* and *exp*, and in particular, it may allow for a decreasing marginal return of experience. We are not claiming that the log-log model offers the most accurate approximation to the underlying function. However, we choose it because it aligns with a concave and monotonic relation between wage and work experience, as suggested by the literature discussed earlier, and it is straightforward for the purposes of our illustration. Additionally, it is important to note that within the proposed framework, it is not possible to identify the effect of education and other time-invariant variables. Consequently, as an illustrative example, we concentrate on examining the impact of experience.

### 5.1.2 Continuous response model

In column (i) of Table 1, for the sake of comparison with our proposal, we present the estimated values of the parameters obtained using the standard LD estimator, which involves applying OLS to the regression model expressed in long differences (i.e., the difference between time  $t$  and  $t_0$ ). This estimator, with a long tradition in panel data econometrics, was initially proposed to address the errors-in-variables problem (Griliches and Hausman, 1986) and has since been applied in various contexts (Hahn et al., 2007; Hanlon and Miscio, 2017; Behaghel et al., 2014; Segú, 2020). However, it assumes a continuous density function and, in the presence of the zero-inflated phenomenon described by Equations (1), (2), and (3), it is generally a biased estimator of  $\pi\theta$  unless  $\pi(\mathbf{x}_t, \beta_t) = \pi$  does not depend on  $\mathbf{x}_t$ . Below, we will provide evidence that in this empirical application, the conditional probability of observing zero is significantly affected by some of the explanatory variables that are considered in the continuous part of the model.

The results are as follows. When considering the standard LD estimator, the estimated coefficient of  $\log(\text{exp}_{i,t})$  is .183 (s.e.=0.037), suggesting a concave monotonic wage–experience relation (i.e., diminishing returns to experience). This result is close to what is obtained by

---

<sup>2</sup>Indeed, while  $\text{exp}_{i,t} = a_i + t$  is perfectly collinear with respect to the time dummies,  $\log(\text{exp}_{i,t})$  is not as  $\log(a_i + t) \neq \log(a_i) + \log(t)$ .

employing the FD estimator, which is equal to 0.191 (s.e.=0.036) and is broadly consistent with the above-cited literature, which mainly exploits cross-sectional data. Comparing this result with that obtained without including the time dummies may provide some interesting insight into the possible bias that arises because of the omission of time-related factors. In that case, the estimated coefficient of  $\log(\text{exp}_{i,t})$  increases up to 0.817 (0.822 for the FD estimator) indicating an almost linear wage–experience relation and suggesting a sizeable omitted common factors bias. When the model does not contain time effects, we can also apply the long difference estimator to a typical Mincer log-level equation that contains experience and its square as regressors instead of the logarithm of experience. In this case, the LD estimator provides estimates of the coefficients of experience and of its square equal to 0.118 and  $-0.0005$ , respectively (0.116 and  $-0.0005$ , for the FD), which suggests an unsatisfactorily increasing exponential relation between wage and experience, thus reinforcing the idea that including time effects in the econometric specification is of crucial empirical relevance.

Nevertheless, even with the inclusion of time dummies, the standard LD estimator, assuming an underlying continuous response, may suffer from bias due to the zero-inflation phenomenon. Empirical evidence supporting this claim is discussed below.

### 5.1.3 Conditional mixture model

The objective now is to recover partial effects (PEs, see (8)) and average partial effects (APEs) of the underlying zero-inflated model, which is inherently non-linear. To achieve this, the vector of unknown parameters  $\theta$  is estimated using the estimator described in (19), referred to as the *subset estimator*. The estimation results are presented in Table 1, column (ii). The conditional probability  $\pi(\mathbf{x}_t, \beta_t)$  and the partial effects from the binary model  $\frac{\partial \pi(\mathbf{x}_t, \beta_t)}{\partial \mathbf{x}_t}$  are obtained through a probit regression model. For simplicity, we assume that the probit regression coefficient is constant in time, that is  $\forall t, \beta_t = \beta$ .<sup>3</sup>

===== Table 1 =====

From the probit regression model, in column (iii), it emerges that  $\log(\text{exp}_{i,t})$  has a negative and significant effect on the conditional probability  $\pi(\mathbf{x}, \beta)$ , i.e., the conditional probability of observing zero (i.e., a null variation in wages). From the probit model, we can also observe that other factors have a significant effect. These factors are *south*, *union*, both positively affecting the

---

<sup>3</sup>We considered a standard probit model supplemented with time averages of the continuous explanatory variables; refer to Wooldridge (2010) for a thorough discussion.

conditional probability, and *edu*, which instead has a negative effect. Estimating Probit Model (iii) not only provides the basis for computing the PEs of the zero-inflated model but also offers valuable insights from an economic perspective. However, due to the specific focus of this work, we refrain from entering into a detailed discussion of the economic interpretation of these results.

We finally compute, using (8), the individual estimated values of the PE related to  $\log(\text{exp})$  based on the zero-inflated approach. The proposed mixture model yields a highly heterogeneous PE of  $\log(\text{exp})$  across cross-sectional units, ranging from -1.073 to 0.191. The estimated APE is 0.039, notably different from the value of 0.183 obtained with the standard LD estimator. The kernel density estimate of the distribution of the individual PEs (see Figure 5) reveals an asymmetric distribution with a mode around 0.169. We also note that approximately 25% of the PEs exhibit negative values, and that the fourth quartile is concentrated in a dense portion of the domain, specifically between 0.190 and 0.191.

===== Figure 5 =====

These results reveal two significant findings: i) a considerable overestimation of the APE when standard approaches, relying on OLS estimators applied after individual differencing over time (FD and LD), are erroneously employed, and ii) the assumption of an underlying continuous response fails to capture the heterogeneity of the PE induced by zero inflation.

## 5.2 Program evaluation of rural development policies in France

### 5.2.1 Description of the programs, variables, and ATEs of interest

We exploit the dataset compiled by Cardot et al. (2019), which contains comprehensive information on French rural policies, employment, and other socio-economic variables. In France, enterprise-zone programs have been implemented to stimulate job creation. These policies offer fiscal incentives to firms situated in economically disadvantaged areas. Specifically aimed at bolstering employment in rural regions, the *Zones de Revitalisation Rurale* (ZRR) program commenced on September 1, 1996, covering the period from 1996 to 2004. At a broader, supranational level, the European Union pursues objectives of territorial cohesion, convergence, and harmonious development across regions through structural funds. Within this framework, the Objective 5B programs (1991–1993 and 1994–1999) allocated financial subsidies to firms and public entities situated in designated “*rural areas in decline*”. A notable feature of both programs is that the selection process of the treated units was clearly not random, and sources of selection on both observables and unobservables are expected to be relevant.

Municipalities serve as the statistical units of analysis, with the dependent variable  $Y_{i,t}$  representing the number of employees at time  $t$ . This variable has been observed over a decade, from 1993 to 2002. Policy variables include ZRR zoning during the period and 5B zoning from 1994 to 1999. Confounding variables are sourced from the French census of 1990, encompassing demographics, education, and work qualifications aggregated at the municipality level. Additionally, satellite images from 1990 provide data on land use, deemed relevant by literature on local employment growth. Pre-treatment covariates are employed to ensure that  $D$  causes  $\mathbf{x}$  and  $Y$  causes  $\mathbf{x}$  do not occur (Lechner, 2011; Lee, 2005). Noteworthy is the inclusion of the initial level of employment as a regressor, which implies assuming unconfoundedness given a lagged outcome. This inclusion avoids an omitted variable bias, which would be particularly relevant if the average outcome of the treated and control groups differ substantially in the first period (Imbens and Wooldridge, 2009), as in this case.

We focus on the assessment of ZRR and 5B as well as their joint effect and thus adopt a framework with  $R = 4$  multiple potential outcomes. These potential outcomes are associated with the potential treatments  $\{0, ZRR, 5B, ZRR\&5B\}$  indicating the program in which each municipality actually participated. The modality 0 indicates that the municipality was not endowed with either policy measure, whereas  $ZRR$  (respectively,  $5B$ ) indicates that the municipality received incentives only from the ZRR initiative (respectively, only from the 5B initiative) and  $ZRR\&5B$  indicates that the municipality received incentives from both ZRR and 5B. Specifically, we focus on the estimation of the following ATEs:  $ATE^{5B}(t, \mathbf{x}) = \mathbb{E}(Y_t^{5B} - Y_t^0 | \mathbf{x}_i)$  and  $ATE^{ZRR\&5B}(t, \mathbf{x}) = \mathbb{E}(Y_t^{ZRR\&5B} - Y_t^0 | \mathbf{x}_i)$ . As for the effect of ZRR, it is notable that only a few municipalities (specifically 722) are treated. Consequently, we focus our attention on the 7014 municipalities that received incentives from both 5B and ZRR. We calculate the differential effect denoted by  $ATE^{ZRR}(t, \mathbf{x}) = \mathbb{E}(Y_t^{ZRR\&5B} - Y_t^{5B} | \mathbf{x}_i)$ , representing the expected difference in outcomes between municipalities participating in both the ZRR and 5B programs and those participating only in the 5B program, conditioned on the covariates  $\mathbf{x}_i$ .

