Monotonic Path-Specific Effects: Application to Estimating Educational Returns

Aleksei Opacic*
Department of Sociology, Harvard University

August 15, 2024

Abstract

Conventional research on educational effects typically either employs a "years of schooling" measure of education, or dichotomizes attainment as a point-in-time treatment. Nevertheless, such a conceptualization of education is at odds with the sequential process by which individuals make educational transitions. In this paper, I propose a causal mediation framework for the study of educational effects on outcomes such as earnings. The framework considers the effect of a given educational transition as operating indirectly, via progression through subsequent transitions, as well as directly, net of these transitions. I demonstrate that the average treatment effect (ATE) of education can be additively decomposed into mutually exclusive components that capture these direct and indirect effects. The decomposition has several special properties which distinguish it from conventional mediation decompositions of the ATE, properties which facilitate less restrictive identification assumptions as well as identification of all causal paths in the decomposition. An analysis of the returns to high school completion in the NLSY97 cohort suggests that the payoff to a high school degree stems overwhelmingly from its direct labor market returns. Mediation via college attendance, completion and graduate school attendance is small because of individuals' low counterfactual progression rates through these subsequent transitions.

Keywords: causal inference, mediation, sequential ignorability, education

*Many thanks to Clem Aeppli, Kosuke Imai, Ian Lundberg, Michael Zanger-Tishler, Yi Zhang, and especially to Xiang Zhou, as well as to members of the Harvard C.A.R.E.S. lab and a reviewer from the Alexander and Diviya Magaro Peer Pre-Review program, for helpful feedback and conversations.

1 Introduction

One of the most resilient social scientific findings across a range of national contexts is the strong association between educational attainment and a variety of life outcomes, including earnings, health, social capital, and family stability (Hout, 2012; Chetty et al., 2023). Conventionally, researchers have taken one of two approaches to evaluating the social and economic returns to education: the first employs a "years of schooling" measure of educational attainment (Angrist and Krueger, 1991, 1992; Kane and Rouse, 1993; Card, 1994; Ashenfelter and Zimmerman, 1997; Card, 1999, 2001; Angrist and Chen, 2011), while the second dichotomizes attainment as a point-in-time treatment. This latter approach has been particularly influential in the study of the impact of postsecondary attainment on earnings, where the treatment considered is often an indicator for whether an individual has attended, or graduated from, college (Brand and Xie, 2010; Carneiro et al., 2011; Zimmerman, 2014; Goodman et al., 2017; Smith et al., 2020; Bleemer, 2022; Mountjoy, 2022).

Despite the important insights this literature has made into establishing the causal effect of educational attainment on important social and economic outcomes, extant work has been inattentive to the sequential process by which people make educational transitions (Mare, 1980).[1] At the end of high school, individuals decide whether or not to enroll in college. Among college enrollees, only 60% receive a BA within six years of initial college entry (Snyder et al., 2016), with an even lower proportion for low-income students and students of color (Eller and DiPrete, 2018; Zhou and Pan, 2023). Moreover, amidst higher educational expansion in the US, college graduates must increasingly choose whether to enter the labor market or to enroll in post-graduate education. Increasingly, therefore, educational attain-

[1] I use the term "educational transition" to refer both to vertical transitions (e.g. enrollment at a secondary or tertiary institution), as well as to the attainment of a qualification at a given level (e.g. high school graduation or BA completion).

ment in the US has become a field of multiple levels with sequential transitions, all of which are independently consequential for individuals' labor market outcomes, and therefore of independent scientific interest.

The sequential nature of educational transitions implies that a causal mediation framework can be employed to study the causal paths by which education's "value-added" occurs. Specifically, we can consider the first transition in a sequence of educational levels of interest as a treatment variable, A, and subsequent transitions as mediators that "transmit" the effects of the treatment and of prior transitions, M_k $(1 \leq k \leq K)$. For example, if we are interested in the total effect of high school completion on earnings, we may ask to what extent this total effect operates indirectly, through the effects of college attendance and college completion (putative mediators) on earnings, or directly, through alternative causal pathways. The insight that the total causal effect of education can be decomposed into its direct and indirect effects opens up a range of important research and policy-oriented questions. For example, tracing to what extent an early-stage educational intervention boosts outcomes such as earnings via its promotion of subsequent educational attainment (its indirect effects), or via earnings directly, would enable policy-makers to discern what drives the intervention's value and to hone subsequent policy (e.g. Hurwitz and Howell, 2014; Sullivan et al., 2019; Castleman et al., 2020; Bird et al., 2021; Dynarski et al., 2021; Black et al., 2023; Turner and Gurantz, 2024). Relatedly, if the early intervention's effects are heterogeneous across demographic groups, assessing the intervention's direct and indirect effects could guide researchers to aspects of the educational experience that either promote or inhibit upward mobility. Nevertheless, prior empirical approaches are not well-suited to answering these questions: a "years-of-schooling" approach captures the direct effect of each additional year of schooling, while the dichotomous approach conflates the direct and

indirect effects.[2]

