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

WORK IN PROGRESS

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 an 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 program.

Keywords: Direct and indirect effects, causal mediation analysis, changes-in-changes, causal mechanisms.

JEL classification: C21.

Addresses for correspondence: Martin Huber, Chair of Applied Econometrics - Evaluation of Public Policies, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland, Martin.Huber@unifr.ch. Mark Schelker, Chair of Public Economics, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland, Mark.Schelker@unifr.ch. Anthony Strittmatter, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstr. 14, 9000 St. Gallen, Switzerland, Anthony.Strittmatter@unisg.ch, www.anthonystrittmatter.com.

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 effects running through a binary mediator. Outcomes are required to be observed both prior 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 outcome is strictly monotonic in unobserved heterogeneity and that the distribution of heterogeneity does not change over time conditional on the treatment and the mediator. 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 how/whether 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) or never one (never takers) irrespective of treatment assignment. Furthermore, we identify the total, direct, and indirect treatment effects on those whose mediator value complies with treatment assignment (compliers). If appropriately weighted, the respective average effects of these populations 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 (2018), Bellani and Bia (2018), and Huber, Lechner, and Strittmatter (2018). 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.

While most studies aim at evaluating direct and indirect effects in the total population, 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 how/whether 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 focussing on subpopulations rather than the 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 heterogeneiy 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 rather than mean 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 program previously analyzed by Vinokur, Price, and Schul (1995), a randomized job training intervention designed to analyze the impact of job training on mental health outcomes. Focussing on individuals with a high depression risk prior to treatment assignment, 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 program 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 small 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 total and indirect effects on those induced to participate by assignment are statistically significantly negative 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 7 provides a simulation study. Section 5 provides an application to Jobs II. Section 6 concludes.

2 Notation and effects

2.1 Average effects

For each individual $i = 1, ..., N$, let D_i denote a binary treatment (e.g., having two children with the same gender) and M_i a binary intermediate variable or mediator that may be a function of D_i (e.g., giving birth to a third child). Furthermore, let T indicate a particular time period: $T = 0$ denotes the baseline period prior to the realisation of D_i and M_i, $T = 1$ the follow up period after measuring D_i and M_i in which the effect of the outcome is evaluated. Finally, let Y_{it} denote the outcome of interest (e.g., income of mother) 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_i and M_i. To define the parameters of interest, we make use of the potential outcome notation, see for instance Rubin (1974), and denote by $Y_{it}(d, m)$ the potential outcome for treatment state $D_i = d$ and mediator state

$M_i = m$ in time $T = t$, with $d, m, t, \in \{0, 1\}$. Furthermore, let $M_i(d)$ 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_i and M_i, 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_i causally precedes M_i. Therefore, D_i and M_i 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 $\Delta_1 = E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(0))]$. That is, the ATE corresponds to the effect of D_i on the outcome that either affects the latter directly (net of any effect on the mediator) or indirectly through an effect on M_i. Indeed, the total ATE can be disentangled into the direct and indirect effects, denoted by $\theta_1(d) = E[Y_{i1}(1, M_i(d)) - Y_{i1}(0, M_i(d))]$ and $\delta_1(d) = E[Y_{i1}(d, M_i(1)) - Y_{i1}(d, M_i(0))]$, by adding and subtracting $Y_{i1}(1, M_i(0))$ or $Y_{i1}(0, M_i(1))$, respectively:

$$\begin{aligned}
\Delta_1 &= E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(0))], \\
&= \underbrace{E[Y_{i1}(1, M_i(1)) - Y_{i1}(1, M_i(0))]}_{=\delta_1(1)} + \underbrace{E[Y_{i1}(1, M_i(0)) - Y_{i1}(0, M_i(0))]}_{=\theta_1(0)}, \\
&= \underbrace{E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(1))]}_{=\theta_1(1)} + \underbrace{E[Y_{i1}(0, M_i(1)) - Y_{i1}(0, M_i(0))]}_{=\delta_1(0)}.
\end{aligned}$$

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_i and M_i such that the 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: $M_i(1) = M_i(0) = 1$) whose mediator is always one, compliers (c: $M_i(1) = 1$, $M_i(0) = 0$) whose mediator corresponds to the treatment value, defiers (de: $M_i(1) = 0$, $M_i(0) = 1$) whose mediator opposes the treatment value, and never-takers (n: $M_i(1) = M_i(0) = 0$) whose mediator is never one. Note that τ cannot be pinned down for any individual, because either $M_i(1)$ or $M_i(0)$ is observed, but never both.

Let $\Delta_1^\tau = E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(0))|\tau_i]$ 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 $M_i(1) = M_i(0) = 0$ for any never-taker, the indirect effect for this group is by definition zero ($\delta_1^n(d) = E[Y_{i1}(d, 0) - Y_{i1}(d, 0)|\tau_i = n] = 0$) and $\Delta_1^n = E[Y_{i1}(1, 0) - Y_{i1}(0, 0)|\tau_i = n] = \theta_1^n(1) = \theta_1^n(0) = \theta_1^n$ equals the direct effect. Correspondingly, because $M_i(1) = M_i(0) = 1$ for any always-taker, the indirect effect for this group is by definition zero ($\delta_1^a(d) = E[Y_{i1}(d, 0) - Y_{i1}(d, 0)|\tau_i = a] = 0$) and $\Delta_1^a = E[Y_{i1}(1, 1) - Y_{i1}(0, 1)|\tau_i = a] = \theta_1^a(1) = \theta_1^a(0) = \theta_1^a$ equals the direct effect. For the compliers, both direct and indirect effects may exist. Note that $M_i(d) = d$ due to the definition of compliers. Therefore, $\theta_1^c(d) = E[Y_{i1}(1, d) - Y_{i1}(0, d)|\tau_i = c]$ and $\delta_1^c(d) = E[Y_{i1}(d, 1) - Y_{i1}(d, 0)|\tau_i = c]$, while $\Delta_1^c = E[Y_{i1}(1, 1) - Y_{i1}(0, 0)|\tau_i = c]$. 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 $D_i = d$ and mediator states $M_i = M_i(d) = m$, which are denoted by $\theta_1^{d,m}(d) = E[Y_{i1}(1, m) - Y_{i1}(0, m)|D_i = d, M_i(d) = m]$. 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, $\theta_1^{1,0}(1) = E[Y_{i1}(1, 0) - Y_{i1}(0, 0)|D_i = 1, M_i(1) = 0]$ is the effect on a mixture of never takers and defiers, as these two groups satisfy $M_i(1) = 0$. Likewhise, $\theta_1^{0,0}(0)$ refers to never takers and compliers (satisfying $M_i(0) = 0$), $\theta_1^{0,1}(0)$ to always takers and defiers (satisfying $M_i(0) = 1$), and $\theta_1^{1,1}(1)$ to always takers and compliers (satisfying $M_i(1) = 1$).

2.2 Quantile effects

We denote by $F_{Y_t(d,m)}(y) = Pr(Y_{it}(d,m) \leq y)$ the cumulative distribution function of $Y_{it}(d,m)$ at outcome level y. Its inverse, $F^{-1}_{Y_t(d,m)}(q) = \inf\{y : F_{Y_{it}(d,m)}(y) \geq q\}$, is the quantile function of $Y_{it}(d,m)$ at rank q. The quantile treatment effects (QTEs) are denoted by $\Delta_1(q) = F^{-1}_{Y_{i1}(1,M_i(1))}(q) - F^{-1}_{Y_{i1}(0,M_i(0))}(q)$. The QTE can be disentangled into the direct quantile treatment effects, denoted by $\theta_1(q,d) = F^{-1}_{Y_{i1}(1,M_i(d))}(q) - F^{-1}_{Y_{i1}(0,M_i(d))}(q)$, and the indirect quantile treatment effects, denoted by $\delta_1(q,d) = F^{-1}_{Y_{i1}(d,M_i(1))}(q) - F^{-1}_{Y_{i1}(d,M_i(0))}(q)$.

The conditional distribution function by type τ_i is $F_{Y_t(d,m)|\tau}(y) = Pr(Y_{it}(d,m) \leq y|\tau_i)$ and the corresponding conditional quantile function is $F^{-1}_{Y_{it}(d,m)|\tau}(q) = \inf\{y : F_{Y_t(d,m)|\tau_i}(y) \geq q\}$ for $\tau_i \in \{a,c,d,n\}$. Using the previously described stratification approach, we can define the QTE conditional on $\tau \in \{a,c,de,n\}$: $\Delta_1^\tau(q) = F^{-1}_{Y_{i1}(1,M_i(1))|\tau}(q) - F^{-1}_{Y_{i1}(0,M_i(0))|\tau}(q)$. For never-takers, the QTEs $\Delta_1^n(q) = F^{-1}_{Y_{i1}(1,0)|n}(q) - F^{-1}_{Y_{i1}(0,0)|n}(q) = \theta_1^n(q)$ equals the direct quantile treatment effects. For any always-takers, the QTEs $\Delta_1^a(q) = F^{-1}_{Y_{i1}(1,1)|a}(q) - F^{-1}_{Y_{i1}(0,1)|a}(q) = \theta_1^a(q)$ equal the direct quantile treatment effects. For compliers, the QTEs are $\Delta_1^c(q) = F^{-1}_{Y_{i1}(1,1)|c}(q) - F^{-1}_{Y_{i1}(0,0)|c}(q)$, while the quantile direct treatment effects are $\theta_1^c(q,d) = F^{-1}_{Y_{i1}(1,d)|c}(q) - F^{-1}_{Y_{i1}(0,d)|c}(q)$ and the indirect quantile treatment effects are $\delta_1^c(q,d) = F^{-1}_{Y_{i1}(d,1)|c}(q) - F^{-1}_{Y_{i1}(d,0)|c}(q)$. Finally, we define the direct quantile treatment effects conditional on specific values $D_i = d$ and mediator states $M_i = M_i(d) = m$, which is

$$\theta_1^{d,m}(q,1) = F^{-1}_{Y_{i1}(1,m)|D_i=d,M_i(1)=m}(q) - F^{-1}_{Y_{i1}(0,m)|D_i=d,M_i(1)=m}(q) \text{ and}$$
$$\theta_1^{d,m}(q,0) = F^{-1}_{Y_{i1}(1,m)|D_i=d,M_i(0)=m}(q) - F^{-1}_{Y_{i1}(0,m)|D_i=d,M_i(0)=m}(q),$$

with $F_{Y_t(d,m)|D_i=d,M_i(d)=m}(y) = Pr(Y_{it}(d,m) \leq y|D_i = d, M_i(d) = m)$.

