Distribution Regression Difference-In-Differences

Iván Fernández-Val[*] Jonas Meier[†] Aico van Vuuren[‡] Francis Vella[§]

February 28, 2024

Abstract

We provide a simple distribution regression estimator for treatment effects in the difference-in-differences (DiD) design. Our procedure is particularly useful when the treatment effect differs across the distribution of the outcome variable. Our proposed estimator easily incorporates the role of covariates and can also be employed to examine whether the treatment affects the joint distribution of multiple outcomes. Our key identifying restriction is that the counterfactual distribution of the treated in the untreated state has no interaction effect between treatment and time. This assumption results in a parallel trend assumption on the transformation of the distribution. We highlight the relationship between our procedure and assumptions with the changes-in-changes estimator of Athey and Imbens (2006). We also provide an empirical example which highlights the utility of our approach.

1 Introduction

The remarkable popularity of the difference-in-difference estimator, inspired by an approach to evaluate the impact of policy interventions on economic outcomes introduced by David Card, is one of the most striking features of empirical work on treatment and policy effects (see, for example, Card 1990, Card and Krueger, 1994). While the methodological

[*]Boston University
[†]Swiss National Bank.
[‡]University of Groningen.
[§]Georgetown University.

innovations in this literature cover a range of issues, including the use of constructed control groups, the staggered timing of treatments, and fuzzy rather than sharp designs, the vast majority of the associated empirical work has been restricted to estimating the mean effect of the treatment on a single economic outcome (see Arkhangelsky and Imbens, 2023 for a recent review article). This seems restricted given that a fuller evaluation of a policy effect generally requires an examination of distributional effects and also a consideration of multiple outcomes. We address each of these issues by providing a simple procedure for estimating distributional treatment effects in the presence of a single treatment but when the outcomes of interest are potentially multivariate.

The initial methodological innovation devoted to providing a distributional approach to difference-in-differences (DiD) estimation is the changes-in-changes procedure of Athey and Imbens (2006). That paper focuses on estimating the counterfactual distribution of the treated group in the absence of treatment and comparing it to the observed distribution in the presence of treatment. Other work has adopted the approach of conducting DiD estimation to explain the impact of treatment at various quantiles of the outcome via the use of quantile regression. These include Callaway and Li (2018, 2019). In contrast, Dube (2019), Goodman-Bacon (2021), and Goodman-Bacon and Schmidt (2020) employ conventional DiD estimation to explore the impact of the treatment at different points of the outcome distribution. Other approaches include comparisons based on actual and constructed counterfactual distributions. Two papers that adopt this approach are Kim and Wooldridge (2023) and Biewen, Fitzenberger, and Rümmele (2022). The former suggests the use of an inverse probability weighting procedure, while the latter employs a distribution regression (DR) approach. This paper also adopts a DR approach to constructing counterfactuals. In contrast to Biewen, Fitzenberger, and Rümmele (2022), who construct the counterfactual distributions via a series of linear probability models, we employ a series of non-linear link functions such as probit or logit models. This has a number of advantages, which we discuss below. In addition, we provide the associated identifying conditions required for this form of the implementation of DR.

While DiD has typically been employed to evaluate the treatment effect on a certain economic outcome there are many instances in which the treatment is likely to have an impact on a number of outcomes. Evaluating how the treatment affects the relationship between these outcomes might also be of economic interest. For example, a change in tax rates on earnings of married couples may affect both the labor supply of the husbands and wives so an analysis of such a change should include the impact on both outcomes as the objects of interest. However, a richer analysis would not only examine the impact on the respective marginal hours distribution of husbands and wives but also the joint distribution of hours. We introduce this type of treatment effect via the bivariate distribution regression (BDR) approach of Fernandez-Val et al (2023). This requires that we first estimate the joint distribution by BDR and then construct the appropriate counterfactual. As in the univariate case, the treatment effects are then constructed via the appropriate comparisons.

The following section introduces the model and provides an analysis of the univariate case. We extend our analysis to include covariates and contrast our approach with the Athey and Imbens (2006) changes-in-changes estimator. Section 3 extends our analysis to the multiple outcome case and section 4 discusses estimation. Section 5 provides an empirical illustration of our methodology by revisiting the Malesky et al. (2014) investigation of the impact of recentralization in Vietnam. Section 6 concludes.

2 Econometric analysis of the univariate case

Consider the standard DiD design with 2 periods, $T \in \{0, 1\}$, and 2 groups, $G \in \{0, 1\}$ in which a binary treatment, $D \in \{0, 1\}$, is administered only to the treatment group with $G = 1$ in the second period $T = 1$. Let Y_0 and Y_1 denote the potential outcomes under the non-treated and treated statuses. The observed outcome is $Y = Y_0(1 - D) + Y_1 D$, which corresponds to Y_0 for both groups at $T = 0$, with Y_0 for $G = 0$ at $T = 1$, and to Y_1 for $G = 1$ at $T = 1$. Note that this implicitly imposes a non-anticipation assumption as we do not distinguish between the outcomes of the treated and non-treated state for $G = 1$ in period $T = 0$.

We are interested in the distributions of the potential outcomes of the treated at $T = 1$, that is $F_{Y_1 \mid G,T}(y \mid 1, 1)$ and $F_{Y_0 \mid G,T}(y \mid 1, 1)$. $F_{Y_1 \mid G,T}(y \mid 1, 1)$ is identified from the observed outcome for $G = 1$ at $T = 1$,

$$F_{Y_1 \mid G,T}(y \mid 1, 1) = F_{Y \mid G,T}(y \mid 1, 1);$$

whereas $F_{Y_0 \mid G,T}(y \mid 1, 1)$ is not identified without further assumptions.

The distribution of Y_0 conditional on G and T can be written as

$$F_{Y_0 \mid G,T}(y \mid g, t) = \Lambda(\alpha(y) + \beta(y)t + \gamma(y)g + \delta(y)gt), \quad y \in \mathbb{R}, \tag{1}$$

where Λ is an invertible CDF such as the logistic, normal or uniform, and $y \mapsto (\alpha(y), \beta(y), \gamma(y), \delta(y))$ is a vector of function-valued parameters.

The representation in (1) does not make any parametric assumption about the underlying distribution of $Y_0 \mid G, T$. The representation is local (*i.e.* is assumed to hold for a specific level of y) and because the dummy variable representation within the parentheses at the right-hand side is saturated it makes no parametric assumption. To understand the flexibility of (1), note that for any representation of Λ, $\alpha(y)$, $\beta(y)$, $\gamma(y)$ and $\delta(y)$ solve[1]

$$\begin{aligned}
\alpha(y) &= \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 0)\right) \\
\beta(y) &= \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 1)\right) - \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 0)\right) \\
\gamma(y) &= \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 1, 0)\right) - \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 0)\right) \\
\delta(y) &= \Lambda^{-1}\left(F_{Y_1 \mid G,T}(y \mid 1, 1)\right) - \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 1, 0)\right) \\
&\quad - \left[\Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 1)\right) - \Lambda^{-1}\left(F_{Y_0 \mid G,T}(y \mid 0, 0)\right)\right]
\end{aligned}$$

Distribution regression was originally developed by Williams and Grizzle (1972) and by Foresi and Peracchi (1995). Inference for distribution regression with continuous outcome variables was developed in Chernozhukov et al. (2013) and Chernozhukov et al. (2020) extend inference to outcome variables which are discrete, mixed discrete or continuous.

We make the following identifying assumptions with respect to the distribution function in (1).

Assumption 1 [No-interaction assumption]**.** *For the distribution function* $F_{Y_0 \mid G,T}(y \mid g, t)$ *introduced in* (1) *we impose* $\delta(y) = 0$ *for all* $y \in \mathbb{R}$*.*

[1] See also Wooldridge (2023) equations (2.6) and (2.7).

To proceed, we now define the support of the random variable $Y_i|G = g, T = t$ with $i, g, t \in 0, 1$ by $\mathcal{Y}_i(G = g, T = t)$.

Assumption 2 [Support conditions]**.** *We have the following restrictions with respect to the supports*

$$\mathcal{Y}_0(G = 1; T = 1) \subseteq \mathcal{Y}_0(G = 0; T = 1) \cup \mathcal{Y}_0(G = 1; T = 0) \cup \mathcal{Y}_0(G = 0; T = 0)$$

Assumption 1 implies that the distribution of the potential outcome Y_0 should not change differently in the second period for the treatment group compared to the control group. We allow a difference between the distributions of the potential outcome Y_0 between the treatment and control group, but this difference should be identical in both periods. This is a parallel trend type assumption on a transformation of the distribution and can be written as

$$\Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,1,1)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,1,0)\right) = \\ \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,1)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right).$$

This assumption is sensitive to the link function and can impose restrictions on the distribution $F_{Y_0\,|\,G,T}$ for some link functions. For example, if Λ is the identity link used in the linear probability model, it imposes strong requirements on the tails of the distributions $F_{Y_0\,|\,G,T}(y\,|\,1,0)$, $F_{Y_0\,|\,G,T}(y\,|\,0,1)$ and $F_{Y_0\,|\,G,T}(y\,|\,0,0)$ to guarantee that $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ is between 0 and 1. Thus, it requires that $F_{Y_0\,|\,G,T}(y\,|\,1,0) \leq 1 + F_{Y_0\,|\,G,T}(y\,|\,0,0) - F_{Y_0\,|\,G,T}(y\,|\,0,1)$, which might be restrictive at the top of the distribution, and $F_{Y_0\,|\,G,T}(y\,|\,1,0) \geq F_{Y_0\,|\,G,T}(y\,|\,0,0) - F_{Y_0\,|\,G,T}(y\,|\,0,1)$, which might be restrictive at the bottom of the distribution.[2] Other links such as the normal or logistic CDFs do not require such restrictions because the transformation expands the range of the distribution. Assumption 1 cannot be empirically verified but when we have multiple observations in the pre-treatment period, it is possible to examine whether the "parallel trends" assumption at least holds pre-treatment.

When the identity function is used for the link function in (1), then Assumption 1 results in models employed in, for example, Almond et al. (2011), Dube (2019), Cengiz

[2]This requirements can be used to develop a specification test for the identity link. We do not pursue this approach in the paper because we do not encourage the use of the linear probability model in practice.

et al. (2019), Goodman-Bancon and Smith (2020), Goodman-Bacon (2021) and Biewen et al. (2022). As previously noted by Blundell et al. (2004) and Wooldridge (2023), using the identity function has as a drawback that the parallel trends assumption might be suspect as the outcome variable is limited and non-linear by nature. This is especially the case when the outcome is near the borders of 0 and 1. That is an increase of, for example, 0.2 in probability over time might be realistic for the control group when the original probability equals 0.5. However, when the treatment group already starts in the first period with a probability of, for example, 0.9, then it is not even possible for the common trends assumption to hold.

Assumption 2 is a restriction of the support of the counterfactual outcome of Y_0 for the treated group in the treated period.

These two assumptions identify $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ because

$$\begin{aligned}
F_{Y_0\,|\,G,T}(y\,|\,1,1) &= \Lambda(\alpha(y)+\beta(y)+\gamma(y)) \\
&= \Lambda\left[\Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,1,0)\right)+\Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,1)\right)-\Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right)\right] \\
&= \Lambda\left[\Lambda^{-1}\left(F_{Y\,|\,G,T}(y\,|\,1,0)\right)+\Lambda^{-1}\left(F_{Y\,|\,G,T}(y\,|\,0,1)\right)-\Lambda^{-1}\left(F_{Y\,|\,G,T}(y\,|\,0,0)\right)\right], \quad (2)
\end{aligned}$$

under the non-anticipation assumption. The support restrictions in Assumption 2 ensure the term in parentheses in (2) is determined. Note that as $\lim_{x\to\infty}\Lambda(x)=1$ and $\lim_{x\to-\infty}\Lambda(x)=0$, our assumptions are sufficient but not necessary.

We present this identification result in the following lemma:

Lemma 1 [Identification with One Outcome]**.** $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ *is identified under Assumptions 1 and 2.*

Proof of Lemma 1. The results follows from equation (2). □

2.1 Inclusion of covariates

Including covariates is appealing as the assumption that $\delta(y)=0$ may be harder to defend when there are either differences in the trend between covariates or when the composition of the treatment group changes over time in terms of observed characteristics; see also Melly and Santangelo (2015). Covariates can be trivially incorporated for identification by making the analysis conditional on them and adding an overlapping support assumption. Specifically, let X be a vector of covariates such that the non-interaction assumption holds conditional on X; see Assumption 3. The distribution of Y_0 conditional on G, T and X can be written as

$$F_{Y_0 \,|\, G,T,X}(y \,|\, g,t,x) = \Lambda(\alpha(y,x) + \beta(y,x)t + \gamma(y,x)g + \delta(y,x)gt), \quad y \in \mathbb{R}, \tag{3}$$

where $(y,x) \mapsto (\alpha(y,x), \beta(y,x), \gamma(y,x), \delta(y,x))$ is a vector of unspecified functions.