In this article, I introduce a causal mediation framework for analyzing the effects of educational transitions. For the setting of K (≥ 1) monotonic mediators, I develop a general formula that decomposes the total effect of any level of education into $K + 1$ monotonic path-specific effects (MPSEs): a direct effect net of K subsequent educational transitions, reflecting the path $A \rightarrow Y$, and K mutually exclusive "continuation" or gross effects, reflecting the paths $A \rightarrow M_1 \rightarrow Y$, $A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$, and $A \rightarrow M_1 \cdots \rightarrow M_K \rightarrow Y$. Most importantly, this decomposition exploits a unique characteristic of this empirical setting, in which mediators are characterized by "monotonicity": that is, where an individual's potential $k + 1$ mediator value is deterministically zero if that individual's kth mediator value is 0. The resultant decomposition of the ATE into $K + 1$ monotonic path-specific effects (MPSEs) can be non-parametrically identified under the assumption of sequential ignorability, which allows for the effect of each educational level to be confounded by a distinct set of (observed) intermediate covariates. I introduce several estimation strategies for my proposed decomposition, including a simple linear model-based regression-with-residuals (RWR) procedure, and a non-parametric estimation strategy based on the efficient influence functions (EIFs) of the target parameters (see Chernozhukov et al., 2017; Kennedy, 2022).

This study makes three main contributions. Within the realm of education research, I draw on important work by Heckman et al. (2018), who present a similar decomposition of the effect of schooling over the early life course, but differs in two important respects. First, I provide nonparametric definitions, identification results, and estimation strategies

[2]A further strand of literature, especially prominent in labor economics, explores labor market returns to horizontal aspects of differentiation within a given educational level (e.g. college selectivity, as well as specific colleges) or college types (e.g. Cohodes and Goodman, 2014; Goodman et al., 2017; Mountjoy and Hickman, 2021; Chetty et al., 2023; Ciocca-Eller, 2023). While my proposed framework prioritizes the effects of different levels of education, I discuss in my concluding remarks how the framework could be extended to accommodate multivariate mediators.

for decomposing the total effect of schooling through its direct and indirect components. Second, my decomposition accommodates the presence of a distinct set of observed intermediate confounders for each transition. While one limitation of my approach is that I assume away the presence of *unobserved* confounders for each transition, I propose a sensitivity analysis that assesses the robustness of the results to unobserved confounding, under a set of simplifying assumptions.

More broadly, my framework speaks to the burgeoning field of causal mediation analysis in the social, economic, and health sciences, targeted at assessing the causal pathways by which a treatment affects an outcome. While prior literature overwhelmingly focuses on single-mediator decompositions of the ATE, a growing body of work examines mediation estimands in settings with multiple mediators (Avin et al., 2005; Albert and Nelson, 2011; VanderWeele and Vansteelandt, 2014; Lin and VanderWeele, 2017; Miles et al., 2017; Steen et al., 2017; Vansteelandt and Daniel, 2017; Miles et al., 2020). In particular, in the case of two causally ordered mediators, Daniel et al. (2015) show that the ATE can be decomposed into multiple path-specific effects (PSEs), and outline the assumptions under which some of these effects are identified. Most recently, Zhou (2022b) generalized this framework to the case of K mediators, establishing a set of identifiable PSEs and introducing several regression-based, weighting, and semiparametric efficient estimators. I extend this literature by examining a special empirical setting where the mediators are monotonic. Compared with traditional mediation-based decompositions, monotonicity faciliates PSE identification under weaker identification assumptions, enables identification of all of the causal paths in question, as opposed to just a strict subset of them, and further permits a finer-grained decomposition. The general decomposition also extends previous literature on mediation under monotonicity which has focused exclusively on the case of a single

mediator (e.g. Zhou, 2022a).

