UNI FR

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

WORKING PAPERS SES

Direct and Indirect Effects based on Changes-in-Changes

Martin Huber,
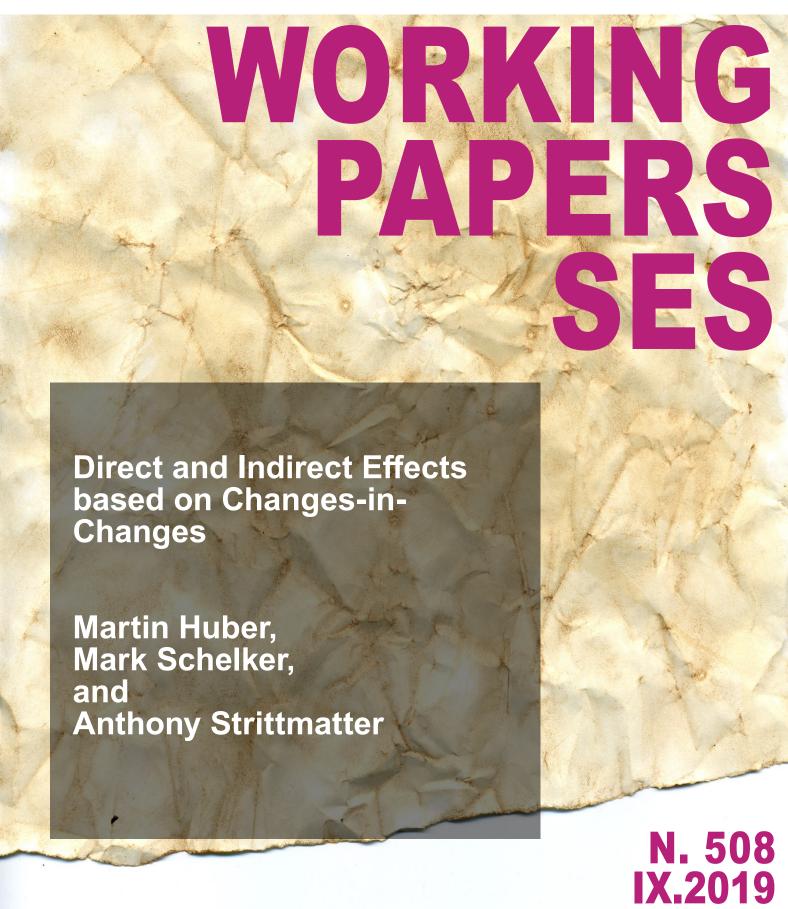## Mark Schelker,
and
Anthony Strittmatter

**N. 508
IX.2019**

Direct and Indirect Effects based on Changes-in-Changes

Martin Huber*, Mark Schelker*, Anthony Strittmatter+

*University of Fribourg, Dept. of Economics

+University of St. Gallen, Swiss Institute for Empirical Economic Research

Abstract: We propose a novel approach for causal mediation analysis based on changes-in-changes assumptions restricting unobserved heterogeneity over time. This allows disentangling the causal effect of a binary treatment on a continuous outcome into an indirect effect operating through a binary intermediate variable (called mediator) and a direct effect running via other causal mechanisms. We identify average and quantile direct and indirect effects for various subgroups under the condition that the outcome is monotonic in the unobserved heterogeneity and that the distribution of the latter does not change over time conditional on the treatment and the mediator. We also provide a simulation study and an empirical application to the Jobs II programme.

1   Introduction

Causal mediation analysis aims at disentangling a total treatment effect into an indirect effect operating through an intermediate variable – commonly referred to as mediator – as well as the direct effect. The latter includes any causal mechanisms not operating through the mediator of interest. Even when the treatment is random, direct and indirect effects are generally not identified by simply controlling for the mediator without accounting for its potential endogeneity, as this likely introduces selection bias, see Robins and Greenland (1992).

This paper suggests a novel identification strategy for causal mediation analysis based on changes-in-changes (CiC) as suggested by Athey and Imbens (2006) for evaluating (total) average and quantile treatment effects. We adapt the approach to the identification of the direct effect and the indirect effect running through a binary mediator. The outcome variable must be continuous and is required to be observed both prior to and after treatment and mediator assignment as it is the case in repeated cross sections or panel data. The key identifying assumptions imply that the continuous outcome is strictly monotonic in unobserved heterogeneity and that the distribution of unobserved heterogeneity does not change over time conditional on the treatment and the mediator (the latter assumption is also known as stationarity). Given appropriate common support conditions, this permits identifying direct effects on subpopulations conditional on the treatment and the mediator states, even if both treatment and mediator assignment are endogenous.

Augmenting the assumptions by random treatment assignment and weak monotonicity of the mediator in the treatment allows for causal mediation analysis in subpopulations defined upon whether and how the mediator reacts to the treatment. Specifically, we show the identification of direct effects among those whose mediator is always one (always-takers in the denomination of Angrist, Imbens, and Rubin, 1996) and never one (never-takers) irrespective of treatment assignment, respectively. Furthermore, we identify the total, direct, and indirect treatment effects on those whose mediator value complies with treatment assignment (compliers). For

1

any set of assumptions, we discuss the identification of both average and quantile direct and indirect effects. We note that if appropriately weighted, the respective average effects among compliers, always-takers, and never-takers add up to the average direct and indirect effects in the population.

Identification in the earlier mediation literature typically relied on linear models for the mediator and outcome equations and often neglected endogeneity issues, see for instance Cochran (1957), Judd and Kenny (1981), and Baron and Kenny (1986). More recent contributions use more general identification approaches based on the potential outcome framework and take endogeneity issues explicitly into consideration. Examples include Robins and Greenland (1992), Pearl (2001), Robins (2003), Petersen, Sinisi, and van der Laan (2006), VanderWeele (2009), Imai, Keele, and Yamamoto (2010), Hong (2010), Albert and Nelson (2011), Imai and Yamamoto (2013), Tchetgen Tchetgen and Shpitser (2012), Vansteelandt, Bekaert, and Lange (2012), and Huber (2014). The vast majority of the literature assumes that the covariates observed in the data are sufficiently rich to control for treatment and mediator endogeneity. Also in empirical economics, there has been an increase in the application of such selection on observables approaches, see for instance Simonsen and Skipper (2006), Flores and Flores-Lagunes (2009), Heckman, Pinto, and Savelyev (2013), Huber (2015), Keele, Tingley, and Yamamoto (2015), Conti, Heckman, and Pinto (2016), Huber, Lechner, and Mellace (2017), Bijwaard and Jones (2019), Bellani and Bia (2018), Huber, Lechner, and Strittmatter (2018), and Doerr and Strittmatter (2019). Comparably few studies in economics develop or apply instrumental variable approaches for disentangling direct and indirect effects, see for instance Frölich and Huber (2017), Powdthavee, Lekfuangfu, and Wooden (2013), Brunello, Fort, Schneeweis, and Winter-Ebmer (2016) and Chen, Chen, and Liu (2017). Our paper provides another, CiC-based identification strategy that neither rests on selection on observables assumptions nor on instrumental variables for the treatment or the mediator.

While most studies aim at evaluating direct and indirect effects in the total pop-

ulation, a smaller strand of the literature uses the principal stratification framework of Frangakis and Rubin (2002) to investigate effects in subpopulations (or principal strata) defined upon whether and how the mediator reacts to the treatment, see Rubin (2004). This approach has been criticized for typically focussing on direct effects on populations whose mediator is constant (i.e. always- and never-takers) rather than decomposing direct and indirect effects on compliers and for considering subpopulations rather than the total population, see VanderWeele (2008) and VanderWeele (2012). Deuchert, Huber, and Schelker (2017) suggest a difference-in-differences (DiD) strategy that alleviates such criticisms. Identification relies on a randomized treatment, monotonicity of the (binary) mediator in the treatment, and particular common trend assumptions on mean potential outcomes across principal strata. The latter imply that mean potential outcomes under specific treatment and mediator states change by the same amount over time across specific subpopulations. Depending on the strength of common trend and effect homogeneity assumptions across principal strata, direct and indirect effects are identified for different subpopulations and under the strongest set of assumptions even for the total population.

Our paper contributes to this literature on principal strata effects, but relies on different identifying assumptions than Deuchert, Huber, and Schelker (2017). While differential time trends across subpopulations are permitted, our approach restricts the conditional distribution of unobserved heterogeneity over time. The two sets of assumptions are not nested and their appropriateness is to be judged in the empirical context at hand. However, both approaches could be used simultaneously for testing the joint validity of the identifying assumptions of either method, in which case both CiC and DiD converge to the same, true average direct and indirect effects. As a further distinction to Deuchert, Huber, and Schelker (2017), our proposed method also permits assessing quantile treatment effects (QTEs) rather than average effects only.

We provide a simulation study in which we compare the CiC to the DiD approach to illustrate our identification results. We also consider an empirical application to

the Jobs II programme previously analysed by Vinokur, Price, and Schul (1995), a randomized job training intervention designed to analyse the impact of job training on labour market and mental health outcomes. We investigate the direct effect of the randomized offer of treatment on a depression index, as well as its indirect effect through actual participation in the programme as mediator. The reason for investigating the direct effect is that treatment assignment could have a motivation or discouragement effect on those randomly offered or not offered the training. We, however, find the direct effect estimates to be close to zero and statistically insignificant and therefore no indication for the violation of the exclusion restriction when using treatment assignment as instrumental variable for actual participation. In contrast, the moderately negative total and indirect effects on those induced to participate by assignment are statistically significant at least at the 10% level in all but one case and very much in line with the estimate obtained by instrumental variable regression.

The remainder of this study is organized as follows. Section 2 introduces the notation and defines the direct and indirect effects of interest. Section 3 presents the assumptions underlying our CiC approach as well as the identification results. Section 4 provides a simulation study. Section 5 provides an application to Jobs II. Section 6 concludes.

2   Notation and effects

2.1   Average effects

Let D denote a binary treatment (e.g., receiving the offer to participate in a training programme) and M a binary intermediate variable or mediator that may be a function of D (e.g., the actual participation in a training programme). Furthermore, let T indicate a particular time period: $T = 0$ denotes the baseline period prior to the realisation of D and M, $T = 1$ the follow up period after measuring D and M in which the effect of the outcome is evaluated. Finally, let Y_t denote the outcome of

4

interest (e.g., health measures) in period $T = t$. Indexing the outcome by the time period $t \in \{0, 1\}$ implies that it is measured both in the baseline period and after the realisation of D and M. To define the parameters of interest, we make use of the potential outcome notation, see for instance Rubin (1974), and denote by $Y_t(d, m)$ the potential outcome for treatment state $D = d$ and mediator state $M = m$ in time $T = t$, with $d, m, t, \in \{0, 1\}$. Furthermore, let $M(d)$ denote the potential mediator as a function of the treatment state $d \in \{0, 1\}$. For notational ease, we will not use any time index for D and M, because either is assumed to be measured at a single point in time between $T = 0$ and $T = 1$, albeit not necessarily at the same point, as D causally precedes M. Therefore, D and M correspond to the actual treatment and mediator status in $T = 1$, while it is assumed that no treatment or mediation takes place in $T = 0$.

Using this notation, the average treatment effect (ATE) in the ex-post period is defined as $\Delta_1 = E[Y_1(1, M(1)) - Y_1(0, M(0))]$. That is, the ATE corresponds to the effect of D on the outcome that either affects the latter directly (net of any effect on the mediator) or indirectly through an effect on M. Indeed, the total ATE can be disentangled into the direct and indirect effects, denoted by $\theta_1(d) = E[Y_1(1, M(d)) - Y_1(0, M(d))]$ and $\delta_1(d) = E[Y_1(d, M(1)) - Y_1(d, M(0))]$, by adding and subtracting $Y_1(1, M(0))$ or $Y_1(0, M(1))$, respectively:

$$
\begin{aligned}
\Delta_1 &= E[Y_1(1, M(1)) - Y_1(0, M(0))], \\
&= \underbrace{E[Y_1(1, M(1)) - Y_1(1, M(0))]}_{=\delta_1(1)} + \underbrace{E[Y_1(1, M(0)) - Y_1(0, M(0))]}_{=\theta_1(0)}, \\
&= \underbrace{E[Y_1(1, M(1)) - Y_1(0, M(1))]}_{=\theta_1(1)} + \underbrace{E[Y_1(0, M(1)) - Y_1(0, M(0))]}_{=\delta_1(0)}.
\end{aligned}
$$

Distinguishing between $\theta_1(1)$ and $\theta_1(0)$ or $\delta_1(1)$ and $\delta_1(0)$, respectively, implies the possibility of interaction effects between D and M such that the direct and indirect effects could be heterogeneous across values $d = 1$ and $d = 0$.

In our approach, we consider the concepts of direct and indirect effects within

specific subpopulations. The latter are either defined conditional on the treatment and mediator values or conditional on potential mediator values under either treatment states, which matches the so-called principal stratum framework of Frangakis and Rubin (2002). As outlined in Angrist, Imbens, and Rubin (1996) in the context of instrumental variable-based identification, any individual i in the population belongs to one of four strata, henceforth denoted by τ, according to their potential mediator status under either treatment state: always-takers (a: $M(1) = M(0) = 1$) whose mediator is always one, compliers (c: $M(1) = 1$, $M(0) = 0$) whose mediator corresponds to the treatment value, defiers (de: $M(1) = 0$, $M(0) = 1$) whose mediator opposes the treatment value, and never-takers (n: $M(1) = M(0) = 0$) whose mediator is never one. Note that τ cannot be pinned down for any individual, because either $M(1)$ or $M(0)$ is observed, but never both.

Let $\Delta_1^\tau = E[Y_1(1, M(1)) - Y_1(0, M(0))|\tau]$ denote the ATE conditional on $\tau \in \{a, c, de, n\}$; $\theta_1^\tau(d)$ and $\delta_1^\tau(d)$ denote the corresponding direct and indirect effects. Because $M(1) = M(0) = 0$ for any never-taker, the indirect effect for this group is by definition zero ($\delta_1^n(d) = E[Y_1(d, 0) - Y_1(d, 0)|\tau = n] = 0$) and $\Delta_1^n = E[Y_1(1, 0) - Y_1(0, 0)|\tau = n] = \theta_1^n(1) = \theta_1^n(0) = \theta_1^n$ equals the direct effect for never-takers. Correspondingly, because $M(1) = M(0) = 1$ for any always-taker, the indirect effect for this group is by definition zero ($\delta_1^a(d) = E[Y_1(d, 1) - Y_1(d, 1)|\tau = a] = 0$) and $\Delta_1^a = E[Y_1(1, 1) - Y_1(0, 1)|\tau = a] = \theta_1^a(1) = \theta_1^a(0) = \theta_1^a$ equals the direct effect for always-takers. For the compliers, both direct and indirect effects may exist. Note that $M(d) = d$ due to the definition of compliers. Accordingly, $\theta_1^c(d) = E[Y_1(1, d) - Y_1(0, d)|\tau = c]$ equals the direct effect for compliers, $\delta_1^c(d) = E[Y_1(d, 1) - Y_1(d, 0)|\tau = c]$ equals the indirect effect for compliers, and $\Delta_1^c = E[Y_1(1, 1) - Y_1(0, 0)|\tau = c]$ equals the total effect for compliers. In the absence of any direct effect, the indirect effects on the compliers are homogeneous, $\delta_1^c(1) = \delta_1^c(0) = \delta_1^c$, and correspond to the local average treatment effect (LATE, e.g., Angrist, Imbens, and Rubin, 1996). Analogous results hold for the defiers.

As already mentioned, we will also consider direct effects conditional on specific

values $D = d$ and mediator states $M = M(d) = m$, which are denoted by $\theta_1^{d,m}(d) = E[Y_1(1,m) - Y_1(0,m)|D = d, M(d) = m]$. These parameters are identified under weaker assumptions than strata-specific effects, but are also less straightforward to interpret, as they refer to mixtures of two strata. For instance, $\theta_1^{1,0}(1) = E[Y_1(1,0) - Y_1(0,0)|D = 1, M(1) = 0]$ is the effect on a mixture of never-takers and defiers, as these two groups satisfy $M(1) = 0$. Likewise, $\theta_1^{0,0}(0)$ refers to never-takers and compliers satisfying $M(0) = 0$, $\theta_1^{0,1}(0)$ to always-takers and defiers satisfying $M(0) = 1$, and $\theta_1^{1,1}(1)$ to always-takers and compliers satisfying $M(1) = 1$.