2.3 Additional notation

It shall be useful for later to define the conditional distribution function by treatment d and mediator status m, $F_{Y_t|D=d,M=m}(y) = Pr(Y_{it} \leq y|D_i = d, M_i = m)$. The corresponding conditional quantile function is $F^{-1}_{Y_t|D=d,M=m}(q) = \inf\{y : F_{Y_{it}|D_i=d,M_i=m}(y) \geq q\}$ for $d, m \in \{0, 1\}$. 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 into the corresponding y' at rank q in period 1.

3 Identification and Estimation

3.1 Identification

We subsequently discuss 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 discussion, however, covariates are not considered for the sake of ease of notation. Our first assumption implies that potential outcomes are characterized by a nonparametric function, denoted by h, that is strictly monotonic in a scalar U_i that reflects unobserved heterogeneity.

Assumption 1: Strict monotonicity of potential outcomes in ubobserved heterogeneity.

The potential outcomes satisfy the following model: $Y_{it}(d, m) = h(d, m, t, U_i)$, with the general function h being strictly increasing in the scalar unobservable U_i for all $d, m, t \in \{0, 1\}$.

Assumption 1 implies that individuals with identical unobserved characteristics U_i have the same potential outcomes $Y_{it}(d, m)$, while higher values of U_i correspond to strictly higher potential outcomes $Y_{it}(d, m)$. Strict monotonicity is automatically satisfied in additively separable models, but Assumption 1 also allows for more flexible non-additive structures that arise in nonparametric economic 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.
$Y_{i0}(d, m) - Y_{i0}(d', m') = 0$, for $d, d', m, m'\{1, 0\}$.

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 actual treatment status does not change.

Furthermore, we assume conditional independence between the unobserved heterogeneity and the time period given the treatment and non-mediator.

Assumption 3: Conditional independence of U_i and T given $D_i = 1, M_i = 0$ or $D_i = 0, M_i = 0$.
(a) $U_i \perp\!\!\!\perp T|D_i = 1, M_i = 0$,
(b) $U_i \perp\!\!\!\perp T|D_i = 0, M_i = 0$.

Under Assumption 3a, the distribution of U_i is allowed to vary by treatment and mediator group, but not over time conditional on $D_i = 1, M_i = 0$. Assumption 3b imposes the same restriction conditional on $D_i = 0, M_i = 0$. We may interpret Assumption 3 as stationarity of U_i within groups defined on D_i and M_i. This assumption is weaker than (and thus implied by) requiring that U_i is constant across T for each individual i. For example, Assumption 3 is satisfied in the fixed effect model $U_i = \eta_i + v_{it}$, with η_i being a time-invariant individual-specific unobservable (fixed effect) and v_{it} 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, $U_i \perp\!\!\!\perp T|D_i = d$, to identify the average treatment effect on the treated, $\varphi_1 = E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(0))|D_i = 1]$ or local average treatment effect, $\varphi_1 = E[Y_{i1}(1, M_i(1)) - Y_{i1}(0, M_i(0))|\tau_i = c]$, respectively. We additionally condition on the mediator status to identify direct and indirect effects.

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

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

(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_i in the population with $D_i = 1, M_i = 0$ is also contained in the population with $D_i = 0, M_i = 0$. Assumption 4b imposes that any value of U_i conditional on $D_i = 0, M_i = 0$ also exists conditional on $D_i = 0, M_i = 0$. Both assumptions together imply that the support of U_i 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 effect under $d = 1$ conditional on $D_i = 1$ and $M_i(1) = 0$ is identified:

$$\theta_1^{1,0}(1) = E[Y_{i1} - Q_{00}(Y_{i0})|D_i = 1, M_i = 0],$$
$$\theta_1^{1,0}(q, 1) = F^{-1}_{Y_{i1}|D_i=1,M_i=0}(q) - F^{-1}_{Q_{00}(Y_{i0})|D_i=1,M_i=0}(q).$$

(b) and Assumption 4b, the average and quantile direct effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 0$ is identified:

$$\theta_1^{0,0}(0) = E[Q_{10}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 0],$$
$$\theta_1^{0,0}(q, 0) = F^{-1}_{Q_{10}(Y_{i0})|D_i=0,M_i=0}(q) - F^{-1}_{Y_{i1}|D_i=0,M_i=0}(q).$$

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_i and T given $D_i = 0, M_i = 1$ or $D_i = 1, M_i = 1$.

(a) $U_i \perp\!\!\!\perp T|D_i = 0, M_i = 1$,

(b) $U_i \perp\!\!\!\perp T|D_i = 1, M_i = 1$.

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

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

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

(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_i in the population with $D_i = 0, M_i = 1$ is also contained in the population with $D_i = 1, M_i = 1$. Assumptions 6b states that any value of U_i conditional on $D_i = 1, M_i = 1$ exists conditional on $D_i = 0, M_i = 1$.

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 effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 1$ is identified:

$$\theta_1^{0,1}(0) = E[Y_{i1} - Q_{11}(Y_{i0})|D_i = 0, M_i = 1],$$
$$\theta_1^{0,1}(q, 0) = F^{-1}_{Y_{i1}|D_i=0,M_i=1}(q) - F^{-1}_{Q_{11}(Y_{i0})|D_i=0,M_i=1}(q).$$

(b) and Assumption 6b, the average and quantile direct effect under $d = 1$ is identified conditional on $D_i = 1$ and $M_i(1) = 1$:

$$\theta_1^{1,1}(1) = E[Q_{01}(Y_{i0}) - Y_{i1}|D_i = 1, M_i = 1],$$
$$\theta_1^{1,1}(q, 1) = F^{-1}_{Q_{01}(Y_{i0})|D_i=1,M_i=1}(q) - F^{-1}_{Y_{i1}|D_i=1,M_i=1}(q).$$

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_i as instrument and M_i 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.

Our next assumption imposes independence between the treatment and the potential post-treatment variables.

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

Assumption 7 implies that there are no confounders jointly affecting the treatment and the mediator and/or outcome and 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.
$Pr(M_i(1) \geq M_i(0)) = 1$.

Assumption 8 is standard in the instrumental variable literature on local average treatment effects when denoting by D_i the instrument and by M_i 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, Assumptions 7 and 8 yield the strata proportions, denoted by $p_{\tau_i} = Pr(\tau_i)$, as functions of the conditional mediator probabilities given the treatment, which we denote by $p_{(m|d)} = Pr(M_i = m|D_i = d)$ for $d, m \in \{0, 1\}$:

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

Furthermore, Assumptions 2, 7, and 8 imply that $\Delta_{0,c} = E[Y_{i0}(1,1) - Y_{i0}(0,0)|c] = E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0] = 0$. Therefore, a rejection of the testable implication $E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0] = 0$ in the data would point to a violation of these assumptions. Furthermore and as pointed out in Deuchert, Huber, and Schelker (2017), the assumptions imply that the differences in average baseline outcomes across always or never-takers and compliers are given by

$$E[Y_{i0}(0,0)|n] - E[Y_{i0}(0,0)|c] = \frac{p_n + p_c}{p_c}[E(Y_{i0}|D_i = 1, M_i = 0) - E(Y_{i0}|D_i = 0, M_i = 0)], \tag{1}$$

$$E[Y_{i0}(0,0)|a] - E[Y_{i0}(0,0)|c] = \frac{p_a + p_c}{p_c}[E(Y_{i0}|D_i = 0, M_i = 1) - E(Y_{i0}|D_i = 1, M_i = 1)], \tag{2}$$

see Appendix A.3.

The additional assumptions of treatment randomization and mediator monotonicity permit identifying total, direct, and indirect effects on compliers, never-takers, 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 effect on the never-takers is identified by:

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

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

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

Furthermore, the inverse of the distributions

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

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

identify the quantile direct effect under $d = 0$ on compliers $\theta_1^c(q,0) = F^{-1}_{Y_1(1,0)|c}(q) - F^{-1}_{Y_1(0,0)|c}(q)$.

Theorem 4: Under Assumptions 1–3, 5, 7-8,

a) and Assumption 6a, the average and quantile direct effect on the always-takers is identified by:

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

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

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

Furthermore, the inverse of the distributions

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

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

identify the quantile direct effect under $d=1$ on compliers $\theta_1^c(q,1) = F^{-1}_{Y_1(1,1)|c}(q) - F^{-1}_{Y_1(0,1)|c}(q)$.