Finally, I also contribute to a growing parallel literature that proposes a range of non-parametric, and semi-parametric efficient estimators for alternative mediation estimands, based on the efficient influence functions (EIFs) of the causal quantities of interest (e.g. Miles et al., 2020; Farbmacher et al., 2022; Zhou, 2022b), as well as to related closely-related work that proposes semi-parametric efficient estimators for dynamic treatment effects (Lewis and Syrgkanis, 2020; Viviano and Bradic, 2021; Bodory et al., 2022).

In the following sections, I first introduce the decomposition for the case of a single intermediate educational transition, before discussing the general case of K intermediate transitions and its identification under the assumption of sequential ignorability (Section 2). In Section 3, I introduce a semiparametric estimation strategy for estimating the proposed decomposition, and in Section 4, I illustrate the proposed framework and methods using data from the National Longitudinal Survey of Youth (NLSY97) cohort. Section 5 concludes.

2 Monotonic Path-Specific Effects

2.1 A Single Intermediate Transition

I first consider the case of a single intermediate educational transition (monotonic mediator). Suppressing subscripts i, let A denote an indicator for high school graduation (the initial educational transition), M_1, an indicator for college attendance (a monotonic mediator or transition), and Y, a binary or continuous outcome of interest such as earnings. A single-transition decomposition thus assesses the educational sequence $A \rightarrow M \rightarrow Y$: high school graduation→college attendance→earnings. In this way, I treat college atten-

dance as a mediator of the total effect of high school graduation on earnings, in relation to which the total effect of high school graduation can be decomposed into an indirect effect (that "flows through" college attendance), and a direct effect (net of college attendance). Following Heckman et al. (2018), I refer to this latter term as the "continuation" value of educational transition A.

Using potential outcomes notation, let $M(a)$ denote an individual's potential value of the mediator if their treatment status were set to a, and let $Y(a, m)$ denote that individual's potential outcome if their treatment and mediator statuses were set to a and m, respectively. I assume, as I formalize in the following section, that sequential transitions are characterized by monotonicity; here, this means that individuals who do not complete high school cannot attend college, or $M(0) = 0$. This sequential nature of educational transitions therefore implies the following set potential outcomes: $\{Y(1), Y(0), Y(1,0), Y(0), Y(1,1)\}$. Further, by the composition assumption, $Y(0) = Y(0, M(0) = Y(0,0)$, and

$$Y(1) = Y(1,0) + M(1)[Y(1,1) - Y(1,0)].$$

Thus, the individual total effect of A on Y can be decomposed as

$$Y(1) - Y(0) = Y(1,0) - Y(0,0) + M(1)[Y(1,1) - Y(1,0)]. \tag{1}$$

Since these individual-level quantities are unidentified, I focus on their population-level analogs. Taking the expectation of Equation 1, we obtain the following decomposition of

the ATE (see Zhou, 2022a):

$$\begin{aligned}\text{ATE} &= \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[Y(1,0) - Y(0,0)] + \mathbb{E}[M(1)]\mathbb{E}[Y(1,1) - Y(1,0)] + \text{cov}[M(1), Y(1,1) - Y(1,0)] \quad (2) \\ &= \underbrace{\Delta_0}_{A \to Y} + \underbrace{\pi_1 \Delta_1 + \eta_1}_{A \to M \to Y}. \quad (3)\end{aligned}$$

Here, Δ_0 and Δ_1 denote the direct effects of the first and intermediate transitions on the outcome, $A \to Y$ and $M \to Y$, respectively, π_1 denotes the total effect of the first transition on the intermediate transition $A \to M$, and η_1 denotes the covariance between the effect of the initial transition on completion of the second and the effect of the second transition on Y. Specifically, η_1 is positive if those who would attend college given high school completion (i.e., $M(1) = 1$) benefit more from college attendance in terms of their later earnings (i.e., have a larger $Y(1,1) - Y(1,0)$) than those who do not (i.e., $M(1) = 0$), and negative if the opposite is true. Meanwhile, the composite term $(\pi_1 \Delta_1 + \eta_1)$ captures the average indirect effect of the treatment via the intermediate transition $(A \to M \to Y)$, comprising the sum of (i) the probability of college enrollment if an individual graduated high school, multiplied by the direct of college enrollment, and (ii) the covariance between college enrollment and its direct effect on earnings.