As for the pre-treatment period  $t_0$ , we set  $t_0 = 1993$ , which is before the introduction of both policies. When setting  $t$ , in principle we could use all of the available information in the data. In particular, by setting  $t = 1994, 1995$  we could conduct placebo tests on ZRR, which was introduced in 1996, and use the remaining time periods,  $t = 1996, \dots, 2002$ , to estimate the temporal treatment effects for ZRR and 5B as well as their interaction, as in Cardot and Musolesi (2020). With the aim of providing an illustration of the proposed approach, we set  $t = 1999$ , which is the last time period under the 5B program.



### 5.2.2 Estimation results and comparison with the continuous response model

In this subsection, we compare the estimated values of the ATEs defined in (31) obtained with the proposed mixture approach with those obtained using a naive method that does not account for the mass at zero and only assumes a continuous response model (Imbens and Wooldridge, 2009). This comparison may offer valuable insights into the magnitude of bias when neglecting the zero-inflation feature of the data. We consider alternative specifications for the regression function (22), which are detailed below. The estimation results are presented in Table 2.

We begin by adopting a common practice in the econometric literature, which involves using a linear specification for the confounding variables and assuming that only the intercept varies between treated groups, while the slope parameters remain constant (Model (i)). This approach extends the DID estimator to account for temporal policy effects and considers linear effects of the initial conditions (Abadie, 2005). We use the same set of variables as in Cardot and Musolesi (2020). Next, we consider more flexible models. In the second model (Model (ii)), due to the strong linearity assumption and the potential for misspecification in the relation between  $Y^r(t)$  for  $r \in \{0, ZRR, 5B, ZRR\&5B\}$  and the regressors, we allow for non-linear effects of the confounding variables. This is achieved by adopting natural cubic regression splines, i.e., piecewise-cubic splines with the constraint that they are linear in their tails beyond the boundary knots. Natural cubic splines are generally preferred to cubic splines because they exhibit fewer edge effects (Harrell Jr, 2015). This also makes the underlying identification conditions less restrictive (Lechner, 2011). Finally, in the third model (Model (iii)), we rely on a linear regression model but assume that both the intercepts and the slope parameters of some confounding variables vary between treated groups (see, for example, Heckman and Hotz, 1989, Eq. 3.9). Following Cardot and Musolesi (2020), we retain only two significant interactions of the policy variable: the first one with the initial level of employment (variable `size`) in the municipality and the second one with its population density (variable `density`).

In order to build confidence intervals, we consider the non-parametric bootstrap approach to approximate the distribution of the conditional counterfactual outcome of each municipality  $i$  having the characteristic  $\mathbf{x}_i$ . We draw  $B = 1000$  bootstrap samples, and for each bootstrap sample  $b$ , with  $b = 1, \dots, B$ , we make the following estimation of the ATE (see (31)):

$$\widehat{\text{ATE}}^{r,b}(t, \mathbf{x}) = \left( \pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}_{r,t}^b) \widehat{\boldsymbol{\theta}}_{r,t}^b - \pi(\mathbf{x}, \widehat{\boldsymbol{\beta}}_{0,t}^b) \widehat{\boldsymbol{\theta}}_{0,t}^b \right)^\top \mathbf{x}$$

Bootstrap confidence intervals are then deduced using the percentile method.

**Average treatment effects** When comparing the proposed conditional mixture model with the naive DID model, it becomes evident from Table 2 that accounting for a mass of observations at zero leads to an increase in the estimated ATEs by approximately 5%–10%. This finding holds true for all three specifications considered (Models (i), (ii), and (iii)), providing robust evidence that accounting for the mass of observations at zero is crucial to prevent a significant underestimation of the average effect of the policies.

===== Table 2 =====

**Distributional treatment effects** The results discussed above conceal another crucial aspect of the proposed mixture model. Despite assuming that only the intercept varies between treated groups, while the slope parameters remain constant, as in Models (i) and (ii), the resulting treatment effects exhibit heterogeneity across individuals according to (26). Distributional treatment effects are reported in Table 3.

===== Table 3 =====

First, concentrating on Models (i) and (ii) reveals significant variation in estimated treatment effects across units when addressing the zero-inflated phenomenon. Specifically, treatment effects for the 99th percentile are often more than twice those of the 1st percentile, contrasting with the estimated treatment effects derived from a continuous response model, which do not vary across units. When examining Model (iii), it becomes apparent that the estimated treatment effects exhibit even greater variation across units compared to Models (i) and (ii). When focusing on Model (iii), it is also interesting to observe a similarity in the distributional treatment effects, as reported in Table 3, when comparing the two estimators: the naive continuous response estimator and our proposal. However, this similarity does not ensure that the two estimators provide similar estimates at the individual (municipal) level.

To illustrate the differences at an individual level between the estimates obtained with the two methods, we construct a new variable defined as the relative change between the treatment effect obtained from the zero-inflated approach ( $\widehat{tez}_i^r$ ) and that obtained from the naive estimator ( $\widehat{ten}_i^r$ ). This variable is defined as  $\widehat{rc}_i^r = (\widehat{tez}_i^r - \widehat{ten}_i^r) / \widehat{ten}_i^r$ . The estimated density functions of  $\widehat{rc}_i^r$ , with bandwidths selected using biased cross-validation, are depicted in Figure 6. For

Models (i) and (ii), all the estimated densities exhibit left-skewed distributions, with the mode around 0.15–0.2. In contrast, for Model (iii), the estimated densities are relatively symmetric, with bimodal shapes observed in two out of three cases. These results underscore that when examining distributional treatment effects, rather than solely focusing on the mean effect, the naive estimator is subject to a considerable bias, the direction of which can be either positive or negative.

===== Figure 6 =====

## 6 Conclusion

In this paper, we introduce a statistical formalization that combines a continuous response regression model with a mass at zero. This framework is designed to address the zero-inflation phenomenon, which may arise when differences over time of the outcome variable are considered. This occurs, for instance, when getting rid of individual effects in panel data or for identification purposes in program evaluation, as well as when directly modeling outcome variation over time for economic reasons.

We first focus attention on unobserved effects panel data models and we provide a mathematical approximation by means of conditional mixtures. Our estimators of the regression coefficients are based on the subset on the subsample of units for which the dependent variable has non-null variations and we derive its asymptotic properties under a specific conditional independence assumption, which is likely to be satisfied in many empirical circumstances. The probability of having no variation over time can be estimated thanks to usual binary regression models, such as probit or logistic regression. We prove the asymptotic normality of the estimator that combines both effects as well as consistency of the empirical bootstrap. We then study difference-in-differences estimation under zero inflation and propose an estimator of the average treatment effect that is proven to be consistent.

We also bring new evidence based both on simulated and real data. The simulated example illustrates the effect of zero inflation on the expected value of the variation of the response variable, and it clearly shows that the zero-inflated phenomenon can produce very different functional relations that depend on the underlying parameters, whereas the linear model fails to provide a

faithful description of the underlying DGP. The simulation study also provides evidence of the effectiveness of non-parametric paired bootstrapping with small samples.

Finally, we revisit two real data examples and apply our statistical methodology to analyze a classical Mincer wage equation, as well as to estimate the ATE of two distinct public policies aimed at boosting rural development in France. In both cases, the estimation results provide additional insights into the effectiveness of our proposed estimator. They also indicate that commonly used regression models, which assume that the response variable is continuous, may introduce a significant bias in average effects. Moreover, assuming an underlying continuous density function fails to capture the heterogeneity of PEs arising from the non-linear nature of the zero-inflation model.

The present work could be extended in many directions. For instance, further studies could explore instrumental variables estimation techniques tailored specifically for situations involving zero inflation. or may focus on more flexible non-parametric regression models. These extensions lie beyond the scope of the present paper and deserve further investigation.

## References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies* 72(1), 1–19.
- Akerlof, G. A. and J. L. Yellen (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics* 105(2), 255–283.
- Álvarez, L. J., E. Dhyne, M. Hoeberichts, C. Kwapil, H. Le Bihan, P. Lünnemann, F. Martins, R. Sabbatini, and P. Vermeulen (2006). Price changes in the euro area and the united states: Some facts from individual consumer price data. *Journal of Economic Perspectives* 20(2), 171–192.
- Arulampalam, W., A. L. Booth, and M. L. Bryan (2004). Microeconomic evidence of nominal wage rigidity in the united kingdom. *The Economic Journal* 114(496), F377–F399.
- Baltagi, B. H. and S. Khanti-Akom (1990). On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics* 5(4), 401–406.
- Behaghel, L., E. Caroli, and M. Roger (2014). Age-biased technical and organizational change, training and employment prospects of older workers. *Economica* 81(322), 368–389.

