I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17264

Distribution Regression Difference-in-Differences

Iván Fernández-Val
Jonas Meier
Aico van Vuuren
Francis Vella

AUGUST 2024

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17264

Distribution Regression Difference-in-Differences

Iván Fernández-Val
Boston University

Jonas Meier
Swiss National Bank

Aico van Vuuren
University of Groningen and IZA

Francis Vella
Georgetown University and IZA

AUGUST 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Distribution Regression Difference-in-Differences

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 covariates and, importantly, can be extended to settings where the treatment potentially 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 a transformation of the distribution. We highlight the relationship between our procedure and assumptions with the changes-in-changes approach of Athey and Imbens (2006). We also reexamine two existing empirical examples which highlight the utility of our approach.

JEL Classification: C10, C21

Keywords: difference-in-differences, treatment effects, distribution regression

Corresponding author:
Francis Vella
Economics Department
Georgetown University
37th and O Streets, NW
Washington, DC 20057
USA

E-mail: fgv@georgetown.edu

1 Introduction

The remarkable popularity of the difference-in-difference (DiD) estimator, inspired by an approach to evaluating the impact of policy interventions on economic outcomes introduced by David Card (see, for example, Card 1990, Card and Krueger, 1994), is one of the most striking features of empirical work on treatment and policy effects. While the methodological innovations in this literature (see Arkhangelsky and Imbens, 2023 for a recent review) include 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 estimated the mean effect of the treatment on a single economic outcome. This seems somewhat limited and a fuller evaluation of a policy treatment would be based on an examination of the marginal and joint distributions of all outcomes it potentially influences. This paper provides a simple procedure for estimating distributional treatment effects in the presence of a single treatment when the outcomes of interest are potentially multivariate.

An initial methodological innovation focusing on distributional effects in DiD estimation is the changes-in-changes procedure of Athey and Imbens (2006), which estimates the counterfactual distribution of the treated group in the absence of treatment to compare with its observed distribution in the presence of treatment. Torous et al. (2024) extend the Athey and Imbens approach (2006) to the multiple outcome setting. Other work has adapted DiD estimation to examine the treatment effects at different quantiles of the outcome via the use of quantile regression. This includes, for example, 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 distributional approaches include Kim and Wooldridge (2023) and Biewen, Fitzenberger, and Rümmele (2022). The former proposes an inverse probability weighting based procedure, while the latter employs a distribution regression (DR) approach. In this paper we also adopt a DR approach to constructing counterfactuals. In contrast to Biewen, Fitzenberger, and Rümmele (2022), who construct the counterfactual distributions via linear probability models, we employ 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-DiD.

While DiD has typically been employed to evaluate the treatment effect on a specified economic outcome, there are many instances in which the treatment may affect multiple outcomes. For example, a change in tax rates on earnings of married couples may affect the hours of work of both husbands and wives. An analysis of such a tax change should include the impact on each of the outcomes. 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. Alternatively, while evaluations of minimum wage laws typically evaluate their impact on employment, they may also affect the wage distribution. We illustrate how this joint effect can be evaluated via the bivariate distribution regression (BDR) approach of Fernandez-Val et al. (2024a). This requires that we first estimate the joint distribution by BDR and then construct the appropriate counterfactual. The treatment effects are obtained via the appropriate comparisons. An alternative to this approach is extending the changes-in-changes procedure to multiple outcomes as is done in Torous et al. (2024).

The following section introduces the model and provides an analysis of the univariate case without covariates. We also extend our analysis to include covariates and contrast our approach with the Athey and Imbens (2006) changes-in-changes procedure. Section 3 extends our analysis to the multiple outcome case and Section 4 discusses estimation. Section 5 provides two empirical illustrations of our methodology. We first revisit the Malesky et al. (2014) investigation of the impact of recentralization in Vietnam. We also employ the data studied in Callaway and Li (2019) and feed it with data from the Bureau of Labor Statistics and the Census in order to explore the impact of changes in the minimum wage on the joint distributions of average wages, poverty rates and unemployment rates. 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 $Y_0 \mid G, T$ since the dummy variable representation within the parentheses on the right-hand side is fully saturated. The parameters of the representation are local as they vary with y. To understand why (1) does not impose any restriction, note that $\alpha(y)$, $\beta(y)$, $\gamma(y)$ and $\delta(y)$ can be defined as the solutions to:[1]

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

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

Assumption 1 [No-interaction]**.**

$$\delta(y) = 0 \textit{ for all } y \in \mathbb{R} \textit{ in } (1).$$

Let $\mathcal{Y}_d(G = g, T = t)$ denote the support of $Y_d\,|\,G = g, T = t$, for $d, g, t \in 0, 1$. We also assume:

Assumption 2 [Support]**.**

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

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. That is, 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:

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

This assumption is sensitive to the link function and imposes restrictions on the distribution $F_{Y_0\,|\,G,T}$ for some link functions. For example, if Λ is the identity link used in the

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

linear probability model as in, for example, Almond et al. (2011), Dube (2019), Cengiz et al. (2019), Goodman-Bacon and Smith (2020), Goodman-Bacon (2021) and Biewen et al. (2022), one needs strong requirements in order to satisfy the parallel trends assumption (Blundell et al., 2004 and Wooldridge, 2023) That is, we need restrictions on the tails of the distribution of $F_{Y_0\,|\,G,T}(y\,|\,1,0)$, $F_{Y_0\,|\,G,T}(y\,|\,0,1)$ 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.[2,3] Link functions such as the normal or logistic CDFs do not require such restrictions since the transformation expands the range of the distribution to the entire real line.

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 holds pre-treatment. 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 $F_{Y_0\,|\,G,T}(y\,|\,1,1)$ since:

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

under Assumption 1. The support restrictions in Assumption 2 ensure that the term inside the squared brackets 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:

[2] These requirements could be used to develop a specification test for the identity link. Roth and Sant'Anna (2023) proposed a test for the sharp hypothesis that $y\mapsto F_{Y_0\,|\,G,T}(y\,|\,1,0)+F_{Y_0\,|\,G,T}(y\,|\,0,1)-F_{Y_0\,|\,G,T}(y\,|\,0,0)$ be weakly increasing, which can be adapted to our setting. We do not pursue this route as we do not encourage the use of the linear probability model.

[3] For example, an increase in 0.2 in probability over time might be realistic for the control group when the initial probability was 0.5. However, if treatment group has a probability of, for example, 0.9, in the first period then it is not possible for the common trends assumption to hold.

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

Proof of Lemma 1. The results follow from equation (2). $\square$

2.1 Inclusion of Covariates

Including covariates is appealing as the assumption that $\delta(y) = 0$ may be harder to defend when there are differences in the trend between covariates and/or the composition of the treatment group changes over time in terms of observed characteristics; see also Melly and Santangelo (2015). Covariates are easily incorporated into the identification result by conditioning 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:

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

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]**.**

$$\delta(y,X) = 0 \text{ *almost surely for all* } y \in \mathbb{R} \text{ *in* } (3).$$

Assumption 4 [Support conditions with Covariates]**.**

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

almost surely.

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

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

under the Assumption 3. The support restrictions in Assumption 4 ensure 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.

Let $\mathcal{X}_{11}$ denote the support of X conditional on $G = 1$ and $T = 1$. The following Lemma states that $F_{Y_0,Z_0\,|\,G,T,X}$ is identified under the previous assumptions.

Lemma 2 [Identification with Covariates]**.** *Under Assumptions 3 and 4,* $(y, x) \mapsto F_{Y_0\,|\,G,T,X}(y\,|\,1, 1, x)$ *is identified on* $(y, x) \in \mathbb{R} \times \mathcal{X}_{11}$.

Proof of Lemma 2. The result follows from equations (4). $\square$

We can then identify $F_{Y_0\,|\,G,T}(y\,|\,1, 1)$ as:

$$F_{Y_0\,|\,G,T}(y\,|\,1,1) = \int_{\mathcal{X}_{11}} F_{Y_0\,|\,G,T,X}(y\,|\,1,1,x)\mathrm{d}F_{X\,|\,G,T}(x\,|\,1,1). \tag{5}$$

2.2 Comparison with Changes-In-Changes (Athey and Imbens, 2006)

As our proposals provide an alternative approach to the changes-in-changes (CiC) procedure of Athey and Imbens (2006), it is useful to contrast their set up and assumptions with ours. CiC assumes that the outcome of an individual without treatment satisfies the relationship $Y_0 = h(U, T)$ for the treatment and control groups, where U is an unobserved and uniformly distributed random variable. It also assumes 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 previous assumptions identify the quantile function of

$F_{Y_0 \mid G,T}(y \mid 1,1)$ as:

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

where it is assumed 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 the distribution function is evaluated (Athey and Imbens, 2006, page 441). Hence, their identification results follow since 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. Their assumptions imply 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:

$$F_{Y_0 \mid G,T}(y \mid g,0) = F_{Y_0 \mid G,T}(h(h^{-1}(y,0),1) \mid g,1). \tag{6}$$

Evaluating (6) 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 (6) 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.[4] They would complicate estimation in our framework as they involve two different levels of Y and the transformation ϕ needs to be estimated.

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

2.3 Comparison with Roth and Sant'Anna (2023)

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

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

for the parallel trends assumption in expectations:

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

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

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

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 both. Our model, however, generally satisfies a different invariance property with respect to strictly monotonic transformations that we specify in subsection 2.4.

2.4 Invariance to Strictly Monotonic Transformations

The DR model in (1) with no-interaction is invariant to strictly monotonic transformations in the 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:

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

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 when h is strictly decreasing. Unlike the parallel trends in expectation, the no-interaction or parallel trends in distribution is invariant to strictly monotonic transformations.[5]

3 Multiple Outcomes

Some settings may feature multiple outcomes which are potentially affected by the treatment. In these situations we might be interested in not only how each of the outcomes are affected by the treatment, but also how the relationship between the outcomes is affected by the treatment. For this it is necessary to identify the joint distribution of the potential outcomes with and without the treatment. We now consider a setting with two outcomes Y and Z and we focus on 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 without the treatment, Y_0 and Z_0, for the treated group $G = 1$ in the post-treatment period $T = 1$. For the sake of illustration we consider two measures of dependence. Namely, Spearman's and Kendall's rank correlation.

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

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

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

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

where we assume that Y_d and Z_d are continuous random variables to obtain the expressions on the right hand side.

As in the univariate case, $F_{Y_1,Z_1\,|\,G,T}(y, z\,|\,1,1)$ is identified by the joint distribution of the observed outcomes, $F_{Y,Z\,|\,G,T}(y, z\,|\,1,1)$, whereas $F_{Y_0,Z_0\,|\,G,T}(y, z\,|\,1,1)$ is not identified from the data. To analyze identification, we use a variation of the local Gaussian representation (LGR) of a bivariate distribution from Chernozhukov, Fernandéz-Val and Luo (2018). Let Φ denote the Gaussian distribution function and $\Phi_2(\cdot,\cdot;\rho)$ denote the

[5] The distributional approach of Kim and Wooldridge (2023) also satisfies this property.

distribution of the bivariate standard normal with parameter ρ. Moreover, Λ is, again, a strictly increasing cumulative distribution function. As we show in Section 4, there is a benefit of using the logistic link function in our univariate analysis. Accordingly, we employ this in our empirical analysis for estimating both the univariate and bivariate effects.