The identifying assumptions with covariates become:

Assumption 3 [No-interaction with Covariates]**.** *For the distribution function $F_{Y_0 \,|\, G,T,X}$ as introduced in* (3) *we impose that $\delta(y,X) = 0$ almost surely for all $y \in \mathbb{R}$.*

Assumption 4 [Support conditions with Covariates]**.** *We have the following restrictions with respect to the supports*

$$\mathcal{Y}_0(G=1;T=1;X) \subseteq \mathcal{Y}_0(G=0;T=1;X) \cup \mathcal{Y}_0(G=1;T=0;X) \cup \mathcal{Y}_0(G=0;T=0;X)$$

almost surely.

These two assumptions identify $F_{Y_0 \,|\, G,T,X}(y \,|\, 1,1,x)$ because

$$\begin{aligned} &F_{Y_0 \,|\, G,T,X}(y \,|\, 1,1,x) = \Lambda(\alpha(y,x) + \beta(y,x) + \gamma(y,x)) \\ &= \Lambda\left[\Lambda^{-1}\left(F_{Y_0 \,|\, G,T,X}(y \,|\, 1,0,x)\right) + \Lambda^{-1}\left(F_{Y_0 \,|\, G,T,X}(y \,|\, 0,1,x)\right) - \Lambda^{-1}\left(F_{Y_0 \,|\, G,T,X}(y \,|\, 0,0,x)\right)\right] \\ &= \Lambda\left[\Lambda^{-1}\left(F_{Y \,|\, G,T,X}(y \,|\, 1,0,x)\right) + \Lambda^{-1}\left(F_{Y \,|\, G,T,X}(y \,|\, 0,1,x)\right) - \Lambda^{-1}\left(F_{Y \,|\, G,T,X}(y \,|\, 0,0,x)\right)\right], \end{aligned} \tag{4}$$

under the non-anticipation assumption. The support restrictions in Assumption 4 make sure that the term between parentheses in (4) is determined. Note that as $\lim_{x\to\infty} \Lambda(x) = 1$ and $\lim_{x\to-\infty} \Lambda(x) = 0$, our assumptions are sufficient but not necessary. Then, we can identify $F_{Y_0 \,|\, G,T}(y \,|\, 1,1)$ as

$$F_{Y_0 \,|\, G,T}(y \,|\, 1,1) = \int F_{Y_0 \,|\, G,T,X}(y \,|\, 1,1,x) \mathrm{d}F_{X \,|\, G,T}(x \,|\, 1,1). \tag{5}$$

We gather this identification result in the following lemma:

Lemma 2 [Identification with Covariates]**.** *Under Assumptions 3 and 4,* $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ *is identified.*

Proof of Lemma 1. The results follows from equations (4) and (5). □

For estimation, we replace the functions $(y,x) \mapsto (\alpha(y,x), \beta(y,x), \gamma(y,x))$ by semi-parametric linear indexes leading to the DR model for the conditional distribution:

$$F_{Y_0\,|\,G,T,X}(y\,|\,g,t,x) = \Lambda(p(x)'\alpha(y) + q(x)'\beta(y)t + r(x)'\gamma(y)g), \quad y \in \mathbb{R}, \tag{6}$$

where $p(x)$, $q(x)$ and $r(x)$ are vectors including the covariates and their transformations, and $y \mapsto (\alpha(y), \beta(y), \gamma(y))$ is a vector of function-valued parameters.

2.2 A comparison of our model with the Changes-In-Changes model (Athey and Imbens, 2006)

The changes-in-changes (CiC) design assumes that the outcome of an individual without treatment satisfies the relationship $Y_0 = h(U,T)$ with an unobserved and uniformly distributed term U. It is assumed that h is strictly increasing in the first term and that the distribution of U is independent of time given the treatment outcome, i.e. $U \perp\!\!\!\perp T|G$. Finally, the support of U for the treated population should be a subset of those of the untreated population. The final assumption implies in terms of the support of the outcomes that

$$\mathcal{Y}_0(G=1,T=0) \subseteq \mathcal{Y}_0(G=0,T=0)$$
$$\mathcal{Y}_0(G=1,T=1) \subseteq \mathcal{Y}_0(G=0,T=1)$$

Their second support restriction is less restrictive than ours but we do not need their first support restriction.

The assumptions of Athey and Imbens (2006) identify the quantile function of $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ as

$$F^{-1}_{Y_0\,|\,G,T}(u\,|\,1,1) = \phi\left(F^{-1}_{Y_0\,|\,G,T}(u\,|\,1,0)\right),$$
$$\phi(y) := F^{-1}_{Y_0\,|\,G,T}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\,|\,0,1\right), \quad u \in \{0,1\},$$

where we assume that Y_0 is continuous with strictly increasing distribution function. The transformation ϕ gives the second period outcome for an individual with an unobserved component u such that $h(u, 0) = y$, with y the location at which we evaluate the distribution function (Athey and Imbens, 2006, page 441). Hence, their identification results follows from the idea that ϕ evaluated in the first period observations of the treatment group is equally distributed as the distribution of the untreated outcome of the treatment group in the second period. It means that their assumptions result in the implicit assumption that the transformation ϕ that maps quantiles of Y_0 from period 0 to period 1 is the same for the treatment and control groups. This condition imposes the following restrictions on the coefficients of the representation of the conditional distribution in (1):

$$\alpha(y) = \alpha(\phi(y)) + \beta(\phi(y)), \quad \gamma(y) = \gamma(\phi(y)) + \delta(\phi(y)).$$

To see this, note that

$$F_{Y_0 \mid G,T}(y \mid g, 0) = F_{Y_0 \mid G,T}(h(h^{-1}(y, 0), 1) \mid g, 1). \tag{7}$$

Evaluating (7) at $g = 0$ and applying $F^{-1}_{Y_0 \mid G,T}(\cdot \mid 0, 1)$ to both sides

$$h(h^{-1}(y, 0), 1) = F^{-1}_{Y_0 \mid G,T}\left(F_{Y_0 \mid G,T}(y \mid 0, 0) \mid 0, 1\right) =: \phi(y).$$

Replacing $\phi(y)$ back in (7) and using the representation (1)

$$\Lambda(\alpha(y) + \gamma(y)g) = \Lambda(\alpha(\phi(y)) + \beta(\phi(y)) + \gamma(\phi(y))g + \delta(\phi(y))g).$$

The restrictions then follow from equalizing the coefficients in both sides.[3] They complicate estimation in our framework as they involve two different levels of Y and the transformation ϕ needs to be estimated.

2.3 Comparison with Roth and Sant'Anna (2023)

Roth and Sant'Anna (2023) derive the condition,

$$F_{Y_0 \mid G,T}(y \mid 1, 1) - F_{Y_0 \mid G,T}(y \mid 1, 0) = F_{Y_0 \mid G,T}(y \mid 0, 1) - F_{Y_0 \mid G,T}(y \mid 0, 0), \quad y \in \mathbb{R},$$

[3]There is only a binding restriction because $\alpha(y) = \alpha(\phi(y)) + \beta(\phi(y))$ holds by definition of $\phi(y)$.

for the parallel trends assumption in expectations,

$$\mathbb{E}(Y_0 \mid G=1, T=1) - \mathbb{E}(Y_0 \mid G=1, T=0) = \mathbb{E}(Y_0 \mid G=0, T=1) - \mathbb{E}(Y_0 \mid G=0, T=0),$$

to be invariant to strictly monotone transformations of Y_0. This condition is different from our no-interaction assumption. Indeed, our DR model with the no-interaction condition does not generally satisfy the parallel trends assumption in expectation because

$$\mathbb{E}(Y_0 \mid G=g, T=1) - \mathbb{E}(Y_0 \mid G=g, T=0) = \int_{-\infty}^{\infty} [\Lambda(\alpha(y) + \gamma(y)g) - \Lambda(\alpha(y) + \beta(y) + \gamma(y)g)] \mathrm{d}y$$

depends on g unless Λ is the identity map, or $\beta(y) = 0$ (no trend) or $\gamma(y) = 0$ (random assignment) for $y \in \mathbb{R}$. Roth and Sant'Anna (2023) show that their condition holds if and only if there are no trends, random assignment or a mixture of the previous two. Our model, however, generally satisfies a different invariance property with respect to strictly monotonic transformations that we specify in Remark 2.4.

2.4 Invariance to Strictly Monotonic Transformations

The DR model in (1) and the no-interaction assumption are invariant to strictly monotonic transformations in a sense that we specify here. If Y_0 follows the DR model and satisfies the no-interaction assumption, then $\tilde{Y}_0 = h(Y_0)$ also follows the DR model and satisfies the no-interaction assumption for any strictly monotonic transformation h. To see this result, note that, if h is strictly increasing,

$$F_{\tilde{Y}_0 \mid G,T,X}(\tilde{y} \mid g, t, x) = \Lambda(\alpha(h^{-1}(\tilde{y})) + \beta(h^{-1}(\tilde{y}))t + \gamma(h^{-1}(\tilde{y}))g) = \Lambda(\tilde{\alpha}(\tilde{y}) + \tilde{\beta}(\tilde{y})t + \tilde{\gamma}(\tilde{y})g),$$

where $\tilde{y} \mapsto h^{-1}(\tilde{y})$ is the inverse function of $y \mapsto h(y)$, $\tilde{\alpha} = \alpha \circ h^{-1}$, $\tilde{\beta} = \beta \circ h^{-1}$ and $\tilde{\gamma} = \gamma \circ h^{-1}$. A similar argument applies to the case where h is strictly decreasing. In other words, unlike the parallel trends in expectation, the no-interaction or parallel trends in distribution is invariant to strictly monotonic transformations.

3 Multiple Outcomes

There are situations where we observe multiple outcomes and are interested in how the treatment affects their relationship. To analyze such situations, we need to identify the joint distribution of the potential outcomes with and without the treatment. Here, we consider the case of two outcomes Y and Z. We are interested in comparing features of the joint distribution of the potential outcomes with the treatment, Y_1 and Z_1, and the joint distribution of the potential outcomes with the treatment, Y_0 and Z_0, for the treated group $G = 1$ in the post-treatment period $T = 1$. For example, Spearman's rank correlation between Y_d and Z_d, $d \in \{0, 1\}$, can be expressed as

$$\rho[Y_d, Z_d \mid G = 1, T = 1] = \mathrm{Corr}[F_{Y_d \mid G,T}(Y_d \mid 1, 1), F_{Z_d \mid G,T}(Z_d \mid 1, 1) \mid G = 1, T = 1] = \\ 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_{Y_d \mid G,T}(y \mid 1, 1) - 1/2][F_{Z_d \mid G,T}(z \mid 1, 1) - 1/2] F_{Y_d, Z_d \mid G,T}(\mathrm{d}y, \mathrm{d}z \mid 1, 1);$$

and Kendall's rank correlation between Y_d and Z_d, $d \in \{0, 1\}$, can be expressed as

$$\tau[Y_d, Z_d \mid G = 1, T = 1] = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_{Y_d, Z_d \mid G,T}(y, z \mid 1, 1) - 1/4] F_{Y_d, Z_d \mid G,T}(\mathrm{d}y, \mathrm{d}z \mid 1, 1),$$

where we have assumed that Y_d and Z_d are continuous random variables.

As in the univariate case, $F_{Y_1, Z_1 \mid G,T}(y, z \mid 1, 1)$ is identified by the joint distribution of the observed outcomes, $F_{Y,Z \mid G,T}(y, z \mid 1, 1)$, whereas $F_{Y_0, Z_0 \mid G,T}(y, z \mid 1, 1)$ is not identified from the data. To analyze identification, we use the local Gaussian representation (LGR) of a bivariate distribution from Chernozhukov, Fernandéz-Val and Luo (2019). By the LGR, $F_{Y_0, Z_0 \mid G,T}$ can be expressed as

$$F_{Y_0, Z_0 \mid G,T}(y, z \mid g, t) = \\ \Phi_2(\alpha_Y(y) + \beta_Y(y)t + \gamma_Y(y)g + \delta_Y(y)gt, \alpha_Z(z) + \beta_Z(z)t + \gamma_Z(z)g + \delta_Z(z)gt; \rho_{Y,Z \mid G,T}(y, z \mid g, t)),$$

where $\Phi_2(\cdot, \cdot; \rho)$ is the CDF of a standard bivariate normal with correlation ρ, and $\rho_{Y,Z \mid G,T}(y, z \mid g, t) = \alpha_{Y,Z}(y, z) + \beta_{Y,Z}(y, z)t + \gamma_{Y,Z}(y, z)g + \delta_{Y,Z}(y, z)gt$. In the LGR, the marginals are represented by

$$F_{Y_0 \mid G,T}(y \mid g, t) = \Phi(\alpha_Y(y) + \beta_Y(y)t + \gamma_Y(y)g + \delta_Y(y)gt),$$

and

$$F_{Z_0\,|\,G,T}(z\,|\,g,t) = \Phi(\alpha_Z(z) + \beta_Z(z)t + \gamma_Z(z)g + \delta_Z(z)gt),$$

where Φ is the CDF of a standard univariate normal.