- Blanchard, O. J. and J. Gali (1992). Wage rigidity in europe: Evidence from a survey of european firms. Technical Report 4155, National Bureau of Economic Research.
- Blanchard, O. J. and L. F. Katz (1991). Wage rigidity: A literature review. *NBER Macroeconomics Annual* 6.
- Bose, A. and S. Chatterjee (2003). Generalized bootstrap for estimators of minimizers of convex functions. *J. Statist. Plann. Inference* 117(2), 225–239.
- Brenner, R. (1980). The role of nominal wage contracts in keynes’ general theory. *History of Political Economy* 12(4), 582–587.
- Calmfors, L. and J. Driffill (1988). Bargaining structure, corporatism and macroeconomic performance. *Economic policy* 3(6), 13–61.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics* 3, 1801–1863.
- Cardot, H. and A. Musolesi (2020). Modeling temporal treatment effects with zero inflated semi-parametric regression models: the case of local development policies in france. *Econometric Reviews* 39, 135–157.
- Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.
- Druant, M., S. Fabiani, G. Kezdi, A. Lamo, F. Martins, R. Sabbatini, and G. Weber (2012). Firms’ price and wage adjustment in europe: Survey evidence on nominal stickiness. *Labour Economics* 19(5), 772–782.
- Friedman, M. (1971). A monetary theory of nominal income. *Journal of Political Economy* 79(2), 323–337.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225(2), 254–277.
- Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of econometrics* 31(1), 93–118.
- Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of econometrics* 140(2), 574–617.

- Haley, J. (1990). Theoretical foundations for sticky wages. *Journal of Economic Surveys* 4(2), 115–155.
- Han, S. (2021). Identification in nonparametric models for dynamic treatment effects. *Journal of Econometrics* 225(2), 132–147.
- Hanlon, W. W. and A. Miscio (2017). Agglomeration: A long-run panel data approach. *Journal of Urban Economics* 99, 1–14.
- Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.
- Heckman, J. and V. Hotz (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., H. Ichimura, J. Smith, and P. Todd (1998, September). Characterizing Selection Bias Using Experimental Data. *Econometrica* 66(5), 1017–1098.
- Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64(4), 605–654.
- Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education* 1, 307–458.
- Henderson, D. J. and A.-C. Souto (2018). An introduction to nonparametric regression for labor economists. *Journal of Labor Research* 39(4), 355–382.
- Hjort, N. L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.
- Imbens, G. W. and J. M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1), 5–86.
- Kahn, S. (1997). Evidence of nominal wage stickiness from microdata. *The American Economic Review* 87(5), 993–1008.
- Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics* 4(3), 165–224.

- Lechner, M. (2015). Treatment effects and panel data. In B. Baltagi (Ed.), *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.
- Lee, M.-j. and C. Kang (2006). Identification for difference in differences with cross-section and panel data. *Economics letters* 92(2), 270–276.
- Lemieux, T. (2006). The “mincer equation” thirty years after schooling, experience, and earnings. In *Jacob Mincer a pioneer of modern labor economics*, pp. 127–145. Springer.
- Lindbeck, A., D. J. Snower, et al. (1989). The insider-outsider theory of employment and unemployment. *MIT Press Books 1*.
- Mincer, J. (1974). *Schooling, experience and earnings*. Columbia University Press for National Bureau of Economic Research, New York.
- Murphy, K. M. and F. Welch (1990). Empirical age-earnings profiles. *Journal of Labor economics* 8(2), 202–229.
- Newey, K. and D. McFadden (1994). Large sample estimation and hypothesis. *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, 2112–2245.
- Pissarides, C. A. (2009). Wage rigidity and job creation. *European Economic Review*.
- Sala-i Martin, X. (1997). I just ran two million regressions. *The American Economic Review* 87(2), 178–183.
- Segú, M. (2020). The impact of taxing vacancy on housing markets: Evidence from france. *Journal of Public Economics* 185, 104079.
- 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.
- Stiglitz, J. E. (1984). Theories of wage rigidity. Technical report, National Bureau of Economic Research.
- Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225, 175–199.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

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.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.



## A Proofs

*Proof.* of Proposition 2.1

First note that, with (3),

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) \Delta \mathbf{x}_{i,t} \right). \quad (35)$$

Under assumption  $(\mathbf{H}_2)$ ,  $(\sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top)$ ,  $i = 1, \dots, n$ , are i.i.d with expectation  $\mathbf{Q}_\pi$ . The Khintchine's weak law of large numbers gives us, as  $n$  tends to infinity,

$$\left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top \right) - \mathbf{Q}_\pi = o_p(1). \quad (36)$$

The application of the continuous mapping theorem (see van der Vaart (1998), Theorem 2.3), together with assumption  $(\mathbf{H}_2)$  which implies that inversion is continuous in a neighborhood of  $\mathbf{Q}_\pi$ , gives us

$$\left( \frac{1}{n} \sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top \right)^{-1} - \mathbf{Q}_\pi^{-1} = o_p(1). \quad (37)$$

We also have that, with the set of assumptions  $(\mathbf{H}_{1,t})$ ,  $t = 1, \dots, T$ ,

$$\begin{aligned} \mathbb{E}[Z_{i,t}(\epsilon_{i,t} - \epsilon_{i,0})\Delta \mathbf{x}_{i,t}] &= \mathbb{E}\left(\mathbb{E}[Z_{i,t}|\mathbf{x}_{i,t}, \mathbf{x}_{i,0}] \mathbb{E}[\epsilon_{i,t} - \epsilon_{i,0}|\mathbf{x}_{i,t}, \mathbf{x}_{i,0}] \Delta \mathbf{x}_{i,t}\right) \\ &= 0 \end{aligned} \quad (38)$$

and with the Khintchine's weak law of large numbers, as  $n$  tends to infinity,

$$\frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) \Delta \mathbf{x}_{i,t} = o_p(1). \quad (39)$$

We deduce, using the continuous mapping theorem, (37) and (39) that

$$\left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \Delta \mathbf{x}_{i,t} \Delta \mathbf{x}_{i,t}^\top \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) \Delta \mathbf{x}_{i,t} \right) = o_p(1)$$

which proves, with decomposition (35), the first point of the proposition.

To get the asymptotic normality of  $\widehat{\boldsymbol{\theta}}$ , note that the random vectors  $(\sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) \Delta \mathbf{x}_{i,t})$ ,  $i = 1, \dots, n$  are i.i.d, with expectation 0 and variance-covariance matrix  $\mathbf{Q}_{Z,\epsilon}$ . We deduce from the central limit theorem that

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) \Delta \mathbf{x}_{i,t} \right) \rightsquigarrow \mathcal{N}(0, \mathbf{Q}_{Z,\epsilon}) \quad (40)$$

and with (35), (37) and Slutsky's Lemma (see van der Vaart (1998), Proposition 2.8),

$$\sqrt{n} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightsquigarrow \mathcal{N}(0, \mathbf{Q}_\pi^{-1} \mathbf{Q}_{Z,\epsilon} \mathbf{Q}_\pi^{-1}).$$

□

*Proof.* of Proposition 2.2 *The proof is based on classical arguments (see Newey and McFadden (1994), Theorem 3.1), and relies on a Taylor expansion of the gradient of the objective function  $\Psi_n$  as well as the conditional independence assumptions  $(\mathbf{H}_{1,t}), t = 1, \dots, T$ . We clearly have, with the additive structure of  $\Psi_n$  given in (14), that the Hessian matrix is block diagonal, since, for  $t = 1, \dots, T$ , and  $\nu \neq t$ ,*

$$\frac{\partial^2 \Psi_n}{\partial \beta_t \partial \theta} = 0 \qquad \frac{\partial^2 \Psi_n}{\partial \beta_t \partial \beta_\nu} = 0.$$

*The gradient of  $\Psi_n$  being equal to zero at  $(\hat{\theta}, \hat{\beta}_1, \dots, \hat{\beta}_T)$ , we thus have*

$$0 = \begin{pmatrix} \frac{\partial \Psi_n}{\partial \theta} \\ \frac{\partial \Psi_n}{\partial \beta_1} \\ \vdots \\ \frac{\partial \Psi_n}{\partial \beta_T} \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \Psi_n}{\partial \theta^\top \partial \theta} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 \Psi_n}{\partial \beta_1^\top \partial \beta_1} & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \Psi_n}{\partial \beta_T^\top \partial \beta_T} \end{pmatrix} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\beta}_1 - \beta_1 \\ \vdots \\ \hat{\beta}_T - \beta_T \end{pmatrix} \quad (41)$$

*where the second order partial derivatives are evaluated componentwise, at points between  $(\hat{\theta}, \hat{\beta}_1, \dots, \hat{\beta}_T)$  and  $(\theta, \beta_1, \dots, \beta_T)$ . On the other hand, we have with (4),*

$$\begin{aligned} \frac{\partial \Psi_n}{\partial \theta} &= -\frac{2}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} \left( \Delta_c Y_{i,t} (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) - (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) (\mathbf{x}_{i,t} - \mathbf{x}_{i,0})^\top \theta \right) \\ &= -\frac{2}{n} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t} (\epsilon_{i,t} - \epsilon_{i,0}) (\mathbf{x}_{i,t} - \mathbf{x}_{i,0}) \end{aligned} \quad (42)$$