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

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

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

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

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

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

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

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

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

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

3.2 Estimation

We apply the sample analogy principle and estimate population moments using sample moments (e.g. Manski, 1988). The most challenging estimation steps are the quantile transformations. The first application of quantile transformation dates back to at least Juhn, Murphy, and Pierce (1991) and current studies using this approach include Chaisemartin and D'Haultfeuille (2018), Wuethrich (2019), and Strittmatter (2019). We follow the estimation approach proposed by Athey and Imbens (2006). We replace the potential outcome distributions by empirical distribution functions. Transformed quantiles can be obtained using a plug-in approach between the empirical quantiles and ranks of the empirical distribution functions. The average effects are estimated by replacing the expected values with the transformed sample averages that incorporate quantile transformation. Athey and Imbens (2006) show that the resulting estimates are $\sqrt{n}$-consistent and asymptotically normal. We use a nonparametric bootstrap approach to calculate the standard errors. Chaisemartin and D'Haultfeuille (2018) show the validity of the bootstrap approach for this estimation. Athey and Imbens (2006) propose different ways to incorporate covariates into the estimation procedure. Melly and Santangelo (2015) provide formal results for a flexible semiparametric estimator that can incorporate covariates.

4 Simulations

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

$$
\begin{aligned}
T &\sim Binom(0.5),\ D \sim Binom(0.5),\ U \sim Unif(-1,1),\ V \sim N(0,1) \text{ independent of each other,} \\
M &= I\{D+U+V>0\}, \qquad Y_T = \Lambda((1+D+M+D\cdot M)\cdot T+U). \qquad (7)
\end{aligned}
$$

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 confounder 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. 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 Assumption 6. Compliers, always takers and never takers satisfy, respectively: $c = I\{U+V \leq 0, 1+U+V>0\}, a = I\{U+V>0\}, n = I\{1+U+V \leq 0\}$.

Table 1: Linear model with random treatment

	Changes-in-Changes							Difference-in-Differences						
	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
	n=1000													
bias	0.00	-0.00	-0.01	-0.01	-0.01	-0.00	-0.01	0.01	-0.00	-0.01	-0.01	0.00	-0.02	0.00
sd	0.11	0.08	0.23	0.10	0.13	0.27	0.27	0.11	0.09	0.14	0.14	0.12	0.19	0.10
rmse	0.11	0.08	0.23	0.10	0.13	0.27	0.27	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	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	0.11	0.04	0.05	0.07	0.12	0.10	0.10
	n=4000													
bias	-0.00	-0.00	0.00	-0.00	-0.01	0.01	0.01	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00
sd	0.06	0.04	0.12	0.05	0.07	0.14	0.14	0.06	0.04	0.07	0.07	0.06	0.10	0.05
rmse	0.06	0.04	0.12	0.05	0.07	0.14	0.14	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	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	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.

In our 1000 simulations, we consider two sample sizes ($n = 1000, 4000$) and investigate the behaviour of our CiC methods as well as the DiD approach of Deuchert, Huber, and Schelker (2017) in both a linear (Λ equal to identity function) and nonlinear outcome model, with Λ being equal to the exponential function. To implement the CiC estimators, we make use of the 'CiC' command in the 'qte' package by Callaway (2016) for the statistical software 'R' with its default values. Table 1 reports the bias, standard deviation ('sd'), root mean squared error ('rmse'), true effect ('true'), and the relative root mean squared error as 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 CiC and DiD 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

	Changes-in-Changes							Difference-in-Differences						
	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
	n=1000													
bias	0.01	-0.14	-0.48	-0.35	-0.11	-0.37	-0.13	-0.27	-8.91	14.42	11.46	-1.49	15.91	2.96
sd	0.48	5.08	8.47	6.20	1.16	8.64	4.23	0.46	2.62	2.58	2.62	0.47	2.61	0.47
rmse	0.48	5.08	8.48	6.21	1.17	8.65	4.23	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	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	0.15	0.14	0.28	0.25	0.33	0.34	0.63
	n=4000													
bias	-0.01	0.01	-0.00	-0.11	-0.07	0.07	0.11	-0.28	-8.79	14.51	11.57	-1.51	16.02	2.94
sd	0.25	2.63	4.37	3.20	0.66	4.44	2.04	0.24	1.28	1.26	1.28	0.25	1.27	0.23
rmse	0.25	2.63	4.37	3.20	0.66	4.44	2.04	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	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	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 nonlinear link function. While the CiC assumptions hold, average common time trends are heterogeneous across complier types such that the DiD approach of Deuchert, Huber, and Schelker (2017) is inconsistent. Accordingly, the biases of the CiC estimates generally approach zero as the sample size increases, while this is not the case for the DiD estimates. CiC yields a lower root mean squared error than the respective DiD estimator in all but one case (namely $\hat{\delta}_1^c(0)$ with $n = 1000$) and its relative attractiveness increases in the sample size due to its lower bias.

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 (7) is replaced by:

$$D = I\{U + Q > 0\},\ Q \sim N(0,1) \text{ independent of } T \text{ and any unobservable,} \tag{8}$$

where Q is 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 CiC-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.

5 Application

Our empirical application is based on the JOBS II dataset by Vinokur and Price (1999). JOBS II is a randomized job training intervention, designed to analyze the

Table 3: Nonlinear model with non-random treatment

	Changes-in-Changes							Difference-in-Differences						
	$\hat{\theta}_1^{0,1}$	$\hat{\theta}_1^{1,0}$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$	$\hat{\theta}_1^n$	$\hat{\theta}_1^a$	$\hat{\Delta}_c$	$\hat{\theta}_1^c(1)$	$\hat{\theta}_1^c(0)$	$\hat{\delta}_1^c(1)$	$\hat{\delta}_1^c(0)$
	n=1000													
bias	0.02	0.13	47.21	40.19	-1.44	48.64	7.02	0.35	19.98	29.00	27.65	0.04	28.96	1.35
sd	0.71	4.56	5.45	4.11	0.75	5.53	2.92	0.67	2.48	2.46	2.48	0.67	2.51	0.45
rmse	0.71	4.56	47.52	40.40	1.62	48.96	7.60	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	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	0.17	0.37	0.56	0.58	0.14	0.61	0.30
	n=4000													
bias	-0.00	0.06	47.38	40.13	-1.53	48.91	7.25	0.34	20.02	28.98	27.65	0.02	28.96	1.33
sd	0.38	2.35	2.84	2.04	0.38	2.86	1.51	0.35	1.22	1.19	1.22	0.35	1.24	0.23
rmse	0.38	2.35	47.47	40.18	1.57	48.99	7.40	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	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	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.

impact of job training on mental health outcomes, see Vinokur, Price, and Schul (1995). The study was conducted in southeastern Michigan, where 2005 job seekers were eligible to participate in a randomized field experiment. In a pre-screening, individuals were classified as having either a high or low depression risk. Since randomization was carried out at this group level, we focus on high depression risk respondents. The treatment consisted of five 4-hours job training seminars conducted in morning sessions during one week. The respondents in 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.[1] The control group received a booklet with information on job search methods.

We analyze the impact of this particular job training intervention on mental health outcomes, namely symptoms of depression, 6 months after the job training. The depression measure is based on a subset of 11 questions on depression symptoms

[1] In comparison to the earlier interventions in the context of the JOBS I program (Caplan, Vinokur, Price, and van Ryn (1989)), the job training sessions focused more strongly on building a sense of mastery, personal control and self- efficacy in job search. Previous research has shown that an increase in this sense of mastery, control and self-efficacy improved observed effort in job search behavior (Eden and Aviram (1993)). Marshall and Lang (1990) show that mastery is a strong predictor of depression symptoms in women. For a detailed discussion of the literature, the exact sampling process, the training program, etc., see Vinokur, Price, and Schul (1995).

of the standard 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, or having thoughts of ending their lives, or experiencing a loss of sexual interest. The questions were asked on a 5-point scale, going from 'not at all' (1) to 'extremely' (5). Depression is measured at a baseline before the random assignment to the job training program, and again 6 weeks and 6 months after the job training. The treatment (D) is the randomized assignment to the job training among our sample of respondents in the category with an elevated depression risk at the baseline. The mediator (M) is an indicator of whether or not a randomly selected respondent participated in the job training. The reason for considering the direct effect of mere treatment assignment is that the latter could have a motivation or discouragement effect on those randomly offered or not offered the training. In the context of outcomes related to mental health, the assignment might therefore operate through other mechanisms than actual participation. 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.

Table 4: 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	-0.03	-0.24	0.09	-0.34	0.01	-0.27	-0.01	-0.26	0.45	0.55
se	0.08	0.10	0.09	0.14	0.08	0.10	0.08	0.13	0.02	0.02
pval	0.72	0.02	0.28	0.01	0.90	0.01	0.88	0.05	0.00	0.00

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

Our sample consists of 1321 observation of respondents with a high risk of depression at the baseline and for which we have non-missing data for D, M, and Y. Given the study design, we do not observe always takers, as members of the control group were not having access to the job training program. From those assigned to the treatment, 45% did not show up and are counted as never takers, while 55% are compliers. Therefore, identification will be based 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 and Assumptions 4b and 6a for the indirect effects on compliers under $d = 1$ ($\delta_1^c(1)$).