Lemma 3 [LGR with non-Normal Marginals]**.** *The joint distribution of two random variables Y and Z conditional on X can be represented by:*

$$F_{Y,Z\,|\,X}(y,z\,|\,x)(y,z\,|\,x) \equiv \Phi_2(\Phi^{-1}(\Lambda(\mu_{Y\,|\,X}(y\,|\,x))), \Phi^{-1}(\Lambda(\mu_{Z\,|\,X}(y\,|\,x))); \rho_{Y,Z\,|\,X}(y,z\,|\,x)),$$

for all y, z, x, where $\mu_{Y\,|\,X}(y\,|\,x) = \Lambda^{-1}(F_{Y\,|\,X}(y\,|\,x))$, $\mu_{Z\,|\,X}(y\,|\,x) = \Lambda^{-1}(F_{Z\,|\,X}(z\,|\,x))$, and $\rho_{Y,Z\,|\,X}(y,z\,|\,x))$ is the unique solution in ρ to the equation:

$$F_{Y,Z\,|\,X}(y,z\,|\,x)(y,z\,|\,x) = \Phi_2(\Phi^{-1}(F_{Y\,|\,X}(y\,|\,x)(y,z\,|\,x)), \Phi^{-1}(F_{Z\,|\,X}(z\,|\,x)(y,z\,|\,x)); \rho).$$

Proof. The proof is identical to the proof of Lemma 2.1 of Chernozhukov, Fernandéz-Val and Luo (2018) using:

$$\Phi^{-1}(\Lambda(\mu_{Y\,|\,X}(y\,|\,x))) = \Phi^{-1}(F_{Y\,|\,X}(y\,|\,x))$$

and

$$\Phi^{-1}(\Lambda(\mu_{Z\,|\,X}(z\,|\,x))) = \Phi^{-1}(F_{Z\,|\,X}(z\,|\,x)).$$

□

The difference between Lemma 3 and the LGR of Chernozhukov, Fernandéz-Val and Luo (2018) is that the marginals are represented by a general link rather than Gaussian links, that is:

$$F_{Y\,|\,X}(y\,|\,x)(y\,|\,x) \equiv \Lambda(\mu_{Y\,|\,X}(y\,|\,x)), \quad F_{Z\,|\,X}(z\,|\,x) \equiv \Lambda(\mu_{Z\,|\,X}(z\,|\,x)).$$

By the LGR, $F_{Y_0,Z_0\,|\,G,T}$ can be expressed as:

$$F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,g,t) \equiv \\ \Phi_2(\Phi^{-1}(\Lambda(\mu_{Y_0\,|\,G,T}(y\,|\,g,t))), \Phi^{-1}(\Lambda(\mu_{Z_0\,|\,G,T}(y\,|\,g,t))); \rho_{Y_0,Z_0\,|\,G,T}(y,z\,|\,g,t)), \quad (7)$$

where $\mu_{Y_0\,|\,G,T}(y\,|\,g,t) = \alpha_Y(y) + \beta_Y(y)t + \gamma_Y(y)g + \delta_Y(y)gt$, $\mu_{Z_0\,|\,G,T}(y\,|\,g,t) = \alpha_Z(z) + \beta_Z(z)t + \gamma_Z(z)g + \delta_Z(z)gt$, and $\rho_{Y,Z\,|\,G,T}(y,z\,|\,g,t) = \alpha_{Y,Z}(y,z) + \beta_{Y,Z}(y,z)t + \gamma_{Y,Z}(y,z)g + \delta_{Y,Z}(y,z)gt$. In the LGR, the marginals are represented by:

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

and

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

We make the following identifying assumptions with respect to the distribution function in (7).

Assumption 5 [Bivariate No-interaction]**.**

$$\delta_Y(y) = \delta_Z(z) = \delta_{Y,Z}(y,z) = 0 \textit{ for all } (y,z) \in \mathbb{R}^2 \textit{ in } (7).$$

Let $\mathcal{YZ}_d(G = g, T = t)$ denote the support of $(Y_d, Z_d)\,|\,G = g, T = t$, for $d, g, t \in 0, 1$. We also assume:

Assumption 6 [Bivariate Support]**.**

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

Lemma 4 [Identification with Two Outcomes]**.** $(y,z) \mapsto F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)$ *is identified on* $\mathbb{R}^2$ *under Assumptions 5 and 6.*

Proof of Lemma 4. Under the assumptions of the Lemma, $\mu_{Y_0\,|\,G,T}(y\,|\,g,t) = \alpha_Y(y) + \beta_Y(y)t + \gamma_Y(y)g$, $\mu_{Z_0\,|\,G,T}(y\,|\,g,t) = \alpha_Z(z) + \beta_Z(z)t + \gamma_Z(z)g$, 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,

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

□

Covariates can be incorporated in a similar fashion as the univariate case. In particular, the LGR of $F_{Y_0,Z_0\,|\,G,T,X}$ becomes:

$$\begin{aligned}&F_{Y_0,Z_0\,|\,G,T,X}(y,z\,|\,g,t,x)\equiv\\&\Phi_2(\Phi^{-1}(\Lambda(\mu_{Y_0\,|\,G,T,X}(y\,|\,g,t,x))),\Phi^{-1}(\Lambda(\mu_{Z_0\,|\,G,T,X}(y\,|\,g,t,x)));\rho_{Y_0,Z_0\,|\,G,T,X}(y,z\,|\,g,t,x)),\end{aligned}\tag{8}$$

where $\mu_{Y_0\,|\,G,T,X}(y\,|\,g,t,x)=\alpha_Y(y,x)+\beta_Y(y,x)t+\gamma_Y(y,x)g+\delta_Y(y,x)gt$, $\mu_{Z_0\,|\,G,T,X}(y\,|\,g,t,x)=\alpha_Z(z,x)+\beta_Z(z,x)t+\gamma_Z(z,x)g+\delta_Z(z,x)gt$, and $\rho_{Y,Z\,|\,G,T,X}(y,z\,|\,g,t,x)=\alpha_{Y,Z}(y,z,x)+\beta_{Y,Z}(y,z,x)t+\gamma_{Y,Z}(y,z,x)g+\delta_{Y,Z}(y,z,x)gt$.

The identifying assumptions with covariates become:

Assumption 7 [Bivariate No-interaction with Covariates]**.**

$$\delta_Y(y,X)=\delta_Z(z,X)=\delta_{Y,Z}(y,z,X)=0 \textit{ almost surely for all } (y,z)\in\mathbb{R}^2 \textit{ in } (8).$$

Assumption 8 [Bivariate Support with Covariates]**.**

$$\begin{aligned}&\mathcal{YZ}_0(G=1;T=1;X)\subseteq\\&\qquad\mathcal{YZ}_0(G=0;T=1;X)\cup\mathcal{YZ}_0(G=1;T=0;X)\cup\mathcal{YZ}_0(G=0;T=0;X),\end{aligned}$$

almost surely.

Lemma 5 [Identification with Two Outcomes and Covariates]**.** *Under Assumptions 7 and 8, $(y,z)\mapsto F_{Y_0,Z_0\,|\,G,T,X}(y,z\,|\,1,1,x)$ is identified on $\mathbb{R}^2\times\mathcal{X}_{11}$.*

The marginalized distribution $F_{Y_0,Z_0\,|\,G,T}(y\,|\,1,1)$ is then identified by

$$F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)=\int_{\mathcal{X}_{11}}F_{Y_0,Z_0\,|\,G,T,X}(y,z\,|\,1,1,x)\mathrm{d}F_{X\,|\,G,T}(x\,|\,1,1).$$

4 Estimation

4.1 Univariate Case

Assume we have a sample $\{(Y_i, X_i, G_i, T_i) : 1 \leq i \leq N\}$ of (Y, X, G, T). For estimation, we replace the functions $(y, x) \mapsto (\alpha(y, x), \beta(y, x), \gamma(y, x))$ in (3) by semiparametric linear indexes leading to the DR model for the conditional distribution:

$$F_{Y_0 \mid G,T,X}(y \mid g, t, x) = \Lambda(p_\alpha(x)'\alpha(y) + p_\beta(x)'\beta(y)t + p_\gamma(x)'\gamma(y)g), \quad y \in \mathbb{R}, \tag{9}$$

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

We implement the DR DiD estimator via the sequence of logit models at each point of the distribution of the outcome variable (Foresi and Peracchi, 1995, Chernozhukov, Fernandez-Val and Melly, 2013). We choose logit because it is the canonical link for binary outcomes allowing for pooled estimation of the distributions of the potential outcomes with and without the treatment (Wooldridge, 2023). Accordingly, we estimate the DR model for the observed outcomes on all observations including those with $D_i = 1$:

$$F_{Y \mid G,T,X}(y \mid g, t, x) = \Lambda(p_\alpha(x)'\alpha(y) + p_\beta(x)'\beta(y)t + p_\gamma(x)'\gamma(y)g + p_\theta(x)'\theta(y)gt), \quad y \in \mathcal{Y}, \tag{10}$$

where $p_\alpha(x)$, $p_\beta(x)$, $p_\gamma(x)$ and $p_\theta(x)$ are vectors including a constant as the first component, and transformations of the covariates, and $\mathcal{Y}$ is a finite grid on $\mathbb{R}$. Let $I_i^y := 1(Y_i \leq y)$ and $\bar{I}_i^y = 1 - I_i^y$.

Algorithm 1 [Univariate Estimator]**.** *1. Estimate the parameters of model* (10) *by distribution regression, that is, for $y \in \mathcal{Y}$,*

$$(\hat{\alpha}(y), \hat{\beta}(y), \hat{\gamma}(y), \hat{\theta}(y)) \in \arg\max_{a,b,c,d} \sum_{i=1}^{N} \ell_i(a, b, c, d),$$

$$\ell_i(a, b, c, d) = I_i^y \log \Lambda(p_\alpha(X_i)'a + p_\beta(X_i)'b\ T_i + p_\gamma(X_i)'c\ G_i + p_\theta(X_i)'d\ G_iT_i)$$
$$+ \bar{I}_i^y \log \Lambda(-p_\alpha(X_i)'a - p_\beta(X_i)'b\ T_i - p_\gamma(X_i)'c\ G_i - p_\theta(X_i)'d\ G_iT_i).$$

2. Construct plug-in estimators of the distributions of the potential outcomes

$$\hat{F}_{Y_0 \mid G,T}(y \mid 1, 1) = \frac{1}{N_{11}} \sum_{i=1}^{N} G_iT_i\ \Lambda(p_\alpha(X_i)'\hat{\alpha}(y) + p_\beta(X_i)'\hat{\beta}(y) + p_\gamma(X_i)'\hat{\gamma}(y)),$$

and

$$\hat{F}_{Y_1\,|\,G,T}(y\,|\,1,1) = \frac{1}{N_{11}}\sum_{i=1}^{N} G_i T_i\ \Lambda(p_\alpha(X_i)'\hat{\alpha}(y) + p_\beta(X_i)'\hat{\beta}(y) + p_\gamma(X_i)'\hat{\gamma}(y) + p_\theta(X_i)'\hat{\theta}(y)),$$

where $N_{11} = \sum_{i=1}^{N} G_i T_i$.

3. *If needed, rearrange the estimates* $y \mapsto \hat{F}_{Y_d\,|\,G,T}(y\,|\,1,1)$ *on* $y \in \mathcal{Y}$*,* $d \in \{0,1\}$*, to make them increasing.*

By the properties of the logistic link, the estimator of $F_{Y_1\,|\,G,T}(y\,|\,1,1)$ is identical to the empirical distribution of Y conditional on $G = 1$ and $T = 1$,

$$\hat{F}_{Y_1\,|\,G,T}(y\,|\,1,1) \equiv \frac{1}{N_{11}}\sum_{i=1}^{N} G_i T_i\ 1(Y_i \leq y).$$

Note that this estimator is therefore invariant to the specification of $p_\theta(x)$. We set $p_\theta(x) = 1$ to speed up computation. Estimators of the functionals of the distributions of potential outcomes such as quantile functions and effects can be constructed using the plug-in principle.

Our algorithm has an advantage over the alternative of estimating p_α, p_β, p_γ using only those observations for which $D_i = 0$ via direct estimation of (10). The distributional treatment effect, *i.e.*

$$\mathbb{E}_X\left[F_{Y_1|G,T,X}(y|1,1,X) - F_{Y_0|G,T,X}(y|1,1,X)\right],$$

equals the average derivative estimator of D_i for the logit model used for distribution regression. This estimate of the average derivative and its standard error are reported by many software packages (Wooldridge, 2024).