*and, with (15),*

$$\frac{\partial \Psi_n}{\partial \beta_t} = -\frac{1}{n} \sum_{i=1}^n (Z_{i,t} \phi_1(\mathbf{x}_{i,t}, \beta_t) + \phi_2(\mathbf{x}_{i,t}, \beta_t)) \mathbf{x}_{i,t} \quad (43)$$

*for some known continuous functions  $\phi_1(.,.)$  and  $\phi_2(.,.)$ . At the true value  $(\theta, \beta)$ , we have  $\mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \right] = 0$  and  $\mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \beta} \right] = 0$ , so that the covariance matrix of  $\frac{\partial \Psi_n}{\partial \theta}$  and  $\frac{\partial \Psi_n}{\partial \beta}$  is equal to  $\mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \frac{\partial \Psi_n}{\partial \beta^\top} \right]$ . Conditioning on  $\mathbf{x}_t$  and  $Z_t$ , for  $t = 1, \dots, T$ , we get*

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \frac{\partial \Psi_n}{\partial \beta^\top} \mid \mathbf{x}_t, Z_t, t = 1, \dots, T \right] &= \mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \mid \mathbf{x}_t, Z_t, t = 1, \dots, T \right] \mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \beta^\top} \mid \mathbf{x}_t, Z_t, t = 1, \dots, T \right] \\ &= 0 \end{aligned} \quad (44)$$

*almost surely. Indeed, under assumption  $(\mathbf{H}_{1,t})$  and decomposition (42), we have*

$$\begin{aligned} \mathbb{E}[Z_t (\epsilon_t - \epsilon_0) (\mathbf{x}_t - \mathbf{x}_0) \mid \mathbf{x}_t, \mathbf{x}_0, Z_t] &= Z_t (\mathbf{x}_t - \mathbf{x}_0) \mathbb{E}[\epsilon_t - \epsilon_0 \mid \mathbf{x}_t, \mathbf{x}_0, Z_t] \\ &= 0 \quad \text{almost surely.} \end{aligned}$$

*Thus*

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \frac{\partial \Psi_n}{\partial \beta^\top} \right] &= \mathbb{E} \left( \mathbb{E} \left[ \frac{\partial \Psi_n}{\partial \theta} \frac{\partial \Psi_n}{\partial \beta^\top} \mid \mathbf{x}_t, Z_t, t = 1, \dots, T \right] \right) \\ &= 0. \end{aligned}$$

*Consequently, the covariance matrix of the score vector is block diagonal and the asymptotic covariance matrix of the estimators is also block diagonal.  $\square$*

*Proof.* of Proposition 2.3.

The first part of the Proposition is a direct consequence of Theorem 2.1 and Theorem 2.4 in Bose and Chatterjee (2003), remarking that if we assume that the link function for  $\pi(\mathbf{x}, \boldsymbol{\beta})$  has a logit or probit shape, the objective function  $\Psi_n(\boldsymbol{\theta}, \boldsymbol{\beta}; (Y_{i,0}, \dots, Y_{i,T}, \mathbf{x}_{i,0}, \dots, \mathbf{x}_{i,T})_{i=1}^n)$  is a convex function, in  $(\boldsymbol{\theta}, \boldsymbol{\beta})$  that is also twice differentiable. The Hessian matrix is positive definite at the true value of the parameter  $(\boldsymbol{\theta}, \boldsymbol{\beta})$  thanks to hypotheses  $(\mathbf{H}_2)$  and  $(\mathbf{H}_{3,t})$   $t = 1, \dots, T$ .

The second part of the proof is a direct consequence of the delta method for bootstrapped estimates (see Theorem 23.9 in van der Vaart (1998)) considering the functions  $\pi(\mathbf{x}, \boldsymbol{\beta})\boldsymbol{\theta}^\top \mathbf{x}$  and  $\frac{\partial}{\partial \mathbf{x}} \pi(\mathbf{x}, \boldsymbol{\beta})\boldsymbol{\theta}^\top \mathbf{x}$ , which are differentiable with respect to  $(\boldsymbol{\theta}, \boldsymbol{\beta})$ .  $\square$

*Proof.* of Proposition 3.2.

We follow the same lines as the proof of Proposition 2.1 and thus omit some details. First note that our estimators  $(\widehat{\boldsymbol{\theta}}_{0,t}, \widehat{\boldsymbol{\theta}}_{r,t}, \widehat{\boldsymbol{\beta}}_{0,t}, \widehat{\boldsymbol{\beta}}_{r,t})$  are defined as the minimizers of the functional

$$\Psi_n(\boldsymbol{\theta}_0, \boldsymbol{\theta}_r, \boldsymbol{\beta}_0, \boldsymbol{\beta}_r) = \Psi_{1,n,t}^0(\boldsymbol{\beta}_0) + \Psi_{1,n,t}^r(\boldsymbol{\beta}_r) + \Psi_{2,n,t}^0(\boldsymbol{\theta}_0) + \Psi_{2,n,t}^r(\boldsymbol{\theta}_r).$$

It is thus straightforward, under hypotheses  $(\mathbf{H}_{4,t})$ ,  $(\mathbf{H}_{5,t})$ , and  $(\mathbf{H}_{6,t})$  to get that the regression parameters are consistent. As  $n$  tends to infinity,  $\widehat{\boldsymbol{\theta}}_{0,t} - \widetilde{\boldsymbol{\theta}}_{0,t} = o_p(1)$  and  $\widehat{\boldsymbol{\theta}}_{r,t} - \widetilde{\boldsymbol{\theta}}_{r,t} = o_p(1)$ .

As far as  $\boldsymbol{\beta}_{0,t}$  and  $\boldsymbol{\beta}_{r,t}$  are concerned, their maximum likelihood estimators do not come from a standard maximum likelihood framework because the number of observations (the sample size),  $n_r(n) = \sum_{i=1}^n D_i^r$  is not deterministic. If  $n_r$  was not random, we would directly get under previous assumptions that the maximum likelihood estimator of  $\boldsymbol{\beta}_{r,t}$  is consistent and asymptotically Gaussian. Note that in our random number of observations case, we have, with expression (28), that for all  $\boldsymbol{\beta}_r \in \mathbb{R}^p$ ,

$$\mathbb{E} [\Psi_{1,n,t}^r(\boldsymbol{\beta}_r)] = -\mathbb{E} \left( Z_t^r \ln \left( \frac{\pi(\mathbf{x}, \boldsymbol{\beta}_r)}{1 - \pi(\mathbf{x}, \boldsymbol{\beta}_r)} \right) + \ln(1 - \pi(\mathbf{x}, \boldsymbol{\beta}_r)) \mid D_t^r = 1 \right) \mathbb{P}[D_t^r = 1]. \quad (45)$$

By assumption  $\mathbf{H}_{5,t}$  we have, given  $D_t^r = 1$ ,

$$\mathbb{P}[Z_t^r = 1 \mid \mathbf{x}] = \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t})$$

so that, with assumption  $(\mathbf{H}_{6,t})$ ,

$$\begin{aligned} \mathbb{E} [\Psi_{1,n,t}^r(\boldsymbol{\beta}_r)] &= -\mathbb{E} \left( \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t}) \ln \left( \frac{\pi(\mathbf{x}, \boldsymbol{\beta}_r)}{1 - \pi(\mathbf{x}, \boldsymbol{\beta}_r)} \right) + \ln(1 - \pi(\mathbf{x}, \boldsymbol{\beta}_r)) \mid D^r = 1 \right) \mathbb{P}[D^r = 1] \\ &> \mathbb{E} [\Psi_{1,n,t}^r(\boldsymbol{\beta}_{r,t})], \end{aligned}$$

for all  $\boldsymbol{\beta}_r \neq \boldsymbol{\beta}_{r,t}$  (see e.g Lemma 2.2 in Newey and McFadden (1994)). We also get, with the strong law of large numbers that for all  $\boldsymbol{\beta}_r \in \mathbb{R}^p$ ,

$$\Psi_{1,n,t}^r(\boldsymbol{\beta}_r) - \mathbb{E} [\Psi_{1,n,t}^r(\boldsymbol{\beta}_r)] \rightarrow 0, \quad \text{almost surely}$$

and we can deduce, by Theorem 2.7 in Newey and McFadden (1994) that the sequence  $\widehat{\boldsymbol{\beta}}_{r,t}$  of minimizers of  $\Psi_{1,n,t}^r$  tends to  $\boldsymbol{\beta}_{r,t}$  almost surely.