2.2 Quantile effects

We denote by $F_{Y_t(d,m)}(y) = \Pr(Y_t(d,m) \leq y)$ the cumulative distribution function of $Y_t(d,m)$ at outcome level y. Its inverse, $F_{Y_t(d,m)}^{-1}(q) = \inf\{y : F_{Y_t(d,m)}(y) \geq q\}$, is the quantile function of $Y_t(d,m)$ at rank q. The total QTE are denoted by $\Delta_1(q) = F_{Y_1(1,M(1))}^{-1}(q) - F_{Y_1(0,M(0))}^{-1}(q)$. The QTE can be disentangled into the direct quantile effects, denoted by $\theta_1(q,d) = F_{Y_1(1,M(d))}^{-1}(q) - F_{Y_1(0,M(d))}^{-1}(q)$, and the indirect quantile effects, denoted by $\delta_1(q,d) = F_{Y_1(d,M(1))}^{-1}(q) - F_{Y_1(d,M(0))}^{-1}(q)$.

The conditional distribution function in stratum τ is $F_{Y_t(d,m)|\tau}(y) = \Pr(Y_t(d,m) \leq y|\tau)$ and the corresponding conditional quantile function is $F_{Y_t(d,m)|\tau}^{-1}(q) = \inf\{y : F_{Y_t(d,m)|\tau}(y) \geq q\}$ for $\tau \in \{a,c,d,n\}$. Using the previously described stratification framework, we define the QTE conditional on $\tau \in \{a,c,de,n\}$: $\Delta_1^\tau(q) = F_{Y_1(1,M(1))|\tau}^{-1}(q) - F_{Y_1(0,M(0))|\tau}^{-1}(q)$. The direct quantile treatment effect among never-takers equals $\Delta_1^n(q) = F_{Y_1(1,0)|n}^{-1}(q) - F_{Y_1(0,0)|n}^{-1}(q) = \theta_1^n(q)$. The direct quantile effect among always-takers equals $\Delta_1^a(q) = F_{Y_1(1,1)|a}^{-1}(q) - F_{Y_1(0,1)|a}^{-1}(q) = \theta_1^a(q)$. The total QTE among compliers equals $\Delta_1^c(q) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$, the direct quantile effect among compliers equals $\theta_1^c(q,d) = F_{Y_1(1,d)|c}^{-1}(q) - F_{Y_1(0,d)|c}^{-1}(q)$, and the indirect quantile effect among compliers equals $\delta_1^c(q,d) = F_{Y_1(d,1)|c}^{-1}(q) - F_{Y_1(d,0)|c}^{-1}(q)$. Finally, we define the direct quantile treatment effects conditional on specific values $D = d$

and mediator states $M = M(d) = m$,

$$\theta_1^{d,m}(q,1) = F^{-1}_{Y_1(1,m)|D=d,M(1)=m}(q) - F^{-1}_{Y_1(0,m)|D=d,M(1)=m}(q) \text{ and}$$

$$\theta_1^{d,m}(q,0) = F^{-1}_{Y_1(1,m)|D=d,M(0)=m}(q) - F^{-1}_{Y_1(0,m)|D=d,M(0)=m}(q),$$

with the quantile function $F^{-1}_{Y_t(d,m)|D=d,M(d)=m}(q) = \inf\{y : F_{Y_t(d,m)|D=d,M(d)=m}(y) \geq q\}$ and the distribution function $F_{Y_t(d,m)|D=d,M(d)=m}(y) = \Pr(Y_t(d,m) \leq y|D = d, M(d) = m)$.

2.3 Observed distribution and quantile transformations

We subsequently define various functions of the observed data required for the identification results. The conditional distribution function of the observed outcome Y_t conditional on treatment value d and mediator state m, is given by $F_{Y_t|D=d,M=m}(y) = \Pr(Y_t \leq y|D = d, M = m)$ for $d, m \in \{0,1\}$. The corresponding conditional quantile function is $F^{-1}_{Y_t|D=d,M=m}(q) = \inf\{y : F_{Y_t|D=d,M=m}(y) \geq q\}$. Furthermore,

$$Q_{dm}(y) := F^{-1}_{Y_1|D=d,M=m} \circ F_{Y_0|D=d,M=m}(y) = F^{-1}_{Y_1|D=d,M=m}(F_{Y_0|D=d,M=m}(y))$$

is the quantile-quantile transform of the conditional outcome from period 0 to 1 given treatment d and mediator status m. This transform maps y at rank q in period 0 ($q = F_{Y_0|D=d,M=m}(y)$) into the corresponding y' at rank q in period 1 ($y' = F^{-1}_{Y_1|D=d,M=m}(q)$).

3 Identification and Estimation

3.1 Identification

This sections discusses the identifying assumptions along with the identification results for the various direct and indirect effects. We note that our assumptions could be adjusted to only hold conditional on a vector of observed covariates. In

this case, the identification results would hold within cells defined upon covariate values. In our main discussion, however, covariates are not considered for the sake of ease of notation. For notational convenience, we maintain throughout that $\Pr(T = t, D = d, M = m) > 0$ for $t, d, m \in \{1, 0\}$, implying that all possible treatment-mediator combinations exist in the population in both time periods. Our first assumption implies that potential outcomes are characterized by a continuous nonparametric function, denoted by h, that is strictly monotonic in a scalar U that reflects unobserved heterogeneity.

Assumption 1: Strict monotonicity of continuous potential outcomes in unobserved heterogeneity.

The potential outcomes satisfy the following model: $Y_t(d, m) = h(d, m, t, U)$, with the general function h being continuous and strictly increasing in the scalar unobservable $U \in \mathbb{R}$ for all $d, m, t \in \{0, 1\}$.

Assumption 1 requires the potential outcomes to be continuous implying that there is a one-to-one correspondence between a potential outcome's distribution and quantile functions, which is a condition for point identification. For discrete potential outcomes, only bounds on the effects could be identified, in analogy to the discussion in Athey and Imbens (2006) for total (rather than direct and indirect) effects. Assumption 1 also implies that individuals with identical unobserved characteristics U have the same potential outcomes $Y_t(d, m)$, while higher values of U correspond to strictly higher potential outcomes $Y_t(d, m)$. Strict monotonicity is automatically satisfied in additively separable models, but Assumption 1 also allows for more flexible non-additive structures that arise in nonparametric models.

The next assumption rules out anticipation effects of the treatment or the mediator on the outcome in the baseline period. This assumption is plausible if assignment to the treatment or the mediator cannot be foreseen in the baseline period, such that behavioral changes affecting the pre-treatment outcome are ruled out.

Assumption 2: No anticipation effect of M and D in the baseline period.

$Y_0(d, m) - Y_0(d', m') = 0$, for $d, d', m, m'\{1, 0\}$.

9

Similarly, Athey and Imbens (2006) and Chaisemartin and D'Haultfeuille (2018) assume the assignment to the treatment group does not affect the potential outcomes as long as the treatment is not yet realized.

Furthermore, we assume conditional independence between unobserved heterogeneity and time periods given the treatment and no mediation.

Assumption 3: Conditional independence of U and T given $D = 1, M = 0$ or $D = 0, M = 0$.

(a) $U \perp\!\!\!\perp T | D = 1, M = 0$,

(b) $U \perp\!\!\!\perp T | D = 0, M = 0$.

Under Assumption 3a, the distribution of U is allowed to vary across groups defined upon treatment and mediator state, but not over time within the group with $D = 1, M = 0$. Assumption 3b imposes the same restriction conditional on $D = 0, M = 0$. Assumption 3 thus imposes stationarity of U within groups defined on D and M. This assumption is weaker than (and thus implied by) requiring that U is constant across T for each individual i. For example, Assumption 3 is satisfied in the fixed effect model $U = \eta + v_t$, with η being a time-invariant individual-specific unobservable (fixed effect) and v_t an idiosyncratic time-varying unobservable with the same distribution in both time periods.

Athey and Imbens (2006) and Chaisemartin and D'Haultfeuille (2018) impose time invariance conditional on the treatment status, $U \perp\!\!\!\perp T | D = d$, to identify the average treatment effect on the treated, $\varphi_1 = E[Y_1(1, M(1)) - Y_1(0, M(0)) | D = 1]$ or local average treatment effect, $\varphi_1 = E[Y_1(1, M(1)) - Y_1(0, M(0)) | \tau = c]$, respectively. We additionally condition on the mediator status to identify direct and indirect effects.

For our next assumption, we introduce some further notation. Let $F_{U|d,m}(u)) = \Pr(U \le u | D = d, M = m)$ be the conditional distribution of U with support $\mathbb{U}_{dm}$.

Assumption 4: Common support given $M = 0$.

(a) $\mathbb{U}_{10} \subseteq \mathbb{U}_{00}$,

(b) $\mathbb{U}_{00} \subseteq \mathbb{U}_{10}$.

Assumption 4a is a common support assumption, implying that any possible value of U in the population with $D = 1, M = 0$ is also contained in the population with $D = 0, M = 0$. Assumption 4b imposes that any value of U conditional on $D = 0, M = 0$ also exists conditional on $D = 1, M = 0$. Both assumptions together imply that the support of U is the same in both populations, albeit the distributions may generally differ.

Assumptions 1 to 3 permit identifying direct effects on mixed populations of never-takers and defiers as well as never-takers and compliers, respectively, as formally stated in Theorem 1.

Theorem 1: Under Assumptions 1–3,

(a) and Assumption 4a, the average and quantile direct effects under $d = 1$ conditional on $D = 1$ and $M(1) = 0$ are identified:

$$\theta_1^{1,0}(1) = E[Y_1 - Q_{00}(Y_0)|D = 1, M = 0],$$
$$\theta_1^{1,0}(q, 1) = F_{Y_1|D=1,M=0}^{-1}(q) - F_{Q_{00}(Y_0)|D=1,M=0}^{-1}(q).$$

(b) and Assumption 4b, the average and quantile direct effects under $d = 0$ conditional on $D = 0$ and $M(0) = 0$ are identified:

$$\theta_1^{0,0}(0) = E[Q_{10}(Y_0) - Y_1|D = 0, M = 0],$$
$$\theta_1^{0,0}(q, 0) = F_{Q_{10}(Y_0)|D=0,M=0}^{-1}(q) - F_{Y_1|D=0,M=0}^{-1}(q).$$

Proof. See Appendix A.

To identify direct effects on further populations, we invoke a conditional independence assumption that is in the spirit of Assumption 3, but refers to different combinations of the treatment and the mediator.

Assumption 5: Conditional independence of U and T given $D = 0, M = 1$ or $D = 1, M = 1$.

(a) $U \perp\!\!\!\perp T|D = 0, M = 1,$

(b) $U \perp\!\!\!\perp T | D = 1, M = 1$.

Under Assumption 5a, the distribution of U is allowed to vary by treatment and mediator group, but not over time conditional on $D = 0, M = 1$. Assumption 5b imposes the same restriction conditional on $D = 1, M = 1$.

Assumption 6 is similar to Assumption 4, but imposes common support conditional on $M = 1$ rather than $M = 0$.

Assumption 6: Common support given $M = 1$.

(a) $\mathbb{U}_{01} \subseteq \mathbb{U}_{11}$,

(b) $\mathbb{U}_{11} \subseteq \mathbb{U}_{01}$.

Assumptions 6a implies that any possible value of U in the population with $D = 0, M = 1$ is also contained in the population with $D = 1, M = 1$. Assumptions 6b states that any value of U conditional on $D = 1, M = 1$ exists conditional on $D = 0, M = 1$.

Theorem 2 shows the identification of the direct effects on mixed populations of always-takers and defiers as well as always-takers and compliers.

Theorem 2: Under Assumptions 1-2, 5,

(a) and Assumption 6a, the average and quantile direct effects under $d = 0$ conditional on $D = 0$ and $M(0) = 1$ are identified:

$$\theta_1^{0,1}(0) = E[Y_1 - Q_{11}(Y_0)|D = 0, M = 1],$$

$$\theta_1^{0,1}(q, 0) = F_{Y_1|D=0,M=1}^{-1}(q) - F_{Q_{11}(Y_0)|D=0,M=1}^{-1}(q).$$

(b) and Assumption 6b, the average and quantile direct effects under $d = 1$ is identified conditional on $D = 1$ and $M(1) = 1$ are identified:

$$\theta_1^{1,1}(1) = E[Q_{01}(Y_0) - Y_1|D = 1, M = 1],$$

$$\theta_1^{1,1}(q, 1) = F_{Q_{01}(Y_0)|D=1,M=1}^{-1}(q) - F_{Y_1|D=1,M=1}^{-1}(q).$$

Proof. See Appendix B.

In the instrumental variable framework, any direct effects of the instrument are typically ruled out by imposing the exclusion restriction, in order to identify the causal effect of an endogenous regressor on the outcome, see for instance Imbens and Angrist (1994). By considering D as instrument and M as endogenous regressor, $\theta_1^{1,0}(1) = \theta_1^{0,0}(0) = \theta_1^{0,1}(0) = \theta_1^{1,1}(1) = 0$ yield testable implications of the exclusion restriction under Assumptions 1-6.

So far, we did not impose exogeneity of the treatment or mediator. In the following, we assume treatment exogeneity by invoking independence between the treatment and the potential post-treatment variables.

Assumption 7: Independence of the treatment and potential mediators/outcomes.
$\{Y_t(d,m), M(d)\} \perp\!\!\!\perp D$, for all $d, m, t, \in \{0, 1\}$.

Assumption 7 implies that there are no confounders jointly affecting the treatment on the one hand and the mediator and/or outcome on the other hand. It is satisfied under treatment randomization as in successfully conducted experiments. This allows identifying the ATE: $\Delta_1 = E[Y_1|D=1] - E[Y_1|D=0]$.

Furthermore, we assume the mediator to be weakly monotonic in the treatment.

Assumption 8: Weak monotonicity of the mediator in the treatment.
$\Pr(M(1) \geq M(0)) = 1$.

Assumption 8 is standard in the instrumental variable literature on local average treatment effects when denoting by D the instrument and by M the endogenous regressor, see Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). It rules out the existence of defiers.

As discussed in the Appendix C, the total ATE $\Delta_1 = E[Y_1|D=1] - E[Y_1|D=0]$ and QTE $\Delta_1(q) = F_{Y_1|D=1}^{-1}(q) - F_{Y_1|D=0}^{-1}(q)$ for the entire population are identified under Assumption 7. Furthermore, Assumptions 7 and 8 yield the strata proportions, denoted by $p_\tau = \Pr(\tau)$, as functions of the conditional mediator probabilities given the treatment, which we denote by $p_{(m|d)} = \Pr(M = m|D = d)$ for $d, m \in \{0, 1\}$

(see Appendix C):

$$p_a = p_{1|0}, p_c = p_{1|1} - p_{1|0} = p_{0|0} - p_{0|1}, p_n = p_{0|1}. \tag{1}$$

Furthermore, Assumptions 2, 7, and 8 imply that (see Appendix C)

$$\Delta_{0,c} = E[Y_0(1,1) - Y_0(0,0)|c] = \frac{E[Y_0|D=1] - E[Y_0|D=0]}{p_{1|1} - p_{1|0}} = 0. \tag{2}$$

Therefore, a rejection of the testable implication $E[Y_0|D=1] - E[Y_0|D=0] = 0$ in the data would point to a violation of these assumptions.

Assumptions 7 and 8 permit identifying additional parameters, namely the total, direct, and indirect effects on compliers, and the direct effects on never- and always-takers, as shown in Theorems 3 to 5. This follows from the fact that defiers are ruled out and that the proportions and potential outcome distributions of the various principal strata are not selective w.r.t. the treatment.