Table 4 presents the estimation results for our identification strategy in the changes-in-changes (CiC) framework and compares them to the DiD strategy of Deuchert, Huber, and Schelker (2017). The estimates of the direct effects on never takers ($\hat{\theta}_1^c(0)$) as well as on compliers ($\hat{\theta}_1^c(0)$) are not statistically significant and close to zero. Hence, we do not find statistical evidence for a direct effect of the mere assignment into the training program on depression outcomes or a violation of the exclusion restriction when using assignment as instrument for participation. In contrast, we find for both CiC and DiD statistically significantly negative indirect effects ($\hat{\delta}_1^c(1)$) among compliers on depression symptoms 6 month after the intervention, that are comparable to the total complier effect $\hat{\Delta}_c$. Both $\hat{\delta}_1^c(1)$ and $\hat{\Delta}_c$ are in fact similar to a two stage least squares regression relying on the exclusion restriction, which yields a local average treatment effect on compliers of -0.25 with a heteroskedasticity-robust standard error of 0.08.

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 identfying 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 program.

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," *forthcoming in the Journal of the Royal Statistical Society: Series A (Statistics in Society).*

Bijwaard, G. E., and A. M. Jones (2018): "An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality," *Empirical Economics*, pp. 1–47.

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*.

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 DP 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 Caseworkers 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 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.

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.

Rubin, D. B. (1974): "Estimating the Causal Effect of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66(5), 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): "Why does the Job Corps increase gender earnings inequality? A translated quantile approach," *Working Paper*.

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 Y. Schul (1995): "Impact of the JOBS Intervention on Unemployed Workers Varying in Risk for Depression," *American Journal of Community Psychology*, 23, 39–74.

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

Appendices

A Proofs

A.1 Proof of Theorem 1

A.1.1 Average direct effect under $d = 1$ conditional on $D_i = 1$ and $M_i(1) = 0$

In the following, we proof that $\theta_1^{1,0}(1) = E[Y_{i1}(1,0) - Y_{i1}(0,0)|D_i = 1, M_i(1) = 0] = E[Y_{i1} - Q_{00}(Y_{i0})|D_i = 1, M_i = 0]$. Using the observational rule, we obtain $E[Y_{i1}(1,0)|D_i = 1, M_i(1) = 0] = E[Y_{i1}|D_i = 1, M_i = 0]$. Accordingly, we have to show that $E[Y_{i1}(0,0)|D_i = 1, M_i(1) = 0] = E[Q_{00}(Y_{i0})|D_i = 1, M_i = 0]$ 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

$$\begin{aligned}
F_{Y_t(d,0)|D=1,M=0}(y) &\overset{A1}{=} Pr(h(d,m,t,U) \leq y|D_i = 1, M_i = 0, T = t), \\
&= Pr(U \leq h^{-1}(d,m,t;y)|D_i = 1, M_i = 0, T = t), \\
&\overset{A3a}{=} Pr(U \leq h^{-1}(d,m,t;y)|D_i = 1, M_i = 0), \\
&= F_{U|10}(h^{-1}(d,m,t;y)),
\end{aligned} \tag{9}$$

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

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

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

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

Combining (10) and (11) yields,

$$h(0,0,1,h^{-1}(0,0,0;y)) = F^{-1}_{Y_1(0,0)|D=1,M=0} \circ F_{Y_0(0,0)|D=1,M=0}(y). \tag{12}$$

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, $E[F^{-1}_{Y_1(0,0)|D=1,M=0} \circ F_{Y_0(0,0)|D=1,M=0}(Y_{i0})|D_i = 1, M_i = 0] = E[Y_{i1}(0,0)|D_i = 1, M_i = 0]$. We could 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 $F^{-1}_{Y_1(0,0)|D=1,M=0} \circ F_{Y_0(0,0)|D=1,M=0}(y)$ under the additional Assumption 3b.

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

$$\begin{aligned} F_{Y_t(d,0)|D=0,M=0}(y) &\overset{A1}{=} Pr(h(d,m,t,U) \leq y|D_i = 0, M_i = 0, T = t), \\ &= Pr(U \leq h^{-1}(d,m,t;y)|D_i = 0, M_i = 0, T = t), \\ &\overset{A2b}{=} Pr(U \leq h^{-1}(d,m,t;y)|D_i = 0, M_i = 0), \\ &= F_{U|00}(h^{-1}(d,m,t;y)), \end{aligned} \tag{13}$$

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

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

$$h(0,0,1,u) = F^{-1}_{Y_1(0,0)|D=0,M=0}(F_{U|00}(u)). \tag{14}$$

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

$$F^{-1}_{U|00}(F_{Y_0(0,0)|D=0,M=0}(y)) = h^{-1}(0,0,0;y). \tag{15}$$

Combining (14) and (15) yields,

$$h(0,0,1,h^{-1}(0,0,0;y)) = F^{-1}_{Y_1(0,0)|D=0,M=0} \circ F_{Y_0(0,0)|D=0,M=0}(y). \tag{16}$$

The left sides of (12) and (16) are equal.

In contrast to (12), (16) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(0,0)|D=0,M=0}(y) = Pr(Y_t(0,0) \leq y|D_i = 0, M_i = 0) = Pr(Y_t \leq y|D_i = 0, M_i = 0)$. Accordingly, we can identify $F^{-1}_{Y_1(0,0)|D=1,M=0} \circ F_{Y_0(0,0)|D=1,M=0}(y)$ by $Q_{00}(y) \equiv F^{-1}_{Y_1|D=0,M=0} \circ F_{Y_0|D=0,M=0}(y)$.

Parsing Y_{i0} through $Q_{00}(\cdot)$ in the treated group without mediator gives

$$\begin{aligned}
E[Q_{00}(Y_{i0})|D_i = 1, M_i = 0] &= E[F^{-1}_{Y_1|D=0,M=0} \circ F_{Y_0|D=0,M=0}(Y_{i0})|D_i = 1, M_i = 0] \\
&= E[F^{-1}_{Y_1(0,0)|D=0,M=0} \circ F_{Y_0(0,0)|D=0,M=0}(Y_{i0}(1,0))|D_i = 1, M_i = 0] \\
&\overset{A1,A3b}{=} E[h(0,0,1,h^{-1}(0,0,0;Y_{i0}(1,0)))|D_i = 1, M_i = 0] \\
&\overset{A2}{=} E[h(0,0,1,h^{-1}(0,0,0;Y_{i0}(0,0)))|D_i = 1, M_i = 0] \\
&\overset{A1,A3a}{=} E[F^{-1}_{Y_1(0,0)|D=1,M=0} \circ F_{Y_0(0,0)|D=1,M=0}(Y_{i0}(0,0))|D_i = 1, M_i = 0] \\
&= E[Y_{i1}(0,0)|D_i = 1, M_i = 0] = E[Y_{i1}(0,0)|D_i = 1, M_i(1) = 0],
\end{aligned} \tag{17}$$

which has data support because of Assumption 4a.

A.1.2 Quantile direct effect under $d = 1$ conditional on $D_i = 1$ and $M_i(1) = 0$

In the following, we proof that $\theta_1^{1,0}(q,1) = F^{-1}_{Y_1(1,0)|D=1,M(1)=0}(q) - F^{-1}_{Y_1(0,0)|D=1,M(1)=0}(q) = F^{-1}_{Y_1|D=1,M=0}(q) - F^{-1}_{Q_{00}(Y_0)|D=1,M=0}(q)$. For this purpose, we have to proof that

$$F_{Y_1(1,0)|D=1,M(1)=0}(y) = F_{Y_1|D=1,M=0}(y) \text{ and} \tag{18}$$

$$F_{Y_1(0,0)|D=1,M(1)=0}(y) = F_{Q_{00}(Y_0)|D=1,M=0}(y), \tag{19}$$

which is sufficient to show that the quantiles are also identified. We can show (18) using the observational rule $F_{Y_1(1,0)|D=1,M(1)=0}(y) = F_{Y_1|D=1,M=0}(y) = E[1\{Y_{i1} \leq y\}|D_i = 1, M_i = 0]$, with $1\{\cdot\}$ being an indicator function.

Using (17), we obtain

$$\begin{aligned}
F_{Q_{00}(Y_0)|D_i=1,M_i=0}(y) &= E[1\{Q_{00}(Y_{i0}) \leq y\}|D_i = 1, M_i = 0] \\
&= E[1\{F^{-1}_{Y_1|D=0,M=0} \circ F_{Y_0|D=0,M=0}(Y_{i0}) \leq y\}|D_i = 1, M_i = 0] \\
&= E[1\{Y_{i1}(0,0) \leq y\}|D_i = 1, M_i = 0] = F_{Y_1(0,0)|D=1,M(1)=0}(y),
\end{aligned} \tag{20}$$

which proofs (19).

A.1.3 Average direct effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 0$

In the following, we proof that $\theta_1^{0,0}(0) = E[Y_{i1}(1,0) - Y_{i1}(0,0)|D_i = 0, M_i(0) = 0] = E[Q_{10}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 0]$. Using the observational rule, we obtain $E[Y_{i1}(0,0)|D_i = 0, M_i(0) = 0] = E[Y_{i1}|D_i = 0, M_i = 0]$. Accordingly, we have to show that $E[Y_{i1}(1,0)|D_i = 0, M_i(0) = 0] = E[Q_{10}(Y_{i0})|D_i = 0, M_i = 0]$ to finish the proof.