For the asymptotic normality, first observe from (22) and (30) that

$$\begin{pmatrix} \widehat{\boldsymbol{\theta}}_{r,t} - \widetilde{\boldsymbol{\theta}}_{r,t} \\ \widehat{\boldsymbol{\theta}}_{0,t} - \widetilde{\boldsymbol{\theta}}_{0,t} \end{pmatrix} = \begin{pmatrix} (\mathbf{Q}_{n,t}^r)^{-1} & 0 \\ 0 & (\mathbf{Q}_{n,t}^0)^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n D_i^r Z_{i,t} \mathbf{x}_i (\epsilon_{i,t}^r - \epsilon_{i,0}) \\ \frac{1}{n} \sum_{i=1}^n D_i^0 Z_{i,t} \mathbf{x}_i (\epsilon_{i,t}^0 - \epsilon_{i,0}) \end{pmatrix} \quad (46)$$

with  $\mathbf{Q}_{n,t}^r = \frac{1}{n} \sum_{i=1}^n D_i^r Z_{i,t} \mathbf{x}_i \mathbf{x}_i^\top$ . The strong law of large numbers gives directly that  $\mathbf{Q}_{n,t}^r - \mathbf{Q}_t^r = o_p(1)$  as  $n$  tends to infinity. The random vectors  $(D_i^0 Z_{i,t} \mathbf{x}_i (\epsilon_{i,t}^0 - \epsilon_{i,0}), D_i^r Z_{i,t} \mathbf{x}_i (\epsilon_{i,t}^r - \epsilon_{i,0}))$ ,

$i = 1, \dots, n$  are i.i.d copies of  $(D^0 Z_t \mathbf{x}(\epsilon_t^0 - \epsilon_0), D^r Z_t \mathbf{x}(\epsilon_t^r - \epsilon_0))$ , and

$$\begin{aligned} \mathbb{E}(D^0 Z_t \mathbf{x}(\epsilon_t^0 - \epsilon_0)) &= \mathbb{E}[\mathbb{E}(D^0 Z_t(\epsilon_t^0 - \epsilon_0) | \mathbf{x}) \mathbf{x}] \\ &= \mathbb{E}[\mathbb{E}(D^0 | \mathbf{x}) \mathbb{E}(Z_t(\epsilon_t^0 - \epsilon_0) | \mathbf{x}) \mathbf{x}] && \text{with } \mathbf{H}_{5,t} \\ &= \mathbb{E}[\mathbb{E}(D^0 | \mathbf{x}) \mathbb{E}(Z_t | \mathbf{x}) \mathbb{E}(\epsilon_t^0 - \epsilon_0 | \mathbf{x}) \mathbf{x}] && \text{with } \mathbf{H}_{4,t} \\ &= 0. \end{aligned} \tag{47}$$

For  $r \neq 0$ , we have

$$\begin{aligned} \text{Cov}((D^0 Z_t \mathbf{x}(\epsilon_t^0 - \epsilon_0)) \mathbf{x}, D^r Z_t(\epsilon_t^r - \epsilon_0) \mathbf{x}) &= \mathbb{E} \left[ D^0 D^r Z_t(\epsilon_t^0 - \epsilon_0)(\epsilon_t^r - \epsilon_0) \mathbf{x} \mathbf{x}^\top \right] \\ &= 0 \end{aligned} \tag{48}$$

because  $D^0 D^r = 0$  almost surely. The Central Limit Theorem and Slutsky's Lemma allow to conclude that

$$\sqrt{n} \left( \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{0,t} \\ \widehat{\boldsymbol{\theta}}_{r,t} \end{pmatrix} - \begin{pmatrix} \widetilde{\boldsymbol{\theta}}_{0,t} \\ \widetilde{\boldsymbol{\theta}}_{r,t} \end{pmatrix} \right) \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Gamma}_{\theta,t}^r),$$

with  $\boldsymbol{\Gamma}_{\theta,t}^r$  a block diagonal covariance matrix.

Note that the asymptotic normality of  $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{r,t} - \boldsymbol{\beta}_{r,t})$  is based on an application of Theorem 3.3 in Newey and McFadden (1994) for probit regression, with asymptotic variance given by

$$\boldsymbol{\Gamma}_t^r = \frac{1}{\mathbb{P}[D^r = 1]} \left[ \mathbb{E} \left( \lambda(\boldsymbol{\beta}_{r,t}^\top \mathbf{x}) \lambda(-\boldsymbol{\beta}_{r,t}^\top \mathbf{x}) \mathbf{x} \mathbf{x}^\top | D^r = 1 \right) \right]^{-1}$$

where  $\lambda(u) = \Phi'(u)/\Phi(u)$ ,  $u \in \mathbb{R}$ . In case of logistic regression, it can be deduced from Theorem 5.1 in Hjort and Pollard (2011) that

$$\boldsymbol{\Gamma}_t^r = \frac{1}{\mathbb{P}[D^r = 1]} \left[ \mathbb{E} \left( \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t})(1 - \pi(\mathbf{x}, \boldsymbol{\beta}_{r,t})) \mathbf{x} \mathbf{x}^\top | D^r = 1 \right) \right]^{-1}.$$

For the joint normality, we have, with (28), that

$$\frac{\partial \Psi_{1,n,t}^r}{\partial \boldsymbol{\beta}_r} = -\frac{1}{n} \sum_{i=1}^n D_i^r (Z_{i,t} \phi_1(\mathbf{x}_i, \boldsymbol{\beta}_r) + \phi_2(\mathbf{x}_i, \boldsymbol{\beta}_r)) \mathbf{x}_i \tag{49}$$

for some known continuous functions  $\phi_1(\cdot, \cdot)$  and  $\phi_2(\cdot, \cdot)$ . We thus get, for  $r$  and  $\kappa$  in  $\{0, 1, \dots, R-1\}$ ,

$$\mathbb{E} \left[ \frac{\partial \Psi_{1,n,t}^r}{\partial \boldsymbol{\beta}_r} \frac{\partial \Psi_{2,n,t}^\kappa}{\partial \boldsymbol{\theta}_\kappa^\top} \right] = \frac{2}{n} \mathbb{E} \left[ D^r D^\kappa (Z_t \phi_1(\mathbf{x}, \boldsymbol{\beta}_r) + \phi_2(\mathbf{x}, \boldsymbol{\beta}_r)) Z_t \left( (Y_t^\kappa - Y_0) - \mathbf{x}^\top \boldsymbol{\theta}_\kappa \right) \mathbf{x} \mathbf{x}^\top \right] \tag{50}$$

Previous expression is clearly equal to 0 when  $r \neq \kappa$  since  $D^r D^\kappa = 0$  for  $r \neq \kappa$ . When  $r = \kappa$ , the covariance evaluated at the true value  $(\widetilde{\boldsymbol{\theta}}_{r,t}, \boldsymbol{\beta}_{r,t})$  is equal

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \Psi_{1,n,t}^r}{\partial \boldsymbol{\beta}_r} \frac{\partial \Psi_{2,n,t}^r}{\partial \boldsymbol{\theta}_r^\top} \right] &= \frac{2}{n} \mathbb{E} \left[ D^r Z_t (Z_t \phi_1(\mathbf{x}, \boldsymbol{\beta}_{r,t}) + \phi_2(\mathbf{x}, \boldsymbol{\beta}_{r,t})) \left( (Y_t^r - Y_0) - \mathbf{x}^\top \widetilde{\boldsymbol{\theta}}_{r,t} \right) \mathbf{x} \mathbf{x}^\top \right] \\ &= \frac{2}{n} \mathbb{E} \left[ D^r Z_t (Z_t \phi_1(\mathbf{x}, \boldsymbol{\beta}_r) + \phi_2(\mathbf{x}, \boldsymbol{\beta}_r)) (\epsilon_t^r - \epsilon_0) \mathbf{x} \mathbf{x}^\top \right] \\ &= 0, \end{aligned} \tag{51}$$

thanks to (23) and assumptions  $(\mathbf{H}_{4,t})$  and  $(\mathbf{H}_{5,t})$ . This implies that the asymptotic covariance matrix is block diagonal.  $\square$

## B Appendix: description of the variables

We present here the variables that were considered in Section 5.2. A detailed description of the definition of these variables as well as some descriptive statistics can be found in the Appendix of Cardot and Musolesi (2020).

The dependent variable  $Y_{i,t}$  corresponds to the number of employees at time  $t$  for municipality  $i$ . 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. By starting from a set of sixteen possible explanatory variables, the final set of variables, which were selected by employing a backward variable selection procedure, contains the following eleven variables:

- **size**  $\equiv Y_{t_0}$  is the initial outcome, i.e the level of employment at  $t_0$ , with  $t_0$  equals to 1993.
- **density**  $\equiv (\text{total population}) / (\text{total surface in terms of } km^2)$ ;
- **income**  $\equiv (\text{net taxable income}) / (\text{total population})$ ;
- **old**  $\equiv (\text{population over 65}) / (\text{total population})$ ;
- **fact**  $\equiv (\text{number of factory workers}) / (\text{total population})$ ;
- **bts**  $\equiv \frac{(\text{number of people with a technical degree called “Brevet de Technicien Supérieur”})}{(\text{total population})}$ ;
- **agri**  $\equiv (\text{farmland surface}) / (\text{total surface})$ ;
- **cult**  $\equiv (\text{cultivated land surface}) / (\text{total surface})$ ;
- **urb**  $\equiv (\text{urban surface}) / (\text{total surface})$ ;
- **ind**  $\equiv (\text{industrial surface}) / (\text{total surface})$ ;
- **ara**  $\equiv (\text{arable surface}) / (\text{total surface})$ ;

where the total surface and the total population should be understood within the considered municipality.