2.2 Generalization to K Intermediate Transitions

I now generalize the approach introduced in the preceding section to the case of K intermediate transitions. As previously, I denote the treatment ("initial transition") of high school graduation by A, and use $M_1, \ldots M_K$ to refer to the K subsequent transitions of interest

("intermediate transitions"), where I assume that all of $M_1, \ldots, M_K$ are binary and that for any $i < j$, M_i temporally precedes M_j. For instance, we may wish to decompose the total effect of high school completion on earnings via college attendance (M_1), college completion (M_2) and graduate school attendance (M_3). Let an overbar denote a vector of variables, such that $\overline{M}_k = (M_1, M_2, \ldots M_k)$ and $\overline{1}_k = (A = 1, M_1 = 1, \ldots, M_{k-1} = 1)$. Further, let $[K]$ denote the set $\{0, 1, \ldots, K\}$. In addition, I denote by X a vector of pretreatment confounders of the effect of $(A, \overline{M}_k)$ on (M_{k+1}, Y), and by $\overline{Z}_k = (Z_1, \ldots Z_k)$ a vector of intermediate confounders that may confound the causal effect of M_k on (M_{k+1}, Y). Using potential outcomes notation, $Y(\overline{1}_k, m_k)$ thus denotes an individual's potential earnings if they completed, possibly contrary to fact, the treatment in addition to $k-1$ intermediate transitions, and then either completed ($m_k = 1$) or did not complete ($m_k = 0$) the kth *intermediate* transition. Similarly, $M_{k+1}(\overline{1}_{k+1})$ denotes an individual's potential value of the $k+1$th intermediate transition were that individual to complete the treatment as well as k prior intermediate transitions. I invoke the following constraint on units' potential transition values:

Assumption 1. *Monotonicity:* $M_{k+1}(M_k = 0) = 0 \forall k \in [K-1], M_0 \equiv A$.

Informally, Assumption 1 (*monotonicity*) states that an individual's potential $k+1$th transition value is deterministically 0 if that individual fails to complete the prior (kth) transition. It is analogous to a one-sided non-compliance assumption within an instrumental variables (IV) framework, which rules out an "always-takers" principal stratum. We can then use this assumption to decompose the ATE of A on Y, which I denote by τ_0. Specifically, let τ_k denote the gross effect of the kth mediator on Y, i.e.,

$$\tau_k = \mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)],$$

let Δ_0 denote the direct effect of A on Y, and let Δ_k denote the direct effect of the kth mediator on Y, i.e.,

$$\Delta_k = \mathbb{E}[Y(\overline{1}_{k+1}, 0) - Y(\overline{1}_k, 0)].$$

To explicate my approach, note that the gross effect of the kth mediator, τ_k, includes not only the direct effect $M_k \rightarrow Y$, net of subsequent educational transitions Δ_k), but also the indirect effects of M_k via subsequent transitions ($M \rightsquigarrow Y$, where a squiggly arrow denotes a combination of multiple paths). This insight motivates us to further decompose τ into its direct and indirect components. Under the composition assumption, τ_k can be decomposed as

$$\tau_k = \Delta_k + \pi_{k+1}\tau_{k+1} + \eta_{k+1}, \tag{4}$$

where

$$\pi_{k+1} = \mathbb{E}[M_{k+1}(\overline{1}_{k+1})],$$
$$\eta_{k+1} = \text{cov}[M_{k+1}(\overline{1}_{k+1}), Y(\overline{1}_{k+2}) - Y(\overline{1}_{k+1}, 0)].$$

For $k = 1, \ldots, K-1$, iteratively substituting equation 4 into the corresponding expression for τ_{k-1} yields

$$\tau_0 = \underbrace{\Delta_0}_{A \rightarrow Y} + \sum_{k=1}^{K} \underbrace{(\Pi_{j=1}^{k}\pi_j)\Delta_k + (\Pi_{j=1}^{k-1}\pi_j)\eta_k}_{\theta_k \triangleq A \rightarrow M_1 \ldots \rightarrow M_k \rightarrow Y}, \tag{5}$$

where $\Delta_K = \tau_K$, i.e. Δ_K is a *gross* or continuation effect, since this latter path is a composite one that contains all residual paths omitted in the decomposition (i.e., through educational transitions subsequent to K, if they exist). Thus, the θ_k terms capture how much of the total effect of high school completion flows through each intermediate transition considered (i.e., via college attendance, via college completion, and via graduate school

attendance), while Δ_0 captures that portion of the total effect that operates directly, net of the K intermediate transitions considered.