First, we use (13) 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 $F^{-1}_{Y_1(1,0)|D=0,M=0}(q)$ to both sides, we have

$$h(1,0,1,u) = F^{-1}_{Y_1(1,0)|D=0,M=0}(F_{U|10}(u)). \tag{21}$$

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

$$F^{-1}_{U|10}(F_{Y_0(1,0)|D=0,M=0}(y)) = h^{-1}(1,0,0;y). \tag{22}$$

Combining (21) and (22) yields,

$$h(1,0,1,h^{-1}(1,0,0;y)) = F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(y). \tag{23}$$

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, $E[F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(Y_{i0})|D_i = 0, M_i = 0] = E[Y_{i1}(1,0)|D_i = 1, M_i = 0]$. We could 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 $F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(y)$ under the additional Assumption 3a.

Second, we use (9) 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

$$h(1,0,1,u) = F^{-1}_{Y_1(1,0)|D=1,M=0}(F_{U|10}(u)). \tag{24}$$

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

$$F^{-1}_{U|10}(F_{Y_0(1,0)|D=1,M=0}(y)) = h^{-1}(1,0,0;y). \tag{25}$$

Combining (24) and (25) yields,

$$h(1,0,1,h^{-1}(1,0,0;y)) = F^{-1}_{Y_1(1,0)|D=1,M=0} \circ F_{Y_0(1,0)|D=1,M=0}(y). \tag{26}$$

In contrast to (23), (26) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(1,0)|D=1,M=0}(y) = Pr(Y_t(1,0) \leq y|D_i = 1, M_i = 0) = Pr(Y_t \leq y|D_i = 1,_i M = 0)$. 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_{i0} through $Q_{10}(\cdot)$ in the non-treated group without mediator gives

$$\begin{aligned} E[Q_{10}(Y_{i0})|D=0,M=0] &= E[F^{-1}_{Y_1|D=1,M=0} \circ F_{Y_0|D=1,M=0}(Y_{i0})|D_i=0,M_i=0] \\ &= E[F^{-1}_{Y_1(1,0)|D=1,M=0} \circ F_{Y_0(1,0)|D=1,M=0}(Y_{i0}(0,0))|D_i=0,M_i=0] \\ &\overset{A1,A3a}{=} E[h(1,0,1,h^{-1}(1,0,0;Y_{i0}(0,0)))|D_i=0,M_i=0] \\ &\overset{A2}{=} E[h(1,0,1,h^{-1}(1,0,0;Y_{i0}(1,0)))|D_i=1,M_i=0] \\ &\overset{A1,A3b}{=} E[F^{-1}_{Y_1(1,0)|D=0,M=0} \circ F_{Y_0(1,0)|D=0,M=0}(Y_{i0}(1,0))|D_i=0,M_i=0] \\ &= E[Y_{i1}(1,0)|D_i=0,M_i=0] = E[Y_{i1}(1,0)|D_i=0,M_i(0)=0]. \end{aligned} \tag{27}$$

which has data support because of Assumption 4b.

A.1.4 Quantile direct effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 0$

In the following, we proof 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 proof that

$$F_{Y_1(1,0)|D=0,M(0)=0}(y) = F_{Q_{10}(Y_0)|D=0,M=0}(y) \text{ and} \tag{28}$$

$$F_{Y_1(0,0)|D=0,M(0)=0}(y) = F_{Y_1|D=0,M=0}(y), \tag{29}$$

which is sufficient to show that the quantiles are also identified. We can show (29) using the observational rule $F_{Y_1(0,0)|D=0,M(0)=0}(y) = F_{Y_1|D=0,M=0}(y) = E[1\{Y_{i1} \leq y\}|D_i = 0, M_i = 0]$.

Using (27), we obtain

$$
\begin{aligned}
F_{Q_{10}(Y_0)|D=0,M=0}(y) &= E[1\{Q_{10}(Y_{i0}) \leq y\}|D_i = 0, M_i = 0] \\
&= E[1\{F^{-1}_{Y_1|D=1,M=0} \circ F_{Y_0|D=1,M=0}(Y_{i0}) \leq y\}|D_i = 0, M_i = 0] \\
&= E[1\{Y_{i1}(1,0) \leq y\}|D_i = 0, M_i = 0] = F_{Y_1(1,0)|D=0,M(0)=0}(y),
\end{aligned}
$$

which proofs (28).

A.2 Proof of Theorem 2

A.2.1 Average direct effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 1$

In the following, we proof that $\theta_1^{0,1}(0) = E[Y_{i1}(1,1) - Y_{i1}(0,1)|D = 0, M(0) = 1] = E[Q_{11}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 1]$. Using the observational rule, we obtain $E[Y_{i1}(0,1)|D_i = 0, M_i(0) = 1] = E[Y_{i1}|D_i = 0, M_i = 1]$. Accordingly, we have to show that $E[Y_{i1}(1,1)|D_i = 0, M_i(0) = 1] = E[Q_{11}(Y_{i0})|D_i = 0, M_i = 1]$ to finish the proof.

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

$$
\begin{aligned}
F_{Y_t(d,0)|D=1,M=0}(y) &\overset{A1}{=} Pr(h(d,m,t,U) \leq y|D_i = 0, M_i = 1, T = t), \\
&= Pr(U \leq h^{-1}(d,m,t;y)|D_i = 0, M_i = 1, T = t), \\
&\overset{A5a}{=} Pr(U \leq h^{-1}(d,m,t;y)|D_i = 0, M_i = 1), \\
&= F_{U|01}(h^{-1}(d,m,t;y)),
\end{aligned}
\tag{30}
$$

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

$$h(1,1,1,u) = F^{-1}_{Y_1(1,1)|D=0,M=1}(F_{U|01}(u)). \tag{31}$$

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

$$F^{-1}_{U|01}(F_{Y_0(1,1)|D=0,M=1}(y)) = h^{-1}(1,1,0;y). \tag{32}$$

Combining (31) and (32) yields,

$$h(1,1,1,h^{-1}(1,1,0;y)) = F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(y). \tag{33}$$

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, $E[F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_{i0})|D_i = 0, M_i = 1] = E[Y_{i1}(1,1)|D_i = 0, M_i = 1]$. We could 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

$$\begin{aligned} F_{Y_t(d,1)|D=1,M=1}(y) &\overset{A1}{=} Pr(h(d,m,t,U) \leq y|D_i = 1, M_i = 1, T = t), \\ &= Pr(U \leq h^{-1}(d,m,t;y)|D_i = 1, M_i = 1, T = t), \\ &\overset{A5b}{=} Pr(U \leq h^{-1}(d,m,t;y)|D_i = 1, M_i = 1), \\ &= F_{U|11}(h^{-1}(d,m,t;y)), \end{aligned} \tag{34}$$

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

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

$$h(1,1,1,u) = F^{-1}_{Y_1(1,1)|D=1,M=1}(F_{U|11}(u)). \tag{35}$$

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

$$F^{-1}_{U|11}(F_{Y_0(1,1)|D=1,M=1}(y)) = h^{-1}(1,1,1;y). \tag{36}$$

Combining (35) and (36) yields,

$$h(1,1,1,h^{-1}(1,1,0;y)) = F^{-1}_{Y_1(1,1)|D=1,M=1} \circ F_{Y_0(1,1)|D=1,M=1}(y). \tag{37}$$

The left sides of (33) and (37) are equal.

In contrast to (33), (37) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(1,1)|D=1,M=1}(y) = Pr(Y_t(1,1) \leq y|D_i = 1, M_i = 1) = Pr(Y_t \leq y|D_i = 1, M_i = 1)$. Accordingly, we can identify $F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(y)$ by $Q_{11}(y) \equiv F^{-1}_{Y_1|D=1,M=1} \circ F_{Y_0|D=1,M=1}(y)$.

Parsing Y_{i0} through $Q_{11}(\cdot)$ in the non-treated group with mediator gives

$$\begin{aligned}
E[Q_{11}(Y_{i0})|D=0,M=1] &= E[F^{-1}_{Y_1|D=1,M=1} \circ F_{Y_0|D=1,M=1}(Y_{i0})|D_i=0,M_i=1] \\
&= E[F^{-1}_{Y_1(1,1)|D=1,M=1} \circ F_{Y_0(1,1)|D=1,M=1}(Y_{i0}(0,1))|D_i=0,M_i=1] \\
&\overset{A1,A4b}{=} E[h(1,1,1,h^{-1}(1,1,0;Y_{i0}(0,1)))|D_i=0,M_i=1] \\
&\overset{A3}{=} E[h(1,1,1,h^{-1}(1,1,0;Y_{i0}(0,0)))|D_i=0,M_i=1] \\
&\overset{A1,A4a}{=} E[F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_{i0}(0,0))|D_i=0,M_i=1] \\
&= E[Y_{i1}(1,1)|D=0,M=1] = E[Y_{i1}(1,1)|D_i=0,M_i(0)=1],
\end{aligned} \tag{38}$$

which has data support because of Assumption 6a.