4.2 Bivariate Case

Assume we have a sample $\{(Y_i, Z_i, X_i, G_i, T_i) : 1 \leq i \leq N\}$ of (Y, Z, X, G, T). For estimation, as in the univariate case, we replace the functions in $\mu_{Y_0\,|\,G,T,X}$, $\mu_{Z_0\,|\,G,T,X}$ and $\rho_{Y,Z\,|\,G,T,X}$ by semiparametric generalized linear indexes leading to a bivariate distribution regression (BDR) model:

$$\mu_{Y_0\,|\,G,T,X}(y\,|\,g,t,x) = p_\alpha(x)'\alpha_Y(y) + p_\beta(x)'\beta_Y(y)t + p_\gamma(x)'\gamma_Y(y)g, \tag{11}$$

$$\mu_{Z_0\,|\,G,T,X}(y\,|\,g,t,x) = q_\alpha(x)'\alpha_Z(z) + q_\beta(x)'\beta_Z(z)t + q_\gamma(x)'\gamma_Z(z)g, \tag{12}$$

and

$$\rho_{Y,Z\,|\,G,T,X}(y,z\,|\,g,t,x) = h(r_\alpha(x)'\alpha_{Y,Z}(y,z) + r_\beta(x)'\beta_{Y,Z}(y,z)t + r_\gamma(x)'\gamma_{Y,Z}(y,z)g), \tag{13}$$

where $p_\alpha(x)$, $p_\beta(x)$, $p_\gamma(x)$, $q_\alpha(x)$, $q_\beta(x)$, $q_\gamma(x)$, $r_\alpha(x)$, $r_\beta(x)$ and $r_\gamma(x)$ are vectors including the covariates and their transformations, and $h(u) = \operatorname{arctanh}(u)$ is the Fisher transformation that enforces $\rho_{Y,Z\,|\,G,T,X}$ to lie in $[-1,1]$.

We estimate all the parameters of $F_{Y_0,Z_0\,|\,G,T}(y,z\,|\,1,1)$ using the bivariate distribution regression estimator of Fernandez-Val et al. (2024a). We employ an imputation method that combines the parameter estimates from the sample of the first period for both groups and the sample of the second period for the untreated group, with the sample of the covariates in the second period for the treated group. The distribution $F_{Y_1,Z_1\,|\,G,T}(y,z\,|\,1,1)$ is estimated using the empirical distribution of Y and Z in the second period for the treated group. Algorithm 2 describes the estimation procedure. Let $\mathcal{Y}$ and $\mathcal{Z}$ be finite grids on $\mathbb{R}$, $I_i^y := 1(Y_i \leq y)$, $\bar{I}_i^y = 1 - I_i^y$, $J_i^z := 1(Z_i \leq z)$, $\bar{J}_i^z = 1 - J_i^z$.

Algorithm 2 [Bivariate Estimator]**.** *1. Estimate the parameters of* (11) *and* (12) *using Algorithm 1 on $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$. Obtain*

$$\hat{m}_i^Y(y) = \Phi^{-1}(\Lambda(p_\alpha(X_i)'\hat{\alpha}_Y(y) + p_\beta(X_i)'\hat{\beta}_Y(y)\ T_i + p_\gamma(X_i)'\hat{\gamma}_Y(y)\ G_i)),$$

and

$$\hat{m}_i^Z(z) = \Phi^{-1}(\Lambda(q_\alpha(X_i)'\hat{\alpha}_Z(z) + q_\beta(X_i)'\hat{\beta}_Z(z)\ T_i + q_\gamma(X_i)'\hat{\gamma}_Z(z)\ G_i)),$$

where $\hat{\alpha}_Y(y)$, $\hat{\beta}_Y(y)$, $\hat{\gamma}_Y(y)$, $\hat{\alpha}_Z(z)$, $\hat{\beta}_Z(z)$ and $\hat{\gamma}_Z(z)$ are the estimates of $\alpha_Y(y)$, $\beta_Y(y)$, $\gamma_Y(y)$, $\alpha_Z(z)$, $\beta_Z(z)$ and $\gamma_Z(z)$ obtained from Algorithm 1.

2. Estimate the parameters of (13) *by BDR, that is, for $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$,*

$$(\hat{\alpha}_{Y,Z}(y,z), \hat{\beta}_{Y,Z}(y,z), \hat{\gamma}_{Y,Z}(y,z)) \in \arg\max_{a,b,c} \sum_{i=1}^{N} (1 - G_i T_i)\ \ell_i(a,b,c),$$

$$\begin{aligned}
\ell_i(a,b,c) = {} & I_i^y J_i^z \log \Phi_2(\hat{m}_i^Y(y), \hat{m}_i^Z(z); h(r_\alpha(X_i)'a + r_\beta(X_i)'b\ T_i + r_\gamma(X_i)'c\ G_i)) \\
& + I_i^y \bar{J}_i^z \log \Phi_2(\hat{m}_i^Y(y), -\hat{m}_i^Z(z); -h(r_\alpha(X_i)'a + r_\beta(X_i)'b\ T_i + r_\gamma(X_i)'c\ G_i)) \\
& + \bar{I}_i^y J_i^z \log \Phi_2(-\hat{m}_i^Y(y), \hat{m}_i^Z(z); -h(r_\alpha(X_i)'a + r_\beta(X_i)'b\ T_i + r_\gamma(X_i)'c\ G_i)) \\
& + \bar{I}_i^y \bar{J}_i^z \log \Phi_2(-\hat{m}_i^Y(y), -\hat{m}_i^Z(z); h(r_\alpha(X_i)'a + r_\beta(X_i)'b\ T_i + r_\gamma(X_i)'c\ G_i)).
\end{aligned}$$

3. *Construct plug-in estimators of the distributions of the potential outcomes*

$$\hat{F}_{Y_0 \mid G,T}(y \mid 1,1) = \frac{1}{N_{11}} \sum_{i=1}^{N} G_i T_i \; \Phi_2(\hat{n}_i^Y(y), \hat{n}_i^Z(z); \hat{n}_i^{Y,Z}(y,z)),$$

and

$$\hat{F}_{Y_1,Z_1 \mid G,T}(y, z \mid 1,1) = \frac{1}{N_{11}} \sum_{i=1}^{N} G_i T_i \; 1(Y_i \leq y, Z_i \leq z),$$

where

$$\hat{n}_i^Y(y) = \Phi^{-1}(\Lambda(p_\alpha(X_i)'\hat{\alpha}_Y(y) + p_\beta(X_i)'\hat{\beta}_Y(y) + p_\gamma(X_i)'\hat{\gamma}_Y(y))),$$

$$\hat{n}_i^Z(z) = \Phi^{-1}(\Lambda(q_\alpha(X_i)'\hat{\alpha}_Z(z) + q_\beta(X_i)'\hat{\beta}_Z(z) + q_\gamma(X_i)'\hat{\gamma}_Z(z))),$$

$$\hat{n}_i^{Y,Z}(y,z) = h(r_\alpha(X_i)'\hat{\alpha}_{Y,Z}(y,z) + r_\beta(X_i)'\hat{\beta}_{Y,Z}(y,z) + r_\gamma(X_i)'\hat{\gamma}_{Y,Z}(y,z))$$

and $N_{11} = \sum_{i=1}^{N} G_i T_i$.

Estimators of the functionals of the joint distributions of potential outcomes such as Spearman's and Kendall's rank correlation coefficients can be constructed using the plug-in principle.

4.3 Bootstrap Inference

The estimators described in Algorithms 1 and 2 can be applied to panel and repeated cross-section data. Here we describe a weighted bootstrap algorithm to perform inference on functions of the distributions of potential outcomes designed for panel data. We focus on this case because it is relevant for the empirical applications in Section 5.

To describe the procedure, we need to introduce an indicator ID_i, $i = 1, \ldots, N$, for the units in the panel. For example, if the sample is sorted by unit and time period, $ID = (1, 1, 2, 2, \ldots, n, n)$, where $n = N/2$. The following algorithm describes the weighted bootstrap procedure to construct joint confidence bands for the distributions of the potential outcomes with and without the treatment in the univariate case. Inference for functionals of the distributions and for the bivariate case can be performed using similar algorithms.

Algorithm 3 [Weighted Bootstrap Inference]. 1. *Choose the number of bootstrap repetitions* B*, e.g.,* $B = 500$ *or* $B = 1,000$.

2. *Draw weights for each unit independent and identically from the standard exponential distribution, independently for the data. Construct a vector of weights* $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)$, *where* $\omega_i = \omega_j$ *if* $ID_i = ID_j$, *and normalize the components of* $\boldsymbol{\omega}$ *to add up to one.*

3. *Estimate the parameters of model* (10) *by weighted distribution regression, that is, for* $y \in \mathcal{Y}$,

$$(\tilde{\alpha}(y), \tilde{\beta}(y), \tilde{\gamma}(y), \tilde{\theta}(y)) \in \arg\max_{a,b,c,d} \sum_{i=1}^{N} \omega_i \ell_i(a, b, c, d),$$

$$\ell_i(a,b,c,d) = I_i^y \log \Lambda(p_\alpha(X_i)'a + p_\beta(X_i)'b\ T_i + p_\gamma(X_i)'c\ G_i + p_\theta(X_i)'d\ G_i T_i)$$
$$+ \bar{I}_i^y \log \Lambda(-p_\alpha(X_i)'a - p_\beta(X_i)'b\ T_i - p_\gamma(X_i)'c\ G_i - p_\theta(X_i)'d\ G_i T_i).$$

4. *Construct plug-in weighted estimators of the distributions of the potential outcomes*

$$\hat{F}^b_{Y_0 \mid G,T}(y \mid 1, 1) = \frac{1}{N_{11}} \sum_{i=1}^{N} \omega_i\ G_i T_i\ \Lambda(p_\alpha(X_i)'\tilde{\alpha}(y) + p_\beta(X_i)'\tilde{\beta}(y) + p_\gamma(X_i)'\tilde{\gamma}(y)),$$

and

$$\hat{F}^b_{Y_1 \mid G,T}(y \mid 1, 1) = \frac{1}{N_{11}} \sum_{i=1}^{N} \omega_i\ G_i T_i\ \Lambda(p_\alpha(X_i)'\tilde{\alpha}(y) + p_\beta(X_i)'\tilde{\beta}(y) + p_\gamma(X_i)'\tilde{\gamma}(y) + p_\theta(X_i)'\tilde{\theta}(y)),$$

where $N_{11} = \sum_{i=1}^{N} \omega_i G_i T_i$.

5. *Repeat steps 1-3 B times to obtain*

$$\left\{ \hat{F}^b_{Y_0 \mid G,T}(y \mid 1, 1), \hat{F}^b_{Y_1 \mid G,T}(y \mid 1, 1) : y \in \mathcal{Y}, 1 \leq b \leq B \right\}.$$

6. *Construct an estimator of the* $(1-\alpha)$*-critical value of the maximal t-statistic,* $\bar{t}_{\mathcal{Y}}(1-\alpha)$, *as the* $(1-\alpha)$*-quantile of* $\{\bar{t}^b_{\mathcal{Y}} : 1 \leq b \leq B\}$, *where*

$$\bar{t}^b_{\mathcal{Y}} = \max_{y \in \mathcal{Y}} \left[\frac{|\hat{F}^b_{Y_0 \mid G,T}(y \mid 1, 1) - \hat{F}_{Y_0 \mid G,T}(y \mid 1, 1)|}{S_0(y)}, \frac{|\hat{F}^b_{Y_1 \mid G,T}(y \mid 1, 1) - \hat{F}_{Y_1 \mid G,T}(y \mid 1, 1)|}{S_1(y)} \right],$$

and $S_d(y)$ *is the interquartile range of* $\left\{ \hat{F}^b_{Y_d \mid G,T}(y \mid 1, 1) : 1 \leq b \leq B \right\}$ *divided by* 1.34896, *the interquartile range of the standard normal distribution, for* $d \in \{0, 1\}$.