Lemma 3 [Identification with Two Outcomes]**.** *If* $\delta_Y(y) = \delta_Z(z) = \delta_{Y,Z}(y,z) = 0$*, then* $F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)$ *is identified.*

Proof of Lemma 3. Under the assumptions of the Lemma

$$F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,g,t) = \Phi_2(\alpha_Y(y)+\beta_Y(y)t+\gamma_Y(y)g, \alpha_Z(z)+\beta_Z(z)t+\gamma_Z(z)g; \rho_{Y,Z\,|\,G,T}(y,z\,|\,g,t)),$$

and $\rho_{Y,Z\,|\,G,T}(y,z\,|\,g,t) = \alpha_{Y,Z}(y,z) + \beta_{Y,Z}(y,z)t + \gamma_{Y,Z}(y,z)g$.

The parameters $\alpha_Y(y)$, $\beta_Y(y)$, $\gamma_Y(y)$, $\alpha_Z(z)$, $\beta_Z(z)$, and $\gamma_Z(z)$ are identified from the marginals of Y and Z, by Lemma 1.

The parameter $\alpha_{Y,Z}(y,z)$ is identified as the solution in α to

$$F_{Y,Z\,|\,G,T}(y,z\,|\,0,0) = \Phi_2(\alpha_Y(y) + \beta_Y(y)t + \gamma_Y(y)g, \alpha_Z(z) + \beta_Z(z)t + \gamma_Z(z)g; \alpha).$$

This solution exists and is unique because the RHS is strictly increasing in α. The parameters $\beta_{Y,Z}(y,z)$ and $\gamma_{Y,Z}(y,z)$ are identified similarly as the solutions in β and γ of

$$F_{Y,Z\,|\,G,T}(y,z\,|\,0,1) = \Phi_2(\alpha_Y(y)+\beta_Y(y)t+\gamma_Y(y)g, \alpha_Z(z)+\beta_Z(z)t+\gamma_Z(z)g; \alpha_{Y,Z}(y,z)+\beta).$$

and

$$F_{Y,Z\,|\,G,T}(y,z\,|\,1,0) = \Phi_2(\alpha_Y(y)+\beta_Y(y)t+\gamma_Y(y)g, \alpha_Z(z)+\beta_Z(z)t+\gamma_Z(z)g; \alpha_{Y,Z}(y,z)+\gamma).$$

Finally,

$$F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1) = \Phi_2(\alpha_Y(y) + \beta_Y(y) + \gamma_Y(y), \alpha_Z(z)\beta_Z(z) + \gamma_Z(z); \alpha_{Y,Z}(y,z)+ \\ \beta_{Y,Z}(y,z) + \gamma_{Y,Z}(y,z)).$$

□

We can estimate the distribution of $F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)$ in a similar way as for the univariate case presented above. That is, we first estimate the parameters $\alpha_Y(y)$, $\beta_Y(y)$, $\gamma_Y(y)$, $\alpha_Z(y)$, $\beta_Z(y)$, $\gamma_Z(y)$, $\alpha_{Y,Z}(y,z)$, $\beta_{Y,Z}(y,z)$, $\gamma_{Y,Z}(y,z)$ using bivariate distribution regression and using the sample of the first period and the sample of the second period for the untreated group. Then, in a second step, we plug in the estimated parameters for the sample of the treated in the second period in order to obtain the estimator of $F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)$. Again, additional regressors can trivially be added to the analysis.

4 Estimation

We start our discussion of estimation for the univariate case and for reasons of illustration only we initially assume that there are no covariates as in equation (1). Estimation is straightforward and can be performed in two steps. See also Wooldridge (2023) for a similar estimation strategy (*i.e.* equations (2.31) and (2.32), page C41.)

Algorithm 1. 1. *Estimate the model parameters* $(\alpha(y), \beta(y), \gamma(y))$ *by distribution regression of the indicator* $1(Y \leq y)$ *on a constant,* T *and* G *for multiple values of* $y \in \mathcal{Y}$*, using all the observations for* $G = 0$ *and the observations for* $G = 1$ *at* $T = 0$*. Denote the estimators as* $(\hat{\alpha}(y), \hat{\beta}(y), \hat{\gamma}(y))$*.*

2. *Construct a plug-in estimator of the distribution of the potential outcomes*

$$\hat{F}_{Y_0\,|\,G,T}(y\,|\,1,1) = \Lambda(\hat{\alpha}(y) + \hat{\beta}(y) + \hat{\gamma}(y)).$$

The distribution $F_{Y_1\,|\,G,T}(y\,|\,1,1)$ can be estimated by the empirical distribution of Y for $G = 1$ at $T = 1$. Estimators of functionals of the distributions of potential outcomes such as quantile functions and effects can be also constructed using the plug-in principle.

As in Wooldridge, we can also sestimate the parameters via Quasi-Maximum Likelihood Estimation. Again, this is most easily explained with no regressors. For this, define the statistic $\theta(y)$ by

$$\begin{aligned}\theta(y) &:= \Lambda^{-1}\left(\mathbb{P}(Y_1 \leq y|G=1,T=1)\right) - \Lambda^{-1}\left(\mathbb{P}(Y_0 \leq y|G=1,T=1)\right) \\ &= \Lambda^{-1}\left(\mathbb{P}(Y_1 \leq y|G=1,T=1)\right) - \alpha(y) - \beta(y) - \gamma(y)\end{aligned}$$

Based on this, we have that the Distributional Treatment Effect (DTE) equals

$$\begin{aligned}\tau_y &:= F_{Y_1\,|\,G,T}(y\,|\,1,1) - F_{Y_0\,|\,G,T}(y\,|\,1,1) \\ &= \Lambda\left(\alpha(y) + \beta(y) + \gamma(y) + \theta(y)\right) - \Lambda\left(\alpha(y) + \beta(y) + \gamma(y)\right)\end{aligned}$$

Moreover we have that:

$$\begin{aligned}
\alpha(y) &= \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right)\\
\beta(y) &= \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,1)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right)\\
\gamma(y) &= \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,1,0)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right)\\
\theta(y) &= \Lambda^{-1}\left(F_{Y_1\,|\,G,T}(y\,|\,1,1)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,1,0)\right)\\
&\quad - \left[\Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,1)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\right)\right]
\end{aligned}$$

We can estimate the probabilities in parentheses by the fractions in the dataset

$$\widehat{F}_{Y_0\,|\,G,T}(y\,|\,0,0) = \frac{1}{N_{gt}} \sum_{i=1,\dots,N;G_i=g,T_i=t} \mathbf{1}(Y_i \le y)$$

And we obtain an estimator of τ_y by

$$\widehat{\tau}_y = \Lambda\left(\widehat{\alpha}(y) + \widehat{\beta}(y) + \widehat{\gamma}(y) + \widehat{\theta}(y)\right) - \Lambda\left(\widehat{\alpha}(y) + \widehat{\beta}(y) + \widehat{\gamma}(y)\right) \qquad (8)$$

Note that the same estimators can be derived by running a distribution regression of $\mathbf{1}(Y_i \le y)$ on T_i, G_i and D_i. Note that this implies that we assume the following relationship for $F_{Y|G,T}(y|g,t)$

$$F_{Y|G,T}(y|g,t) = \Lambda\left(\alpha(y) + \beta(y)t + \gamma(y)g + \theta(y)gt\right)$$

Hence, we obtain the following algorithm

Algorithm 2. 1. *Estimate the parameters $\alpha(y)$, $\beta(y)$, $\gamma(y)$ and the statistic $\theta(y)$ by distribution regression of the indicator $1(Y \le y)$ on a constant, T and G for multiple values of $y \in \mathcal{Y}$, using all the observations.*

2. *Estimate $F_{Y_1\,|\,G,T}(y\,|\,1,1)$ and $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ by the plug-in estimators*

$$\begin{aligned}
\widehat{F}_{Y_1\,|\,G,T}(y\,|\,1,1) &= \Lambda\left(\widehat{\alpha}(y) + \widehat{\beta}(y) + \widehat{\gamma}(y) + \widehat{\theta}(y)\right)\\
\widehat{F}_{Y_0\,|\,G,T}(y\,|\,1,1) &= \Lambda\left(\widehat{\alpha}(y) + \widehat{\beta}(y) + \widehat{\gamma}(y)\right)
\end{aligned}$$

The DFE τ_y can be estimated using the plug-in estimator defined in (8).

Note that both algorithms do not necessarily result in numerically identical estimates. An exception arises when we use the identity function or the standard logistic function for the link function Λ for which cases we obtain results that are identical.

Algorithm 1 can be easily extended by using equation (6) instead of (1). The extension of Algorithm 2 is somewhat more involved. For this, define the statistic $\theta(y,x)$ as

$$\begin{aligned}\theta(y,x) &= \Lambda^{-1}\left(F_{Y_1\,|\,G,T,X}(y\,|\,g,t,x)\right) - \Lambda^{-1}\left(F_{Y_0\,|\,G,T,X}(y\,|\,g,t,x)\right)\\ &= \Lambda^{-1}\left(F_{Y_1\,|\,G,T,X}(y\,|\,g,t,x)\right) - \alpha(y) - \beta(y) - \gamma(y) - \pi(y)x.\end{aligned}$$

Hence the DTE conditional on x can be defined by

$$\begin{aligned}\tau_y &= F_{Y_1\,|\,G,T,X}(y\,|\,1,1,x) - F_{Y_0\,|\,G,T,X}(y\,|\,1,1,x)\\ &= \Lambda\left(\alpha(y) + \beta(y) + \gamma(y) + \pi(y)x + \theta(y,x)\right) - \Lambda\left(\alpha(y) + \beta(y) + \pi(y)x + \gamma(y)\right).\end{aligned}$$

As above, we can define a quasi-distribution regression for $F_{Y|G,T}(y|g,t)$ by assuming

$$F_{Y|G,T,X}(y|g,t,x) = \Lambda\left(\alpha(y) + \beta(y)t + \gamma(y)g + \pi x + \theta(y,x)gt\right).$$

where the statistic $\theta(y,x)$ is a function of x, for example

$$\theta(y,x) = \theta_0 + \theta_1 x.$$

Hence, algorithm 2 can be trivially extended by this assumption. Again, both algorithms will not produce numerically identical estimates unless the link function is the identity function or the standard logistic function.

5 Empirical application

5.1 Description of the original empirical exercise

Malesky et al. (2014) investigate the impact of recentralization by looking at a case study in Vietnam. Because of dissatisfaction of the measures of decentralization taken in the early 1990s, Vietnam decided to change their political system in 2007. In particular, they decided to take out one political layer from the decision making process. That is, Vietnam has four layers of the political process: the central government, the provinces (63 in total), the districts (696 in total), and the communes (more than 11,000 in total).[4] The idea was to abolish the political process at the districts (which are governed by the

[4] The total population of Vietnam was 84.76 million in 2007.

so called Districts People Council or DPC). Instead of introducing this change in the system immediately, the Vietnam government decided to first apply an experiment for only ten provinces (with 99 districts). Malesky et al. (2014) use this experiment for their empirical analysis. Note that the experiment was not random, but was decided by the central government to be stratified based on regions and subregions as well as on rural versus urban areas and by socioeconomic and public administration performance of the provinces. The decision to start this experiment was made in 2008 and the abolishment of the DPC in the treatment districts started in 2009.