2.3 Identification

To identify the causal effects of interest, I rely on a series of sequential ignorability assumptions. While most closely associated with the dynamic treatment effects literature, which rely on observing a complete set of time-varying confounders in order to identify longitudinal effects (see e.g. Lewis and Syrgkanis, 2020; Viviano and Bradic, 2021; Bodory et al., 2022), these assumptions can be transferred to a mediation context, given the fact that the mediators of interest are all causally ordered. As will be discussed in the following section, sequential ignorability identification assumptions are distinct from - and in fact weaker than - the assumptions typically employed in studies of causal mediation.

Before proceeding, I introduce the following shorthands. Let $M_0 \triangleq A$ and $M_k = \emptyset \forall k < 0$. In order to estimate the decomposition shown in Equation 5, it suffices to identify the expectation of two types of composite counterfactuals ($Y(\overline{1}_k, m_{k+1})$ and $M_{k+1}(\overline{1}_{k+1})$), as well as covariance terms of the form $\text{cov}[M_{k+1}(\overline{1}_{k+1}), Y(\overline{1}_{k+2}) - Y(\overline{1}_{k+1}, 0)]$ $\forall k \in [K-1]$. I invoke the following three assumptions:

Assumption 2. *Consistency: for any unit, if $A = a$, $Y = Y(a)$; if $(A, \overline{M}_k) = \{\overline{1}_k, m_k\}$, then $Y = Y(\overline{1}_k, m_k)$ $\forall k \in [K]$, and if $(A, \overline{M}_k) = \overline{1}_{k+1}$, then $M_{k+1} = M_{k+1}(\overline{1}_{k+1}) \forall m_{k+1} \in \{0,1\}, \forall k \in [K-1]$.*

Assumption 3. *Sequential ignorability: $(M(1), Y(a)) \perp\!\!\!\perp A|X$; $Y(\overline{1}_k, m_k) \perp\!\!\!\perp \overline{M}_k|X, \overline{Z}_k, \overline{M}_{k-1}$ and $M_{k+1}(\overline{1}_{k+1}) \perp\!\!\!\perp \overline{M}_k|X, \overline{Z}_k, \overline{M}_{k-1}, \forall k \in \{1, \ldots, K\}, \forall m_k \in \{0,1\}, M_0 \equiv A$.*

Assumption 4. *Positivity: $p_{A|X}(a|x) > \epsilon > 0$, $p_{M_k|X,A,\overline{Z}_k,\overline{M}_{k-1}}(m_k|x, a, \overline{z}_k, \overline{m}_{k-1}) > \epsilon > 0$ $\forall k \in [K]$.*

Assumption 2 (*consistency*) states that a unit's observed outcome equals its potential outcome under a given treatment sequence. Assumption 3 (*sequential ignorability*) is the no unmeasured confounding assumption for the treatment and all mediators. It is considered plausible when sufficient pre-treatment and intermediate covariates $(X, \overline{Z}_K)$ are collected. Finally, Assumption 4 (*positivity*) requires that treatment and mediator assignment is not deterministic. Under Assumptions 2-4, $\mathbb{E}[Y(\overline{1}_k, m_k)]$ and $\mathbb{E}[M_{k+1}(\overline{1}_{k+1})]$ are identified, respectively, as

$$\mathbb{E}[Y(\overline{1}_k, m_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k] \big[\prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1}) \big] dP(x) \tag{6}$$

$$\mathbb{E}[M_{k+1}(\overline{1}_{k+1})] = \int_x \int_{\overline{z}_k} \mathbb{E}[M_{k+1}|x, \overline{z}_K, \overline{1}_{k+1}] \big[\prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1}) \big] dP(x) \tag{7}$$

For a proof of the above formulas, see Robins (1986). The covariance (η_k) components in the decomposition are then identified as the "residual" terms such as in Equation 4, which follows directly from the fact that all other components in these equations are identified. Thus, for $k \in \{1, \ldots K\}$, I can identify η_k as

$$\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k.$$

2.4 A Comparison with Conventional Mediation Analysis with Multiple Causally Ordered Mediators

The above decomposition has an analog in the context of a mediation–based decomposition of the ATE with multiple ordered mediators, but differs from conventional mediation analysis in important ways. To illustrate the differences, consider a binary treat-

ment, A, an outcome of interest, Y, and a vector of pretreatment covariates, X, and let $M_1, M_2, \dots M_K$ denote K causally ordered mediators, assuming that for any $i < j, M_i$ precedes M_j, as above. Moreover, let an overbar denote a vector of variables, so that $\bar{M}_k = (M_1, M_2, \dots M_k)$,$\bar{m}_k = (m_1, m_2, \dots m_k)$, and $\bar{a}_k = (a_1, a_2, \dots a_k)$. Using the potential outcomes notation as above, I can define the following expectation of a nested counterfactual,