A.2.2 Quantile direct effect under $d = 0$ conditional on $D_i = 0$ and $M_i(0) = 1$

In the following, we proof that $\theta_1^{0,1}(q,0) = F^{-1}_{Y_1(1,1)|D=0,M(0)=1}(q) - F^{-1}_{Y_1(0,1)|D=0,M(0)=1}(q) = F^{-1}_{Q_{11}(Y_0)|D=0,M=1}(q) - F^{-1}_{Y_1|D=0,M=1}(q)$. For this purpose, we have to proof that

$$F_{Y_1(1,1)|D=0,M(0)=1}(y) = F_{Q_{11}(Y_0)|D=0,M=1}(y) \text{ and} \tag{39}$$

$$F_{Y_1(0,1)|D=0,M(0)=1}(y) = F_{Y_1|D=0,M=1}(y), \tag{40}$$

which is sufficient to show that the quantiles are also identified. We can show (40) using the observational rule $F_{Y_1(0,1)|D=0,M(0)=1}(y) = F_{Y_1|D=0,M=1}(y) = E[1\{Y_{i1} \leq y\}|D_i = 0, M_i = 1]$.

Using (38), we obtain

$$\begin{aligned} F_{Q_{11}(Y_0)|D=0,M=1}(y) &= E[1\{Q_{11}(Y_{i0}) \leq y\}|D_i = 0, M_i = 1] \\ &= E[1\{F^{-1}_{Y_1|D=1,M=1} \circ F_{Y_0|D=1,M=1}(Y_{i0}) \leq y\}|D_i = 0, M_i = 1] \\ &= E[1\{Y_{i1}(1,1) \leq y\}|D_i = 0, M_i = 0] = F_{Y_1(1,1)|D=0,M(0)=1}(y), \end{aligned} \tag{41}$$

which proofs (39).

A.2.3 Average direct effect under $d = 1$ conditional on $D_i = 1$ and $M_i(1) = 1$

In the following, we proof that $\theta_1^{1,1}(1) = E[Y_{i1}(1,1) - Y_{i1}(0,1)|D_i = 1, M_i(1) = 1] = E[Y_{i1} - Q_{01}(Y_{i0})|D_i = 1, M_i = 1]$. Using the observational rule, we obtain $E[Y_{i1}(1,1)|D_i = 1, M_i(1) = 1] = E[Y_{i1}|D_i = 1, M_i = 1]$. Accordingly, we have to show that $E[Y_{i1}(0,1)|D_i = 1, M_i(1) = 1] = E[Q_{01}(Y_{i0})|D_i = 1, M_i = 1]$ to finish the proof.

Using (34), we 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

$$h(0,1,1,u) = F^{-1}_{Y_1(0,1)|D=1,M=1}(F_{U|11}(u)). \tag{42}$$

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

$$F^{-1}_{U|11}(F_{Y_0(0,1)|D=1,M=1}(y)) = h^{-1}(0,1,0;y). \tag{43}$$

Combining (42) and (43) yields,

$$h(0,1,1,h^{-1}(0,1,0;y)) = F^{-1}_{Y_1(0,1)|D=1,M=1} \circ F_{Y_0(0,1)|D=1,M=1}(y). \tag{44}$$

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, $E[F^{-1}_{Y_1(1,1)|D=0,M=1} \circ F_{Y_0(1,1)|D=0,M=1}(Y_{i0})|D_i = 0, M_i = 1] = E[Y_{i1}(1,1)|D_i = 0, M_i = 1]$. We could 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.

Using (30), we 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

$$h(0,1,1,u) = F^{-1}_{Y_1(0,1)|D=0,M=1}(F_{U|01}(u)). \tag{45}$$

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

$$F^{-1}_{U|01}(F_{Y_0(0,1)|D=0,M=1}(y)) = h^{-1}(0,1,1;y). \tag{46}$$

Combining (45) and (46) yields,

$$h(0,1,1,h^{-1}(0,1,0;y)) = F^{-1}_{Y_1(0,1)|D=0,M=1} \circ F_{Y_0(0,1)|D=0,M=1}(y). \tag{47}$$

The left sides of (44) and (47) are equal.

In contrast to (44), (47) contains only distributions that can be identified from observable data. In particular, $F_{Y_t(0,1)|D=0,M=1}(y) = Pr(Y_t(0,1) \leq y|D_i = 0, M_i = 1) = Pr(Y_t \leq y|D_i = 0, M_i = 1)$. 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_{i0} through $Q_{01}(\cdot_i)$ in the treated group with mediator gives

$$\begin{aligned}
E[Q_{01}(Y_{i0})|D=1,M=1] &= E[F^{-1}_{Y_1|D=0,M=1} \circ F_{Y_0|D=0,M=1}(Y_{i0})|D_i=1,M_i=1] \\
&= E[F^{-1}_{Y_1(0,1)|D=0,M=1} \circ F_{Y_0(0,1)|D=0,M=1}(Y_{i0}(1,1))|D_i=1,M_i=1] \\
&\overset{A1,A4b}{=} E[h(0,1,1,h^{-1}(0,1,0;Y_{i0}(1,1)))|D_i=1,M_i=1] \\
&\overset{A3}{=} E[h(0,1,1,h^{-1}(0,1,0;Y_{i0}(0,1)))|D_i=1,M_i=1] \\
&\overset{A1,A4a}{=} E[F^{-1}_{Y_1(0,1)|D=1,M=1} \circ F_{Y_0(0,1)|D=1,M=1}(Y_{i0}(0,1))|D_i=1,M_i=1] \\
&= E[Y_{i1}(0,1)|D_i=1,M_i=1] = E[Y_{i1}(0,1)|D_i=0,M_i(0)=1],
\end{aligned} \tag{48}$$

which has data support under Assumption 6b.

A.2.4 Quantile direct effect under $d=1$ conditional on $D_i=1$ and $M_i(1)=1$

In the following, we proof that $\theta_1^{1,1}(q,1) = F^{-1}_{Y_1(1,1)|D=1,M(1)=1}(q) - F^{-1}_{Y_1(0,1)|D=1,M(1)=1}(q) = F^{-1}_{Y_1|D=1,M=1}(q) - F^{-1}_{Q_{01}(Y_0)|D=1,M=1}(q)$. For this purpose, we have to proof that

$$F_{Y_1(1,1)|D=1,M(1)=1}(y) = F_{Y_1|D=1,M=1}(y) \text{ and} \tag{49}$$

$$F_{Y_1(0,1)|D=1,M(1)=1}(y) = F_{Q_{01}(Y_0)|D=1,M=1}(y), \tag{50}$$

which is sufficient to show that the quantiles are also identified. We can show (49) using the observational rule $F_{Y_1(1,1)|D=1,M(1)=1}(y) = F_{Y_1|D=1,M=1}(y) = E[1\{Y_{i1} \leq$

$y\}|D_i = 1, M_i = 1]$.

Using (48), we obtain

$$
\begin{aligned}
F_{Q_{01}(Y_0)|D=1,M=1}(y) &= E[1\{Q_{01}(Y_{i0}) \leq y\}|D_i = 1, M_i = 1] \\
&= E[1\{F^{-1}_{Y_1|D=0,M=1} \circ F_{Y_0|D=0,M=1}(Y_{i0}) \leq y\}|D_i = 1, M_i = 1] \\
&= E[1\{Y_{i1}(0,1) \leq y\}|D_i = 1, M_i = 0] = F_{Y_1(0,1)|D=1,M(1)=1}(y),
\end{aligned}
$$

which proofs (50).

A.3 Proof of equations (1) and (2)

We denote by $p_\tau = Pr(\tau_i)$ the share of a particular type in the population and by $p_{m|d} = Pr(M_i = m|D_i = d)$ the conditional probability of a particular mediator state given the treatment, with $d, m \in \{1, 0\}$. By Assumption 7, the share of a type τ_i conditional on D_i corresponds to p_τ (in the population), as D_i is randomly assigned. Likewise, $E[Y_{it}(d,m)|\tau_i, D_i = 1] = E[Y_{it}(d,m)|\tau_i, D_i = 0] = E[Y_{it}(d,m)|\tau_i]$ due to the independence of D_i and the potential outcomes as well as the types τ_i (which are a deterministic function of $M_i(d)$). It follows that conditioning on D_i is not required on the right hand side of the following equation, which expresses the mean outcome conditional $D_i = 0$ and $M_i = 0$ as weighted average of the mean potential outcomes of compliers and never-takers:

$$
E[Y_{it}|D_i = 0, M_i = 0] = \frac{p_n}{p_n + p_c} E[Y_{it}(0,0)|\tau_i = n] + \frac{p_c}{p_n + p_c} E[Y_{it}(0,0)|\tau_i = c]. \quad (51)
$$

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

$$
E[Y_{it}(0,0)|\tau_i = n] - E[Y_{it}(0,0)|\tau_i = c] = \frac{p_n + p_c}{p_c} \{E[Y_{it}(0,0)|\tau_i = n] - E[Y_{it}|D_i = 0, M_i = 0]\}. \quad (52)
$$

Next, we consider observations with $D_i = 1$ and $M_i = 0$, which might consist of both never-takers and defiers, as $M_i(1) = 0$ 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:

$$E[Y_{it}|D_i = 1, M_i = 0] \overset{A7,A8}{=} E[Y_{it}(1,0)|\tau_i = n]. \tag{53}$$

Furthermore, by Assumption 2,

$$E[Y_{i0}(0,0)|\tau_i = n] \overset{A2}{=} E[Y_{i0}(1,0)|\tau_i = n] \overset{A7,A8}{=} E[Y_{i0}|D_i = 1, M_i = 0].$$

It follows that when considering (52) in period $T = 0$, $E[Y_0(0,0)|n]$ on the right hand side of the equation may be replaced by $E[Y_0|D = 1, M = 0]$:

$$E[Y_{i0}(0,0)|\tau_i = n] - E[Y_{i0}(0,0)|\tau_i c] = \frac{p_n + p_c}{p_c} \{E[Y_{i0}|D_1 = 1, M_1 = 0] - E[Y_{i0}|D_i = 0, M_i = 0]\}.$$

This finishes the proof of equation (1).