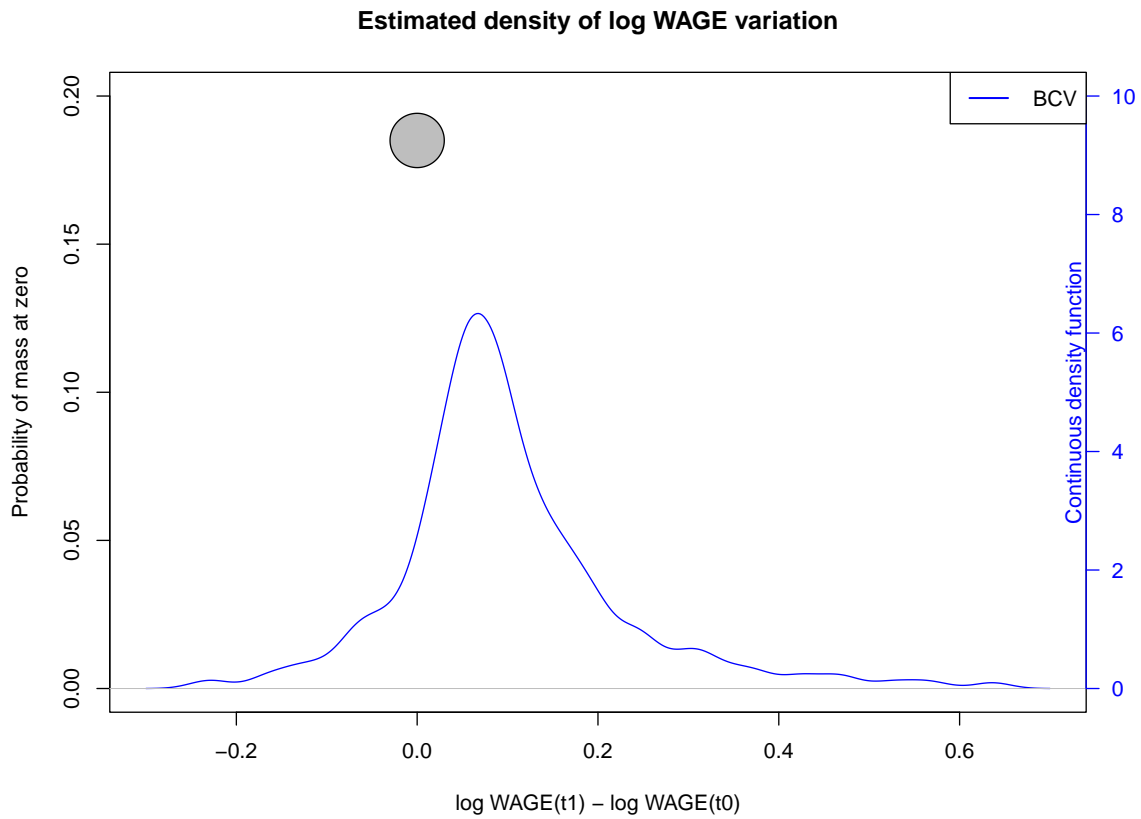


Figure 1: The estimated distribution of  $\log(WAGE_{i,1}) - \log(WAGE_{i,0})$ . The probability of a mass at zero is estimated by the proportion of observations such that  $\log(WAGE_{i,1}) - \log(WAGE_{i,0}) = 0$  (indicated by the circle). We also consider a continuous density estimation of  $\log(WAGE_{i,1}) - \log(WAGE_{i,0}) \neq 0$  thanks to a kernel estimator; BCV: biased cross-validation (see Sheather, 2004; Silverman, 1986).

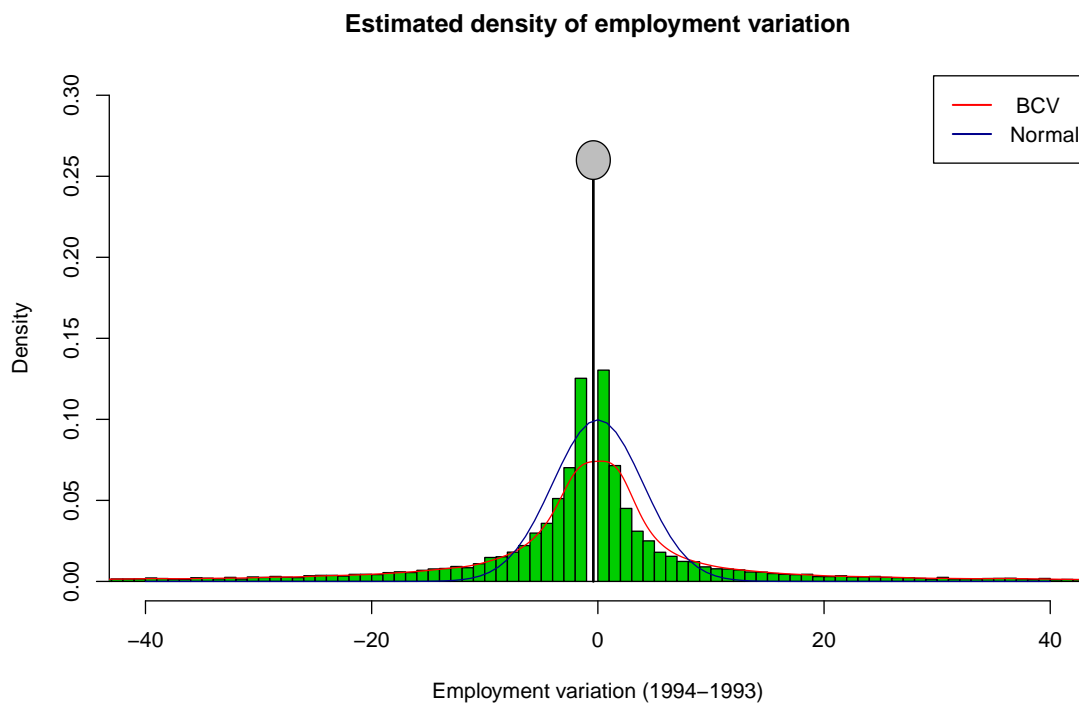


Figure 2: The estimated distribution of  $EMP_{i,1} - EMP_{i,0}$  with  $t_0 = 1993$ . The probability of observing no variation is estimated by the proportion of observations such that  $EMP_{i,1} - EMP_{i,0} = 0$ . The vertical bars represent the probability of observing a given value when  $EMP_{i,1} - EMP_{i,0} \neq 0$ . We also consider a continuous density estimation of  $EMP_{i,1} - EMP_{i,0} \neq 0$  thanks to a kernel estimator; BCV: biased cross-validation (see Sheather, 2004; Silverman, 1986).