Malesky et al. (2014) use the following specification for their analysis:

$$Y_{it} = \alpha + \beta T_t + \gamma G_i + \theta G_i T_t + X_{it}\pi + U_{it}$$

where Y_{it} is the outcome variable for period t of commune i. T_t is a dummy variable that equals one in the treated period while G_i is a dummy variable that equals one in the case that commune i belongs to a treated district. Finally, X_{it} is a set of control variables for commune i and in period t. Malesky et al. (2014) use the log surface area of the commune, the log of the commune population density, whether the commune belongs to a national level city and they use region dummies (8 regions in total). For reasons of data availability, Malesky et al. (2014) only use rural communes and they use two years of observation: 2008 and 2010 (they use 2006 for robustness checks). They use 30 different outcome variables to investigate the impact of the abolishment of the political layer which can be subdivided into 6 categories: Infrastructure index, agricultural services index, health services index, education index, communications index, and household business development index. Due to the fact that most variables are dummy variables, we can only use eight of their original outcome variables: (1) proportion of households supported crop, (2) proportion of households supported agricultural extension, (3) proportion of households supported agriculture tax exemption, (4) the number of visits of agricultural extension staff, (5) proportion of households supported healthcare fee, (6) proportion of households supported tuition fee, (7) proportion of households supported credit, and (8) proportion of households supported business tax exemption. The first four are based on the category agricultural services index, the fifth is related to the category health services index, the sixth related to the education index and the last two are related to the household business development index.

5.2 Our analysis for the univariate case

For our empirical analysis, we use the following specification for $F_{Y_{0,i}|G_i,T_i,X_{it}}(\cdot|g,t,x)$

$$F_{Y_{0,i}|G_i,T_i,X_{it}}(y|g_t,t_t,x_{it}) = \Lambda(\alpha(y)+\beta(y)t_t+\gamma(y)g_i+x_{it}\pi(y))$$

where we use the same control variables as in Malesky et al. (2014). Hence, we can estimate the counterfactual distribution using

$$\widehat{F}_{Y_{0,i}|G_i,T_i,X_{it}}(y|1,1,x_{i,1}) = \Lambda(\widehat{\alpha}(y)+\widehat{\beta}(y)+\widehat{\gamma}(y)+x_{i1}\widehat{\pi}(y))$$

where $\widehat{\alpha}(y)$, $\widehat{\beta}(y)$, $\widehat{\gamma}(y)$, and $\widehat{\pi}(y)$ are estimated by using distribution regression at y. We can estimate the unconditional distribution using

$$F_{Y_{0,i}|G_i,T_i}(y|1,1) = \int_{\mathcal{X}(1,1)} F_{Y_{0,i}|G_i,T_i,X_{it}}(y|1,1,x_{i1})dF_{X_{it}|G,T}(x|1,1)$$

Hence, our estimator becomes

$$\widehat{F}_{Y_{0,i}|G_i,T_i}(y|1,1) = \frac{1}{N_{11}}\sum_{i:G_i=1,T_i=1}\widehat{F}_{Y_{0,i,t}|G_i,T_i,X_{i1}}(y|1,1,X_{i,1})$$

where N_{11} is the total number of observations for which $G_i=1, T_i=1$. We can estimate the quantile treatment effects by inverting the estimated distributions of $F_{Y_{0,i,t}|G_i,T_i}(y|1,1)$ and $F_{Y_{1,i,t}|G_i,T_i}(y|1,1)$, where we estimate $F_{Y_{1,i,t}|G_i,T_i}(y|1,1)$ by using the empirical distribution. In particular we use

$$\widehat{F}^{-1}_{Y_{j,i,t}|G_i,T_i}(q|1,1) = \inf\{y : \widehat{F}_{Y_{j,i,t}|G_i,T_i}(y|1,1) \leq q\} \quad j=0,1$$

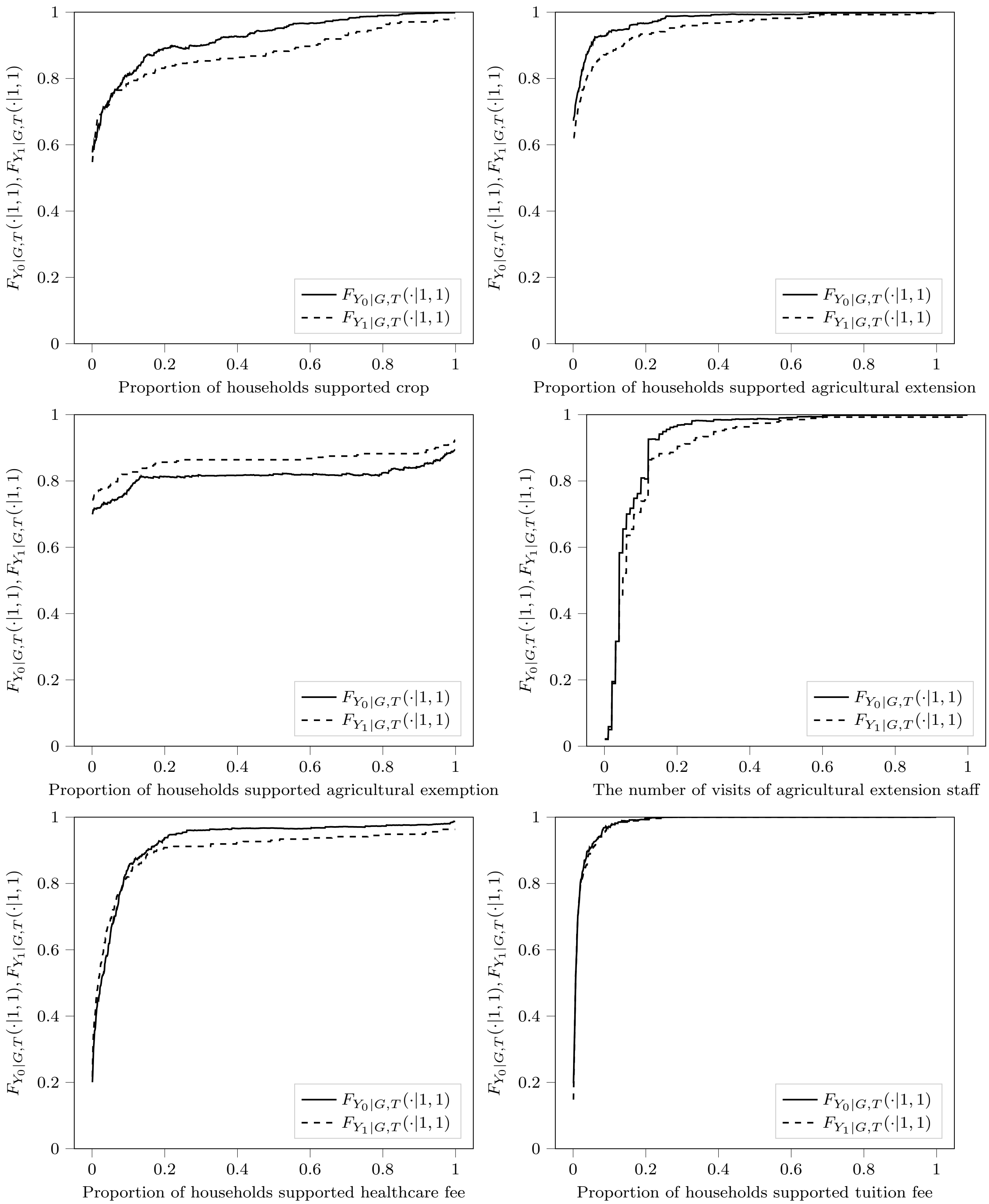
Results of our empirical exercise are in Figure 1. The quantile treatment effects are listed in Table 1. We estimate the quantile treatment effects by simply inverting the estimated distribution functions. As in Malesky et al. (2014), we correct the confidence intervals for clustering at the province level. That is, we use the Bayesian bootstrap and draw the same exponential weight for all observations that belong to the same province (see also Chernozukov et al., 2020). In order to construct the confidence bounds by following steps 1-4 of Algorithm 1 of Chernozhukov et al. (2020) but we use directly the quantile treatment effects rather than the estimated distributions. Note that this is allowed as long as we assume our outcome variable to be continuously distributed. For some of the variables, there is a lot of bunching at zero. For example, for the variable "Proportion of households supported crop" 49.93 percent of the observations equal zero and it equals 54.04 percent for the treatment group in the treatment period while it equals 50.79 for the control group. This implies that the quantile treatment effect is by definition equal to zero up and until the median and as such the impact can only come from the higher quantiles. To distinguish these cases from the cases in which there was a real zero impact, we use dots at these places.

Nevertheless, even if we abstract from the impact of these zeros, both Figure 1 and Table 1 show that there is a lot of treatment heterogeneity and that the mean impacts published in Malesky et al. (2014) are mainly a result of impacts at the top of the distribution. This is most clear from the outcome variable "The number of visits of agricultural extension staff" which has only a substantial impact at Q3 and D9. To some extent, this is also true for the negative value of the variable "Proportion of households supported credit".

Malesky et al. (2014) also use additional data for the year 2006 to check the common trends assumption. They perform this check by simply looking at the results of a DiD design for the periods 2006 and 2008 making 2008 the placebo treatment period. That is, one would not expect any treatment effect in the period before the introduction of the treatment. Results of this robustness check are in Figure 2 and Table 2. As in Malesky et al. (2014), we do find some substantial differences in Figure 2 between the distribution of Y_0 and Y_1. However, as is shown in Table 2 these differences are not significant and they are generally in the opposite direction as found in the original results. For example, for the proportion of households supported crop (the first figure in Figures 1 and 2), the distribution of $F_{Y_1|G,T}(\cdot|1,1)$ is in Figure 2 generally to the left of the distribution of $F_{Y_1|G,T}(\cdot|1,1)$ while the relationship is opposite in Figure 1.

Note that in a standard linear DiD design, checking the common trend assumption as presented above is identical to checking the value of the interaction term in the period(s)

Figure 1: Results of the empirical application.



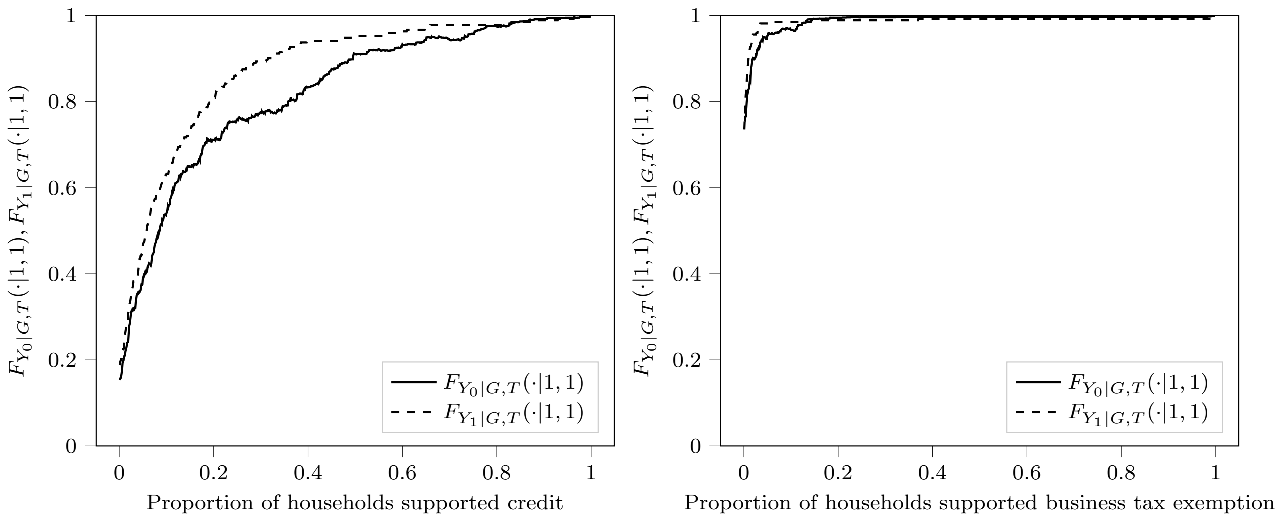
	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	0.0514	·	·	·	-0.003	0.3074
crop	(0.011,0.091)	(·, ·)	(·, ·)	(·, ·)	(-0.096,0.09)	(-0.005,0.62)
Proportion of households supported	0.0214	·	·	·	0.0118	0.0876
agricultural extension	(0.005,0.037)	(·, ·)	(·, ·)	(·, ·)	(-0.006,0.03)	(0.001,0.175)
Proportion of households supported	-0.0427	·	·	·	-0.0687	·
agricultural exemption	(-0.099,0.013)	(·, ·)	(·, ·)	(·, ·)	(-0.79,0.653)	(·, ·)
The number of visits of agricultural	0.0203	-0.0001	0.0	0.01	0.0299	0.08
extension staff	(-0.001,0.042)	(-0.0,0.0)	(-0.009,0.009)	(0.0,0.02)	(-0.009,0.069)	(0.0,0.16)
Proportion of households supported	0.0389	·	-0.001	-0.0077	-0.0073	0.026
healthcare fee	(0.008,0.07)	(·, ·)	(-0.003,0.001)	(-0.019,0.003)	(-0.042,0.027)	(-0.348,0.4)
Proportion of households supported	0.0016	·	0.0005	-0.0001	-0.0001	0.0063
tuition fee	(-0.002,0.005)	(·, ·)	(-0.001,0.002)	(-0.002,0.002)	(-0.006,0.006)	(-0.008,0.021)
Proportion of households supported	-0.0535	·	-0.008	-0.0276	-0.0709	-0.1804
credit	(-0.073,-0.034)	(·, ·)	(-0.015,-0.001)	(-0.045,-0.01)	(-0.204,0.062)	(-0.269,-0.092)
Proportion of households supported	0.0012	·	·	·	·	-0.0143
business tax exemption	(-0.009,0.011)	(·, ·)	(·, ·)	(·, ·)	(·, ·)	(-0.029,-0.0)

Table 1: Quantile treatment effects with 90 percent confidence intervals based on Bayesian weights. Confidence intervals correct for clustering at the province level.