Theorem 3: Under Assumptions 1–3, 7-8,

a) and Assumption 4a, the average and quantile direct effects on never-takers are identified:

$$\theta_1^n = \theta_1^{1,0}(1) \text{ and } \theta_1^n(q) = \theta_1^{1,0}(q,1).$$

b) and Assumption 4, the average direct effect under $d = 0$ on compliers is identified:

$$\theta_1^c(0) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} \theta_1^{0,0}(0) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} \theta_1^{1,0}(1).$$

Furthermore, the potential outcome distributions under $d = 0$ on compliers are identified:

$$F_{Y_1(1,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Q_{10}(Y_0)|D=0,M=0}(y) \\ - \frac{p_{0|1}}{p_{0|0} - p_{0|1}c} F_{Y_1|D=1,M=0}(y), \tag{3}$$

$$F_{Y_1(0,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Y_1|D=0,M=0}(y)$$
$$- \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Q_{00}(Y_0)|D=1,M=0}(y). \tag{4}$$

Therefore, the direct quantile effect under $d = 0$ on compliers, $\theta_1^c(q, 0) = F_{Y_1(1,0)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$, is identified.

Proof. See Appendix D.

Theorem 4: Under Assumptions 1–2, 5, 7-8,

a) and Assumption 6a, the average and quantile direct effects on always-takers are identified:

$$\theta_1^a = \theta_1^{0,1}(0) \text{ and } \theta_1^a(q) = \theta_1^{0,1}(q, 0).$$

b) and Assumption 6, the average direct effect under $d = 1$ on compliers is identified:

$$\theta_1^c(1) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} \theta_1^{1,1}(1) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} \theta_1^{0,1}(0).$$

Furthermore, the potential outcome distributions under $d = 1$ for compliers are identified:

$$F_{Y_1(1,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} F_{Y_1|D=1,M=1}(y)$$
$$- \frac{p_{1|0}}{p_{1|1} - p_{1|0}} F_{Q_{11}(Y_0)|D=0,M=1}(y), \tag{5}$$

$$F_{Y_1(0,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} F_{Q_{01}(Y_0)|D=1,M=1}(y)$$
$$- \frac{p_{1|0}}{p_{1|1} - p_{1|0}} F_{Y_1|D=0,M=1}(y). \tag{6}$$

Therefore, the direct quantile effect under $d = 1$ on compliers $\theta_1^c(q, 1) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(0,1)|c}^{-1}(q)$ is identified.

Proof. See Appendix E.

Theorem 5: Under Assumptions 1-3, 5, 7-8,

a) and Assumptions 4a, 6a, the total average treatment effect on compliers is identified:

$$\Delta_1^c = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_1|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D = 0, M = 1]$$
$$- \frac{p_{0|0}}{p_{1|1} - p_{1|0}} E[Y_1|D = 0, M = 0] + \frac{p_{0|1}}{p_{1|1} - p_{1|0}} E[Q_{00}(Y_0)|D = 1, M = 0].$$

Furthermore, the total quantile treatment effect on compliers $\Delta_1^c(q) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$ is identified using the inverse of (5) and (4).

b) and Assumptions 4a, 6b, the average indirect effect under $d = 0$ on compliers is identified:

$$\delta_1^c(0) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Y_1|D = 0, M = 1]$$
$$- \frac{p_{0|0}}{p_{1|1} - p_{1|0}} E[Y_1|D = 0, M = 0] + \frac{p_{0|1}}{p_{1|1} - p_{1|0}} E[Q_{00}(Y_0)|D = 1, M = 0].$$

Furthermore, the quantile indirect effect under $d = 0$ on compliers $\delta_1^c(q, 0) = F_{Y_1(0,1)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$ is identified using the inverse of (6) and (4).

c) and Assumptions 4b, 6a, the average indirect effect under $d = 1$ on compliers is identified:

$$\delta_1^c(1) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_1|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D = 0, M = 1]$$
$$- \frac{p_{0|0}}{p_{1|1} - p_{1|0}} E[Q_{00}(Y_0)|D = 0, M = 0] + \frac{p_{0|1}}{p_{1|1} - p_{1|0}} E[Y_1|D = 1, M = 0].$$

Furthermore, the quantile indirect effect under $d = 1$ on compliers $\delta_1^c(q, 1) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(1,0)|c}^{-1}(q)$ is identified using the inverse of (5) and (3).

Proof. See Appendix F.

3.2 Estimation

As in Assumption 5.1 of Athey and Imbens (2006), we assume standard regularity conditions, namely that conditional on $T = t$, $D = d$, and $M = m$, Y is a random draw from that subpopulation defined in terms of $t, d, m \in \{1, 0\}$. Furthermore, the outcome in the subpopulations required for the identification results of interest must have compact support and a density that is bounded from above and below as well as continuously differentiable. Denote by N the total sample size across both periods and all treatment-mediator combinations and by $i \in \{1, ..., N\}$ an index for the sampled subject, such that (Y_i, D_i, M_i, T_i) correspond to sample realizations of the random variables (Y, D, M, T).

The total, direct, and indirect effects may be estimated using the sample analogy principle, which replaces population moments with sample moments (e.g. Manski, 1988). For instance, any conditional mediator probability given the treatment, $\Pr(M = m | D = d)$, is to be replaced by an estimate thereof in the sample, $\frac{\sum_{i=1}^{N} I\{M_i = m, D_i = d\}}{\sum_{i=1}^{N} I\{D_i = d\}}$. A crucial step is the estimation of the quantile-quantile transforms. The application of such quantile transformations dates at least back to Juhn, Murphy, and Pierce (1991), see also Chaisemartin and D'Haultfeuille (2018), Wüthrich (2019), and Strittmatter (2019) for recent applications. First, it requires estimating the conditional outcome distribution, $F_{Y_t | D=d, M=m}(y)$, by the conditional empirical distribution $\hat{F}_{Y_t | D=d, M=m}(y) = \frac{1}{\sum_{i=1}^{n} I\{D_i = d, M_i = m, T_i = t\}} \sum_{i: D_i = d, M_i = m, T_i = t} I\{Y_i \leq y\}$. Second, inverting the latter yields the empirical quantile function $\hat{F}_{Y_t | D=d, M=m}^{-1}(q)$. The empirical quantile-quantile transform is then obtained by

$$\hat{Q}_{dm}(y) = \hat{F}_{Y_1 | D=d, M=m}^{-1}(\hat{F}_{Y_0 | D=d, M=m}(y)).$$

This permits estimating the average and quantile effects of interest. Average effects are estimated by replacing any (conditional) expectations with the corresponding sample averages in which the estimated quantile-quantile transforms enter as plug-

in estimates. Taking $\theta_1^{1,0}$ (see Theorem 1) as an example, an estimate thereof is

$$
\begin{aligned}
\hat{\theta}_1^{1,0}(1) \;=\; & \frac{1}{\sum_{i=1}^n I\{D_i = 1, M_i = 0, T_i = 1\}} \sum_{i:D_i=1,M_i=0,T_i=1} Y_i \\
& - \frac{1}{\sum_{i=1}^n I\{D_i = 1, M_i = 0, T_i = 0\}} \sum_{i:D_i=1,M_i=0,T_i=0} \hat{Q}_{00}(Y_i).
\end{aligned}
$$

Likewhise, quantile effects are estimated based on the empirical quantiles.

For the estimation of total ATE and QTE, Athey and Imbens (2006) show that the resulting estimators are $\sqrt{N}$-consistent and asymptotically normal, see their Theorems 5.1 and 5.3. These properties also apply to our context when splitting the sample into subgroups based on the values of a binary treatment and mediator (rather than the treatment only). For instance, the implications of Theorem 1 in Athey and Imbens (2006) when considering subsamples with $D = 1$ and $D = 0$ carry over to considering subsamples with $D = 1, M = 0$ and $D = 0, M = 0$ for estimating the average direct effect on never-takers. In contrast to Athey and Imbens (2006), however, some of our identification results include the conditional mediator probabilities $\Pr(M = m | D = d)$. As the latter are estimated with $\sqrt{N}$-consistency, too, it follows that the resulting effect estimators are again $\sqrt{N}$-consistent and asymptotically normal. We use a non-parametric bootstrap approach to calculate the standard errors. Chaisemartin and D'Haultfeuille (2018) show the validity of the bootstrap approach for such kind of estimators, which follows from their asymptotic normality.

For the case that identifying assumptions to only hold conditional on observed covariates, denoted by X, estimation must be adapted to allow for control variables. Following a suggestion by Athey and Imbens (2006) in their Section 5.1, basing estimation on outcome residuals in which the association of X and Y has been purged by means of a regression is consistent under the additional assumption that the effects of D and M are homogeneous across covariates. As an alternative, Melly and Santangelo (2015) propose a flexible semiparametric estimator that does not impose such a homogeneity-in-covariates assumption and show $\sqrt{N}$-consistency and asymptotic normality.

4 Simulations

To shape the intuition for our identification results, this section presents a brief simulation based on the following data generating process (DGP):

$$T \sim Binom(0.5), \ D \sim Binom(0.5), \ U \sim Unif(-1,1), \ V \sim N(0,1)$$

independent of each other, and

$$M = I\{D + U + V > 0\}, \qquad Y_T = \Lambda((1 + D + M + D \cdot M) \cdot T + U).$$

Treatment D as well as the observed time period T are randomized, while the mediator-outcome association is confounded due to the unobserved time constant heterogeneity U. The potential outcome in period 1 is given by $Y_1(d, M(d')) = \Lambda((1 + d + M(d') + d \cdot M(d')) + U)$, where Λ denotes a link function. If the latter corresponds to the identity function, our model is linear and implies a homogeneous time trend T equal to 1. If Λ is nonlinear, the time trend is heterogeneous, which invalidates the common trend assumption of difference-in-differences models. M is not only a function of D and U, but also of the unobserved random term V, which guarantees common support w.r.t. U, see Assumptions 4 and 6. Compliers, always-takers, and never-takers satisfy, respectively: $c = I\{U + V \leq 0, 1 + U + V > 0\}$, $a = I\{U + V > 0\}$, and $n = I\{1 + U + V \leq 0\}$.

In the simulations with 1,000 replications, we consider two sample sizes ($N = 1,000, 4,000$) and investigate the behaviour of our change-in-changes methods as well as the difference-in-differences approach of Deuchert, Huber, and Schelker (2017) in both a linear (Λ equal to identity function) and nonlinear outcome model where Λ equals the exponential function. To implement the change-in-changes estimators in the simulations as well as the application in Section 5, we make use of the 'cic' command in the `qte` R-package by Callaway (2016) with its default values.

Table 1 reports the bias, standard deviation ('sd'), root mean squared error

Table 1: Linear model with random treatment

	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
A. Changes-in-Changes							
N=1,000							
bias	0.00	-0.00	-0.01	-0.01	-0.01	-0.00	-0.01
sd	0.11	0.08	0.23	0.10	0.13	0.27	0.27
rmse	0.11	0.08	0.23	0.10	0.13	0.27	0.27
true	1.00	2.00	3.00	2.00	1.00	2.00	1.00
relr	0.11	0.04	0.08	0.05	0.13	0.14	0.27
N=4,000							
bias	-0.00	-0.00	0.00	-0.00	-0.01	0.01	0.01
sd	0.06	0.04	0.12	0.05	0.07	0.14	0.14
rmse	0.06	0.04	0.12	0.05	0.07	0.14	0.14
true	1.00	2.00	3.00	2.00	1.00	2.00	1.00
relr	0.06	0.02	0.04	0.02	0.07	0.07	0.14
B. Difference-in-Differences							
N=1,000							
bias	0.01	-0.00	-0.01	-0.01	0.00	-0.02	0.00
sd	0.11	0.09	0.14	0.14	0.12	0.19	0.10
rmse	0.11	0.09	0.14	0.14	0.12	0.19	0.10
true	1.00	2.00	3.00	2.00	1.00	2.00	1.00
relr	0.11	0.04	0.05	0.07	0.12	0.10	0.10
N=4,000							
bias	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00
sd	0.06	0.04	0.07	0.07	0.06	0.10	0.05
rmse	0.06	0.04	0.07	0.07	0.06	0.10	0.05
true	1.00	2.00	3.00	2.00	1.00	2.00	1.00
relr	0.06	0.02	0.02	0.04	0.06	0.05	0.05

Note: 'bias', 'sd', and 'rmse' provide the bias, standard deviation, and root mean squared error of the respective estimator. 'true' and 'relr' are the respective true effect as well as the root mean squared error relative to the true effect.

('rmse'), true effect ('true'), and the relative root mean squared error in percent of the true effect ('relr') of the respective estimators of θ_1^n, θ_1^a, Δ_c, $\theta_1^c(1)$, $\theta_1^c(0)$, $\delta_1^c(1)$, and $\delta_1^c(0)$ for the linear model. In this case, the identifying assumptions underlying both the change-in-changes (Panel A.) and difference-in-differences (Panel B.) estimators are satisfied. Specifically, the homogeneous time trend on the individual level satisfies any of the common trend assumptions in Deuchert, Huber, and Schelker (2017), while the monotonicity of Y in U and the independence of T and U satisfies the key assumptions of this paper. For this reason any of the estimates

in Table 1 are close to being unbiased and appear to converge to the true effect at the parametric rate when comparing the results for the two different sample sizes.

Table 2: Nonlinear model with random treatment

	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
A. Change-in-Changes							
N=1,000							
bias	0.01	-0.14	-0.48	-0.35	-0.11	-0.37	-0.13
sd	0.48	5.08	8.47	6.20	1.16	8.64	4.23
rmse	0.48	5.08	8.48	6.21	1.17	8.65	4.23
true	3.49	68.09	52.42	47.70	4.72	47.70	4.72
relr	0.14	0.07	0.16	0.13	0.25	0.18	0.90
N=4,000							
bias	-0.01	0.01	-0.00	-0.11	-0.07	0.07	0.11
sd	0.25	2.63	4.37	3.20	0.66	4.44	2.04
rmse	0.25	2.63	4.37	3.20	0.66	4.44	2.04
true	3.49	68.09	52.45	47.73	4.72	47.73	4.72
relr	0.07	0.04	0.08	0.07	0.14	0.09	0.43
B. Difference-in-Differences							
N=1,000							
bias	-0.27	-8.91	14.42	11.46	-1.49	15.91	2.96
sd	0.46	2.62	2.58	2.62	0.47	2.61	0.47
rmse	0.53	9.29	14.65	11.76	1.56	16.12	2.99
true	3.49	68.09	52.42	47.70	4.72	47.70	4.72
relr	0.15	0.14	0.28	0.25	0.33	0.34	0.63
N=4,000							
bias	-0.28	-8.79	14.51	11.57	-1.51	16.02	2.94
sd	0.24	1.28	1.26	1.28	0.25	1.27	0.23
rmse	0.37	8.88	14.57	11.64	1.53	16.07	2.95
true	3.49	68.09	52.45	47.73	4.72	47.73	4.72
relr	0.11	0.13	0.28	0.24	0.32	0.34	0.62

Note: 'bias', 'sd', and 'rmse' provide the bias, standard deviation, and root mean squared error of the respective estimator. 'true' and 'relr' are the respective true effect as well as the root mean squared error relative to the true effect.

Table 2 provides the results for the exponential outcome model, in which the time trend is heterogeneous and interacts with U through the nonlinear link function. While the change-in-changes assumptions hold (Panel A.), average time trends are heterogeneous across complier types such that the difference-in-differences approach (Panel B.) of Deuchert, Huber, and Schelker (2017) is inconsistent. Accordingly, the biases of the change-in-changes estimates generally approach zero as the

sample size increases, while this is not the case for the difference-in-differences estimates. Change-in-changes yields a lower root mean squared error than the respective difference-in-differences estimator in all but one case (namely $\hat{\delta}_1^c(0)$ with $N = 1,000$) and its relative attractiveness increases in the sample size due to its lower bias.