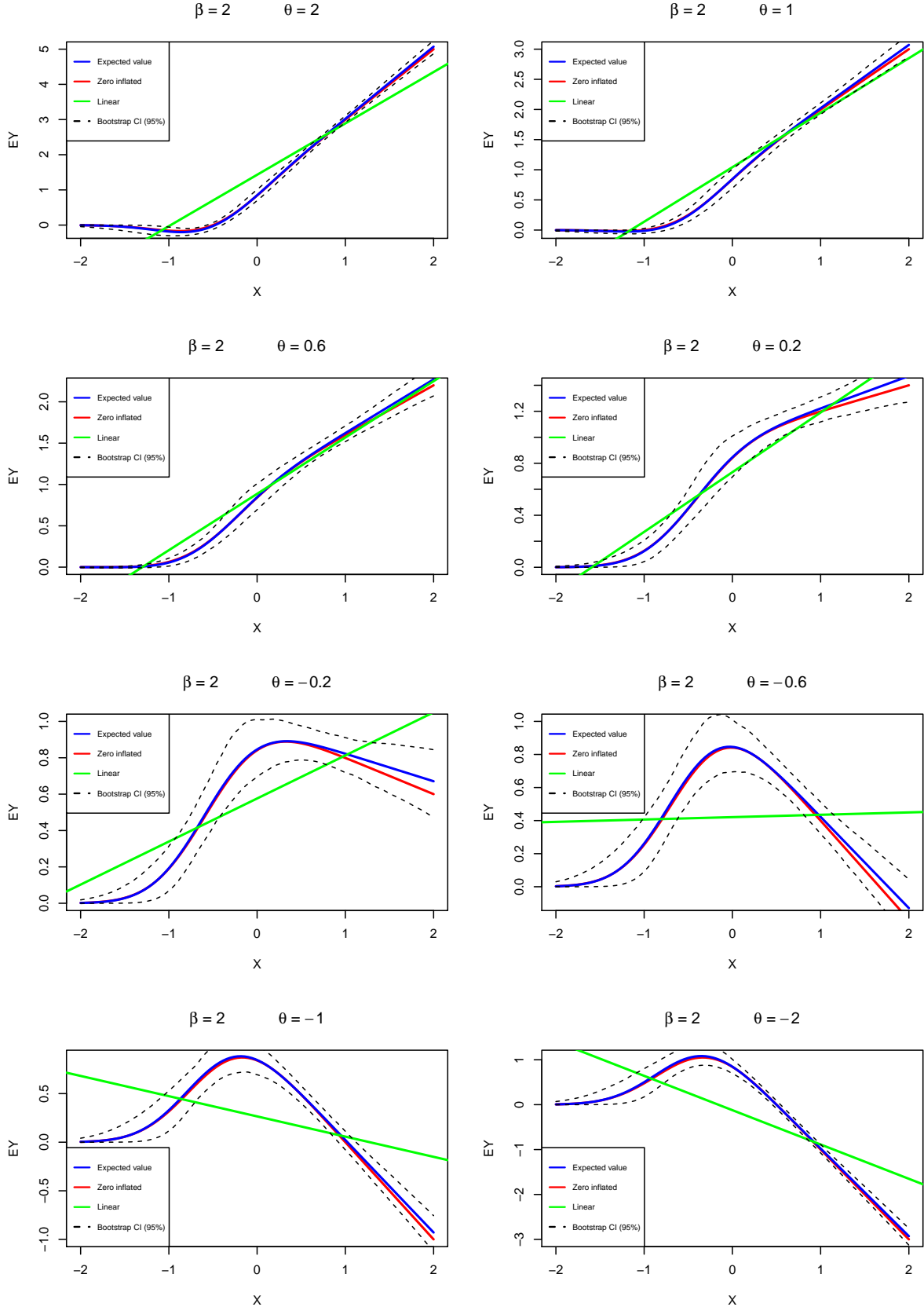


Figure 3: Simulation. Sample size  $n = 200$ . The vertical axis represents the conditional expected value  $\mathbb{E}[Y_{i,1} - Y_{i,0} | x_{i,0}, x_{i,1}] = \pi(x_{i,1} - x_{i,0}, \beta_0, \beta) \times (\theta_1 - \theta_0 + \theta(x_{i,1} - x_{i,0}))$  where  $\pi(x_{i,1} - x_{i,0}, \beta_0, \beta) = \mathbb{P}[\beta_0 + \beta(x_{i,1} - x_{i,0}) + \nu > 0]$  for different values for  $\theta$ , with  $\theta \in \{2, 1, 0.6, 0.2, -0.2, -0.6, -1, -2\}$ , and  $\beta = 2$ . Bootstrap confidence intervals are built by considering the percentile approach over 1000 replications.



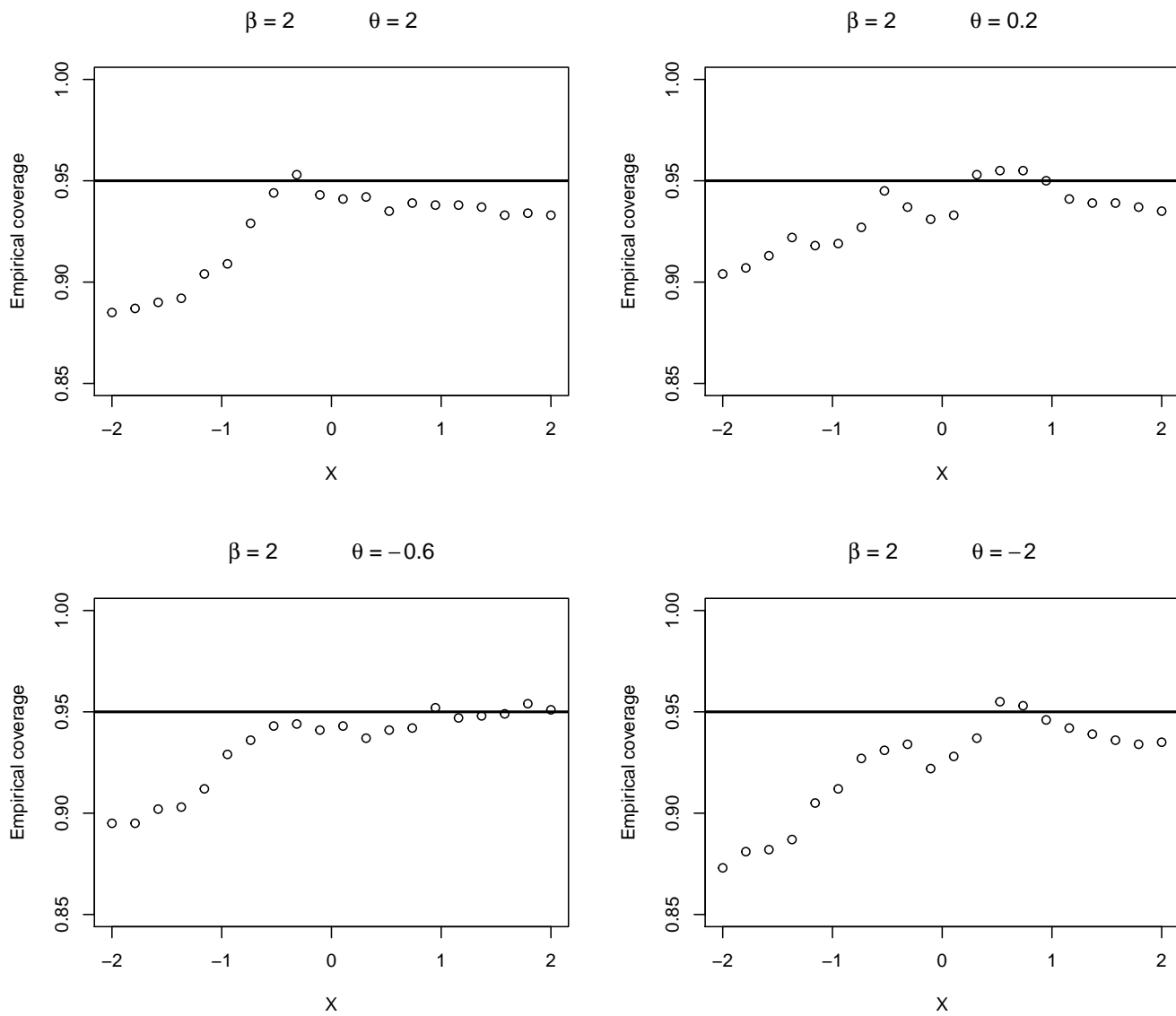


Figure 4: Empirical coverage for the conditional expected value  $\mathbb{E}[Y_{i,1} - Y_{i,0} | x_{i,0}, x_{i,1}]$ , with a nominal level of  $1 - \alpha = 0.95$ , for different values of  $\theta$ , with  $\theta \in \{2, 1, 0.6, 0.2, -0.2, -0.6, -1, -2\}$ , and  $\beta = 2$ . The sample size is  $n = 200$ .

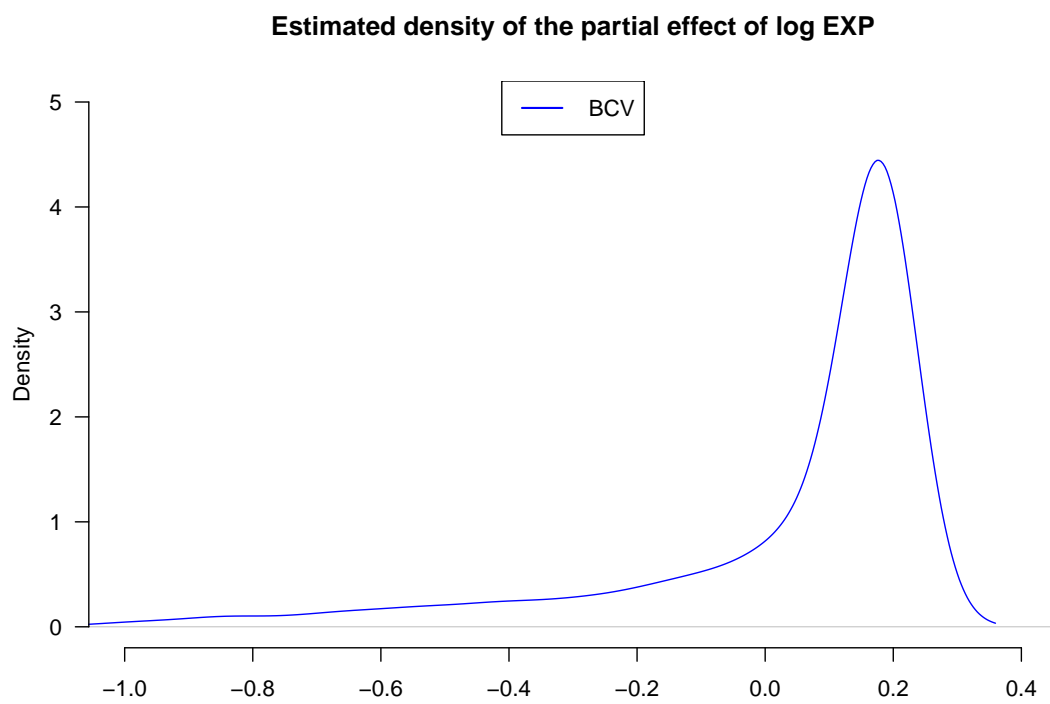


Figure 5: The estimated distribution of the partial effect of  $\log(EXP)$  in the wage equation from the zero-inflated model. Bandwidth selected using biased cross-validation.