	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	-0.0164	·	·	·	-0.0569	-0.0246
crop	(-0.077,0.044)	(·, ·)	(·, ·)	(·, ·)	(-0.119,0.006)	(-0.684,0.635)
Proportion of households supported	-0.0007	·	·	·	·	0.0117
agricultural extension	(-0.012,0.01)	(·, ·)	(·, ·)	(·, ·)	(·, ·)	(-0.014,0.038)
Proportion of households supported	0.0296	·	·	·	-0.0052	·
agricultural exemption	(-0.041,0.1)	(·, ·)	(·, ·)	(·, ·)	(-0.577,0.566)	(·, ·)
The number of visits of agricultural	-0.0002	·	·	·	-0.0001	-0.0001
extension staff	(-0.026,0.026)	(·, ·)	(·, ·)	(·, ·)	(-0.02,0.02)	(-0.08,0.079)
Proportion of households supported	-0.0038	·	·	0.0079	-0.0221	-0.0241
healthcare fee	(-0.037,0.029)	(·, ·)	(·, ·)	(-0.004,0.02)	(-0.043,-0.001)	(-0.135,0.087)
Proportion of households supported	0.0012	·	0.0002	0.0018	0.0033	0.006
tuition fee	(-0.003,0.006)	(·, ·)	(-0.001,0.001)	(-0.0,0.004)	(-0.001,0.008)	(-0.017,0.028)
Proportion of households supported	0.0053	·	-0.0011	-0.0004	-0.0013	-0.0239
credit	(-0.014,0.025)	(·, ·)	(-0.014,0.012)	(-0.023,0.022)	(-0.059,0.056)	(-0.149,0.101)
Proportion of households supported	-0.0058	·	·	·	0.002	0.0142
business tax exemption	(-0.021,0.009)	(·, ·)	(·, ·)	(·, ·)	(-0.001,0.005)	(-0.004,0.032)

Table 2: Quantile treatment effects – robustness check for 2006 and 2008 – parallel trends.

Figure 1: Results of the empirical application (continued).



before the treatment. That is not true in our non-linear design, but we can still perform an additional check looking at the coefficient value of the interaction term. That is, we can estimate the general representation as presented in (1) for all observations in the periods 2006 and 2008. Results of this exercise are presented in Figure 3. Generally, we find that $\delta(y)$ is not significantly different from zero but there are some regions in the distribution of some of the outcome variables where there is a significant difference. For example, for the outcome variable the "Proportion of households supported agricultural exemption", we find a significant difference in between 0.3 and 0.9 of the outcome values.

As a further robustness check, we also interacted the covariates with the time and treatment dummy variables. We interact regions with time but we cannot interact regions with treatment as this will result in perfect multicollinearity due to the setup of the program. The results of our exercise are in Figure 4 and the quantile treatment effects are reported in 5.

5.3 Comparison with the changes-in-changes estimation

For the changes-in-changes estimation, we note as in Athey and Imbens (2006) that the distribution of $Y_0|G = 1, T = 1$ equals the distribution of $\varphi(Y_0|G = 1, T = 0)$. Hence,

Figure 2: Results of robustness check using the years 2006 and 2008.

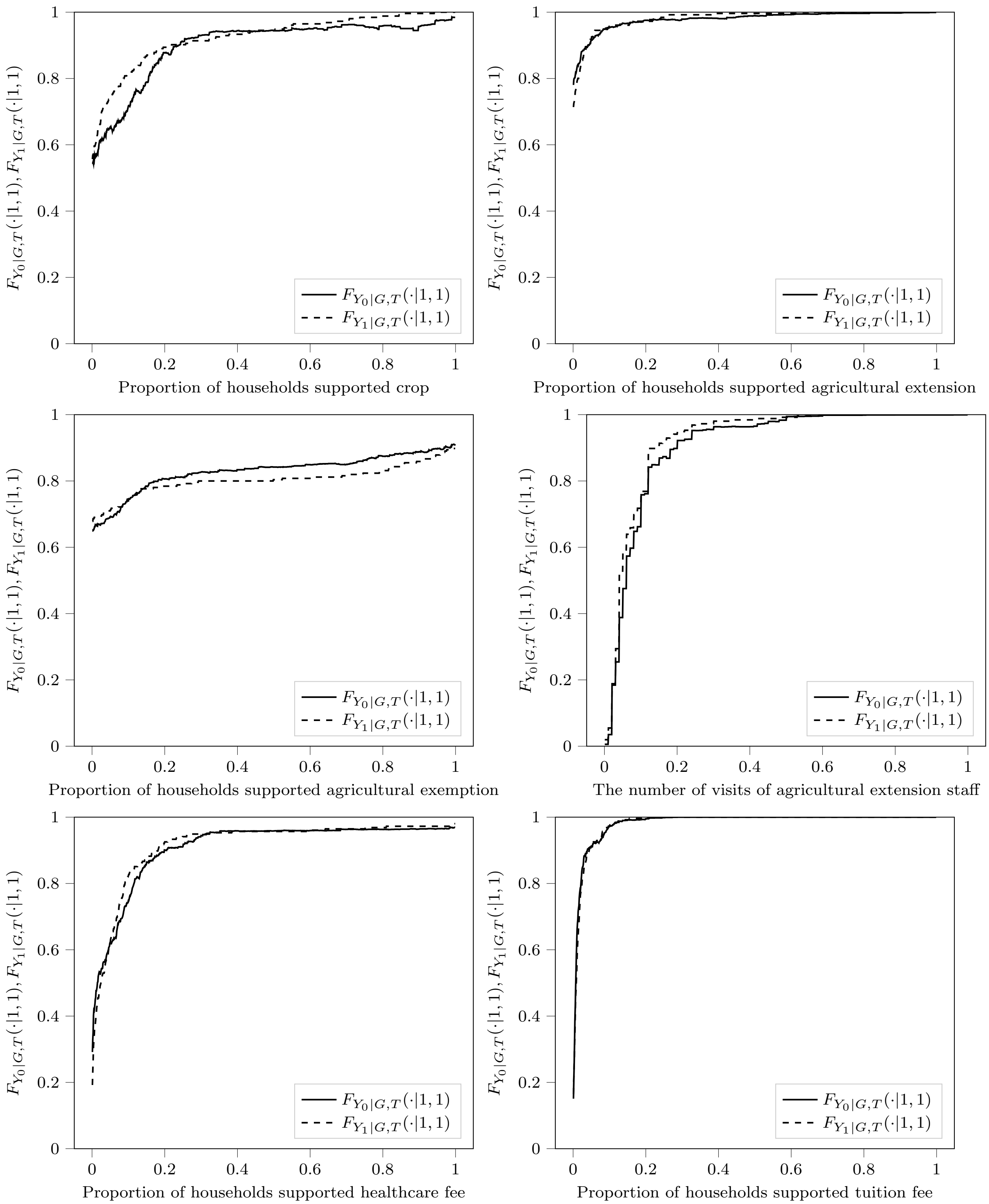
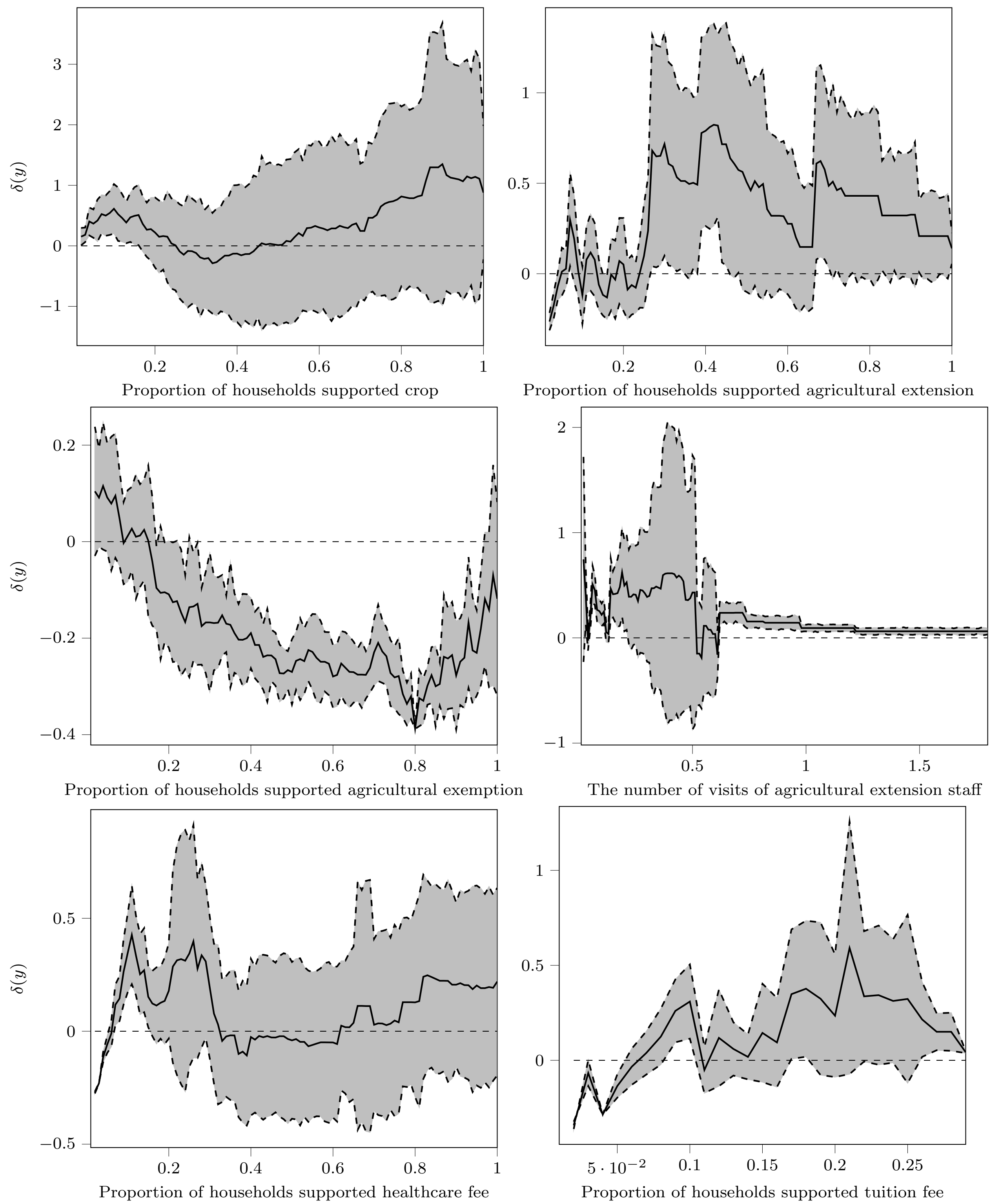


Figure 3: Results of the robustness check to investigate whether $\delta(y)$ of equation (1) equals zero in the period before the treatment.