7. *Construct the* $(1-\alpha)$*-confidence bands as*

$$CB_{1-\alpha}[F_{Y_d \mid G,T}(\cdot \mid 1, 1)] = \hat{F}_{Y_d \mid G,T}(y \mid 1, 1) \pm \bar{t}_{\mathcal{Y}}(1-\alpha) S_d(y), \quad d \in \{0, 1\}.$$

Remark 1 (Empirical Bootstrap). *Empirical bootstrap can be implemented by drawing the weights in step 1 from a multinomial distribution with values* $1, \ldots, n$ *and equal probabilities* $1/n$.

5 Empirical applications

We illustrate our proposed estimators via an examination of data from two existing empirical investigations.[6]. The first is the Malesky et al. (2014) investigation of the impact of recentralization in Vietnam. This paper is useful for our purposes as it considers a single treatment and multiple outcomes. Moreover, Malesky et al. (2014) only report mean effects. We examine whether these mean effects are informative of the heterogeneity across the outcomes' distributions. We also investigate whether the treatment affects the bivariate distribution of different combinations of outcomes. In a second example we explore the impact of increases in the mandatory minimum wage on average weekly wages, and unemployment and poverty rates by examining an extension of the data in Callaway and Li, (2019). We estimate the impact of the treatment on each of these outcome distributions and also the bivariate distributions of some combinations of the outcomes.

5.1 The impact of recentralization (Malesky, et al. 2014)

5.1.1 Description of the original empirical exercise

Malesky et al. (2014) investigate the impact of recentralization via a case study in Vietnam. Due to the dissatisfaction with decentralization measures taken in the early 1990s, Vietnam changed their political system in 2007 by eliminating one political layer from the decision making process. More explicitly, 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).[7] The change involved abolishing the political process at the districts (which are governed by the so called Districts People Council or

[6]We thank Brant Callaway, Tong Li and Pedro Sant'Anna for providing us with the data

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

DPC). Prior to introducing this change, the Vietnam government experimented with 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 abolition of the DPC in the treatment districts started in 2009.

Malesky et al. (2014) employ the following specification;

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

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 if the 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 region dummies (8 regions in total). For reasons of data availability, Malesky et al. (2014) only use rural communes and two years of observations: 2008 and 2010 (they use 2006 for robustness checks). They use 30 different outcome variables to investigate the impact of the abolition of the political layer. As most of their outcome variables are indicator variables, we only employ the following 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.

5.1.2 Our analysis for the univariate case

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

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

and use the same control variables as in Malesky et al. (2014). We 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 via distribution regression at y. We estimate the unconditional distribution using

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

Our estimator then 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 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 inverting the estimated distribution functions. As in Malesky et al. (2014), we correct the confidence intervals for clustering at the province level. We use the Bayesian bootstrap and draw the same exponential weight for all observations belonging to the same province (see also Chernozukov et al., 2020). We 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. This is allowed provided we assume the outcome variable is continuously distributed. For some outcome variables, there is substantial bunching at zero. For example, for the variable "Proportion of households supported crop" 49.93 percent of the observations equal zero. It also has this value for 54.04 percent for the treatment group in the treatment period and 50.79 for the control group. This implies that the quantile treatment effect is by definition equal to zero up to the median and the

Figure 1: Results of the empirical application.

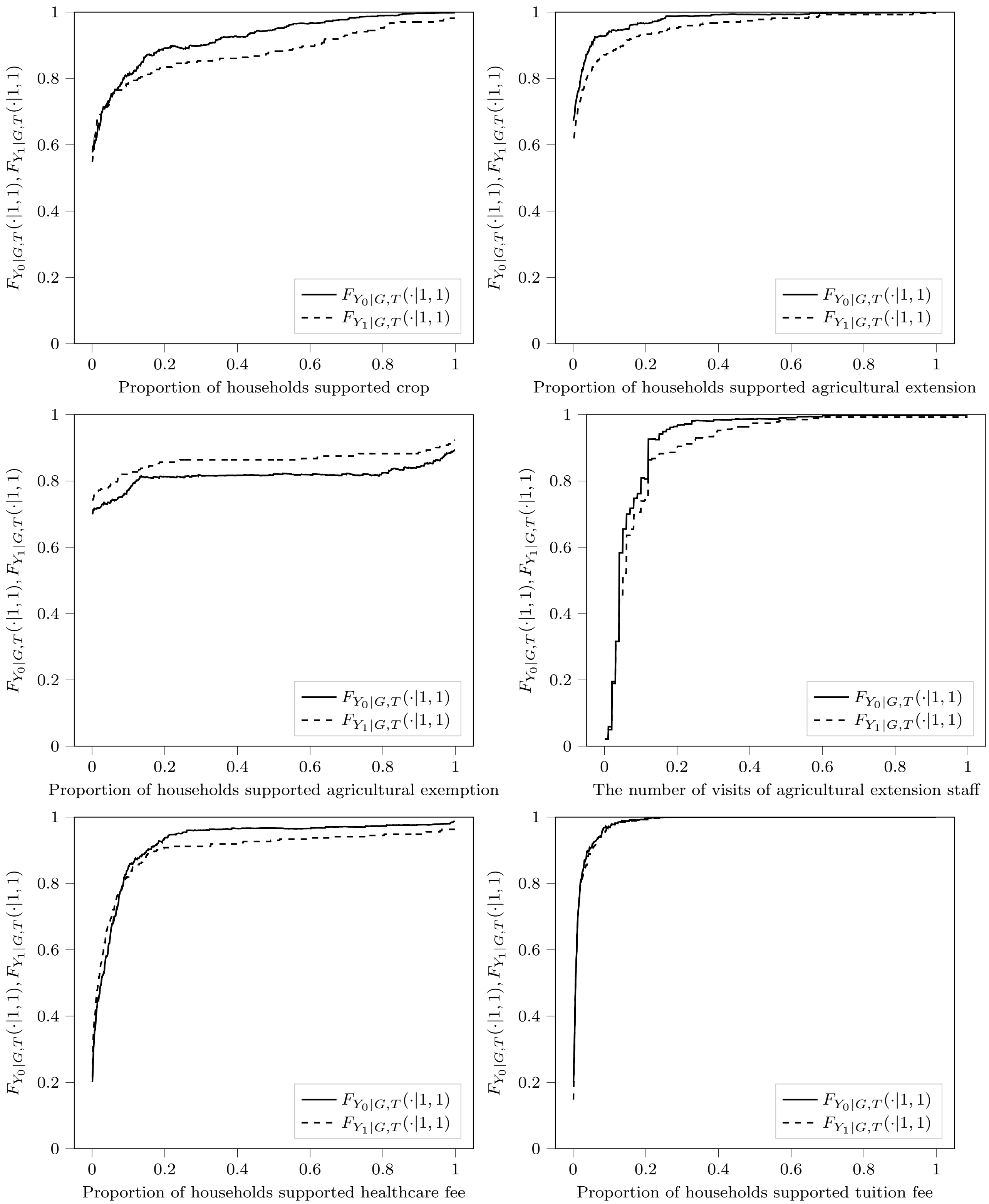
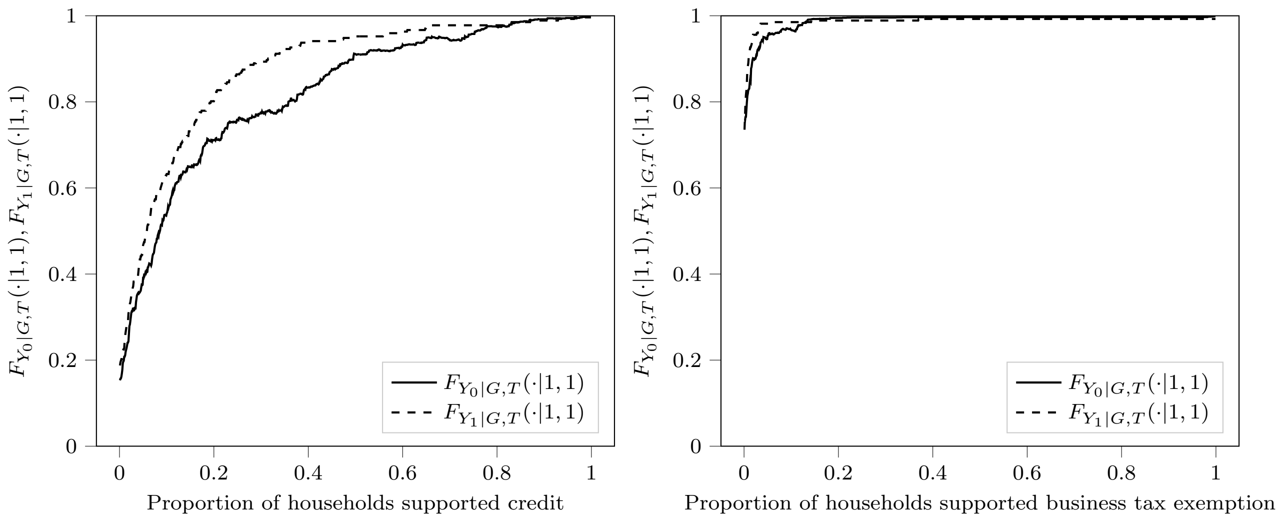


Figure 1: Results of the empirical application (continued).



impact can only occur at the higher quantiles. Note, that we employ dots to distinguish this from cases in which there is an estimated zero impact.

Nevertheless, even after ignoring these zeros, both Figure 1 and Table 1 reveal a great deal of treatment heterogeneity and that the mean impacts reported in Malesky et al. (2014) are primarily generated from 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. A similar conclusion can be drawn for the outcome variable “Proportion of households supported credit”.

Malesky et al. (2014) use data for the year 2006 to check the common trends assumption. They examine the results of a DiD design for the periods 2006 and 2008 in which 2008 is the placebo treatment period. That is, one would not expect any treatment effect in the period before the introduction of the treatment. The results of this exercise are in Figure 2 and Table 2. Similar to Malesky et al. (2014), we do find some substantial differences in Figure 2 between the distribution of Y_0 and Y_1. However, Table 2 indicates these differences are not statistically significant and that 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

	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 corrected for clustering at the province level.

Figure 2 generally to the left of the distribution of $F_{Y_1|G,T}(\cdot|1,1)$ while the relationship is the opposite in Figure 1.

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) before the treatment. This is not true for the non-linear design. AICO: THIS IS NOT SO CLEAR. However we perform an additional check by examining the estimated coefficient value of the interaction term. That is, we estimate the general representation 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 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. These results are shown in Figure 4 and the quantile treatment effects are reported in Table 5.