Table 3: Nonlinear model with non-random treatment

	$\hat{\theta}_1^{0,1}$	$\hat{\theta}_1^{1,0}$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
A. Change-in-Changes							
N=1,000							
bias	0.02	0.13	47.21	40.19	-1.44	48.64	7.02
sd	0.71	4.56	5.45	4.11	0.75	5.53	2.92
rmse	0.71	4.56	47.52	40.40	1.62	48.96	7.60
true	4.41	54.19	52.42	47.70	4.72	47.70	4.72
relr	0.16	0.08	0.91	0.85	0.34	1.03	1.61
N=4,000							
bias	-0.00	0.06	47.38	40.13	-1.53	48.91	7.25
sd	0.38	2.35	2.84	2.04	0.38	2.86	1.51
rmse	0.38	2.35	47.47	40.18	1.57	48.99	7.40
true	4.40	54.18	52.45	47.73	4.72	47.73	4.72
relr	0.09	0.04	0.90	0.84	0.33	1.03	1.57
B. Difference-in-Differences							
N=1,000							
bias	0.35	19.98	29.00	27.65	0.04	28.96	1.35
sd	0.67	2.48	2.46	2.48	0.67	2.51	0.45
rmse	0.75	20.14	29.11	27.76	0.67	29.07	1.43
true	4.41	54.19	52.42	47.70	4.72	47.70	4.72
relr	0.17	0.37	0.56	0.58	0.14	0.61	0.30
N=4,000							
bias	0.34	20.02	28.98	27.65	0.02	28.96	1.33
sd	0.35	1.22	1.19	1.22	0.35	1.24	0.23
rmse	0.49	20.06	29.01	27.68	0.35	28.99	1.35
true	4.40	54.18	52.45	47.73	4.72	47.73	4.72
relr	0.11	0.37	0.55	0.58	0.07	0.61	0.29

Note: 'bias', 'sd', and 'rmse' provide the bias, standard deviation, and root mean squared error of the respective estimator. 'true' and 'relr' are the respective true effect as well as the root mean squared error relative to the true effect.

In our final simulation design, we maintain the exponential outcome model but assume D to be selective w.r.t. U rather than random. To this end, the treatment model in (4) is replaced by $D = I\{U + Q > 0\}$, with the independent variable $Q \sim N(0,1)$ being an unobserved term. Under this violation of Assumption 7,

complier shares and effects are no longer identified, which is confirmed by the simulation results presented in Table 3. The bias in the change-in-changes based total, direct, and indirect effects on compliers do not vanish as the sample size increases. Furthermore, under non-random assignment of D (while maintaining monotonicity of M in D), the never-takers' and always-takers' respective distributions of U differ across treatment. Therefore, average direct effects among the total of never or always-takers, respectively, are not identified. Yet, $\theta_1^{1,0}$, which is still identified by the same estimator as before, yields the direct effect among never-takers with $D = 1$ (as defiers do not exist). Likewise, $\theta_1^{0,1}$ corresponds to the direct effect on always-takers with $D = 0$. Indeed, the results in Table 3 suggest that both parameters are consistently estimated with the change-in-changes model (Panel A.).

5 Application

Our empirical application is based on the JOBS II data by Vinokur and Price (1999). JOBS II was a randomized job training intervention in the US, designed to analyse the impact of job training on labour market and mental health outcomes, see Vinokur, Price, and Schul (1995). It was a modified version of the earlier JOBS programme, which had been found to improve labour market outcomes such as job satisfaction, motivation, earnings, and job stability, see Caplan, Vinokur, Price, and van Ryn (1989) and Vinokur, van Ryn, Gramlich, and Price (1991), as well as mental health, see Vinokur, Price, and Caplan (1991). According to the results of Vinokur, Price, and Schul (1995), the JOBS II programme increased reemployment rates and improved mental health outcomes, especially for participants having an elevated risk of depression. The JOBS interventions had an important impact in the academic literature (see e.g. Wanberg, 2012, Liu, Huang, and Wang, 2014) and the methodology was implemented in field experiments in Finland (Vuori, Silvonen, Vinokur, and Price, 2002, Vuori and Silvonen, 2005) and the Netherlands (Brenninkmeijer and Blonk, 2011), suggesting positive effects on labour market integration in either case.

The JOBS II intervention was conducted in south-eastern Michigan, where 2,464 job seekers were eligible to participate in a randomized field experiment, see Vinokur and Price (1999). In a baseline period prior to programme assignment, individuals responded to a screening questionnaire that collected pre-treatment information on mental health. Based on the latter, individuals were classified as having either a high or low depression risk and those with a high risk were oversampled before the training was randomly assigned.[1] The job training consisted of five 4-hours seminars conducted in morning sessions during one week between March 1 and August 7, 1991. Members of the treatment group who participated in at least four of the five sessions received USD 20. Each of the standardized training sessions consisted, among other aspects, of the learning and practicing of job search and problem-solving skills.[2] The control group received a booklet with information on job search methods (Vinokur, Price, and Schul, 1995, p. 44-49).

We analyse the impact of job training on mental health, namely symptoms of depression 6 months after training participation. The health outcome (Y) is based on a 11-items index of depression symptoms of the Hopkins Symptom Checklist. For example, respondents were asked how much they were bothered by symptoms such as crying easily, feeling lonely, feeling blue, feeling hopeless, having thoughts of ending their lives, or experiencing a loss of sexual interest. The questions were coded on a 5-point scale, going from 'not at all' (1) to 'extremely' (5), and summarized in a depression variable that consists of the average across all questions.

One-sided non-compliance with the random assignment is a major issue in JOBS

---

[1]In the JOBS II intervention, randomization was followed by yet another questionnaire sent out two weeks before the actual job training, see Vinokur, Price, and Schul (1995), which also provided information on whether an individual had been assigned the training. Consequently, the data collected in that questionnaire must be considered post-treatment as they could be affected by learning the assignment. Therefore, we rely on the earlier screening data as the relevant pre-treatment period prior to random programme assignment.

[2]When compared to the earlier JOBS programme (Caplan, Vinokur, Price, and van Ryn, 1989), the job training sessions of JOBS II focused more strongly on building a sense of mastery, personal control and self- efficacy in job search. Previous research had suggested that an increase in this sense of mastery, control and self-efficacy improved observed effort in job search behaviour (Eden and Aviram, 1993). Results in Marshall and Lang (1990) suggest that mastery is a strong predictor of depression symptoms among women. For a detailed discussion of the literature, the exact sampling process, the training programme, and further aspects, see Vinokur, Price, and Schul (1995).

II. While the study design rules out always-takers because members of the control group did not have access to the job training programme, 45% of those assigned to training in our data did not participate and are therefore never-takers, the remaining 55% are compliers. In order to avoid selection bias w.r.t actual participation, the original JOBS II study by Vinokur, Price, and Schul (1995) analysed the total effect of the policy (i.e. the intention-to-treat effect), including those who, despite receiving an offer to participate, did not take part in the job training. In contrast, we use our methodology to separate the direct effect of mere training assignment, which is our treatment D, from the indirect effect operating through actual training participation, which is our mediator M, among compliers.[3] We also consider the direct effect on never-takers, which likely differs from that on the compliers. While being offered (or not offered) the job training might affect compliers' mental health by inducing motivation/enthusiasm (or discouragement), it may not have the same effect among never-takers, who do not attend such seminars whatsoever.

More concisely, we base identification on Theorem 3a with Assumption 4a for the average direct effect on never-takers, θ_1^n, on Theorem 3b with Assumption 4 for the direct effect on compliers under $d = 0$, $\theta_1^c(0)$, and Theorem 5 with Assumptions 4b and 6a for the indirect effect on compliers under $d = 1$, $\delta_1^c(1)$. None of these approaches requires the presence of always-takers in the sample. We also note that if random assignment operated through other mechanisms than actual participation in any of the subpopulations as it may appear reasonable in the context of mental health outcomes, this would violate the exclusion restriction when using assignment as instrumental variable for actual participation in a two stage least squares regression. Given that our identifying assumptions hold, our approach can therefore be used to statistically test the exclusion restriction.

Our evaluation sample consists of a total of 3,360 observations in the pre-treatment and post-mediator periods with non-missing information for D, M, and Y. It is an

---

[3]Imai, Keele, and Tingley (2010) analyse Jobs II in a mediation context as well, but consider a different mediator, namely job search self-efficacy, and a different identification strategy based on selection on observables.

Table 4: Descriptive statistics on outcomes in pre- and post-mediator periods

	pre-treatment $(T = 0)$		post-mediator $(T = 1)$	
	sample size	mean	sample size	mean
overall	1,796	1.86	1,564	1.73
		(0.58)		(0.67)
$D = 0$	551	1.87	486	1.78
		(0.59)		(0.70)
$D = 1$	1,245	1.86	1,078	1.70
		(0.57)		(0.66)
mean diff		0.01		0.08
pval		0.74		0.03
SD		1.63		11.79

Note: Standard deviations are in parentheses. 'mean diff', 'pval', and 'SD' are the mean difference, its p-value, and the standardized difference, respectively.

unbalanced panel due to attrition of roughly 13% of the initial respondents between the two periods. Table 4 provides summary statistics for the outcome in the total sample as well as by treatment group over time. We verify whether randomization was successful by comparing the outcome means of the treatment and control groups in the pre-treatment period $(T = 0)$ just prior to the randomization of D. The small difference of 0.01 is not statistically significant according to a two sample t-test. Furthermore, the standardized difference test suggested by Rosenbaum and Rubin (1985) yields a value of just 1.68 and is thus far below 20, a threshold frequently chosen for indicating problematic imbalances across treatment groups. To test for potential attrition bias we also consider these statistics in the pre-treatment period exclusively among the panel cases that remain in the sample in the post-mediator period (not reported in Table 4). The p-value of the t-test amounts to 0.52 and the standardized difference of 3.5 is low such that attrition bias does not appear to be a concern. We therefore do not find statistical evidence for a violation of the random assignment of D in our sample. Table 4 also reports the mean difference in outcomes in the post-mediator period $(T = 1)$ 6 months after participation, which is an estimate for the total (or intention-to-treat) effect of D. The difference of 0.08 is statistically significant at the 5% level.

Table 5 presents the estimation results based on our CiC approach and the DiD

Table 5: Empirical results for Jobs II

	Changes-in-Changes				Difference-in-Differences				Type shares	
	$\hat{\theta}_1^n$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\theta}_1^n$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{p}(n)$	$\hat{p}(c)$
est	-0.04	-0.11	0.06	-0.17	-0.03	-0.12	-0.06	-0.06	0.45	0.55
se	0.05	0.06	0.05	0.08	0.05	0.06	0.05	0.07	0.01	0.01
pval	0.40	0.06	0.26	0.04	0.52	0.03	0.21	0.43	0.00	0.00

Note: 'est', 'se', and 'pval' provide the effect estimate, standard error, and p-value of the respective estimator. $\hat{p}(n)$ and $\hat{p}(c)$ are the estimated never-taker and complier shares. Standard errors are based on cluster bootstrapping the effects 1999 times where clustering is on the respondent level.

strategy of Deuchert, Huber, and Schelker (2017) when (linearly) controlling for the gender of respondents in either case. Standard errors rely on cluster bootstrapping the direct and indirect effects 1999 times, where clustering is on the respondent level. The CiC and DiD estimates of the direct effects on never-takers, $\hat{\theta}_1^n(0)$, as well as on compliers, $\hat{\theta}_1^c(0)$, are not statistically significant at conventional levels. Hence, we do not find statistical evidence for a direct effect of the mere assignment into the training programme on the depression outcome, which would point to a violation of the exclusion restriction when using assignment as instrument for participation. In contrast, we find for both CiC and DiD negative total effects among compliers $\hat{\Delta}_c$ that are statistically significant at least at the 10% level. In the case of CiC, also the negative indirect effect among compliers, $\hat{\delta}_1^c(1)$, is significant at the 5% level, while this is not in the case for DiD. By and large, our results point to a moderately negative treatment effect on depressive symptoms through actual programme participation, rather than through other (i.e. direct) mechanisms. The CiC estimates $\hat{\delta}_1^c(1)$ and $\hat{\Delta}_c$ are in fact rather similar to the result of a two stage least squares regression relying on the exclusion restriction by using D as instrument for M. The latter approach yields a local average treatment effect on compliers in the post-mediator period of -0.14 with a heteroskedasticity-robust standard error of 0.07 (significant at the 5% level).

6 Conclusion

We proposed a novel identification strategy for causal mediation analysis with repeated cross sections or panel data based on changes-in-changes (CiC) assumptions that are related but yet different to Athey and Imbens (2006) considering total treatment effects. Strict monotonicity of outcomes in unobserved heterogeneity and distributional time invariance of the latter within groups defined on treatment and mediator states are key assumptions for identifying direct effects within these groups. Additionally assuming random treatment assignment and weak monotonicity of the mediator in the treatment permits identifying direct effects on never-takers and always-takers as well as total, direct, and indirect effects on compliers. We also provided a brief simulation study and an empirical application to the Jobs II programme.

References

ALBERT, J. M., AND S. NELSON (2011): "Generalized causal mediation analysis," *Biometrics*, 67, 1028–1038.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472.

ATHEY, S., AND G. W. IMBENS (2006): "Identification and Inference in Nonlinear Difference-In-Difference Models," *Econometrica*, 74, 431–497.

BARON, R. M., AND D. A. KENNY (1986): "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51, 1173–1182.

BELLANI, L., AND M. BIA (2018): "The long-run effect of childhood poverty and the mediating role of education," *Journal of the Royal Statistical Society: Series A*, 182, 37–68.

BIJWAARD, G. E., AND A. M. JONES (2019): "An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality," *Empirical Economics*, 57, 129–175.

BRENNINKMEIJER, V., AND R. W. BLONK (2011): "The effectiveness of the JOBS program among the long-term unemployed: a randomized experiment in the Netherlands," *Health Promotion International*, 27, 220–229.

BRUNELLO, G., M. FORT, N. SCHNEEWEIS, AND R. WINTER-EBMER (2016): "The Causal Effect of Education on Health: What is the Role of Health Behaviors?," *Health Economics*, 25, 314–336.

CALLAWAY, B. (2016): "Quantile Treatment Effects in R: The qte Package," *working paper, Temple University, Philadelphia*.

CAPLAN, R. D., A. D. VINOKUR, R. H. PRICE, AND M. VAN RYN (1989): "Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss," *Journal of Applied Psychology*, 74, 759–769.

CHAISEMARTIN, C., AND X. D'HAULTFEUILLE (2018): "Fuzzy Differences-in-Differences," *Review of Economic Studies*, 85, 999–1028.

CHEN, S. H., Y. C. CHEN, AND J. T. LIU (2017): "The impact of family composition on educational achievement," *Journal of Human Resources*, 0915-7401R1.

COCHRAN, W. G. (1957): "Analysis of Covariance: Its Nature and Uses," *Biometrics*, 13, 261–281.

CONTI, G., J. J. HECKMAN, AND R. PINTO (2016): "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour," *The Economic Journal*, 126, F28–F65.

DEUCHERT, E., M. HUBER, AND M. SCHELKER (2017): "Direct and indirect effects based on difference-in-differences with an application to political preferences

BIJWAARD, G. E., AND A. M. JONES (2019): "An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality," *Empirical Economics*, 57, 129–175.

BRENNINKMEIJER, V., AND R. W. BLONK (2011): "The effectiveness of the JOBS program among the long-term unemployed: a randomized experiment in the Netherlands," *Health Promotion International*, 27, 220–229.

BRUNELLO, G., M. FORT, N. SCHNEEWEIS, AND R. WINTER-EBMER (2016): "The Causal Effect of Education on Health: What is the Role of Health Behaviors?," *Health Economics*, 25, 314–336.

CALLAWAY, B. (2016): "Quantile Treatment Effects in R: The qte Package," *working paper, Temple University, Philadelphia*.

CAPLAN, R. D., A. D. VINOKUR, R. H. PRICE, AND M. VAN RYN (1989): "Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss," *Journal of Applied Psychology*, 74, 759–769.

CHAISEMARTIN, C., AND X. D'HAULTFEUILLE (2018): "Fuzzy Differences-in-Differences," *Review of Economic Studies*, 85, 999–1028.

CHEN, S. H., Y. C. CHEN, AND J. T. LIU (2017): "The impact of family composition on educational achievement," *Journal of Human Resources*, 0915-7401R1.

COCHRAN, W. G. (1957): "Analysis of Covariance: Its Nature and Uses," *Biometrics*, 13, 261–281.

CONTI, G., J. J. HECKMAN, AND R. PINTO (2016): "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour," *The Economic Journal*, 126, F28–F65.