	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	0.0465	·	·	·	-0.009	0.3068
crop	(-0.002,0.095)	(·, ·)	(·, ·)	(·, ·)	(-0.119,0.101)	(0.005,0.608)
Proportion of households supported	0.0204	·	·	·	0.0047	0.0832
agricultural extension	(0.002,0.039)	(·, ·)	(·, ·)	(·, ·)	(-0.014,0.024)	(-0.004,0.171)
Proportion of households supported	-0.0461	·	·	·	-0.0664	-0.0528
agricultural exemption	(-0.102,0.01)	(·, ·)	(·, ·)	(·, ·)	(-0.748,0.615)	(-0.789,0.683)
The number of visits of agricultural	0.0197	-0.0001	-0.0002	0.0098	0.0194	0.0507
extension staff	(-0.002,0.041)	(-0.0,0.0)	(-0.01,0.009)	(0.0,0.02)	(-0.011,0.05)	(-0.028,0.13)
Proportion of households supported	0.0331	·	-0.0009	-0.0049	-0.0071	0.0237
healthcare fee	(0.0,0.066)	(·, ·)	(-0.002,0.0)	(-0.016,0.006)	(-0.039,0.025)	(-0.419,0.466)
Proportion of households supported	0.001	·	0.0004	-0.0005	-0.0014	0.0036
tuition fee	(-0.003,0.005)	(·, ·)	(-0.001,0.002)	(-0.003,0.002)	(-0.009,0.006)	(-0.009,0.017)
Proportion of households supported	-0.0574	·	-0.0089	-0.028	-0.0716	-0.1808
credit	(-0.079,-0.036)	(·, ·)	(-0.019,0.001)	(-0.05,-0.006)	(-0.2,0.057)	(-0.307,-0.054)
Proportion of households supported	0.0008	·	·	·	·	-0.0153
business tax exemption	(-0.008,0.01)	(·, ·)	(·, ·)	(·, ·)	(·, ·)	(-0.028,-0.002)

Table 3: Quantile treatment effects without using additional control variables in the analysis.

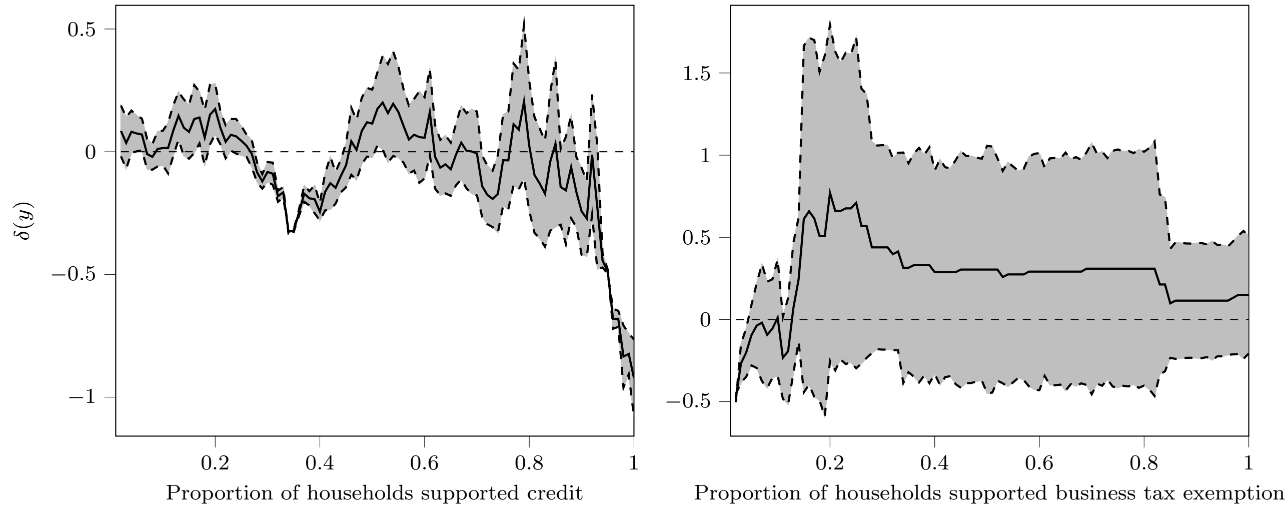
	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	-0.0162	·	·	·	-0.0597	-0.0206
crop	(-0.084,0.052)	(·,·)	(·,·)	(·,·)	(-0.152,0.033)	(-0.762,0.72)
Proportion of households supported	-0.0041	·	·	·	·	0.0007
agricultural extension	(-0.015,0.007)	(·,·)	(·,·)	(·,·)	(·,·)	(-0.032,0.033)
Proportion of households supported	0.0248	·	·	·	-0.0021	·
agricultural exemption	(-0.058,0.107)	(·,·)	(·,·)	(·,·)	(-0.61,0.606)	(·,·)
The number of visits of agricultural	-0.0163	-0.0001	-0.0003	-0.0193	-0.0003	-0.05
extension staff	(-0.032,-0.001)	(-0.0,0.0)	(-0.01,0.009)	(-0.029,-0.009)	(-0.02,0.019)	(-0.119,0.019)
Proportion of households supported	-0.0072	·	·	0.0094	-0.0212	-0.0236
healthcare fee	(-0.045,0.031)	(·,·)	(·,·)	(-0.011,0.029)	(-0.051,0.009)	(-0.133,0.085)
Proportion of households supported	0.0008	·	-0.0001	0.0014	0.0026	0.0051
tuition fee	(-0.005,0.007)	(·,·)	(-0.002,0.001)	(-0.001,0.003)	(-0.003,0.008)	(-0.022,0.032)
Proportion of households supported	0.0008	·	-0.0055	-0.0005	-0.0024	-0.0307
credit	(-0.025,0.026)	(·,·)	(-0.021,0.01)	(-0.026,0.025)	(-0.072,0.067)	(-0.168,0.107)
Proportion of households supported	-0.0061	·	·	·	0.0019	0.014
business tax exemption	(-0.024,0.012)	(·,·)	(·,·)	(·,·)	(-0.002,0.006)	(-0.009,0.037)

Table 4: Quantile treatment effects without using additional control variables in the analysis – robustness check for 2006 and 2008 – parallel trends.

	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	0.0451	·	·	·	0.0042	0.3457
crop	(0.003,0.087)	(·,·)	(·,·)	(·,·)	(-0.092,0.101)	(0.022,0.669)
Proportion of households supported	0.0233	·	·	·	0.0122	0.0912
agricultural extension	(0.007,0.04)	(·,·)	(·,·)	(·,·)	(-0.004,0.028)	(-0.011,0.194)
Proportion of households supported	-0.0554	·	·	·	-0.0911	·
agricultural exemption	(-0.116,0.005)	(·,·)	(·,·)	(·,·)	(-0.847,0.665)	(·,·)
The number of visits of agricultural	0.02	-0.0001	-0.0001	0.0099	0.0193	0.0798
extension staff	(-0.008,0.048)	(-0.0,0.0)	(-0.009,0.009)	(0.0,0.02)	(-0.02,0.058)	(0.0,0.159)
Proportion of households supported	0.0276	·	-0.0002	-0.0017	0.0005	0.0469
healthcare fee	(-0.001,0.056)	(·,·)	(-0.002,0.001)	(-0.013,0.01)	(-0.034,0.035)	(-0.453,0.546)
Proportion of households supported	0.0003	·	0.0002	-0.0004	-0.0016	0.0038
tuition fee	(-0.004,0.004)	(·,·)	(-0.001,0.001)	(-0.004,0.003)	(-0.01,0.006)	(-0.01,0.018)
Proportion of households supported	-0.0484	·	-0.0077	-0.0271	-0.0702	-0.1806
credit	(-0.079,-0.018)	(·,·)	(-0.022,0.007)	(-0.053,-0.001)	(-0.215,0.075)	(-0.328,-0.033)
Proportion of households supported	-0.0016	·	·	·	·	-0.0152
business tax exemption	(-0.012,0.009)	(·,·)	(·,·)	(·,·)	(·,·)	(-0.032,0.001)

Table 5: Quantile treatment effects using interaction terms between the covariates and the time and treatment dummy variables.

Figure 3: Results of the robustness check to investigate whether $\delta(y)$ of equation (1) equals zero in the period before the treatment.(continued).



we can obtain an estimator of the distribution of $Y_0|G=1,T=1$ by using the empirical distribution function of the random variable

$$\mathbb{Q}_{\widehat{F}_{Y_0|G,T}(Y_0|G=1,T=0|0,1)}(Y_0|G=0,T=0).$$

Hence, we can estimate the distribution function of $F_{Y_0|G=1,T=1}$ in point y for our changes-in-changes estimator using the following steps:

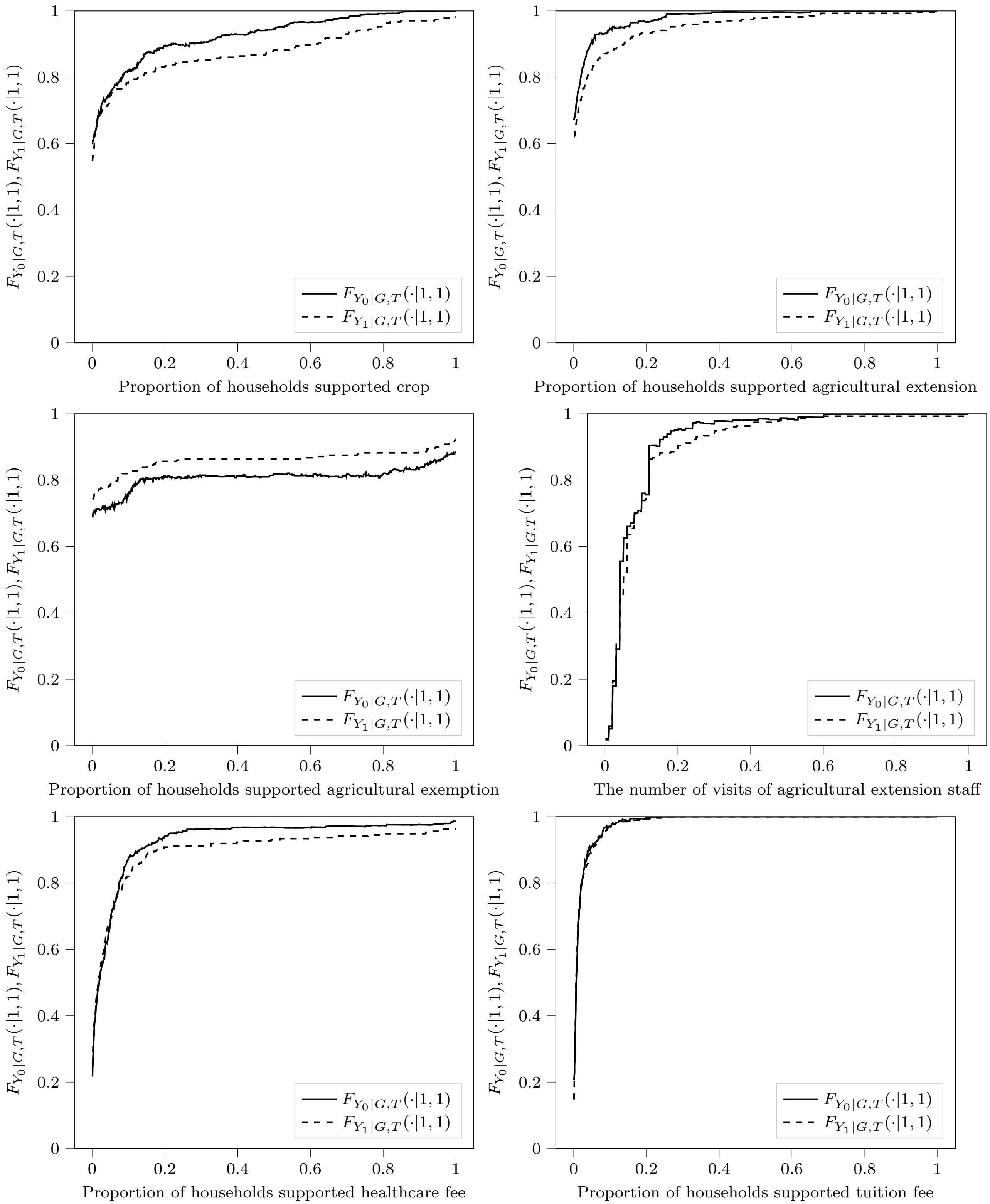
1. For every observation of Y_0 of the subsample of $G=1,T=0$ estimate the empirical distribution function of the subsample for which $G=0,T=0$.

2. For every computed empirical distribution function of step 1 estimate the corresponding quantile of the subsample for which $G=0,T=1$.

3. For all the obtained quantiles from step 2, compute the empirical distribution function in y.

The distribution of $F_{Y_0|G=1,T=1}$ can be estimated using the empirical distribution function. One can obtain the quantile treatment effect by inverting the distribution at desired levels of the distribution.