Similarly to (51) for the never-takers and compliers, consider the mean outcome given $Z_i = 1$ and $D_i = 1$, which is made up by always-takers and compliers (the types with $M_i(1) = 1$)

$$E[Y_{it}|D_i = 1, M_i = 1] = \frac{p_a}{p_a + p_c} E[Y_{it}(1,1)|\tau_i = a] + \frac{p_c}{p_a + p_c} E[Y_{it}(1,1)|\tau_i = c]. \tag{54}$$

After some rearrangements we obtain

$$E[Y_{it}(1,1)|\tau_i = a] - E[Y_{it}(1,1)|\tau_i = c] = \frac{p_a + p_c}{p_c} \{E[Y_{it}(1,1)|\tau_i = a] - E[Y_{it}|D_i = 1, M_i = 1]\}. \tag{55}$$

By Assumptions 7 and 8,

$$E[Y_{it}|D_i = 0, M_i = 1] = E[Y_{it}(0,1)|\tau_i = a]. \tag{56}$$

Now consider (55) for period $T = 0$, and note that by Assumption 2, $E[Y_{i0}(1,1)|\tau_i = a] = E[Y_{i0}(0,0)|\tau_i = a] = E[Y_{i0}(0,1)|\tau_i = a]$ and $E[Y_{i0}(1,1)|\tau_i = c] = E[Y_{i0}(0,0)|\tau_i =$

$c]$, we obtain

$$E[Y_{i0}(0,0)|\tau_i = a] - E[Y_{i0}(0,0)|\tau_i = c] = E[Y_{i0}(0,1)|\tau_i = a] - E[Y_{i0}(1,1)|\tau_i = c],$$
$$= \frac{p_a + p_c}{p_c} \left\{ E[Y_{i0}|D_i = 0, M_i = 1] - E[Y_{i0}|D_i = 1, M_i = 1] \right\}.$$

This finishes the proof of equation (2).

A.4 Proof of Theorem 3

A.4.1 Average direct effect on the never-takers

In the following, we proof that $\theta_1^n = E[Y_{i1}(1,0) - Y_{i1}(0,0)|\tau_i = n] = E[Y_{i1} - Q_{00}(Y_{i0})|D_i = 1, M_i = 0]$. From (53), we obtain the first ingredient $E[Y_{i1}(1,0)|\tau_i = n] = E[Y_{i1}|D_i = 1, M_i = 0]$. Furthermore, from (17) we have $E[Q_{00}(Y_{i0})|D_i = 1, M_i = 0] = E[Y_{i1}(0,0)|D_i = 1, M_i(1) = 0]$. Under Assumption 7 and 8,

$$\begin{aligned} E[Y_{i1}(0,0)|D_i = 1, M_i(1) = 0] &\overset{A7}{=} E[Y_{i1}(0,0)|D_i = 1, \tau_i = n] \\ &\overset{A8}{=} E[Y_{i1}(0,0)|\tau_i = n]. \end{aligned} \tag{57}$$

A.4.2 Quantile direct effect on the never-takers

We proof that $\theta_1^n(q) = F^{-1}_{Y_1(1,0)|\tau=n}(q) - F^{-1}_{Y_1(0,0)|\tau=n}(q) = F^{-1}_{Y_1|D=1,M=0}(q) - F^{-1}_{Q_{00}(Y_0)|D=1,M=0}(q)$. This requires showing that

$$F_{Y_1(1,0)|\tau=n}(y) = F_{Y_1|D=1,M=0}(y) \text{ and} \tag{58}$$

$$F_{Y_1(0,0)|\tau=n}(y) = F_{Q_{00}(Y_0)|D=1,M=0}(y). \tag{59}$$

Under Assumptions 7 and 8,

$$\begin{aligned} F_{Y_t|D=1,M=0}(y) &= E[1\{Y_{it} \leq y\}|D_i = 1, M_i = 0] \\ &\overset{A7,A8}{=} E[1\{Y_{it}(1,0) \leq y\}|\tau_i = n] \\ &= F_{Y_t(1,0)|\tau=n}(y), \end{aligned} \tag{60}$$

which proofs (58). From (20), we have $F_{Q_{00}(Y_0)|D=1,M=0}(y) = F_{Y_1(0,0)|D=1,M(1)=0}(y) = E[1\{Y_{i1}(0,0) \leq y\}|D_i = 1, M_i(1) = 0]$. Under Assumption 7 and 8,

$$\begin{aligned} E[1\{Y_{i1}(0,0) \leq y\}|D_i = 1, M_i(1) = 0] &\overset{A7,A8}{=} E[1\{Y_{i1}(0,0) \leq y\}|\tau_i = n] \\ &= F_{Y_1(0,0)|\tau=n}(y), \end{aligned} \tag{61}$$

which proofs (59).

A.4.3 Average direct effect under $d = 0$ on compliers

In the following, we proof that

$$\begin{aligned} \theta_1^c(0) =& E[Y_{i1}(1,0) - Y_{i1}(0,0)|\tau_i = c], \\ =& \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{10}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 0] - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_{i1} - Q_{00}(Y_{i0})|D_i = 1, M_i = 0]. \end{aligned}$$

Plugging (57) in (51) under $T = 1$, we obtain

$$E[Y_{i1}|D_i = 0, M_i = 0] = \frac{p_n}{p_n + p_c} E[Q_{00}(Y_{i0})|D_i = 1, M_i = 0] + \frac{p_c}{p_n + p_c} E[Y_{i1}(0,0)|\tau_i = c].$$

This allows identifying

$$E[Y_{i1}(0,0)|\tau_i = c] = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Y_{i1}|D_i = 0, M_i = 0] - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Q_{00}(Y_{i0})|D_i = 1, M_i = 0]. \tag{62}$$

Accordingly, we have to show the identification of $E[Y_1(1,0)|c]$ to finish the proof. From (27) we have $E[Y_{i1}(1,0)|D_i = 0, M_i = 0] = E[Q_{10}(Y_{i0})|D_i = 0, M_i = 0]$. Applying the law of iterative expectations, gives

$$\begin{aligned} E[Y_{i1}(1,0)|D_i = 0, M_i = 0] =& \frac{p_n}{p_n + p_c} E[Y_{i1}(1,0)|D_i = 0, M_i = 0, \tau_i = n] \\ &+ \frac{p_c}{p_n + p_c} E[Y_{i1}(1,0)|D_i = 0, M_i = 0, \tau_i = c], \\ \overset{A7}{=}& \frac{p_n}{p_n + p_c} E[Y_{i1}(1,0)|\tau_i = n] + \frac{p_c}{p_n + p_c} E[Y_{i1}(1,0)|\tau_i = c]. \end{aligned}$$

After some rearrangements and using (53), we obtain

$$E[Y_{i1}(1,0)|\tau_i = c] = \frac{p_n + p_c}{p_c} E[Q_{10}(Y_{i0})|D_i = 0, M_i = 0] - \frac{p_n}{p_c} E[Y_{i1}|D_i = 1, M_i = 0].$$

This gives

$$E[Y_{i1}(1,0)|\tau_i = c] = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{10}(Y_{i0})|D_i = 0, M_i = 0] - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_{i1}|D_i = 1, M_i = 0] \tag{63}$$

using $p_n = Pr(M_i = 0|D_i = 1) = p_{0|1}$, and $p_c + p_n = Pr(M_i = 0|D_i = 0) = p_{0|0}$.

A.4.4 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 proofs that $\theta_1^c(q,0) = F^{-1}_{Y_1(1,0)|c}(q) - F^{-1}_{Y_1(0,0)|c}(q)$ is identified.