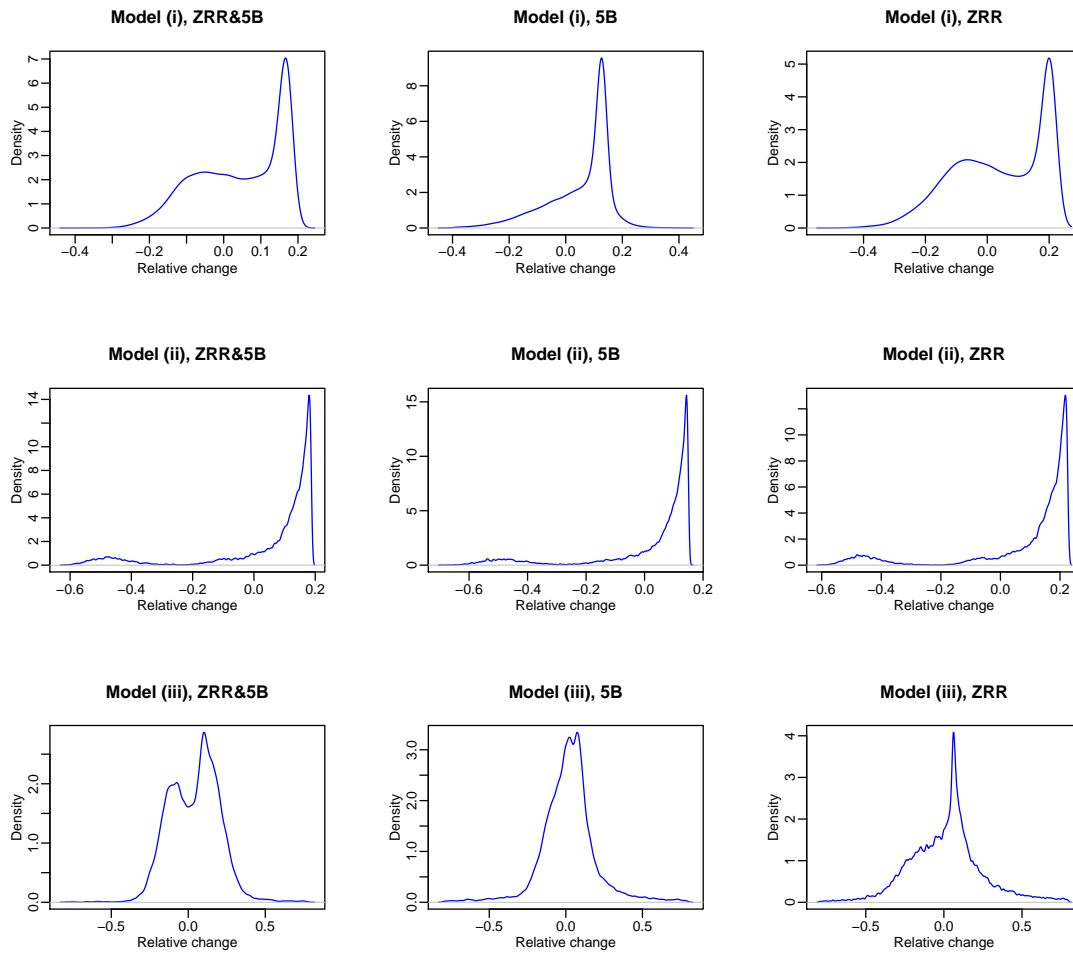


Figure 6: The estimated distribution of the relative change between the treatment effect obtained using the zero-inflated approach and the one obtained adopting the naive estimator. Bandwidth value selected using biased cross-validation.

	CONTINUOUS RESPONSE	CONDITIONAL MIXTURE MODEL		
	(i)	(ii)	(iii)	
		Continuous part	Discrete part	
	LD - whole sample	LD - subset sample	CRE Probit - Pooled MLE	
	Coefficient	Coefficient	Coefficient	APE
log(exp)	0.183*** (0.037)	0.191*** (0.036)	-4.708*** ( 1.111)	-0.270*** (0.060)
log(wks)	0.026 (0.024)	0.027 (0.024)	-0.507 (0.639)	-0.029 (0.036)
occ	-0.017 (0.022)	-0.016 (0.023)	0.169 (0.122)	0.009 (0.006)
ind	0.044* (0.026)	0.045* (0.026)	0. 111 (0.104)	0.006 (0.005)
south	-0.058 (0.079)	-0.060 (0.080)	0.454*** (0.115)	0.0260*** (0.006)
smsa	-0.064 (0.042)	-0.066 (0.042)	-0.010 (0.122)	-0.010 (0.006)
ms	-0.056* (0.029)	-0.056* (0.029)	-0.394* (0.218)	-0.022* (0.012)
union	0.053** (0.027)	0.051* (0.027)	0.385*** ( 0.142)	0.022*** (0.008)
fem			-0.300 (0.277)	-0.0172 (0.016)
blk			0.076 (0.230)	0.004 (0.013)
edu			-0.061** (0.024)	- 0.003** (0.001)

All specifications include a full set of time dummies.

The standard errors of the estimated coefficients (in brackets) are robust to arbitrary serial correlation.

The standard errors of the APEs in the CRE probit model are obtained using the delta method.

\*\*\*, \*\*, \*: significant at 1%, 5%, and 10% level, respectively.

Table 1: Wage equation

	CONTINUOUS RESPONSE MODEL			CONDITIONAL MIXTURE MODEL		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
ATE <sup>ZRR&amp;5B</sup>	2.021 [0.664-3.303]	3.001 [1.688-4.358]	2.955 [1.105-4.828]	2.110 [0.710-3.401]	3.134 [1.860-4.557]	3.016 [1.160-4.941]
ATE <sup>5B</sup>	0.896 [-0.452-2.331]	1.571 [0.213-2.981]	0.781 [-0.530-2.144]	0.943 [-0.400-2.356]	1.604 [0.245-2.997]	0.821 [-0.485-2.180]
ATE <sup>ZRR</sup>	1.125 [-0.194-2.318]	1.430 [0.173-2.785]	2.174 [0.187- 4.140]	1.167 [-0.154-2.387]	1.530 [0.271-2.935]	2.195 [0.205-4.201]

Model (i): DID with linear regression function.

Model (ii): DID with natural cubic regression splines.

Model (iii): DID with linear regression function and policy interaction with density and sie.

Between brackets: 95% confidence bands computed by nonparametric bootstrap (percentile method)

Table 2: Average treatment effects

CONTINUOUS RESPONSE MODEL															
Model	(i)					(ii)					(iii)				
Percentile	1	25	50	75	99	1	25	50	75	99	1	25	50	75	99
ZRR&5B	2.021	2.021	2.021	2.021	2.021	3.001	3.001	3.001	3.001	3.001	-4.721	2.289	2.786	3.558	12.384
5B	0.896	0.896	0.896	0.896	0.896	1.571	1.571	1.571	1.571	1.571	-7.951	0.604	1.350	1.724	2.99
ZRR	1.125	1.125	1.125	1.125	1.125	1.430	1.430	1.430	1.430	1.430	-7.391	0.566	1.380	2.866	20.336
CONDITIONAL MIXTURE MODEL															
ZRR&5B	1.576	1.916	2.148	2.349	2.368	1.382	3.142	3.371	3.470	3.524	-5.167	2.269	2.269	3.594	12.963
5B	0.628	0.889	0.987	1.011	1.101	0.683	1.605	1.724	1.774	1.800	-8.072	0.746	1.352	1.655	3.056
ZRR	0.782	1.036	1.180	1.338	1.357	0.683	1.536	1.646	1.696	1.724	-7.830	0.651	1.356	2.774	20.894
CONDITIONAL MIXTURE MODEL vs. CONTINUOUS RESPONSE MODEL (Relative change in the treatment effect)															
ZRR&5B	-0.220	-0.052	0.063	0.162	0.172	-0.539	0.046	0.123	0.156	0.174	-0.356	-0.088	0.056	0.155	0.678
5B	-0.299	-0.007	0.101	0.128	0.229	-0.564	0.022	0.097	0.130	0.146	-1.937	-0.077	0.022	0.104	2.135
ZRR	-0.303	-0.078	0.049	0.190	0.206	-0.522	0.074	0.151	0.186	0.205	-2.279	-0.136	0.037	0.1431	2.831

Model (i): DID with linear regression function.

Model (ii): DID with cubic regression splines.

Model (iii): DID with linear regression function and policy interaction with density and size.

Table 3: Distributional treatment effects