5.1.3 Comparison with the changes-in-changes estimation

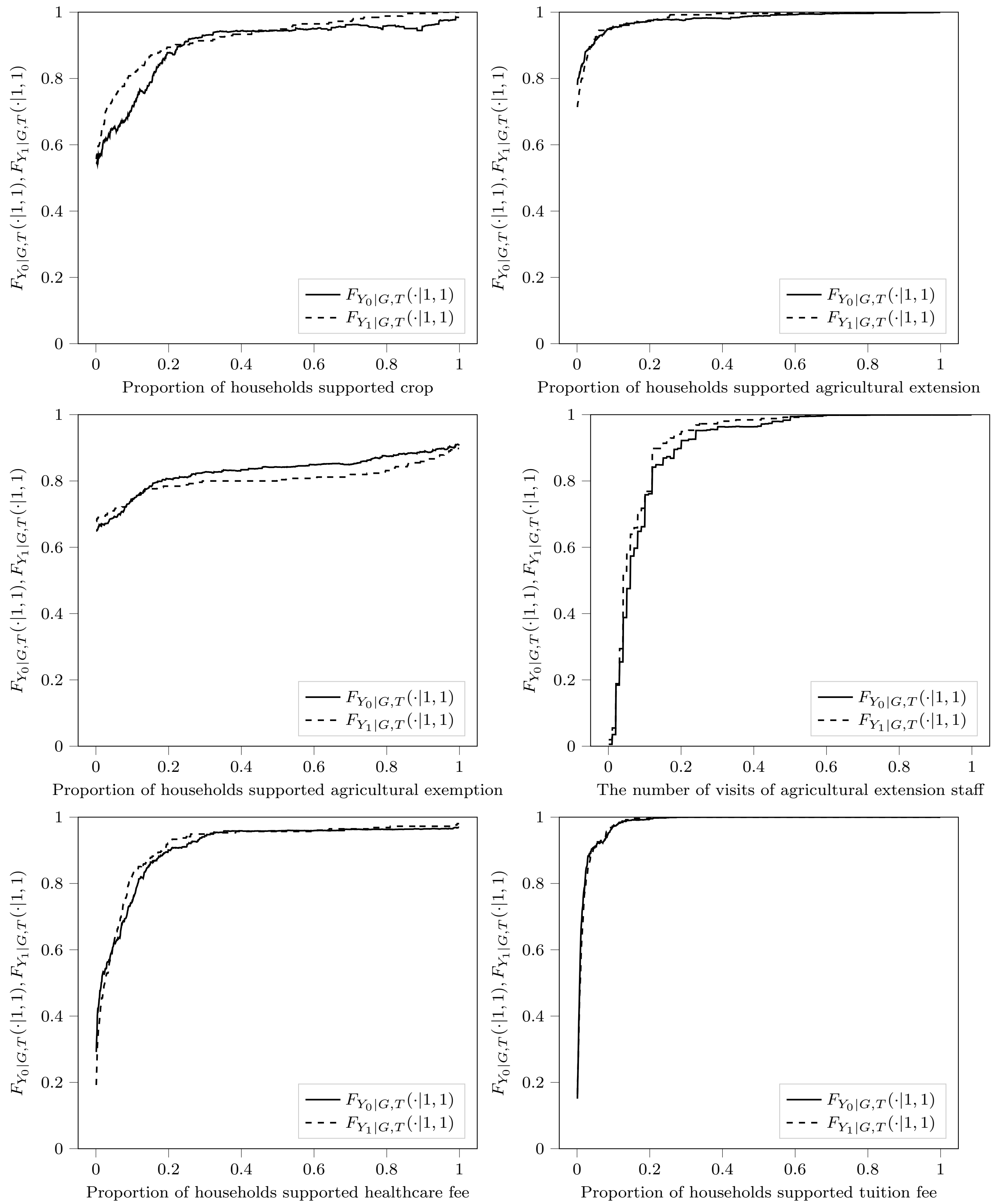
For the changes-in-changes estimation, we note as in Athey and Imbens (2006) that the distribution of $Y_0|G=1,T=1$ equals the distribution of $\varphi(Y_0)|G=1,T=0$, where $\phi(y)$ is defined as in Section 2.2, i.e. $\phi(y) := F^{-1}_{Y_0\,|\,G,T}\left(F_{Y_0\,|\,G,T}(y\,|\,0,0)\,|\,0,1\right)$. We then obtain an estimator of the distribution of $Y_0|G=1,T=1$ by using the empirical distribution function of the random variable:

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

We estimate the distribution function of $F_{Y_0|G=1,T=1}$ at 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

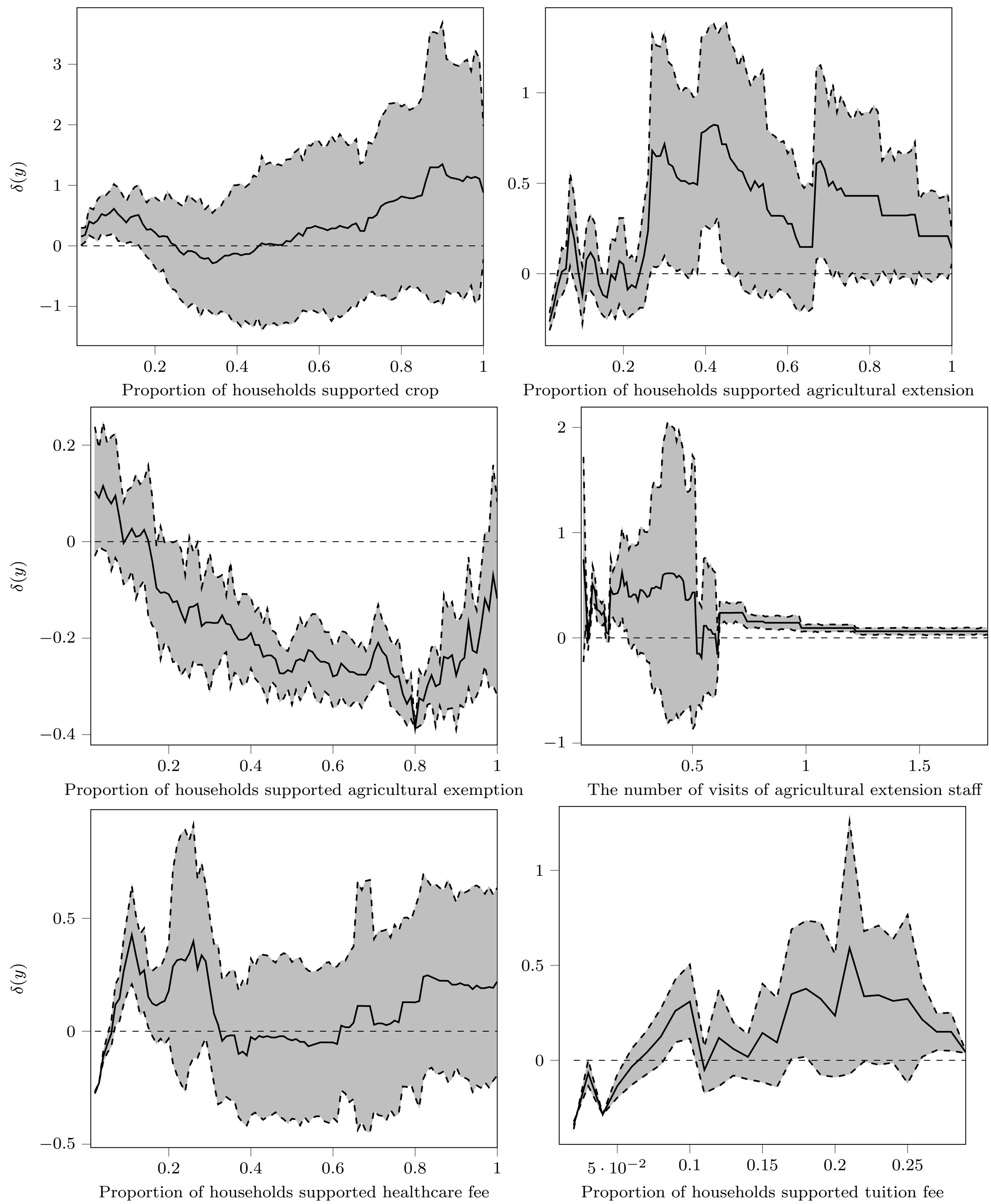
Figure 2: Results of robustness check using the years 2006 and 2008.



	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 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.

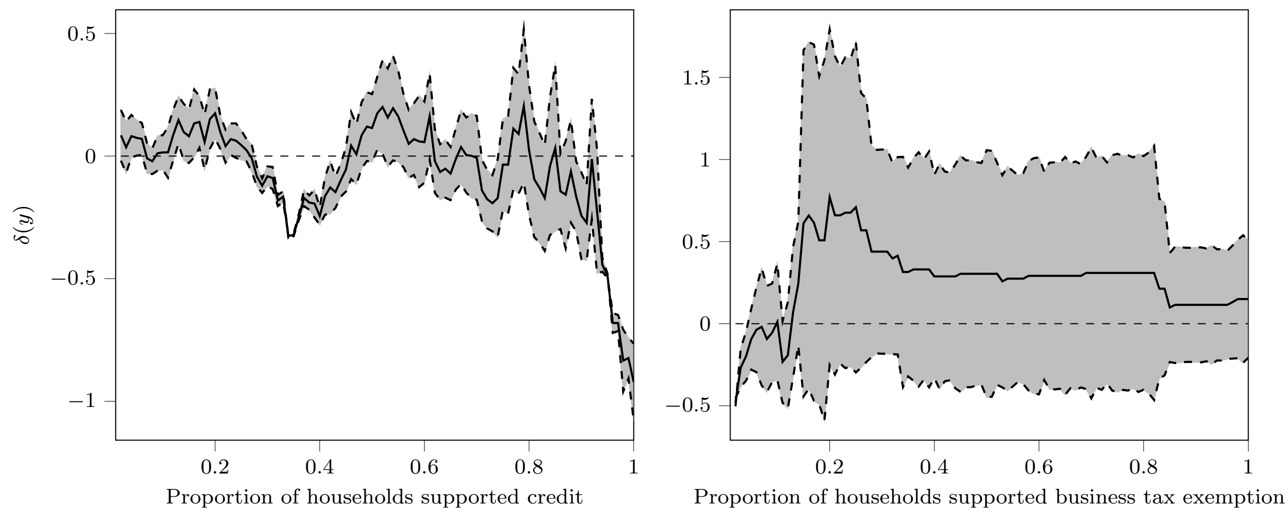
	Mean	0.1	0.25	0.5	0.75	0.9
Proportion of households supported	-0.0162	·	·	·	-0.0597	-0.0206
crop	(-0.084,0.052)	(·, ·)	(·, ·)	(·, ·)	(-0.152,0.033)	(-0.762,0.72)
Proportion of households supported	-0.0041	·	·	·	·	0.0007
agricultural extension	(-0.015,0.007)	(·, ·)	(·, ·)	(·, ·)	(·, ·)	(-0.032,0.033)
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).



distribution function of the subsample for which $G = 0, T = 0$.

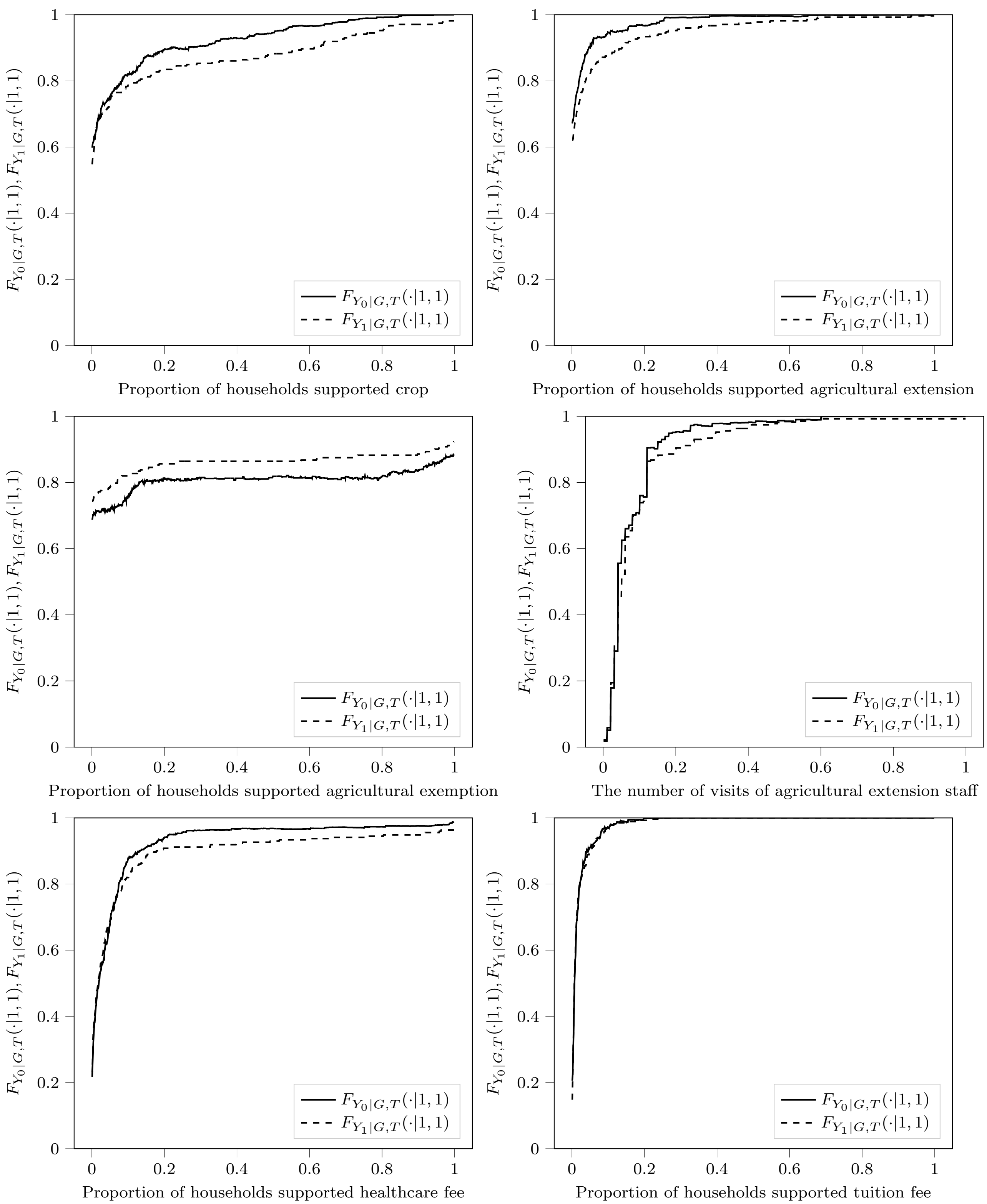
2. For every computed empirical distribution function of step 1 estimate the corresponding quantile of the subsample for which $G = 0, T = 1$.