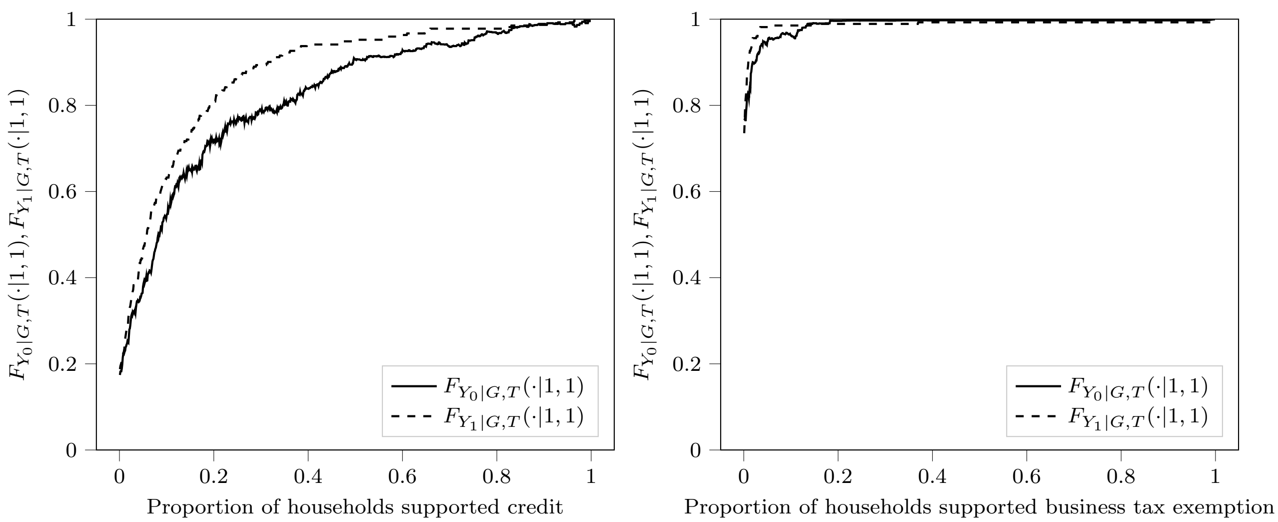
Figure 4: Results of the empirical application using interaction terms between the covariates and the time and treatment dummy variable.



	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	0.0484	·	·	·	-0.0087	0.3863
crop	(0.003,0.093)	(·,·)	(·,·)	(·,·)	(-0.114,0.097)	(0.001,0.772)
Proportion of households supported	0.0115	·	·	·	0.0025	0.0763
agricultural extension	(-0.008,0.031)	(·,·)	(·,·)	(·,·)	(-0.019,0.024)	(-0.011,0.164)
Proportion of households supported	-0.0489	·	·	·	-0.0648	·
agricultural exemption	(-0.107,0.009)	(·,·)	(·,·)	(·,·)	(-0.697,0.567)	(·,·)
The number of visits of agricultural	0.0145	-0.01	-0.0097	0.0098	0.0064	0.0497
extension staff	(-0.01,0.039)	(-0.02,-0.0)	(-0.019,-0.0)	(-0.002,0.021)	(-0.021,0.034)	(-0.02,0.12)
Proportion of households supported	0.0286	·	-0.0009	-0.0066	0.0007	0.0353
healthcare fee	(-0.001,0.059)	(·,·)	(-0.003,0.001)	(-0.017,0.004)	(-0.025,0.026)	(-0.417,0.488)
Proportion of households supported	0.0013	·	0.0005	-0.0003	0.0006	0.0045
tuition fee	(-0.002,0.005)	(·,·)	(-0.001,0.002)	(-0.003,0.002)	(-0.007,0.008)	(-0.008,0.017)
Proportion of households supported	-0.0559	·	-0.0076	-0.0275	-0.0892	-0.1857
credit	(-0.08,-0.032)	(·,·)	(-0.018,0.003)	(-0.05,-0.005)	(-0.206,0.028)	(-0.298,-0.074)
Proportion of households supported	-0.0004	·	·	·	-0.0011	-0.0158
business tax exemption	(-0.008,0.007)	(·,·)	(·,·)	(·,·)	(-0.007,0.005)	(-0.029,-0.002)

Table 6: Quantile treatment using changes-in-changes.

Figure 4: Results of the empirical application using interaction terms between the covariates and the time and treatment dummy variable (continued).



5.4 Results for the bivariate case

For our bivariate analysis we only look at the outcomes of the variables "Proportion of households supported credit" and the "Proportion of households supported healthcare fee". The outcomes of these variables have relatively little bunching at integer values which makes them more interesting for our empirical analysis.

We present the results of the counterfactual and the actual distribution in Figure 7. From this figure, it is possible to see that the joint distribution has changed due to the treatment and that the distribution of the treated population has shifted to the upper-left corner. However, it is difficult to see whether this is not merely a result of the changes in the marginal distributions. Therefore, we also present results using Kendall's tau which equals

$$\tau = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{sgn}(x_i - x_j)\text{sgn}(y_i - y_j)$$

We can directly calculate the Kendall's tau for the joint distribution of the treated sample in the second period when treated as this is observed from the data. For the counterfactual distribution of the treated sample in the second period when not treated, we first sample from the estimated distribution. That is, we sample a value of Y using our estimator

Figure 5: Results without using additional control variables in the analysis.

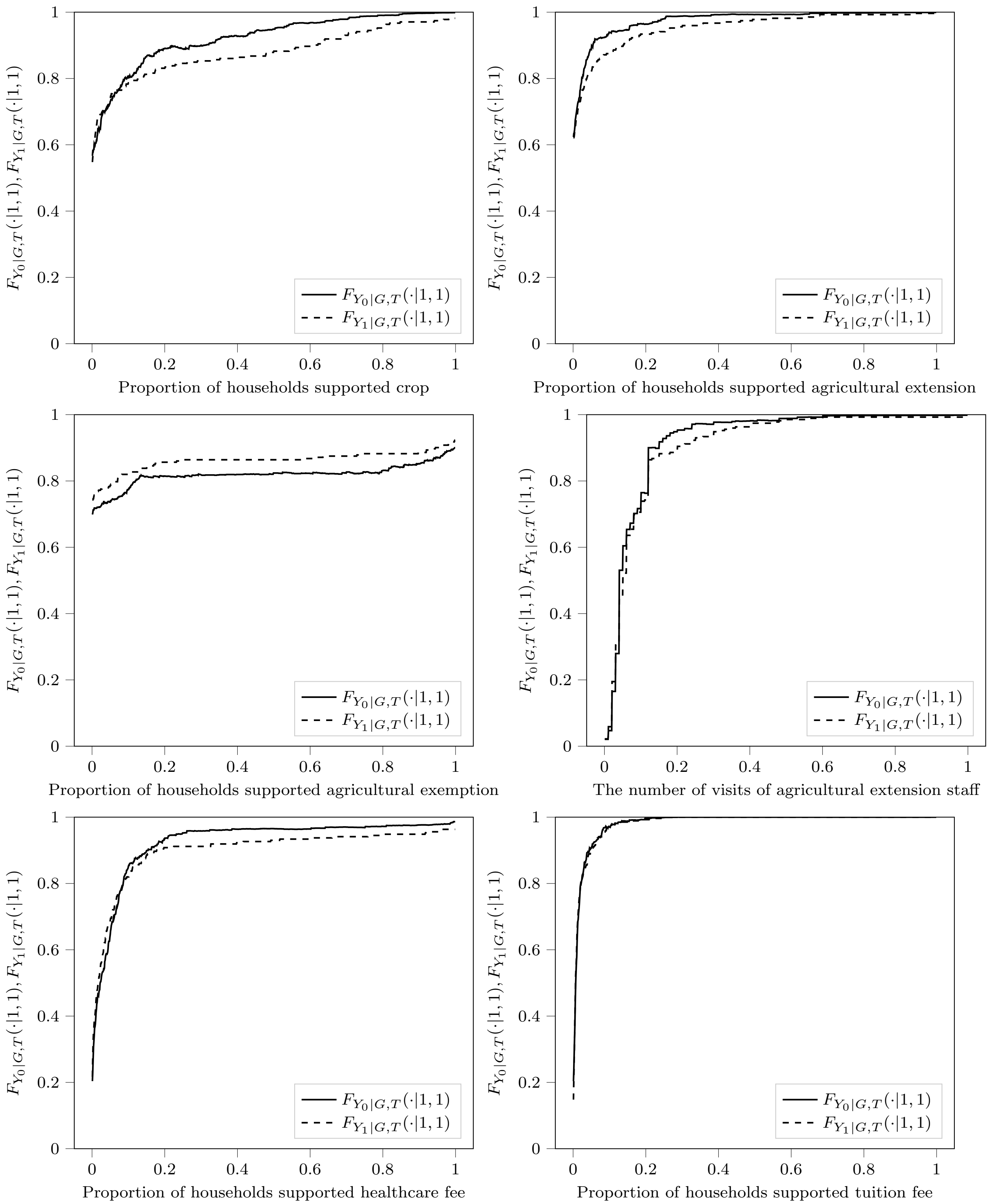
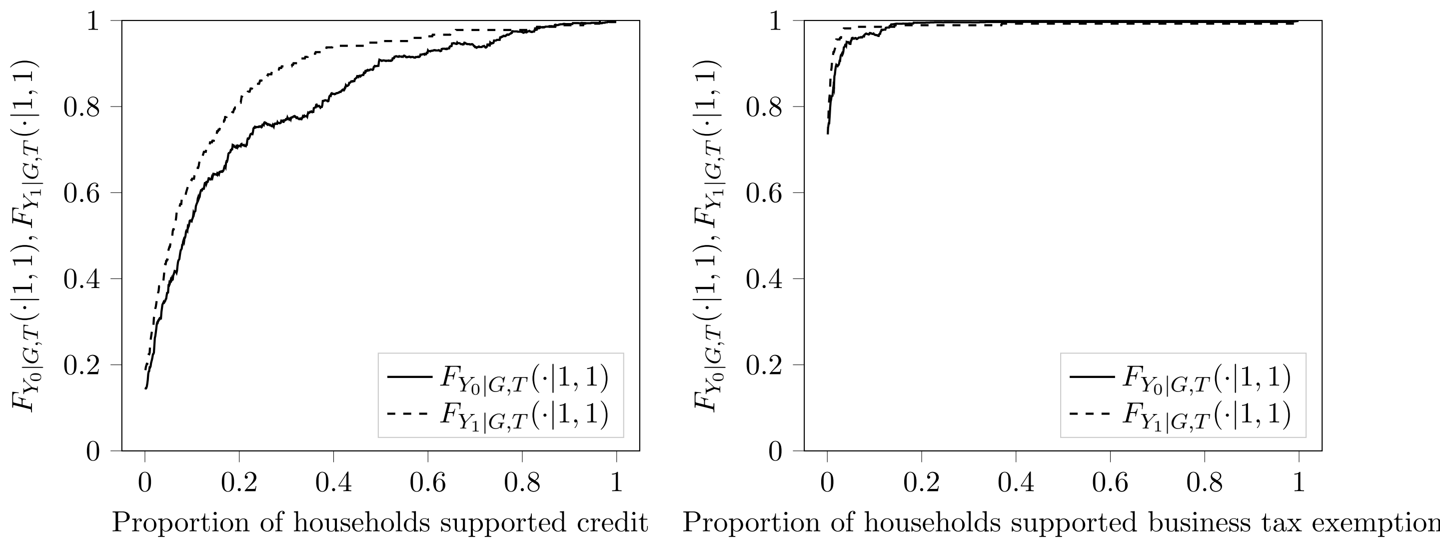


Figure 5: Results without using additional control variables in the analysis. (continued).



of its marginal distribution described for the univariate case above. Then, we sample Z from the conditional distribution of $Z|Y$ which can be obtained using our estimates for the bivariate case. We find that the Kendall's tau based on this procedure gives a value of 0.1253 with a 95-percent confidence interval from 0.0989 to 0.1518. This implies that there is a positive correlation between the two outcomes in the districts. For the observed distribution of the treated group we find a value equal to 0.2463 with a 95-percent confidence interval from 0.2224 to 0.2703. This implies that the correlation between the two outcomes has increased significantly due to the treatment.