DEUCHERT, E., M. HUBER, AND M. SCHELKER (2017): "Direct and indirect effects based on difference-in-differences with an application to political preferences

following the Vietnam draft lottery," *forthcoming in the Journal of Business & Economic Statistics.*

DOERR, A., AND A. STRITTMATTER (2019): "Identifying causal channels of policy reforms with multiple treatments and different types of selection," *working paper, University of St. Gallen.*

EDEN, D., AND A. AVIRAM (1993): "Self-efficacy training to speed reemployment: Helping people to help themselves," *Journal of Applied Psychology*, 78, 352–360.

FLORES, C. A., AND A. FLORES-LAGUNES (2009): "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness," *IZA Discussion Paper No. 4237.*

FRANGAKIS, C., AND D. RUBIN (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

FRÖLICH, M., AND M. HUBER (2017): "Direct and Indirect Treatment Effects – Causal Chains and Mediation Analysis with Instrumental Variables," *Journal of the Royal Statistical Society: Series B*, 79, 1645–1666.

HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103, 2052–2086.

HONG, G. (2010): "Ratio of mediator probability weighting for estimating natural direct and indirect effects," in *Proceedings of the American Statistical Association, Biometrics Section*, p. 2401–2415. Alexandria, VA: American Statistical Association.

HUBER, M. (2014): "Identifying causal mechanisms (primarily) based on inverse probability weighting," *Journal of Applied Econometrics*, 29, 920–943.

HUBER, M. (2015): "Causal pitfalls in the decomposition of wage gaps," *Journal of Business and Economic Statistics*, 33, 179–191.

HUBER, M., M. LECHNER, AND G. MELLACE (2017): "Why Do Tougher Case-workers Increase Employment? The Role of Program Assignment as a Causal Mechanism," *The Review of Economics and Statistics*, 99, 180–183.

HUBER, M., M. LECHNER, AND A. STRITTMATTER (2018): "Direct and indirect effects of training vouchers for the unemployed," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181, 441–463.

IMAI, K., L. KEELE, AND D. TINGLEY (2010): "A General Approach to Causal Mediation Analysis," *Psychological Methods*, 15, 309–334.

IMAI, K., L. KEELE, AND T. YAMAMOTO (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51–71.

IMAI, K., AND T. YAMAMOTO (2013): "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments," *Political Analysis*, 21, 141–171.

IMBENS, G. W., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

JUDD, C. M., AND D. A. KENNY (1981): "Process Analysis: Estimating Mediation in Treatment Evaluations," *Evaluation Review*, 5, 602–619.

JUHN, C., K. M. MURPHY, AND B. PIERCE (1991): "Accounting for the Slowdown in Black-White Wage Convergencee," in *Workers and their Wages: Changing Patterns in the United States*, ed. by M. H. Kosters, pp. 107–143. American Enterprise Insitute, Washigton.

KEELE, L., D. TINGLEY, AND T. YAMAMOTO (2015): "Identifying mechanisms behind policy interventions via causal mediation analysis," *Journal of Policy Analysis and Management*, 34, 937–963.

LIU, S., J. L. HUANG, AND M. WANG (2014): "Effectiveness of Job Search Interventions: A Meta-Analytic Review," *Psychological Bulletin*, 140, 1009–1041.

Manski, C. F. (1988): *Analog Estimation Methods in Econometrics*. Chapman & Hall, New York.

Marshall, G. N., and E. L. Lang (1990): "Optimism, self-mastery, and symptoms of depression in women professionals," *Journal of Personality and Social Psychology*, 59, 132–139.

Melly, B., and G. Santangelo (2015): "The Changes-in-Changes Model with Covariates," *Working Paper*.

Pearl, J. (2001): "Direct and indirect effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–420, San Francisco. Morgan Kaufman.

Petersen, M. L., S. E. Sinisi, and M. J. van der Laan (2006): "Estimation of Direct Causal Effects," *Epidemiology*, 17, 276–284.

Powdthavee, N., W. N. Lekfuangfu, and M. Wooden (2013): "The Marginal Income Effect of Education on Happiness: Estimating the Direct and Indirect Effects of Compulsory Schooling on Well-Being in Australia," *IZA Discussion Paper No. 7365*.

Robins, J. M. (2003): "Semantics of causal DAG models and the identification of direct and indirect effects," in *In Highly Structured Stochastic Systems*, ed. by P. Green, N. Hjort, and S. Richardson, pp. 70–81, Oxford. Oxford University Press.

Robins, J. M., and S. Greenland (1992): "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155.

Rosenbaum, P. R., and D. B. Rubin (1985): "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score.," *The American Statistician*, 39, 33–38.

RUBIN, D. B. (1974): "Estimating the Causal Effect of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66, 688–701.

RUBIN, D. B. (2004): "Direct and Indirect Causal Effects via Potential Outcomes," *Scandinavian Journal of Statistics*, 31, 161–170.

SIMONSEN, M., AND L. SKIPPER (2006): "The Costs of Motherhood: An Analysis Using Matching Estimators," *Journal of Applied Econometrics*, 21, 919–934.

STRITTMATTER, A. (2019): "Heterogeneous Earnings Effects of the Job Corps by Gender Earnings: A Translated Quantile Approach," *forthcoming in Labour Economics*.

TCHETGEN TCHETGEN, E. J., AND I. SHPITSER (2012): "Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis," *The Annals of Statistics*, 40, 1816–1845.

VANDERWEELE, T. J. (2008): "Simple relations between principal stratification and direct and indirect effects," *Statistics & Probability Letters*, 78, 2957–2962.

VANDERWEELE, T. J. (2009): "Marginal Structural Models for the Estimation of Direct and Indirect Effects," *Epidemiology*, 20, 18–26.

VANDERWEELE, T. J. (2012): "Comments: Should Principal Stratification Be Used to Study Mediational Processes?," *Journal of Research on Educational Effectiveness*, 5, 245–249.

VANSTEELANDT, S., M. BEKAERT, AND T. LANGE (2012): "Imputation Strategies for the Estimation of Natural Direct and Indirect Effects," *Epidemiologic Methods*, 1, 129–158.

VINOKUR, A. D., AND R. H. PRICE (1999): "Jobs II Preventive Intervention for Unemployed Job Seekers, 1991-1993," *Inter-university Consortium for Political and Social Research*.

VINOKUR, A. D., R. H. PRICE, AND R. D. CAPLAN (1991): "From field experiments to program implementation: Assessing the potential outcomes of an experimental intervention program for unemployed persons," *American Journal of Community Psychology*, 19, 543–562.

VINOKUR, A. D., R. H. PRICE, AND Y. SCHUL (1995): "Impact of the JOBS Intervention on Unemployed Workers Varying in Risk for Depression," *American Journal of Community Psychology*, 23, 39–74.

VINOKUR, A. D., M. VAN RYN, E. M. GRAMLICH, AND R. H. PRICE (1991): "From field experiments to program implementation: Assessing the potential outcomes of an experimental intervention program for unemployed persons," *Journal of Applied Psychology*, 76, 213–219.

VUORI, J., AND J. SILVONEN (2005): "The benefits of a preventive job search program on re-employment and mental health at 2-year follow-up," *Journal of Occupational and Organizational Psychology*, 78, 43–52.

VUORI, J., J. SILVONEN, A. D. VINOKUR, AND R. H. PRICE (2002): "The Työhön Job Search Program in Finland: Benefits for the Unemployed With Risk of Depression or Discouragement," *Journal of Occupational Health Psychology*, 7, 5–19.

WANBERG, C. R. (2012): "The Individual Experience of Unemployment," *Annual Review of Psychology*, 63, 369–396.

WÜTHRICH, K. (2019): "A comparison of two quantile models with endogeneity," *forthcoming in Journal of Business and Economic Statistics*.

Appendices

A   Proof of Theorem 1

A.1   Average direct effect under d = 1 conditional on D = 1 and M(1) = 0

In the following, we prove that $\theta_1^{1,0}(1) = E[Y_1(1,0) - Y_1(0,0)|D = 1, M_i(1) = 0] = E[Y_1 - Q_{00}(Y_0)|D = 1, M = 0]$. Using the observational rule, we obtain $E[Y_1(1,0)|D = 1, M(1) = 0] = E[Y_1|D = 1, M = 0]$. Accordingly, we have to show that $E[Y_1(0,0)|D = 1, M(1) = 0] = E[Q_{00}(Y_0)|D = 1, M = 0]$ to finish the proof.

Denote the inverse of $h(d, m, t, u)$ by $h^{-1}(d, m, t; y)$, which exists because of the strict monotonicity required in Assumption 1. Under Assumptions 1 and 3a, the conditional potential outcome distribution function equals

$$
\begin{aligned}
F_{Y_t(d,0)|D=1,M=0}(y) &\overset{A1}{=} \Pr(h(d, m, t, U) \leq y|D = 1, M = 0, T = t), \\
&= \Pr(U \leq h^{-1}(d, m, t; y)|D = 1, M = 0, T = t), \\
&\overset{A3a}{=} \Pr(U \leq h^{-1}(d, m, t; y)|D = 1, M = 0), \\
&= F_{U|10}(h^{-1}(d, m, t; y)),
\end{aligned}
\tag{A.1}
$$

for $d, d' \in \{0, 1\}$. We use these quantities in the following.

First, evaluating $F_{Y_1(0,0)|D=1,M=0}(y)$ at $h(0, 0, 1, u)$ gives

$$
F_{Y_1(0,0)|D=1,M=0}(h(0, 0, 1, u)) = F_{U|10}(h^{-1}(0, 0, 1; h(0, 0, 1, u))) = F_{U|10}(u).
$$

Applying $F_{Y_1(0,0)|D=1,M=0}^{-1}(q)$ to both sides, we have

$$
h(0, 0, 1, u) = F_{Y_1(0,0)|D=1,M=0}^{-1}(F_{U|10}(u)).
\tag{A.2}
$$

Second, for $F_{Y_0(0,0)|D=1,M=0}(y)$ we have

$$F_{U|D=1,M=0}^{-1}(F_{Y_0(0,0)|D=1,M=0}(y)) = h^{-1}(0,0,0;y). \tag{A.3}$$

Combining (A.2) and (A.3) yields,

$$h(0,0,1,h^{-1}(0,0,0;y)) = F_{Y_1(0,0)|D=1,M=0}^{-1} \circ F_{Y_0(0,0)|D=1,M=0}(y). \tag{A.4}$$

Note that $h(0,0,1,h^{-1}(0,0,0;y))$ maps the period 1 (potential) outcome of an individual with the outcome y in period 0 under non-treatment without the mediator. Accordingly, $E[F_{Y_1(0,0)|D=1,M=0}^{-1} \circ F_{Y_0(0,0)|D=1,M=0}(Y_0)|D = 1, M = 0] = E[Y_1(0,0)|D = 1, M = 0]$. We can identify $F_{Y_0(0,0)|D=1,M=0}(y)$ under Assumption 2, but we cannot identify $F_{Y_1(0,0)|D=1,M=0}(y)$. However, we show in the following that we can identify the overall quantile-quantile transform $F_{Y_1(0,0)|D=1,M=0}^{-1} \circ F_{Y_0(0,0)|D=1,M=0}(y)$ under the additional Assumption 3b.

Under Assumptions 1 and 3b, the conditional potential outcome distribution function equals

$$\begin{aligned}
F_{Y_t(d,0)|D=0,M=0}(y) &\overset{A1}{=} \Pr(h(d,m,t,U) \le y|D = 0, M = 0, T = t), \\
&= \Pr(U \le h^{-1}(d,m,t;y)|D = 0, M = 0, T = t), \\
&\overset{A3b}{=} \Pr(U \le h^{-1}(d,m,t;y)|D = 0, M = 0), \\
&= F_{U|00}(h^{-1}(d,m,t;y)),
\end{aligned} \tag{A.5}$$

for $d, d' \in \{0,1\}$. We repeat similar steps as above. First, evaluating $F_{Y_1(0,0)|D=0,M=0}(y)$ at $h(0,0,1,u)$ gives

$$F_{Y_1(0,0)D=0,M=0}(h(0,0,1,u)) = F_{U|00}(h^{-1}(0,0,1;h(0,0,1,u))) = F_{U|00}(u).$$

Applying $F_{Y_1(0,0)|D=0,M=0}^{-1}(q)$ to both sides, we have

$$h(0,0,1,u) = F_{Y_1(0,0)|D=0,M=0}^{-1}(F_{U|00}(u)). \tag{A.6}$$

Second, for $F_{Y_0(0,0)|D=0,M=0}(y)$ we have

$$F_{U|00}^{-1}(F_{Y_0(0,0)|D=0,M=0}(y)) = h^{-1}(0,0,0;y). \qquad \text{(A.7)}$$

Combining (A.6) and (A.7) yields,

$$h(0,0,1,h^{-1}(0,0,0;y)) = F_{Y_1(0,0)|D=0,M=0}^{-1} \circ F_{Y_0(0,0)|D=0,M=0}(y). \qquad \text{(A.8)}$$

The left sides of (A.4) and (A.8) are equal. In contrast to (A.4), (A.8) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(0,0)|D=0,M=0}(y) = \Pr(Y_t(0,0) \leq y|D = 0, M = 0) = \Pr(Y_t \leq y|D = 0, M = 0)$. Accordingly, we can identify $F_{Y_1(0,0)|D=1,M=0}^{-1} \circ F_{Y_0(0,0)|D=1,M=0}(y)$ by $Q_{00}(y) \equiv F_{Y_1|D=0,M=0}^{-1} \circ F_{Y_0|D=0,M=0}(y)$.

Parsing Y_0 through $Q_{00}(\cdot)$ in the treated group without mediator gives

$$
\begin{aligned}
&E[Q_{00}(Y_0)|D = 1, M = 0] \\
&\qquad = E[F_{Y_1|D=0,M=0}^{-1} \circ F_{Y_0|D=0,M=0}(Y_0)|D = 1, M = 0], \\
&\qquad = E[F_{Y_1(0,0)|D=0,M=0}^{-1} \circ F_{Y_0(0,0)|D=0,M=0}(Y_0(1,0))|D = 1, M = 0], \\
&\qquad \overset{A1,A3b}{=} E[h(0,0,1,h^{-1}(0,0,0;Y_0(1,0)))|D = 1, M = 0], \qquad \text{(A.9)} \\
&\qquad \overset{A2}{=} E[h(0,0,1,h^{-1}(0,0,0;Y_0(0,0)))|D = 1, M = 0], \\
&\qquad \overset{A1,A3a}{=} E[F_{Y_1(0,0)|D=1,M=0}^{-1} \circ F_{Y_0(0,0)|D=1,M=0}(Y_0(0,0))|D = 1, M = 0], \\
&\qquad = E[Y_1(0,0)|D = 1, M = 0] = E[Y_1(0,0)|D = 1, M(1) = 0],
\end{aligned}
$$

which has data support because of Assumption 4a.

A.2 Quantile direct effect under $d = 1$ conditional on $D = 1$ and $M(1) = 0$

In the following, we prove that

$$\theta_1^{1,0}(q, 1) = F_{Y_1(1,0)|D=1,M(1)=0}^{-1}(q) - F_{Y_1(0,0)|D=1,M(1)=0}^{-1}(q),$$

$$= F_{Y_1|D=1,M=0}^{-1}(q) - F_{Q_{00}(Y_0)|D=1,M=0}^{-1}(q).$$

For this purpose, we have to show that

$$F_{Y_1(1,0)|D=1,M(1)=0}(y) = F_{Y_1|D=1,M=0}(y) \text{ and} \tag{A.10}$$

$$F_{Y_1(0,0)|D=1,M(1)=0}(y) = F_{Q_{00}(Y_0)|D=1,M=0}(y), \tag{A.11}$$

which is sufficient to show that the quantiles are also identified. We can show (A.10) using the observational rule $F_{Y_1(1,0)|D=1,M(1)=0}(y) = F_{Y_1|D=1,M=0}(y) = E[1\{Y_1 \leq y\}|D = 1, M = 0]$, with $1\{\cdot\}$ being the indicator function.

Using (A.9), we obtain

$$F_{Q_{00}(Y_0)|D=1,M=0}(y)$$
$$= E[1\{Q_{00}(Y_0) \leq y\}|D = 1, M = 0],$$
$$= E[1\{F_{Y_1|D=0,M=0}^{-1} \circ F_{Y_0|D=0,M=0}(Y_0) \leq y\}|D = 1, M = 0], \tag{A.12}$$
$$= E[1\{Y_1(0, 0) \leq y\}|D = 1, M = 0],$$
$$= F_{Y_1(0,0)|D=1,M(1)=0}(y),$$

which proves (A.11).