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

The distribution of $F_{Y_1|G=1,T=1}$ can be estimated using the empirical distribution function. One obtains the quantile treatment effect by inverting the distribution at the desired levels of the distribution.

As we consider our estimator for the univariate outcome as a simpler alternative to the CiC approach we contrast our results with those from that approach. We only do so for the no covariates case as the CiC estimator is more difficult to implement in the presence of covariates. The results are reported in Table 3 and Figure 5. The results are reassuring as they are very similar to those in Table 6 and Figure 6 for CiC. The results are similar in terms of both magnitude and statistical significance.

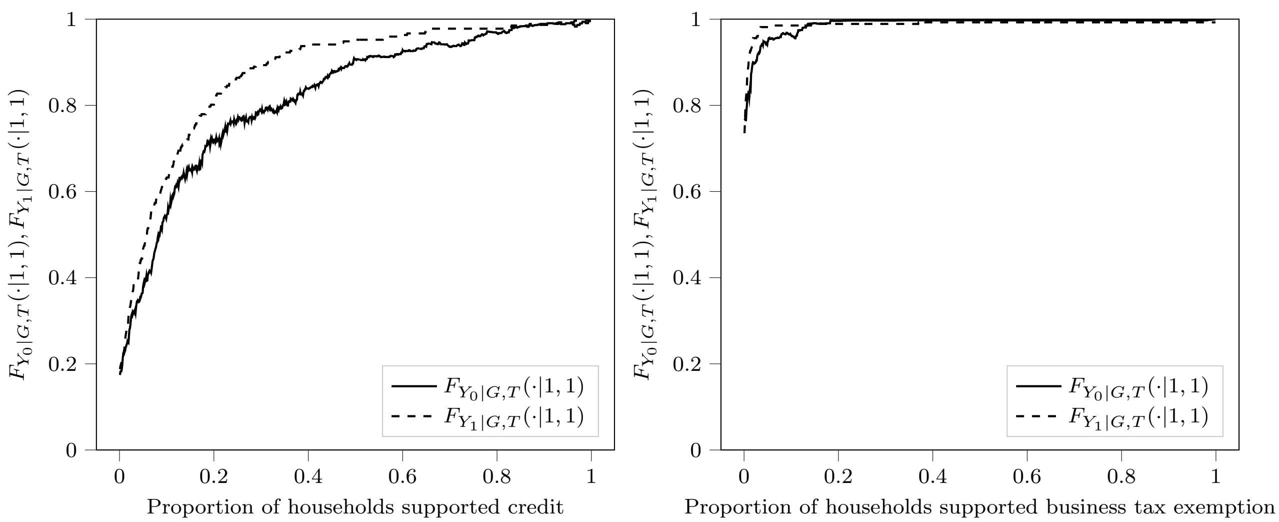
Figure 4: Results of the empirical application using interaction terms between the covariates and the time and treatment dummy variable.



	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).



5.1.4 Results for bivariate outcomes

For the bivariate analysis we consider the outcome variables "Proportion of households supported credit" and the "Proportion of households supported healthcare fee" as these variables have relatively little bunching at integer values. The counterfactual and actual distributions are shown in Figure 10.

The figure reveals 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. We also present results using Kendall's tau given as:

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

The Kendall's tau for the joint distribution of the treated sample in the second period when treated can directly be calculated from the observed 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 estimate of its marginal distribution from above. We then sample Z from the conditional distribution

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

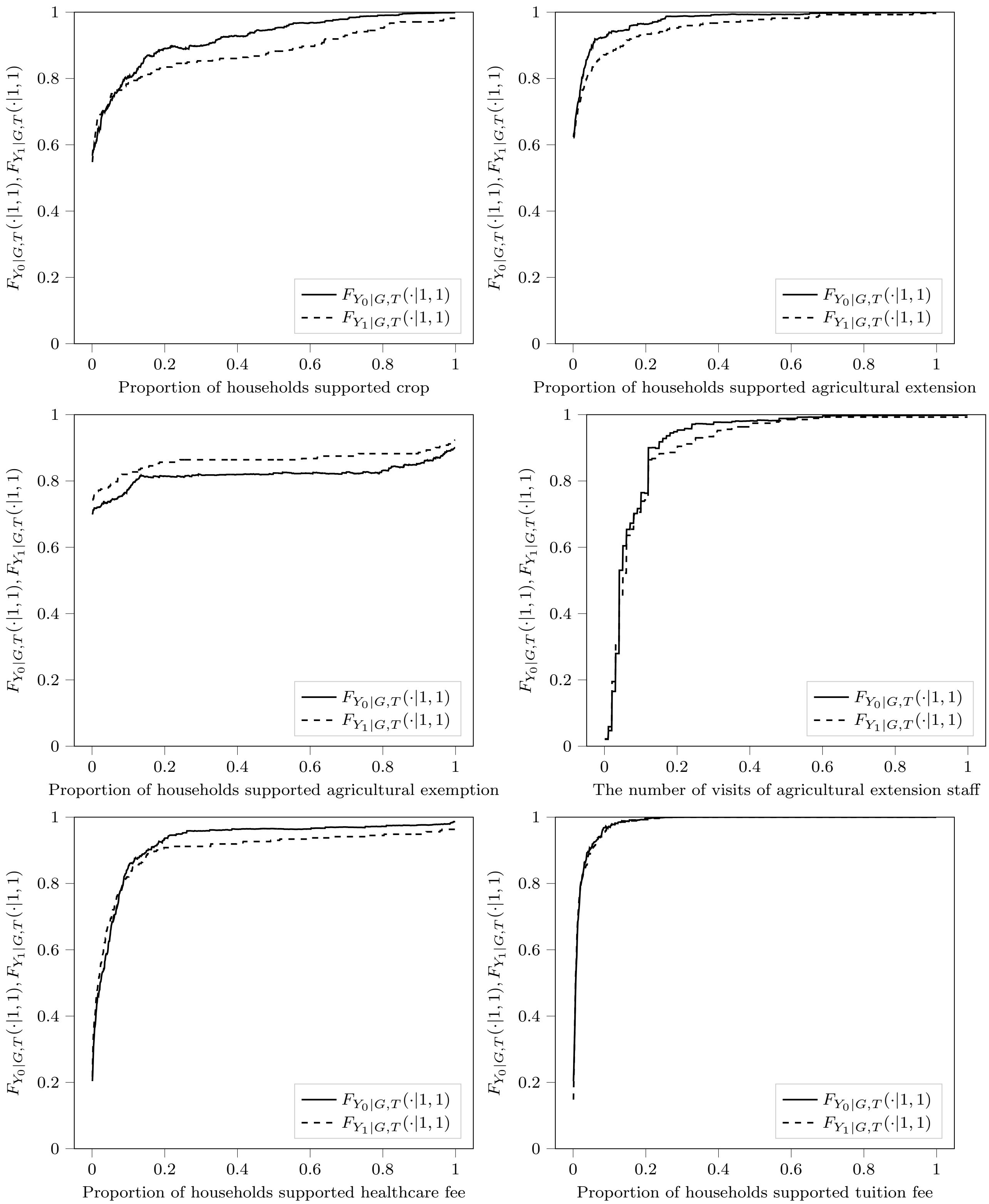
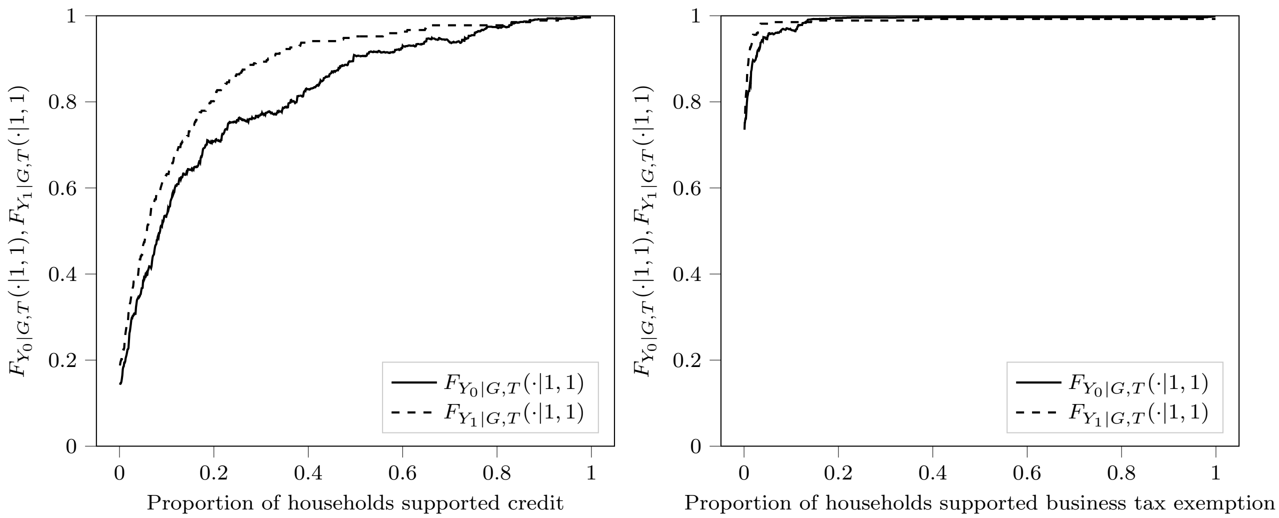


Figure 5: Results without using additional control variables in the analysis. (continued).



of $Z|Y$ which can be obtained using our estimates for the bivariate case. The estimated Kendall's tau from this procedure is 0.1253 with a 95-percent confidence interval from 0.0989 to 0.1518. This implies a positive correlation between the two outcomes in the districts. For the observed distribution of the treated group we obtain a 0.2463 with a 95-percent confidence interval from 0.2224 to 0.2703. This implies that the treatment has statistically significantly increased the correlation between the two outcomes.

5.2 The impact of increases in the mandatory minimum wage

The impact of increases in the mandatory minimum wage has been the focus of many empirical investigations. Some consider their impact on (un)employment (Card and Krueger, 1994, Dube et al., 2010, Cengiz et al, 2019, Callaway and Li, 2019, Torous et al, 2024), while others focus on how they affect income levels (Dube, 2019) or the poverty rate (MaCurdy, 2015). The vast majority of these studies consider the mean of these outcome variables and only investigate the respective outcomes separately. Torous et al. (2024) is an exception as it examines distributional effects and the relationship between part-time and full-time employment. We now consider the joint impact of changes in the mandatory minimum wage on wages and both the unemployment and poverty rates.