From (28), 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

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

Using (58) and rearranging the equation gives,

$$F_{Y_1(1,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Q_{10}(Y_0)|D=0,M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Y_1|D=1,M=0}(y). \tag{64}$$

In analogy to (51), the outcome distribution under $D_i = 0$ and $M_i = 0$ 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(59) and rearranging the equation gives

$$F_{Y_1(0,0)|\tau=c}(y) = \frac{p_{0|0}}{p_{0|0} - p_{0|1}} F_{Y_1|D=0,M=0}(y) - \frac{p_{0|1}}{p_{0|0} - p_{0|1}} F_{Q_{00}(Y_0)|D=1,M=0}(y). \tag{65}$$

A.5 Proof of Theorem 4

A.5.1 Average direct effect on the always-takers

In the following, we proof that $\theta_1^a = E[Y_{i1}(1,1) - Y_{i1}(0,1)|\tau_i = a] = E[Q_{11}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 1]$. From (56), we obtain the first ingredient $E[Y_1(0,1)|a] = E[Y_1|D = 0, M = 1]$. Furthermore, from (38) we have $E[Q_{11}(Y_{i0})|D_i = 0, M_i = 1] = E[Y_{i1}(1,1)|D_i = 0, M_i(0) = 1]$. Under Assumption 7 and 8,

$$\begin{aligned} E[Y_{i1}(1,1)|D_i = 0, M_i(0) = 1] &\overset{A7}{=} E[Y_{i1}(1,1)|D_i = 0, \tau_i = a] \\ &\overset{A8}{=} E[Y_{i1}(1,1)|\tau_i = a]. \end{aligned} \tag{66}$$

A.5.2 Quantile direct effect on the always-takers

We proof that $\theta_1^a(q) = F^{-1}_{Y_{i1}(1,1)|\tau=a}(q) - F^{-1}_{Y_{i1}(0,1)|\tau=a}(q) = F^{-1}_{Q_{11}(Y_0)|D=0,M=1}(q) - F^{-1}_{Y_1|D=0,M=1}(q)$. This requires showing that

$$F_{Y_{i1}(1,1)|\tau=a}(y) = F_{Q_{11}(Y_0)|D=0,M=1}(y) \text{ and} \tag{67}$$

$$F_{Y_{i1}(0,1)|\tau=a}(y) = F_{Y_1|D=0,M=1}(y). \tag{68}$$

Under Assumptions 7 and 8,

$$\begin{aligned} F_{Y_t|D=0,M=1}(y) &= E[1\{Y_{it} \leq y\}|D_i = 0, M_i = 1] \\ &\overset{A7,A8}{=} E[1\{Y_{it}(0,1) \leq y\}|\tau_i = a] \\ &= F_{Y_t(0,1)|\tau=a},(y). \end{aligned} \tag{69}$$

which proofs (68). From (41), we have $F_{Q_{11}(Y_0)|D=0,M=1}(y) = F_{Y_1(1,1)|D=0,M(0)=1}(y) = E[1\{Y_{i1}(1,1) \leq y\}|D_i = 0, M_i(0) = 1]$. Under Assumption 7 and 8,

$$
\begin{aligned}
E[1\{Y_{i1}(1,1) \leq y\}|D_i = 0, M_i(0) = 1] &\overset{A7,A8}{=} E[1\{Y_{i1}(1,1) \leq y\}|\tau_i = a] \\
&= F_{Y_{i1}(1,1)|\tau=a}(y),
\end{aligned} \tag{70}
$$

which proofs (67).

A.5.3 Average direct effect under $d = 1$ on compliers

In the following, we proof that

$$
\begin{aligned}
\theta_1^c(1) =& E[Y_{i1}(1,1) - Y_{i1}(0,1)|\tau_i = c], \\
=& \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_{i1} - Q_{01}(Y_{i0})|D_i = 1, M_i = 1] \\
& - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_{i0}) - Y_{i1}|D_i = 0, M_i = 1].
\end{aligned}
$$

Plugging (66) in (54), we obtain

$$
E[Y_{i1}|D_i = 1, M_i = 1] = \frac{p_a}{p_a + p_c} E[Q_{11}(Y_{i0})|D_i = 0, M_i = 1] + \frac{p_c}{p_a + p_c} E[Y_{i1}(1,1)|\tau_i = c].
$$

This allows identifying

$$
E[Y_{i1}(1,1)|\tau_i = c] = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_{i1}|D_i = 1, M_i = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_{i0})|D_i = 0, M_i = 1]. \tag{71}
$$

From (48) we have $E[Y_{i1}(0,1)|D_i = 1, M_i = 1] = E[Q_{01}(Y_{i0})|D_i = 1, M_i = 1]$. Applying the law of iterative expectations, gives

$$
\begin{aligned}
E[Y_{i1}(0,1)|D_i = 1, M_i = 1] =& \frac{p_a}{p_a + p_c} E[Y_{i1}(0,1)|D_i = 1, M_i = 1, \tau_i = a] \\
& + \frac{p_c}{p_a + p_c} E[Y_{i1}(0,1)|D_i = 1, M_i = 1, \tau_i = c], \\
\overset{A7}{=}& \frac{p_a}{p_a + p_c} E[Y_{i1}(0,1)|\tau_i = a] + \frac{p_c}{p_a + p_c} E[Y_{i1}(0,1)|\tau_i = c].
\end{aligned}
$$

After some rearrangements and using (56), we obtain

$$E[Y_{i1}(0,1)|\tau_i = c] = \frac{p_a + p_c}{p_c} E[Q_{01}(Y_{i0})|D_i = 1, M_i = 1] - \frac{p_a}{p_c} E[Y_{i1}|D_i = 0, M_i = 1].$$

This gives

$$E[Y_{i1}(0,1)|\tau_i = c] = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Q_{01}(Y_{i0})|D_i = 1, M_i = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Y_{i1}|D_i = 0, M_i = 1] \tag{72}$$

with $p_a = Pr(M_i = 1|D_i = 0) = p_{1|0}$, and $p_c + p_a = Pr(M_i = 1|D_i = 1) = p_{1|1}$.

A.5.4 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 proofs that $\theta_1^c(q,1) = F^{-1}_{Y_1(1,1)|c}(q) - F^{-1}_{Y_1(0,1)|c}(q)$ is identified.

In analogy to (54), the outcome distribution under $D_i = 0$ and $M_i = 0$ 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(67) and rearranging the equation gives

$$F_{Y_1(1,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1} - p_{1|0}} F_{Y_1|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} F_{Q_{11}(Y_0)|D=0,M=1}(y). \tag{73}$$

From (50), 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 (68) and rearranging the equation gives,

$$
F_{Y_1(0,1)|\tau=c}(y) = \frac{p_{1|1}}{p_{1|1}-p_{1|0}}F_{Q_{01}(Y_0)|D=1,M=1}(y) - \frac{p_{1|0}}{p_{1|1}-p_{1|0}}F_{Y_1|D=0,M=1}(y). \tag{74}
$$

A.6 Proof of Theorem 5

A.6.1 Average treatment effect on the compliers

In (62) and (71), we show that

$$
\begin{aligned}
\theta_1^c =&E[Y_{i1}(1,1)-Y_1(0,0)|\tau_i=c],\\
=&\frac{p_{1|1}}{p_{1|1}-p_{1|0}}E[Y_{i1}|D_i=1,M_i=1] - \frac{p_{1|0}}{p_{1|1}-p_{1|0}}E[Q_{11}(Y_{i0})|D_i=0,M_i=1]\\
&-\frac{p_{0|0}}{p_{0|0}-p_{0|1}}E[Y_{i1}|D_i=0,M_i=0] + \frac{p_{0|1}}{p_{0|0}-p_{0|1}}E[Q_{00}(Y_{i0})|D_i=1,M_i=0].
\end{aligned}
$$

A.6.2 Quantile treatment effect on the compliers

In (73) and (65), we show that $F_{Y_1(1,1)|c}(y)$ and $F_{Y_1(0,0)|c}(y)$ are identified. Accordingly, $\Delta_1^c(q) = F^{-1}_{Y_1(1,1)|c}(q) - F^{-1}_{Y_1(0,0)|c}(q)$ is identified.

A.6.3 Average indirect effect under $d = 0$ on compliers

In (62) and (72), we show that

$$
\begin{aligned}
\delta_1^c(0) =&E[Y_{i1}(0,1)-Y_{i1}(0,0)|\tau_i=c],\\
=&\frac{p_{1|1}}{p_{1|1}-p_{1|0}}E[Q_{11}(Y_{i0})|D_i=1,M_i=1] - \frac{p_{1|0}}{p_{1|1}-p_{1|0}}E[Y_{i1}|D_i=0,M_i=1]\\
&-\frac{p_{0|0}}{p_{0|0}-p_{0|1}}E[Y_{i1}|D_i=0,M_i=0] + \frac{p_{0|1}}{p_{0|0}-p_{0|1}}E[Q_{00}(Y_{i0})|D_i=1,M_i=0].
\end{aligned}
$$

A.6.4 Quantile indirect effect under $d = 0$ on compliers

In (74) and (65), we show that $F_{Y_1(0,1)|c}(y)$ and $F_{Y_1(0,0)|c}(y)$ are identified. Accordingly, $\delta_1^c(q,0) = F^{-1}_{Y_1(0,1)|c}(q) - F^{-1}_{Y_1(0,0)|c}(q)$ is identified.

A.6.5 Average indirect effect under $d = 1$ on compliers

In (71) and (63), we show that

$$\begin{aligned}
\delta_1^c(1) =& E[Y_{i1}(1,1) - Y_{i1}(1,0)|\tau_i = c], \\
=& \frac{p_{1|1}}{p_{1|1} - p_{1|0}} E[Y_{i1}|D_i = 1, M_i = 1] - \frac{p_{1|0}}{p_{1|1} - p_{1|0}} E[Q_{11}(Y_{i0})|D_i = 0, M_i = 1] \\
&- \frac{p_{0|0}}{p_{0|0} - p_{0|1}} E[Q_{00}(Y_{i0})|D_i = 0, M_i = 0] + \frac{p_{0|1}}{p_{0|0} - p_{0|1}} E[Y_{i1}|D_i = 1, M_i = 0].
\end{aligned}$$

A.6.6 Quantile indirect effect under $d = 1$ on compliers

In (73) and (64), we show that $F_{Y_1(1,1)|c}(y)$ and $F_{Y_1(1,0)|c}(y)$ are identified. Accordingly, $\delta_1^c(q,1) = F^{-1}_{Y_1(1,1)|c}(q) - F^{-1}_{Y_1(1,0)|c}(q)$ is identified.