A.3 Average direct effect under $d = 0$ conditional on $D = 0$ and $M(0) = 0$

In the following, we show that $\theta_1^{0,0}(0) = E[Y_1(1, 0) - Y_1(0, 0)|D = 0, M(0) = 0] = E[Q_{10}(Y_0) - Y_1|D = 0, M = 0]$. Using the observational rule, we obtain

$E[Y_1(0,0)|D=0, M(0)=0] = E[Y_1|D=0, M=0]$. Accordingly, we have to show that $E[Y_1(1,0)|D=0, M(0)=0] = E[Q_{10}(Y_0)|D=0, M=0]$ to finish the proof.

First, we use (A.5) to evaluate $F_{Y_1(1,0)|D=0,M=0}(y)$ at $h(1,0,1,u)$

$$F_{Y_1(1,0)|D=0,M=0}(h(1,0,1,u)) = F_{U|10}(h^{-1}(1,0,1;h(1,0,1,u))) = F_{U|10}(u).$$

Applying $F_{Y_1(1,0)|D=0,M=0}^{-1}(q)$ to both sides, we have

$$h(1,0,1,u) = F_{Y_1(1,0)|D=0,M=0}^{-1}(F_{U|10}(u)). \tag{A.13}$$

Second, for $F_{Y_0(1,0)|D=0,M=0}(y)$ we have

$$F_{U|10}^{-1}(F_{Y_0(1,0)|D=0,M=0}(y)) = h^{-1}(1,0,0;y), \tag{A.14}$$

using (A.5). Combining (A.13) and (A.14) yields,

$$h(1,0,1,h^{-1}(1,0,0;y)) = F_{Y_1(1,0)|D=0,M=0}^{-1} \circ F_{Y_0(1,0)|D=0,M=0}(y). \tag{A.15}$$

Note that $h(1,0,1,h^{-1}(1,0,0;y))$ maps the period 1 (potential) outcome of an individual with the outcome y in period 0 under treatment without the mediator. Accordingly, $E[F_{Y_1(1,0)|D=0,M=0}^{-1} \circ F_{Y_0(1,0)|D=0,M=0}(Y_0)|D=0, M=0] = E[Y_1(1,0)|D=1, M=0]$. We can identify $F_{Y_0(1,0)|D=0,M=0}(y)$ under Assumption 2, but we cannot identify $F_{Y_1(1,0)|D=0,M=0}(y)$. However, we show in the following that we can identify the overall quantile-quantile transform $F_{Y_1(1,0)|D=0,M=0}^{-1} \circ F_{Y_0(1,0)|D=0,M=0}(y)$ under the additional Assumption 3a.

First, we use (A.1) to evaluate $F_{Y_1(1,0)|D=1,M=0}(y)$ at $h(1,0,1,u)$

$$F_{Y_1(1,0)|D=10,M=0}(h(1,0,1,u)) = F_{U|10}(h^{-1}(1,0,1;h(1,0,1,u))) = F_{U|10}(u).$$

Applying $F^{-1}_{Y_1(1,0)|D=1,M=0}(q)$ to both sides, we have

$$h(1,0,1,u) = F^{-1}_{Y_1(1,0)|D=1,M=0}(F_{U|10}(u)). \tag{A.16}$$

Second, for $F_{Y_0(1,0)|D=0,M=0}(y)$ we have

$$F^{-1}_{U|10}(F_{Y_0(1,0)|D=1,M=0}(y)) = h^{-1}(1,0,0;y), \tag{A.17}$$

using (A.1). Combining (A.16) and (A.17) yields,

$$h(1,0,1,h^{-1}(1,0,0;y)) = F^{-1}_{Y_1(1,0)|D=1,M=0} \circ F_{Y_0(1,0)|D=1,M=0}(y). \tag{A.18}$$

The left sides of (A.15) and (A.18) are equal. In contrast to (A.15), (A.18) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(1,0)|D=1,M=0}(y) = \Pr(Y_t(1,0) \leq y|D = 1, M = 0) = \Pr(Y_t \leq y|D = 1, M = 0)$. Accordingly, we can identify $F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(y)$ by $Q_{10}(y) \equiv F^{-1}_{Y_1|D=1,M=0} \circ F_{Y_0|D=1,M=0}(y)$.

Parsing Y_0 through $Q_{10}(\cdot)$ in the non-treated group without mediator gives

$$\begin{aligned}
E[Q_{10}(Y_0)|D = 0, M = 0] \\
&= E[F^{-1}_{Y_1|D=1,M=0} \circ F_{Y_0|D=1,M=0}(Y_0)|D = 0, M = 0], \\
&= E[F^{-1}_{Y_1(1,0)|D=1,M=0} \circ F_{Y_0(1,0)|D=1,M=0}(Y_0(0,0))|D = 0, M = 0], \\
&\overset{A1,A3a}{=} E[h(1,0,1,h^{-1}(1,0,0;Y_0(0,0)))|D = 0, M = 0], \\
&\overset{A2}{=} E[h(1,0,1,h^{-1}(1,0,0;Y_0(1,0)))|D = 1, M = 0], \\
&\overset{A1,A3b}{=} E[F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(Y_0(1,0))|D = 0, M = 0], \\
&= E[Y_1(1,0)|D = 0, M = 0] = E[Y_1(1,0)|D = 0, M(0) = 0],
\end{aligned} \tag{A.19}$$

which has data support because of Assumption 4b.

A.4 Quantile direct effect under $d = 0$ conditional on $D = 0$ and $M(0) = 0$

In the following, we prove that

$$\theta_1^{0,0}(q, 0) = F^{-1}_{Y_1(1,0)|D=0,M(0)=0}(q) - F^{-1}_{Y_1(0,0)|D=0,M(0)=0}(q),$$
$$= F^{-1}_{Q_{10}(Y_0)|D=0,M=0}(q) - F^{-1}_{Y_1|D=0,M=0}(q).$$

For this purpose, we have to show that

$$F_{Y_1(1,0)|D=0,M(0)=0}(y) = F_{Q_{10}(Y_0)|D=0,M=0}(y) \text{ and} \tag{A.20}$$

$$F_{Y_1(0,0)|D=0,M(0)=0}(y) = F_{Y_1|D=0,M=0}(y), \tag{A.21}$$

which is sufficient to show that the quantiles are also identified. We can show (A.21) using the observational rule $F_{Y_1(0,0)|D=0,M(0)=0}(y) = F_{Y_1|D=0,M=0}(y) = E[1\{Y_1 \leq y\}|D = 0, M = 0]$.

Using (A.19), we obtain

$$\begin{aligned}
F_{Q_{10}(Y_0)|D=0,M=0}(y) & \\
&= E[1\{Q_{10}(Y_0) \leq y\}|D = 0, M = 0], \\
&= E[1\{F^{-1}_{Y_1|D=1,M=0} \circ F_{Y_0|D=1,M=0}(Y_0) \leq y\}|D = 0, M = 0], \\
&= E[1\{Y_1(1, 0) \leq y\}|D = 0, M = 0], \\
&= F_{Y_1(1,0)|D=0,M(0)=0}(y),
\end{aligned}$$

which proves (A.20).

B Proof of Theorem 2

B.1 Average direct effect under $d = 0$ conditional on $D = 0$ and $M(0) = 1$

In the following, we show that $\theta_1^{0,1}(0) = E[Y_1(1,1) - Y_1(0,1)|D = 0, M(0) = 1] = E[Q_{11}(Y_0) - Y_1|D = 0, M = 1]$. Using the observational rule, we obtain $E[Y_1(0,1)|D = 0, M(0) = 1] = E[Y_1|D = 0, M = 1]$. Accordingly, we have to show that $E[Y_1(1,1)|D = 0, M(0) = 1] = E[Q_{11}(Y_0)|D = 0, M = 1]$ to finish the proof.

Under Assumptions 1 and 5a, the conditional potential outcome distribution function equals

$$
\begin{aligned}
F_{Y_t(d,0)|D=1,M=0}(y) &\overset{A1}{=} \Pr(h(d,m,t,U) \leq y|D = 0, M = 1, T = t), \\
&= \Pr(U \leq h^{-1}(d,m,t;y)|D = 0, M = 1, T = t), \\
&\overset{A5a}{=} \Pr(U \leq h^{-1}(d,m,t;y)|D = 0, M = 1), \\
&= F_{U|01}(h^{-1}(d,m,t;y)),
\end{aligned}
\tag{B.1}
$$

for $d, d' \in \{0, 1\}$. We use these quantities in the following.

First, evaluating $F_{Y_1(1,1)|D=0,M=1}(y)$ at $h(1,1,1,u)$ gives

$$
F_{Y_1(1,1)|D=0,M=1}(h(1,1,1,u)) = F_{U|01}(h^{-1}(1,1,1;h(1,1,1,u))) = F_{U|01}(u).
$$

Applying $F_{Y_1(1,1)|D=0,M=1}^{-1}(q)$ to both sides, we have

$$
h(1,1,1,u) = F_{Y_1(1,1)|D=0,M=1}^{-1}(F_{U|01}(u)).
\tag{B.2}
$$

Second, for $F_{Y_0(1,1)|D=0,M=1}(y)$ we have

$$
F_{U|01}^{-1}(F_{Y_0(1,1)|D=0,M=1}(y)) = h^{-1}(1,1,0;y).
\tag{B.3}
$$

Combining (B.2) and (B.3) yields,

$$h(1, 1, 1, h^{-1}(1, 1, 0; y)) = F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(y). \qquad \text{(B.4)}$$

Note that $h(1, 1, 1, h^{-1}(1, 1, 0; y))$ maps the period 1 (potential) outcome of an individual with the outcome y in period 0 under treatment with the mediator. Accordingly, $E[F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_0)|D = 0, M = 1] = E[Y_1(1,1)|D = 0, M = 1]$. We can identify $F_{Y_0(1,1)|D=0,M=1}(y) = F_{Y_0|D=0,M=1}(y)$ under Assumption 2, but we cannot identify $F_{Y_1(1,1)|D=0,M=1}(y)$. However, we show in the following that we can identify the overall quantile-quantile transform $F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(y)$ under the additional Assumption 5b.

Under Assumptions 1 and 5b, the conditional potential outcome distribution function equals

$$
\begin{aligned}
F_{Y_t(d,1)|D=1,M=1}(y) \;&\overset{A1}{=}\; \Pr(h(d, m, t, U) \le y|D = 1, M = 1, T = t), \\
&= \Pr(U \le h^{-1}(d, m, t; y)|D = 1, M = 1, T = t), \\
&\overset{A5b}{=}\; \Pr(U \le h^{-1}(d, m, t; y)|D = 1, M = 1), \\
&= F_{U|11}(h^{-1}(d, m, t; y)),
\end{aligned}
\qquad \text{(B.5)}
$$

for $d, d' \in \{0, 1\}$. We repeat similar steps as above. First, evaluating $F_{Y_1(1,1)|D=1,M=1}(y)$ at $h(1, 1, 1, u)$ gives

$$F_{Y_1(1,1)|D=1,M=1}(h(1, 1, 1, u)) = F_{U|11}(h^{-1}(1, 1, 1; h(1, 1, 1, u))) = F_{U|11}(u).$$

Applying $F^{-1}_{Y_1(1,1)|D=1,M=1}(q)$ to both sides, we have

$$h(1, 1, 1, u) = F^{-1}_{Y_1(1,1)|D=1,M=1}(F_{U|11}(u)). \qquad \text{(B.6)}$$

Second, for $F_{Y_0(1,1)|D=1,M=1}(y)$ we have

$$F^{-1}_{U|11}(F_{Y_0(1,1)|D=1,M=1}(y)) = h^{-1}(1, 1, 1; y). \qquad \text{(B.7)}$$

Combining (B.6) and (B.7) yields,

$$h(1, 1, 1, h^{-1}(1, 1, 0; y)) = F_{Y_1(1,1)|D=1,M=1}^{-1} \circ F_{Y_0(1,1)|D=1,M=1}(y). \qquad \text{(B.8)}$$

The left sides of (B.4) and (B.8) are equal. In contrast to (B.4), (B.8) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(1,1)|D=1,M=1}(y) = \Pr(Y_t(1,1) \le y|D = 1, M = 1) = \Pr(Y_t \le y|D = 1, M = 1)$. Accordingly, we can identify $F_{Y_1(1,1)|D=0,M=1}^{-1} \circ F_{Y_0(1,1)|D=0,M=1}(y)$ by $Q_{11}(y) \equiv F_{Y_1|D=1,M=1}^{-1} \circ F_{Y_0|D=1,M=1}(y)$.

Parsing Y_0 through $Q_{11}(\cdot)$ in the non-treated group with mediator gives

$$\begin{aligned}
&E[Q_{11}(Y_0)|D = 0, M = 1] \\
&\quad = E[F_{Y_1|D=1,M=1}^{-1} \circ F_{Y_0|D=1,M=1}(Y_0)|D = 0, M = 1], \\
&\quad = E[F_{Y_1(1,1)|D=1,M=1}^{-1} \circ F_{Y_0(1,1)|D=1,M=1}(Y_0(0,1))|D = 0, M = 1], \\
&\quad \overset{A1,A5b}{=} E[h(1, 1, 1, h^{-1}(1, 1, 0; Y_0(0, 1)))|D = 0, M = 1], \qquad \text{(B.9)} \\
&\quad \overset{A2}{=} E[h(1, 1, 1, h^{-1}(1, 1, 0; Y_0(0, 0)))|D = 0, M = 1], \\
&\quad \overset{A1,A5a}{=} E[F_{Y_1(1,1)|D=0,M=1}^{-1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_0(0, 0))|D = 0, M = 1], \\
&\quad = E[Y_1(1, 1)|D = 0, M = 1] = E[Y_1(1, 1)|D = 0, M(0) = 1],
\end{aligned}$$

which has data support because of Assumption 6a.

B.2 Quantile direct effect under d = 0 conditional on D = 0 and M(0) = 1

In the following, we show that

$$\begin{aligned}
\theta_1^{0,1}(q, 0) &= F_{Y_1(1,1)|D=0,M(0)=1}^{-1}(q) - F_{Y_1(0,1)|D=0,M(0)=1}^{-1}(q), \\
&= F_{Q_{11}(Y_0)|D=0,M=1}^{-1}(q) - F_{Y_1|D=0,M=1}^{-1}(q).
\end{aligned}$$

For this purpose, we have to prove that

$$F_{Y_1(1,1)|D=0,M(0)=1}(y) = F_{Q_{11}(Y_0)|D=0,M=1}(y) \text{ and} \tag{B.10}$$

$$F_{Y_1(0,1)|D=0,M(0)=1}(y) = F_{Y_1|D=0,M=1}(y), \tag{B.11}$$

which is sufficient to show that the quantiles are also identified. We can show (B.11) using the observational rule $F_{Y_1(0,1)|D=0,M(0)=1}(y) = F_{Y_1|D=0,M=1}(y) = E[1\{Y_1 \leq y\}|D = 0, M = 1]$.

Using (B.9), we obtain

$$F_{Q_{11}(Y_0)|D=0,M=1}(y)$$
$$= E[1\{Q_{11}(Y_0) \leq y\}|D = 0, M = 1],$$
$$= E[1\{F_{Y_1|D=1,M=1}^{-1} \circ F_{Y_0|D=1,M=1}(Y_0) \leq y\}|D = 0, M = 1], \tag{B.12}$$
$$= E[1\{Y_1(1,1) \leq y\}|D = 0, M = 0],$$
$$= F_{Y_1(1,1)|D=0,M(0)=1}(y),$$

which proves (B.10).

B.3 Average direct effect under d = 1 conditional on D = 1 and M(1) = 1

In the following, we show that $\theta_1^{1,1}(1) = E[Y_1(1,1) - Y_1(0,1)|D = 1, M(1) = 1] = E[Y_1 - Q_{01}(Y_0)|D = 1, M = 1]$. Using the observational rule, we obtain $E[Y_1(1,1)|D = 1, M(1) = 1] = E[Y_1|D = 1, M = 1]$. Accordingly, we have to show that $E[Y_1(0,1)|D = 1, M(1) = 1] = E[Q_{01}(Y_0)|D = 1, M = 1]$ to finish the proof.