6 Conclusion

We provide a relatively simple distribution regression based estimator to implement the evaluation of treatment effects in a difference-in-difference setting. As our approach provides counterfactual distributions we are able to explore the impact of the treatment at different quantiles of the distribution of the outcome variable. For both the univariate and multivariate cases we provide the identifying assumption and the associated estimation algorithms. We provide an empirical example which revisits an existing study. Our

Figure 6: Results of changes-in-changes

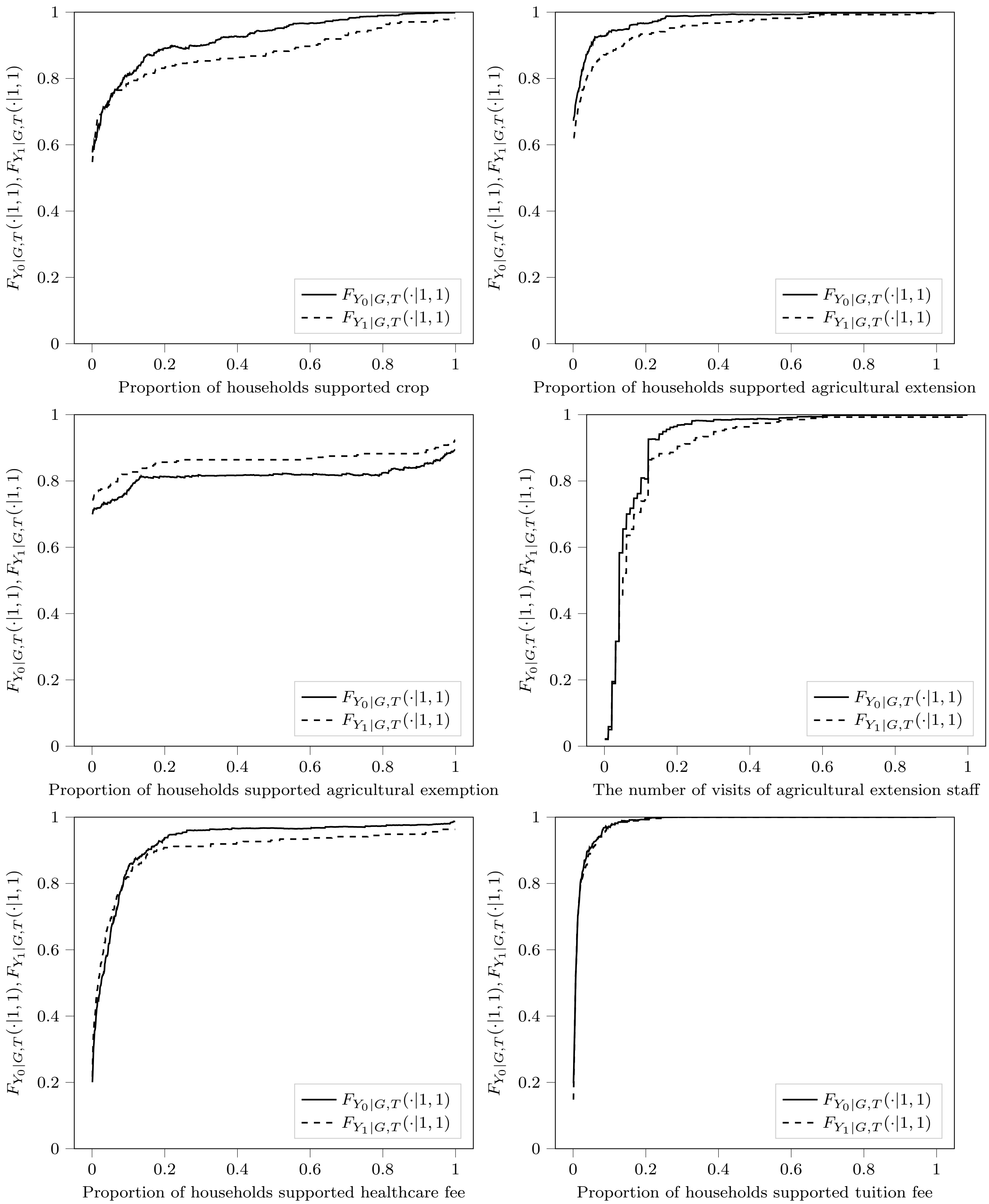
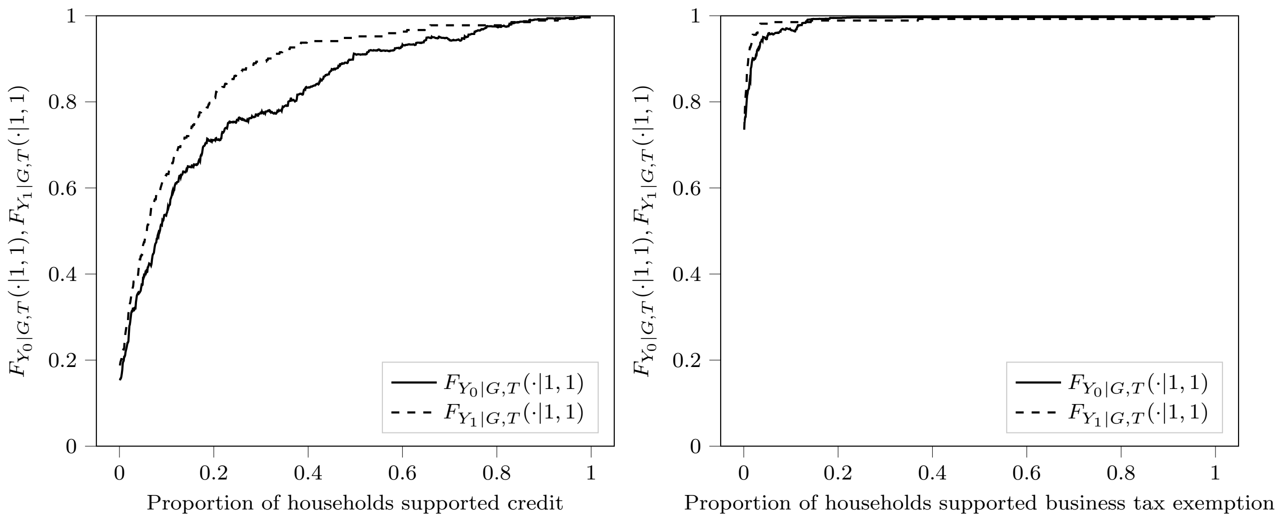


Figure 6: Results of changes-in-changes (continued).



empirical analysis highlights the utility of various aspects of our approach.

Our analysis can easily be extended to the case of multiple time periods and more than two outcomes. We can also extend our distributional regression framework to use time and unit weights as in the synthetic difference-in-difference estimation method of Arkhangelsky et al. (2021). We leave these extensions to future research (e.g. Fernández-Val et al., 2024).

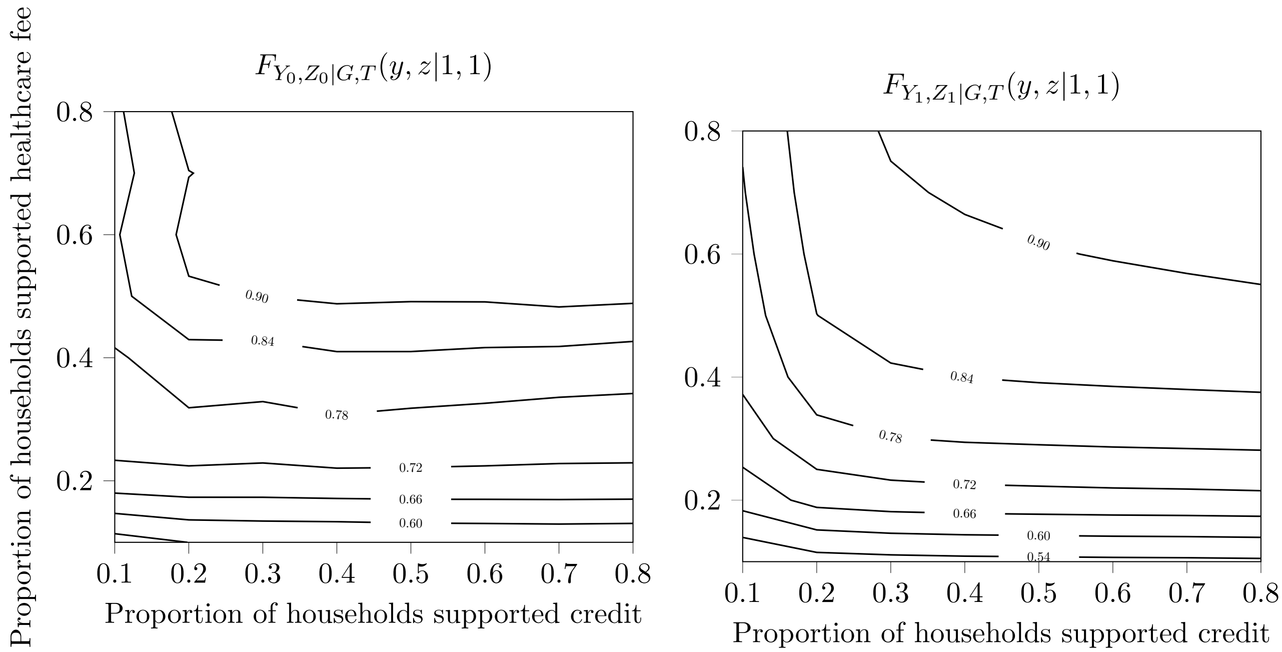
References

Almond, D., H.W. Hoynes, and D.W. Schanzenbach (2011), "Inside the war on poverty: the impact of food stamps on birth outcomes", *Review of Economics and Statistics* **93**, 387–403.

Arkhangelsky, D. and G. Imbens (2023), "Causal Models for longitudinal and panel Data: a survey", working paper, Stanford University.

Arkhangelsky, D. S. Athey, D.A. Hirshberg, G.W. Imbens, and S. Wager (2021), "Synthetic difference-in-differences, *American Economic Review* **111**, 4088–18.

Figure 7: Results of 2-dimensional effects



Athey, S. and G.D. Imbens (2006), "Identification and inference in nonlinear difference-in-differences models", *Econometrica* **74** 431–97.

Biewen, M., M. Rümmele, and B. Fitzenberger (2022), "Using distribution regression Difference-in-Differences to evaluate the effects of a minimum wage introduction on the distribution of hourly wages and hours worked, working paper, Nürnberg.

Blundell, R., C. Meghir, M. Costa Dias and J. Van Reenen (2004), "Evaluating the employment impact of a mandatory job search program", *Journal of the European Economic Association* **2**,569–606.

Card, D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market" *Industrial and Labor Relations Review,* **43**, 245—57.

Card, D. and A.B. Krueger (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania" *American Eco-*

nomic Review **84**, 772—93.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019), "The effect of minimum wages on low-wage jobs ", *Quarterly Journal of Economics* **134**, 1405–54.

Chernozhukov, V., I. Fernández-Val, and Melly B. (2013), "Inference on counterfactual distributions", *Econometrica*, **81**, 2205–68.

Chernozhukov, V., I. Fernández-Val, and S. Luo (2019), "Distribution regression with sample selection, with an application to wage decompositions in the UK", working paper, MIT, Cambridge (MA).

Chernozhukov, V. I. Fernández-Val, B. Melly, and K. Wüthrich (2020), "Generic inference on quantile and quantile effect functions for discrete outcomes", *Journal of the American Statistical Association* **115**, 123–37.

Dube, A. (2019), "Minimum wages and the distribution of family incomes", *American Economic Journal: Applied Economics* **11**, 268—304.

Fernández-Val I., A. van Vuuren and F. Vella (2024) "Distributional synthetic difference-in-differences", working paper, Boston University.

Fernández-Val I., J. Meier, A. van Vuuren and F. Vella (2023) "A bivariate distribution regression model for intergenerational mobility", working paper, Boston University.

Foresi, S. and F. Peracchi (1995), "The conditional distribution of excess returns: an empirical analysis", *Journal of the American Statistical Association*, **90**, 451–66.

Goodman-Bacon, A. (2021), "The long-run effects of childhood insurance coverage: medicaid implementation, adult health, and labor market outcomes, *American Economic Review* **111**, 2550-93.

Goodman-Bacon, A. and L. Schmidt (2020), "Federalizing benefits: The introduction of supplemental security income and the size of the safety net.", *Journal of Public Economics* **185**, 104174.

Kim, D. and J.M. Wooldridge (2023), “Difference-in-differences estimator of quantile treatment effect on the treated ”, working paper, Michigan State University.

Malesky, E.J., C.V. Nguyen, and A. Trahn (2014), “The Impact of recentralization on public services: A difference-in-differences analysis of the abolition of elected councils in Vietnam”, *American Political Science Review* **108** 144–68.

Melly, B. and Santangelo (2015), “The changes-in-changes model with covariates”, working paper, Bern University.

Roth, J. and P.H.C. Sant’Anna (2023), “When Is parallel trends sensitive to functional form?”, *Econometrica* **91** 737–47.

Wooldridge, J.M. (2023), “Simple approaches to nonlinear difference-in-differences with panel data”, *Econometric Journal* **26** C31–66.

Williams, O. D. and J.E. Grizzle (1972), “Analysis of contingency tables having ordered response categories”, *Journal of the American Statistical Association* **67**, 55–63.