Figure 6: Results of changes-in-changes

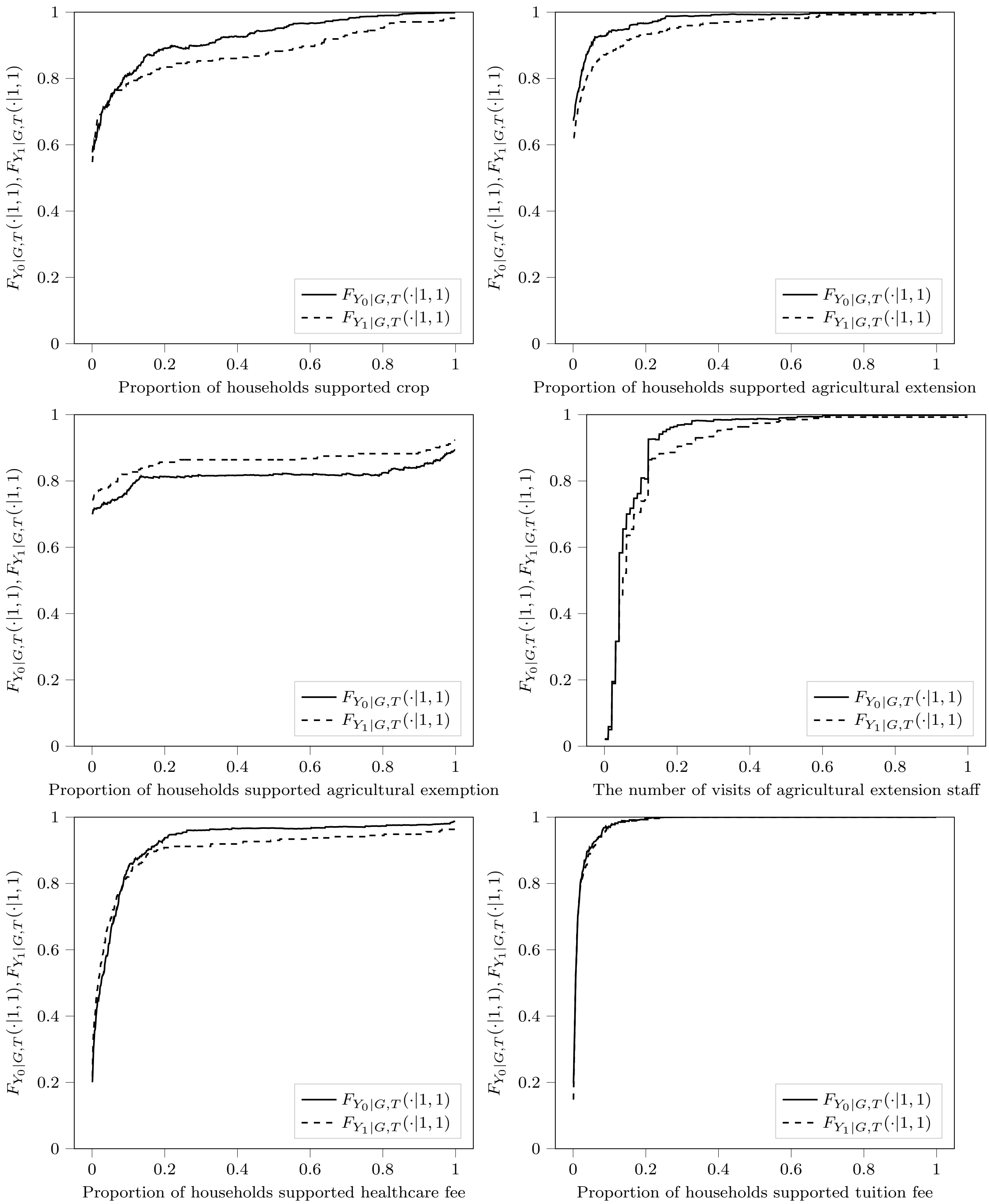
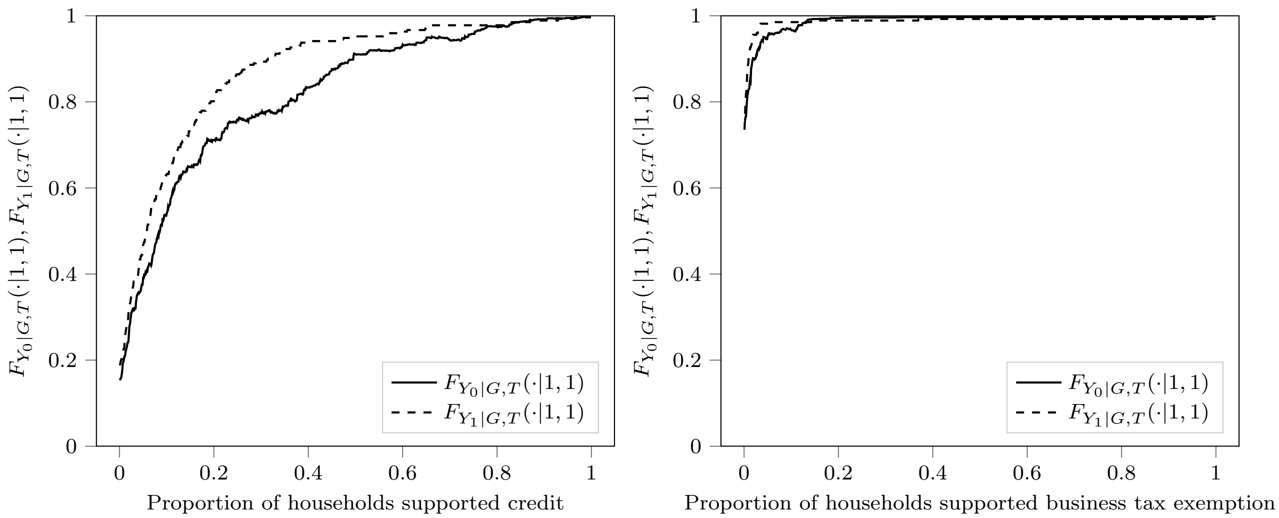


Figure 6: Results of changes-in-changes (continued).



We follow Callaway and Li (2019) and investigate a change in the mandatory state minimum wage in the 11 U.S. states which had their mandatory minimum wage below the federal minimum wage at the beginning of 2007. We employ these states as the treated group. Before this change there were an additional 22 states in which the state minimum wage was lower than the federal but we follow Callaway and Li (2019) and drop New Hampshire and Pennsylvania. This leaves a control group of 20 states.

Similar to Callaway and Li (2019), we first estimate the impact of the change in the state-level minimum wage on the county level unemployment rates. Callaway and Li (2019) employ the Local Area Unemployment Statistics Database from the Bureau of Labor Statistics (BLS) and control variables on population and income from the 2000 County Data Book. We use county data on the percentage of African Americans, the percentage of high school graduates, the percentage of college graduates, the log of the total population, the poverty rate and the log median income (for 1997).

The results using our method are shown in the first row of Table 7 and are quantitatively similar to those of Callaway and Li (2019). The impact on the unemployment rate is negative for those counties that have lower levels of the unemployment rate but it is positive for those counties that have higher unemployment rates. The degree of hetero-

Table 7: Quantile treatment effects of the increase in the mandatory minimum wage. The unemployment rate and poverty rate are measured in percentages.

	Mean	0.1	0.25	0.5	0.75	0.9
Unemployment rate	0.1242	-0.3182	-0.0886	0.1824	0.3398	0.3191
	(-0.104,0.353)	(-0.779,0.143)	(-0.524,0.346)	(-0.186,0.551)	(-0.188,0.868)	(-0.145,0.783)
(log) Average weekly	0.0015	-0.0153	-0.0124	-0.0157	-0.009	-0.001
wage	(-0.015,0.018)	(-0.044,0.013)	(-0.034,0.009)	(-0.042,0.011)	(-0.048,0.03)	(-0.04,0.038)
Poverty rate	0.1748	0.3	0.3	0.0374	0.0	0.0
	(-0.051,0.4)	(-0.301,0.901)	(-0.204,0.804)	(-0.489,0.564)	(-0.504,0.504)	(-0.612,0.612)

geneity in our results is smaller than Callaway and Li (2019). For example, they find an impact of -0.44 at the first decile compared to our estimate of -0.32. This results in our failure to reject the null hypothesis of no impact in any part of the distribution, whereas their impact at the bottom decile is marginally statistically significant.

We also merge the Callaway and Li (2019) data with average weekly wage data taken from the Quarterly Census of Employment and Wages (QCEW) sample from the BLS. Note that we employ data for the first quarter of the year 2006 and 2007. We acknowledge that the average weekly wage is not the ideal feature of the wage distribution to examine for this question but other wage measures were not available. However, the monopsony wages literature has frequently argued that the entire wage distribution may shift due to a change in the mandatory minimum wage (e.g. Van den Berg, 2003). This would have implications for the average wage. Using the same specification as for the unemployment rate equation we examine the impact on average wages and the results are reported in the second row of Table 7. The impact on the average weekly wage is small and statistically insignificant.

We also examine poverty rates from the state and county estimates published by the Census for 2006 and 2007. A shortcoming of these data is that poverty rates are reported for the whole of 2007. As reported by Callaway and Li (2019), the federal minimum wage increased at the end of July 2007 and this change may have an impact on the poverty rates in our control states. Under this proviso we explore the impact of the change in the minimum wage. We continue to use the same specification and the results are in the third row of Table 7. The point estimates suggests that the impact on the poverty rate appears somewhat larger than that on average weekly wages. In addition, the impact is larger for those counties that have low poverty rates. However, there is no evidence of a statistically significant impact.

Our evidence suggests that the univariate distributions are not significantly affected by the change in the mandatory minimum wage. To investigate whether the joint distributions are affected, Figures 7 to 9 report the 2-dimensional effects noting that the left-hand figures show the counterfactual distributions and the right-hand side show the observed distributions. As it is difficult to reach clear conclusions from these figures, Table 8 reports the changes in the Kendall's τ and Spearman's correlation index for the different pairs of outcome variables noting that the latter is computed as:

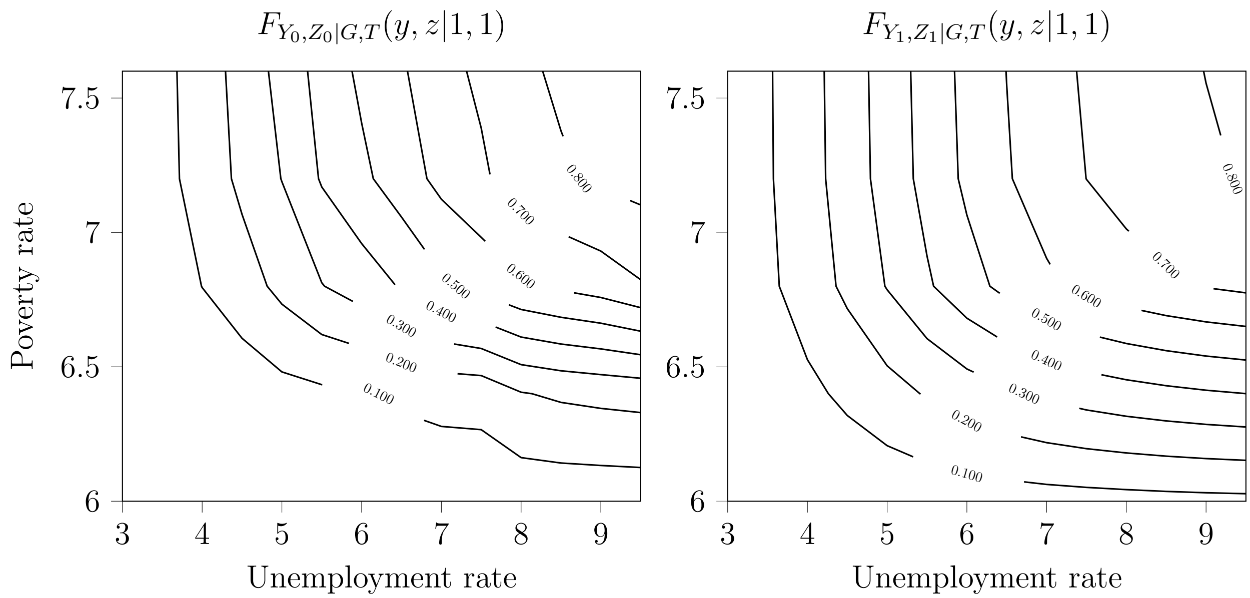
$$\rho_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where d_i is the difference in ranks between the two variables y_i and z_i. Note that we do not report the Spearman's correlation value for the previous empirical example due to the degree of bunching at certain values in the data.