First, using (B.5) to evaluate $F_{Y_1(0,1)|D=1,M=1}(y)$ at $h(0,1,1,u)$ gives

$$F_{Y_1(0,1)|D=1,M=1}(h(0,1,1,u)) = F_{U|11}(h^{-1}(0,1,1;h(0,1,1,u))) = F_{U|11}(u).$$

Applying $F^{-1}_{Y_1(0,1)|D=1,M=1}(q)$ to both sides, we have

$$h(0,1,1,u) = F^{-1}_{Y_1(0,1)|D=1,M=1}(F_{U|11}(u)). \tag{B.13}$$

Second, for $F_{Y_0(0,1)|D=0,M=1}(y)$ we obtain

$$F^{-1}_{U|11}(F_{Y_0(0,1)|D=1,M=1}(y)) = h^{-1}(0,1,0;y), \tag{B.14}$$

using (B.5). Combining (B.13) and (B.14) yields,

$$h(0,1,1,h^{-1}(0,1,0;y)) = F^{-1}_{Y_1(0,1)|D=1,M=1} \circ F_{Y_0(0,1)|D=1,M=1}(y). \tag{B.15}$$

Note that $h(0,1,1,h^{-1}(0,1,0;y))$ maps the period 1 (potential) outcome of an individual with the outcome y in period 0 under non-treatment with the mediator. Accordingly, $E[F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_0)|D=0,M=1] = E[Y_1(1,1)|D=0,M=1]$. We can identify $F_{Y_0(1,1)|D=0,M=1}(y) = F_{Y_0|D=0,M=1}(y)$ under Assumption 2, but we cannot identify $F_{Y_1(1,1)|D=0,M=1}(y)$. However, we show in the following that we can identify the overall quantile-quantile transform $F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(y)$ under the additional Assumption 5a.

First, using (B.1) to evaluate $F_{Y_1(0,1)|D=0,M=1}(y)$ at $h(0,1,1,u)$ gives

$$F_{Y_1(0,1)|D=0,M=1}(h(0,1,1,u)) = F_{U|01}(h^{-1}(0,1,1;h(0,1,1,u))) = F_{U|01}(u).$$

Applying $F^{-1}_{Y_1(0,1)|D=0,M=1}(q)$ to both sides, we have

$$h(0,1,1,u) = F^{-1}_{Y_1(0,1)|D=0,M=1}(F_{U|01}(u)). \tag{B.16}$$

Second, for $F_{Y_0(0,1)|D=0,M=1}(y)$ we obtain

$$F^{-1}_{U|01}(F_{Y_0(0,1)|D=0,M=1}(y)) = h^{-1}(0,1,1;y), \tag{B.17}$$

46

using (B.1). Combining (B.16) and (B.17) yields,

$$h(0, 1, 1, h^{-1}(0, 1, 0; y)) = F^{-1}_{Y_1(0,1)|D=0,M=1} \circ F_{Y_0(0,1)|D=0,M=1}(y). \qquad \text{(B.18)}$$

The left sides of (B.15) and (B.18) are equal. In contrast to (B.15), (B.18) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(0,1)|D=0,M=1}(y) = \Pr(Y_t(0,1) \le y | D = 0, M = 1) = \Pr(Y_t \le y | D = 0, M = 1)$. Accordingly, we can identify $F^{-1}_{Y_1(0,1)|D=1,M=1} \circ F_{Y_0(0,1)|D=1,M=1}(y)$ by $Q_{01}(y) \equiv F^{-1}_{Y_1|D=0,M=1} \circ F_{Y_0|D=0,M=1}(y)$.

Parsing Y_0 through $Q_{01}(\cdot)$ in the treated group with mediator gives

$$
\begin{aligned}
E[Q_{01}(Y_0)|D = 1, M = 1] \\
&= E[F^{-1}_{Y_1|D=0,M=1} \circ F_{Y_0|D=0,M=1}(Y_0)|D = 1, M = 1], \\
&= E[F^{-1}_{Y_1(0,1)|D=0,M=1} \circ F_{Y_0(0,1)|D=0,M=1}(Y_0(1,1))|D = 1, M = 1], \\
&\stackrel{A1,A5a}{=} E[h(0, 1, 1, h^{-1}(0, 1, 0; Y_0(1, 1)))|D = 1, M = 1], \qquad\qquad \text{(B.19)} \\
&\stackrel{A2}{=} E[h(0, 1, 1, h^{-1}(0, 1, 0; Y_0(0, 1)))|D = 1, M = 1], \\
&\stackrel{A1,A5b}{=} E[F^{-1}_{Y_1(0,1)|D=1,M=1} \circ F_{Y_0(0,1)|D=1,M=1}(Y_0(0, 1))|D = 1, M = 1], \\
&= E[Y_1(0, 1)|D = 1, M = 1] = E[Y_1(0, 1)|D = 0, M(0) = 1],
\end{aligned}
$$

which has data support under Assumption 6b.

B.4 Quantile direct effect under $d = 1$ conditional on $D = 1$ and $M(1) = 1$

In the following, we show that

$$
\begin{aligned}
\theta_1^{1,1}(q, 1) &= F^{-1}_{Y_1(1,1)|D=1,M(1)=1}(q) - F^{-1}_{Y_1(0,1)|D=1,M(1)=1}(q), \\
&= F^{-1}_{Y_1|D=1,M=1}(q) - F^{-1}_{Q_{01}(Y_0)|D=1,M=1}(q).
\end{aligned}
$$

For this purpose, we have to prove that

$$F_{Y_1(1,1)|D=1,M(1)=1}(y) = F_{Y_1|D=1,M=1}(y) \text{ and} \tag{B.20}$$

$$F_{Y_1(0,1)|D=1,M(1)=1}(y) = F_{Q_{01}(Y_0)|D=1,M=1}(y), \tag{B.21}$$

which is sufficient to show that the quantiles are also identified. We can show (B.20) using the observational rule $F_{Y_1(1,1)|D=1,M(1)=1}(y) = F_{Y_1|D=1,M=1}(y) = E[1\{Y_1 \leq y\}|D = 1, M = 1]$.

Using (B.19), we obtain

$$
\begin{aligned}
F_{Q_{01}(Y_0)|D=1,M=1}(y) \\
&= E[1\{Q_{01}(Y_0) \leq y\}|D = 1, M = 1], \\
&= E[1\{F^{-1}_{Y_1|D=0,M=1} \circ F_{Y_0|D=0,M=1}(Y_0) \leq y\}|D = 1, M = 1], \\
&= E[1\{Y_1(0,1) \leq y\}|D = 1, M = 0], \\
&= F_{Y_1(0,1)|D=1,M(1)=1}(y),
\end{aligned}
$$

which proves (B.21).

C   Proof of equations (1) and (2)

$\Delta_1 = E[Y_1|D = 1] - E[Y_1|D = 0]$ and quantile treatment effect $\Delta_1(q) = F^{-1}_{Y_1|D=1}(q) - F^{-1}_{Y_1|D=0}(q)$

The average total effect for the entire population is identified by,

$$
\begin{aligned}
\Delta_1 &= E[Y_1(1, M(1))] - E[Y_1(0, M(0))], \\
&\overset{A7}{=} E[Y_1(1, M(1))|D = 1] - E[Y_1(0, M(0))|D = 0], \\
&= E[Y_1|D = 1] - E[Y_1|D = 0],
\end{aligned}
$$

where the first equality is the definition of Δ_1, the second equality hold by Assumption 7, and the last equality holds by the observational rule.

We define the conditional distribution $F_{Y_1|D=d}(y) = \Pr(Y_1 \leq y|D = d)$ and

$F_{Y_1|D=d}^{-1}(q) = \inf\{y : F_{Y_1|D=d}(y) \geq q\}$. We can show the identification of the total QTE for the entire population $\Delta_1(q) = F_{Y_1|D=1}^{-1}(q) - F_{Y_1|D=0}^{-1}(q)$ when we show that $F_{Y_1(1,M(1))}(y) = F_{Y_1|D=1}(y)$ and $F_{Y_1(0,M(0))}(y) = F_{Y_1|D=0}(y)$. Using Assumtpion 7 and the observational rule gives,

$$
\begin{aligned}
F_{Y_1(1,M(1))}(y) &= \Pr(Y_1(1, M(1)) \leq y), \\
&\overset{A7}{=} \Pr(Y_1(1, M(1)) \leq y|D = 1), \\
&= \Pr(Y_1 \leq y|D = 1) = F_{Y_1|D=1}(y),
\end{aligned}
$$

and

$$
\begin{aligned}
F_{Y_1(0,M(0))}(y) &= \Pr(Y_1(0, M(0)) \leq y), \\
&\overset{A7}{=} \Pr(Y_1(0, M(0)) \leq y|D = 0), \\
&= \Pr(Y_1 \leq y|D = 0) = F_{Y_1|D=0}(y),
\end{aligned}
$$

which finishes the proof.

By Assumption 7, the share of a type τ conditional on D corresponds to p_τ (in the population), as D is randomly assigned. This implies that $p_{1|1} = p_a + p_c$, $p_{1|0} = p_a + p_{de}$, $p_{0|1} = p_n + p_{de}$, and $p_{0|0} = p_n + p_c$. Under Assumption 8, $p_{de} = 0$, which finishes the proof of (1).

Furthermore, $E[Y_t(d, m)|\tau, D = 1] = E[Y_t(d, m)|\tau, D = 0] = E[Y_t(d, m)|\tau]$ due to the independence of D and the potential outcomes as well as the types τ (which are a deterministic function of $M(d)$) under Assumption 7. It follows that conditioning on D is not required on the right hand side of the following equation, which expresses the mean outcome conditional $D = 0$ and $M = 0$ as weighted average of the mean potential outcomes of compliers and never-takers:

$$
\begin{aligned}
&E[Y_t|D = 0, M = 0] \\
&= \frac{p_n}{p_n + p_c} E[Y_t(0,0)|\tau = n] + \frac{p_c}{p_n + p_c} E[Y_t(0,0)|\tau = c].
\end{aligned}
\tag{C.1}
$$

Only compliers and never-takers satisfy $M(0) = 0$ and thus make up the group with $D = 0$ and $M = 0$. After some rearrangements we obtain

$$
\begin{aligned}
& E[Y_t(0,0)|\tau = n] - E[Y_t(0,0)|\tau = c] \\
& = \frac{p_n + p_c}{p_c} \{E[Y_t(0,0)|\tau = n] - E[Y_t|D = 0, M = 0]\}.
\end{aligned}
\tag{C.2}
$$

Next, we consider observations with $D = 1$ and $M = 0$, which might consist of both never-takers and defiers, as $M(1) = 0$ for both types. However, by Assumption 8, defiers are ruled out, such that the mean outcome given $D_1 = 1$ and $M_1 = 0$ is determined by never-takers only:

$$
E[Y_t|D = 1, M = 0] \overset{A7,A8}{=} E[Y_t(1,0)|\tau = n].
\tag{C.3}
$$

Furthermore, by Assumption 2,

$$
E[Y_0(0,0)|\tau = n] \overset{A2}{=} E[Y_0(1,0)|\tau = n] \overset{A7,A8}{=} E[Y_0|D = 1, M = 0].
$$

Similarly to (C.1) for the never-takers and compliers, consider the mean outcome given $Z = 1$ and $D = 1$, which is made up by always-takers and compliers (the types with $M(1) = 1$)

$$
\begin{aligned}
& E[Y_t|D = 1, M = 1] \\
& = \frac{p_a}{p_a + p_c} E[Y_t(1,1)|\tau = a] + \frac{p_c}{p_a + p_c} E[Y_t(1,1)|\tau = c].
\end{aligned}
\tag{C.4}
$$

After some rearrangements we obtain

$$
\begin{aligned}
& E[Y_t(1,1)|\tau = a] - E[Y_t(1,1)|\tau = c] \\
& = \frac{p_a + p_c}{p_c} \{E[Y_t(1,1)|\tau = a] - E[Y_t|D = 1, M = 1]\}.
\end{aligned}
\tag{C.5}
$$

By Assumptions 7 and 8,

$$
E[Y_t|D = 0, M = 1] = E[Y_t(0,1)|\tau = a].
\tag{C.6}
$$

Now consider (C.5) for period $T = 0$, and note that by Assumption 2, $E[Y_0(1,1)|\tau = a] = E[Y_0(0,0)|\tau = a] = E[Y_0(0,1)|\tau = a]$ and $E[Y_0(1,1)|\tau = c] = E[Y_0(0,0)|\tau = c]$.

Combining (C.4), (C.6), and the law of iterative expectations (LIE) gives

$E[Y_0|D = 1]$

$\quad \overset{LIE}{=} E[Y_0|D = 1, M = 1] \cdot p_{1|1} + E[Y_0|D = 1, M = 0] \cdot p_{0|1},$

$\quad = E[Y_0(1,1)|\tau = c] \cdot p_c + E[Y_0(1,1)|\tau = a] \cdot p_a + E[Y_0(1,0)|\tau = n] \cdot p_n,$

$\quad \overset{A2}{=} E[Y_0(1,1)|\tau = c] \cdot p_c + E[Y_0(1,1)|\tau = a] \cdot p_a + E[Y_0(0,0)|\tau = n] \cdot p_n.$

Likewise, combining (C.1) and (C.3) gives

$E[Y_0|D = 0]$

$\quad \overset{LIE}{=} E[Y_0|D = 0, M = 1] \cdot p_{1|0} + E[Y_0|D_0 = 1, M = 0] \cdot p_{0|0},$

$\quad = E[Y_0(0,1)|\tau = a] \cdot p_a + E[Y_0(0,0)|\tau = c] \cdot p_c + E[Y_0(0,0)|\tau = n] \cdot p_n,$

$\quad \overset{A2}{=} E[Y_0(1,1)|\tau = a] \cdot p_a + E[Y_0(0,0)|\tau = c] \cdot p_c + E[Y_0(0,0)|\tau = n] \cdot p_n.$

Accordingly,

$$\frac{E[Y_0|D = 1] - E[Y_0|D = 0]}{p_{1|1} - p_{1|0}} = E[Y_0(1,1)|\tau = c] - E[Y_0(0,0)|\tau = c] \overset{A2}{=} 0,$$

which proves (2). Accordingly, $E[Y_0|D = 1] - E[Y_0|D = 0] = 0$ is a testable implication of Assumption 2, 7, and 8.

D   Proof of Theorem 3

D.1   Average direct effect on the never-takers

In the following, we show that $\theta_1^n = E[Y_1(1,0) - Y_1(0,0)|\tau = n] = E[Y_1 - Q_{00}(Y_0)|D = 1, M = 0]$. From (C.3), we obtain the first ingredient $E[Y_1(1,0)|\tau = n] = E[Y_1|D = 1, M = 0]$. Furthermore, from (A.9) we have $E[Q_{00}(Y_0)|D = 1, M = 0] = E[Y_1(0,0)|D =$

$1, M(1) = 0]$. Under Assumption 7 and 8,

$$E[Y_1(0,0)|D = 1, M(1) = 0] \overset{A7}{=} E[Y_1(0,0)|D = 1, \tau = n]$$
$$\overset{A8}{=} E[Y_1(0,0)|\tau = n]. \tag{D.1}$$

D.2   Quantile direct effect on the never-takers

We prove that

$$\theta_1^n(q) = F_{Y_1(1,0)|\tau=n}^{-1}(q) - F_{Y_1(0,0)|\tau=n}^{-1}(q),$$
$$= F_{Y_1|D=1,M=0}^{-1}(q) - F_{Q_{00}(Y_0)|D=1,M=0}^{-1}(q).$$

This requires showing that

$$F_{Y_1(1,0)|\tau=n}(y) = F_{Y_1|D=1,M=0}(y) \text{ and} \tag{D.2}$$

$$F_{Y_1(0,0)|\tau=n}(y) = F_{Q_{00}(Y_0)|D=1,M=0}(y). \tag{D.3}$$

Under Assumptions 7 and 8,

$$F_{Y_t|D=1,M=0}(y) = E[1\{Y_t \le y\}|D = 1, M = 0]$$
$$\overset{A7,A8}{=} E[1\{Y_t(1,0) \le y\}|\tau = n] \tag{D.4}$$
$$= F_{Y_t(1,0)|\tau=n}(y),$$

which proves (D.2). From (A.12), we have

$$F_{Q_{00}(Y_0)|D=1,M=0}(y) = F_{Y_1(0,0)|D=1,M(1)=0}(y) = E[1\{Y_1(0,0) \le y\}|D = 1, M(1) = 0].$$

Under Assumption 7 and 8,

$$E[1\{Y_1(0,0) \le y\}|D = 1, M(1) = 0] \overset{A7,A8}{=} E[1\{Y_1(0,0) \le y\}|\tau = n]$$
$$= F_{Y_1(0,0)|\tau=n}(y), \tag{D.5}$$

which proves (D.3).