In the absence of treatment the estimates of Kendall's τ and Spearman's correlation index for the unemployment rate and the average weekly wage are -0.1095 and -0.1316 respectively. Following the increase in the mandatory minimum wage, this negative relationship becomes weaker, with the corresponding estimate values of -0.0350 and -0.052. However, these changes are not statistically significantly different from zero at a 95 percent significance level. In contrast, the unemployment rate and the poverty rate have a positive correlation prior to the treatment with the respective estimates of the correlation being 0.256 and 0.329. This correlation becomes somewhat stronger after the increase of the mandatory minimum wage with estimates of 0.312 and 0.453. Once again the changes are not statistically significant.

Finally, we examine the correlation between the average weekly wage and the poverty rate. This relationship is also negative prior to treatment with the two estimates of the correlation being -0.3070 and -.413. This correlation also becomes weaker after treatment, increasing to -0.1762 and -.261, and in contrast to the earlier results these changes are statistically significant, or marginally insignificant, at the 95 percent significance level. While we leave a fuller explanation of this result to future work, this is an encouraging

Figure 7: Results of 2-dimensional effects of the unemployment rate and the (log) average weekly wage



result for those who support the increase in the minimum wage. One might have expected that the increase in minimum wage would result in higher average wages which result in higher unemployment and greater number of individuals in poverty. While the first relationship between wages and unemployment is consistent with the first row of Table 8, the second relationship is not supported by the data. This result also highlights the importance of our approach. An examination of the univariate distributions suggests there is no response to treatment. However, the changes in these correlation values suggests that the bivariate relationships are sensitive to the treatment. This provides greater insight into the treatment effects and the mechanisms underlying them.

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

Figure 8: Results of 2-dimensional effects of the unemployment rate and the poverty rate

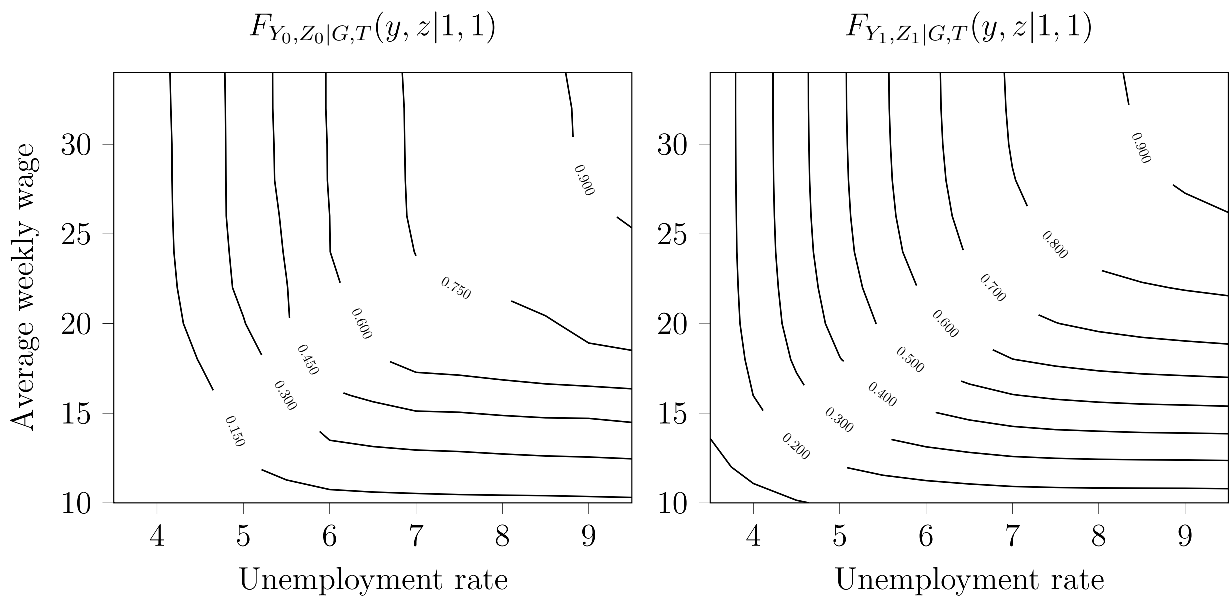


Figure 9: Results of 2-dimensional effects of the (log) average weekly wage and the poverty rate

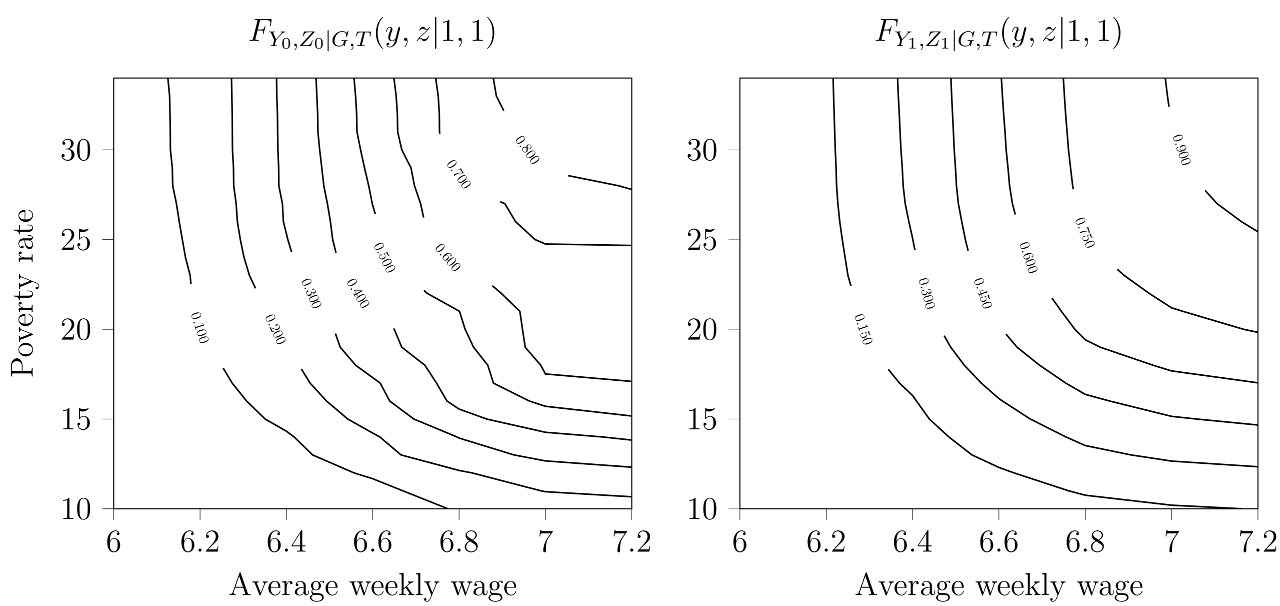
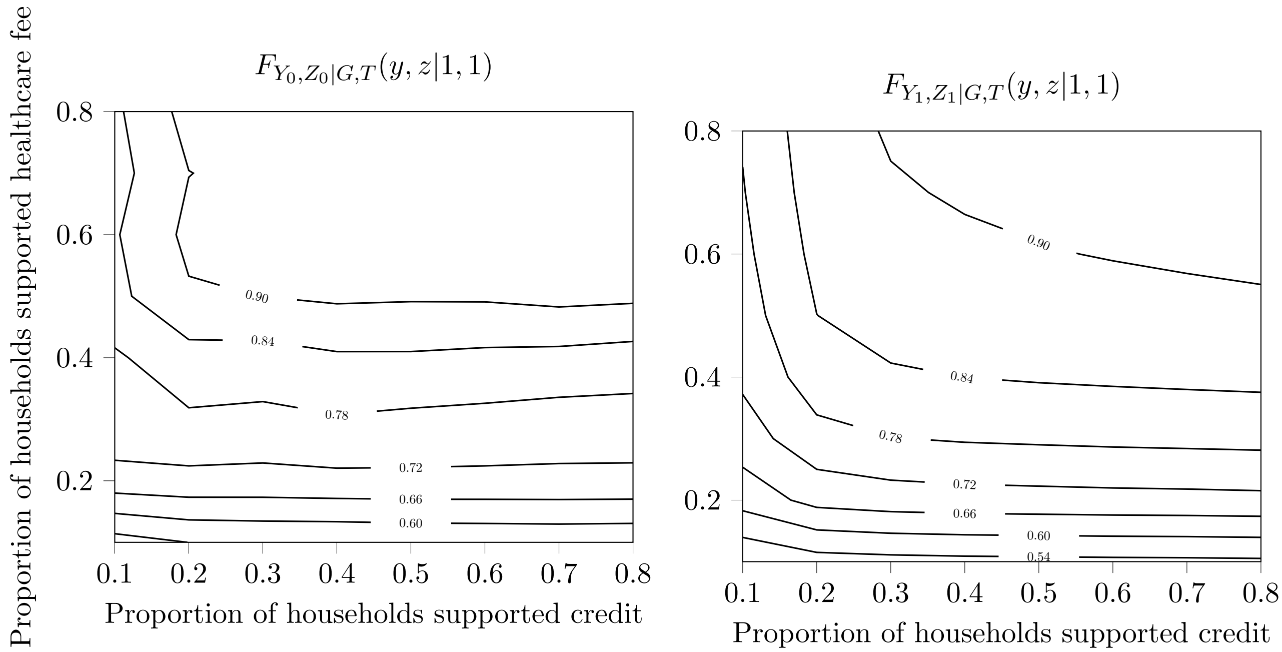


Table 8: Results of Kendall's τ for the treatment group when treated compared to not treated (with 95 percent confidence intervals between parentheses).

	Treatment	Without treatment	Difference
Kendall's τ			
Unemployment rate and (log) average weekly wage	-0.0350	-0.1095	0.0745
	(-0.0571, -0.0129)	(-0.1791, -0.0400)	(-0.0021, 0.1511)
Unemployment rate and poverty rate	0.3119	0.2564	0.0555
	(0.2957, 0.3281)	(0.0184, 0.4944)	(-0.1810, 0.2919)
(log) Average weekly wage and poverty rate	-0.1762	-0.3070	0.1307
	(-0.1950, -0.1574)	(-0.4308, -0.1832)	(0.0060, 0.2554)
Spearman's correlation index			
Unemployment rate and (log) average weekly wage	-0.0522	-0.1316	0.0793
	(-0.0855, -0.0190)	(-0.2134, -0.0497)	(-0.0123, 0.1709)
Unemployment rate and poverty rate	0.4525	0.3287	0.1238
	(0.4301, 0.4749)	(-0.0004, 0.6577)	(-0.2032, 0.4508)
(log) Average weekly wage and poverty rate	-0.2611	-0.4131	0.1520
	(-0.2883, -0.2340)	(-0.5822, -0.2441)	(-0.0144, 0.3184)

Figure 10: Results of 2-dimensional effects



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 two empirical example which revisits an existing studies and which highlight 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., 2024b).

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.

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.

Callaway, B. and T. Li (2019), "Quantile treatment effects in difference in differences models with panel data" *Quantitative Economics* **10**, 1579–1618.

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 Economic 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., J. Meier, A. van Vuuren and F. Vella (2024a) "Bivariate distribution regression with an application to intergenerational mobility", working paper, Boston University.

Fernández-Val I., J. Meier, A. van Vuuren and F. Vella (2024b) "Distributional synthetic difference-in-differences", 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.

MaCurdy, T. (2015), "How effective is the minimum wage at supporting the poor?", *Journal of Political Economy* **123**, 497–545.

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.

Torous, W., F. Gunsilius, and P. Rigollet (2024), "An optimal transport approach to estimating causal effects via nonlinear difference-in-differences, working paper", University of California, Berkeley.

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.