D.3 Average direct effect under d = 0 on compliers

In the following, we show that

$$\theta_1^c(0) = E[Y_1(1,0) - Y_1(0,0)|\tau = c],$$

$$= \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{10}(Y_0) - Y_1|D = 0, M = 0]$$

$$- \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_1 - Q_{00}(Y_0)|D = 1, M = 0].$$

Plugging (D.1) in (C.1) under $T = 1$, we obtain

$$E[Y_1|D = 0, M = 0] = \frac{p_n}{p_n + p_c} E[Q_{00}(Y_0)|D = 1, M = 0]$$

$$+ \frac{p_c}{p_n + p_c} E[Y_1(0,0)|\tau = c].$$

This allows identifying

$$E[Y_1(0,0)|\tau = c] = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Y_1|D = 0, M = 0]$$

$$- \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Q_{00}(Y_0)|D = 1, M = 0]. \qquad \text{(D.6)}$$

Accordingly, we have to show the identification of $E[Y_1(1,0)|c]$ to finish the proof. From (A.19) we have $E[Y_1(1,0)|D = 0, M = 0] = E[Q_{10}(Y_0)|D = 0, M = 0]$. Applying the law of iterative expectations, gives

$$E[Y_1(1,0)|D = 0, M = 0] = \frac{p_n}{p_n + p_c} E[Y_1(1,0)|D = 0, M = 0, \tau = n]$$

$$+ \frac{p_c}{p_n + p_c} E[Y_1(1,0)|D = 0, M = 0, \tau = c],$$

$$\stackrel{A7}{=} \frac{p_n}{p_n + p_c} E[Y_1(1,0)|\tau = n] + \frac{p_c}{p_n + p_c} E[Y_1(1,0)|\tau = c].$$

After some rearrangements and using (C.3), we obtain

$$E[Y_1(1,0)|\tau = c] = \frac{p_n + p_c}{p_c} E[Q_{10}(Y_0)|D = 0, M = 0] - \frac{p_n}{p_c} E[Y_1|D = 1, M = 0].$$

This gives

$$E[Y_1(1,0)|\tau = c] = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{10}(Y_0)|D = 0, M = 0]$$
$$- \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_1|D = 1, M = 0], \tag{D.7}$$

using $p_n = p_{0|1}$, and $p_c + p_n = p_{0|0}$.

D.4 Quantile direct effect under d = 0 on compliers

We show that

$$F_{Y_1(1,0)|\tau = c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Q_{10}(Y_0)|D=0, M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}c} F_{Y_1|D=1, M=0}(y) \text{ and}$$

$$F_{Y_1(0,0)|\tau = c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Y_1|D=0, M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Q_{00}(Y_0)|D=1, M=0}(y),$$

which proves that $\theta_1^c(q, 0) = F_{Y_1(1,0)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$ is identified.

From (A.20), we have $F_{Y_1(1,0)|D=0, M(0)=0}(y) = F_{Q_{10}(Y_0)|D=0, M=0}(y)$. Applying the law of iterative expectations gives

$$F_{Y_1(1,0)|D=0, M(0)=0}(y) = \frac{p_n}{p_n + p_c} F_{Y_1(1,0)|D=0, M(0)=0, \tau=n}(y)$$
$$+ \frac{p_c}{p_n + p_c} F_{Y_1(1,0)|D=0, M(0)=0, \tau=c}(y),$$
$$\overset{A7}{=} \frac{p_n}{p_n + p_c} F_{Y_1(1,0)|\tau=n}(y) + \frac{p_c}{p_n + p_c} F_{Y_1(1,0)|\tau=c}(y).$$

Using (D.2) and rearranging the equation gives,

$$F_{Y_1(1,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Q_{10}(Y_0)|D=0, M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Y_1|D=1, M=0}(y). \tag{D.8}$$

In analogy to (C.1), the outcome distribution under $D = 0$ and $M = 0$ equals:

$$F_{Y_1|D=0, M=0}(y) = \frac{p_n}{p_n + p_c} F_{Y_1(0,0)|\tau=n}(y) + \frac{p_c}{p_n + p_c} F_{Y_1(0,0)|\tau=c}(y).$$

Using (D.3) and rearranging the equation gives

$$F_{Y_1(0,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Y_1|D=0,M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Q_{00}(Y_0)|D=1,M=0}(y). \quad \text{(D.9)}$$

E   Proof of Theorem 4

E.1   Average direct effect on the always-takers

In the following, we show that $\theta_1^a = E[Y_1(1,1) - Y_1(0,1)|\tau = a] = E[Q_{11}(Y_0) - Y_1|D = 0, M = 1]$. From (C.6), we obtain the first ingredient $E[Y_1(0,1)|a] = E[Y_1|D = 0, M = 1]$. Furthermore, from (B.9) we have $E[Q_{11}(Y_0)|D = 0, M = 1] = E[Y_1(1,1)|D = 0, M(0) = 1]$. Under Assumption 7 and 8,

$$\begin{aligned}
E[Y_1(1,1)|D = 0, M(0) = 1] &\overset{A7}{=} E[Y_1(1,1)|D = 0, \tau = a] \\
&\overset{A8}{=} E[Y_1(1,1)|\tau = a].
\end{aligned} \quad \text{(E.1)}$$

E.2   Quantile direct effect on the always-takers

We prove that

$$\begin{aligned}
\theta_1^a(q) &= F_{Y_1(1,1)|\tau=a}^{-1}(q) - F_{Y_1(0,1)|\tau=a}^{-1}(q), \\
&= F_{Q_{11}(Y_0)|D=0,M=1}^{-1}(q) - F_{Y_1|D=0,M=1}^{-1}(q).
\end{aligned}$$

This requires showing that

$$F_{Y_1(1,1)|\tau=a}(y) = F_{Q_{11}(Y_0)|D=0,M=1}(y) \text{ and} \quad \text{(E.2)}$$

$$F_{Y_1(0,1)|\tau=a}(y) = F_{Y_1|D=0,M=1}(y). \quad \text{(E.3)}$$

Under Assumptions 7 and 8,

$$
\begin{aligned}
F_{Y_t|D=0,M=1}(y) &= E[1\{Y_t \le y\}|D=0, M=1] \\
&\stackrel{A7,A8}{=} E[1\{Y_t(0,1) \le y\}|\tau=a] \\
&= F_{Y_t(0,1)|\tau=a,}(y).
\end{aligned}
\tag{E.4}
$$

which proves (E.3). From (B.12), we have

$$
F_{Q_{11}(Y_0)|D=0,M=1}(y) = F_{Y_1(1,1)|D=0,M(0)=1}(y) = E[1\{Y_1(1,1) \le y\}|D=0, M(0)=1].
$$

Under Assumption 7 and 8,

$$
\begin{aligned}
E[1\{Y_1(1,1) \le y\}|D=0, M(0)=1] &\stackrel{A7,A8}{=} E[1\{Y_1(1,1) \le y\}|\tau=a] \\
&= F_{Y_1(1,1)|\tau=a}(y),
\end{aligned}
\tag{E.5}
$$

which proves (E.2).

E.3   Average direct effect under d = 1 on compliers

In the following, we show that

$$
\begin{aligned}
\theta_1^c(1) =& E[Y_1(1,1) - Y_1(0,1)|\tau=c], \\
=& \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_1 - Q_{01}(Y_0)|D=1, M=1] \\
& - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0) - Y_1|D=0, M=1].
\end{aligned}
$$

Plugging (E.1) in (C.4), we obtain

$$
\begin{aligned}
E[Y_1|D=1, M=1] =& \ \frac{p_a}{p_a + p_c} E[Q_{11}(Y_0)|D=0, M=1] \\
& + \frac{p_c}{p_a + p_c} E[Y_1(1,1)|\tau=c].
\end{aligned}
$$

This allows identifying

$$E[Y_1(1,1)|\tau = c] = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_1|D=1, M=1]$$
$$- \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D=0, M=1]. \tag{E.6}$$

From (B.19) we have $E[Y_1(0,1)|D=1, M=1] = E[Q_{01}(Y_0)|D=1, M=1]$. Applying the law of iterative expectations, gives

$$E[Y_1(0,1)|D=1, M=1] = \frac{p_a}{p_a + p_c} E[Y_1(0,1)|D=1, M=1, \tau=a]$$
$$+ \frac{p_c}{p_a + p_c} E[Y_1(0,1)|D=1, M=1, \tau=c],$$
$$\stackrel{A7}{=} \frac{p_a}{p_a + p_c} E[Y_1(0,1)|\tau=a] + \frac{p_c}{p_a + p_c} E[Y_1(0,1)|\tau=c].$$

After some rearrangements and using (C.6), we obtain

$$E[Y_1(0,1)|\tau = c] = \frac{p_a + p_c}{p_c} E[Q_{01}(Y_0)|D=1, M=1] - \frac{p_a}{p_c} E[Y_1|D=0, M=1].$$

This gives

$$E[Y_1(0,1)|\tau = c] = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Q_{01}(Y_0)|D=1, M=1]$$
$$- \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Y_1|D=0, M=1], \tag{E.7}$$

with $p_a = p_{1|0}$, and $p_c + p_a = p_{1|1}$.

E.4 Quantile direct effect under d = 1 on compliers

We show that

$$F_{Y_1(1,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} F_{Y_1|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} F_{Q_{11}(Y_0)|D=0,M=1}(y) \text{ and}$$

$$F_{Y_1(0,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} F_{Q_{01}(Y_0)|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} F_{Y_1|D=0,M=1}(y),$$

which proves that $\theta_1^c(q,1) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(0,1)|c}^{-1}(q)$ is identified.

In analogy to (C.4), the outcome distribution under $D = 0$ and $M = 0$ equals:

$$F_{Y_1|D=1,M=1}(y) = \frac{p_a}{p_a + p_c}F_{Y_1(1,1)|\tau=a}(y) + \frac{p_c}{p_a + p_c}F_{Y_1(1,1)|\tau=c}(y).$$

Using (E.2) and rearranging the equation gives

$$F_{Y_1(1,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}}F_{Y_1|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}}F_{Q_{11}(Y_0)|D=0,M=1}(y). \quad \text{(E.8)}$$

From (B.21), we have $F_{Y_1(0,1)|D=1,M(1)=1}(y) = F_{Q_{01}(Y_0)|D=1,M=1}(y)$. Applying the law of iterative expectations gives

$$\begin{aligned}F_{Y_1(0,1)|D=1,M(1)=1}(y) =& \frac{p_a}{p_a + p_c}F_{Y_1(0,1)|D=1,M(1)=1,\tau=a}(y)\\ &+ \frac{p_c}{p_a + p_c}F_{Y_1(0,1)|D=1,M(1)=1,\tau=c}(y),\\ \overset{A7}{=}& \frac{p_a}{p_a + p_c}F_{Y_1(0,1)|\tau=a}(y) + \frac{p_c}{p_a + p_c}F_{Y_1(0,1)|\tau=c}(y).\end{aligned}$$

Using (E.3) and rearranging the equation gives,

$$F_{Y_1(0,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}}F_{Q_{01}(Y_0)|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}}F_{Y_1|D=0,M=1}(y). \quad \text{(E.9)}$$

F Proof of Theorem 5

F.1 Average treatment effect on the compliers

In (D.6) and (E.6), we show that

$$\begin{aligned}\theta_1^c =& E[Y_1(1,1) - Y_1(0,0)|\tau = c],\\ =& \frac{p_{1|1}}{p_{1|1} - p_{1|0}}E[Y_1|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}}E[Q_{11}(Y_0)|D = 0, M = 1]\\ &- \frac{p_{0|0}}{p_{0|0} - p_{0|1}}E[Y_1|D = 0, M = 0] + \frac{p_{0|1}}{p_{0|0} - p_{0|1}}E[Q_{00}(Y_0)|D = 1, M = 0].\end{aligned}$$

F.2 Quantile treatment effect on the compliers

In (E.8) and (D.9), we show that $F_{Y_1(1,1)|c}(y)$ and $F_{Y_1(0,0)|c}(y)$ are identified. Accordingly, $\Delta_1^c(q) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$ is identified.

F.3 Average indirect effect under $d = 0$ on compliers

In (D.6) and (E.7), we show that

$$
\begin{aligned}
\delta_1^c(0) =& E[Y_1(0,1) - Y_1(0,0)|\tau = c], \\
=& \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Y_1|D = 0, M = 1] \\
& - \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Y_1|D = 0, M = 0] + \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Q_{00}(Y_0)|D = 1, M = 0].
\end{aligned}
$$

F.4 Quantile indirect effect under $d = 0$ on compliers

In (E.9) and (D.9), we show that $F_{Y_1(0,1)|c}(y)$ and $F_{Y_1(0,0)|c}(y)$ are identified. Accordingly, $\delta_1^c(q, 0) = F_{Y_1(0,1)|c}^{-1}(q) - F_{Y_1(0,0)|c}^{-1}(q)$ is identified.

F.5 Average indirect effect under $d = 1$ on compliers

In (E.6) and (D.7), we show that

$$
\begin{aligned}
\delta_1^c(1) =& E[Y_1(1,1) - Y_1(1,0)|\tau = c], \\
=& \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_1|D = 1, M = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_0)|D = 0, M = 1] \\
& - \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{00}(Y_0)|D = 0, M = 0] + \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_1|D = 1, M = 0].
\end{aligned}
$$

F.6 Quantile indirect effect under $d = 1$ on compliers

In (E.8) and (D.8), we show that $F_{Y_1(1,1)|c}(y)$ and $F_{Y_1(1,0)|c}(y)$ are identified. Accordingly, $\delta_1^c(q, 1) = F_{Y_1(1,1)|c}^{-1}(q) - F_{Y_1(1,0)|c}^{-1}(q)$ is identified.

Authors

Martin HUBER
University of Fribourg, Department of Economics Chair of Applied Econometrics - Evaluation of Public Policies,
Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
Phone: +41 26 300 8255; Email: martin.huber@unifr.ch; Website: http://www.unifr.ch/appecon/en/

Mark SCHELKER
University of Fribourg, Department of Economics, Chair of Public Economics, Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
Phone: + +41 26 300 8269; Email: mark.schelker@unifr.ch; Website: http://www.unifr.ch/finpub/en/team/mark-schelker

Anthony STRITTMATTER
University of St. Gallen, Department of Economics, Swiss Institute for Empirical Economic Research, Varnbüelstrasse 14, 9000
St. Gallen, Switzerland. Phone: +41 71 224 23 05 Email: anthony.strittmatter@unisg.ch ; Website: www.anthonystrittmatter.com

Abstract

We propose a novel approach for causal mediation analysis based on changes-in-changes assumptions restricting unobserved heterogeneity over time. This allows disentangling the causal effect of a binary treatment on a continuous outcome into an indirect effect operating through a binary intermediate variable (called mediator) and a direct effect running via other causal mechanisms. We identify average and quantile direct and indirect effects for various subgroups under the condition that the outcome is monotonic in the unobserved heterogeneity and that the distribution of the latter does not change over time conditional on the treatment and the mediator. We also provide a simulation study and an empirical application to the Jobs II programme.

Citation proposal

Jel Classification

C21

Keywords

Direct effects; indirect effects; mediation analysis; changes-in-changes; causal mechanisms; treatment effects.

Working Papers SES collection

Last published

Catalogue and download links

http://www.unifr.ch/ses/wp
http://doc.rero.ch/collection/WORKING_PAPERS_SES

Publisher