$$\psi_{a\bar{a}_k} \triangleq \mathbb{E}\left[Y(a, \bar{M}_k(\bar{a}_k))\right],$$

where $\bar{M}_k(\bar{a}_k) \triangleq \left(\bar{M}_{k-1}(\bar{a}_{k-1}), M_k(a_k, \bar{M}_{k-1}(\bar{a}_{k-1}))\right), \forall k \in [K]$. Under Pearl's (2009) nonparametric structural equation model (NPSEM), Zhou (2022b) demonstrates that the ATE of A on Y can be decomposed into $K+1$ identifiable PSEs corresponding to each of the causal paths $A \to Y$ and $A \to M_k \rightsquigarrow Y$ $(k \in [K])$:

$$\text{ATE} = \psi_{\bar{1}} - \psi_{\bar{0}} = \underbrace{\psi_{1,\bar{0}_K,} - \psi_{\bar{0}_{K+1}}}_{A \to Y} + \sum_{k=1}^{K} \underbrace{\left(\psi_{\bar{1}_{k+1},\underline{0}_{k+1}} - \psi_{\bar{1}_k,\underline{0}_{k+1}}\right)}_{A \to M_k \rightsquigarrow Y}. \tag{8}$$

This decomposition holds algebraically when Assumption 1 does not hold (i.e., when the mediators are not monotonic). In contrast, the monotonic characteristic of the proposed decomposition leads to several important differences. First, the PSE decomposition of the ATE in general mediation settings is not algebraically unique, and thus the PSEs defined under alternative decompositions will differ if the effects of the treatment and each mediator vary across levels of the other mediators. In fact, depending on the order in which the paths $A \to Y$ and $A \to M_k \rightsquigarrow Y$ are considered, there are $(K+1)!$ identifiable different ways of decomposing the ATE; the decomposition shown in Equation 8 is just one such decomposition. Consider the case of two causally dependent mediators. In this setting, the

causal pathway $A \to M_2 \rightsquigarrow Y$ can be defined with respect to four different combinations of levels of the treatment and first mediator: under (i) $a = 1$ and $M_1(1)$, (ii) $a = 1$ and $M_1(0)$, (iii) $a = 0$ and $M_1(1)$, or (iv) $a = 0$ and $M_1(0)$. By contrast, as a direct consequence of monotonicity, the MPSE decomposition is the unique PSE decomposition of the ATE.

Second, for general PSE decompositions of the ATE, the set of identifiable decompositions is merely a small subset of the total number of decompositions that hold algebraically (see Avin et al., 2005). In particular, the identifiable decomposition does not enable us to disentangle the mediating effects of M_k that are direct (net of subsequent mediators) and indirect (through different combinations of subsequent mediators). For example, in the case of two causally dependent mediators, to assess the mediating role of M_1, only the composite path $A \to M_1 \rightsquigarrow Y = (A \to M_1 \to Y) + (A \to M_1 \to M_2 \to Y)$ is identified. By contrast, mediator monotonicity permits a finer-grained decomposition of the ATE: each PSE is identified. In the case of two causally dependent mediators, for example, the causal path $A \to M_2 \to Y$ is zero, and as a result, each of the paths $A \to Y$, $A \to M_1 \to Y$ and $A \to M_1 \to M_2 \to Y$ are identifiable. Figure 1 illustrates the causal pathways defined and identified under the proposed decomposition in the case of two monotonic mediators.

Finally, the sequential ignorability assumption required to identify the MPSE decomposition is weaker than those required to identify a generic PSE decomposition of the ATE. Specifically, the latter requires Pearl's (2009) non-parametric structural equation model (NPSEM), which stipulates that $\left(M_{k+1}\left(a_{k+1}, \bar{m}_k\right), \ldots M_K\left(a_K, \bar{m}_{K-1}\right), Y\left(a_{K+1}, \bar{m}_K\right)\right) \perp\!\!\!\perp M_k\left(a_k, \bar{m}^*_{k-1}\right) \mid X, A, \bar{M}_{k-1}, \forall k \in [K]$. This assumption, sometimes referred to as the "cross-world" independence assumption, is stronger than the sequential ignorability assumption (3) required to identify the MPSE decomposition since it rules out the existence of confounders of the mediators, be they observed or unobserved. By contrast, the MPSE

decomposition identification results accommodate *observed* intermediate confounding without altering the substance of the decomposition.

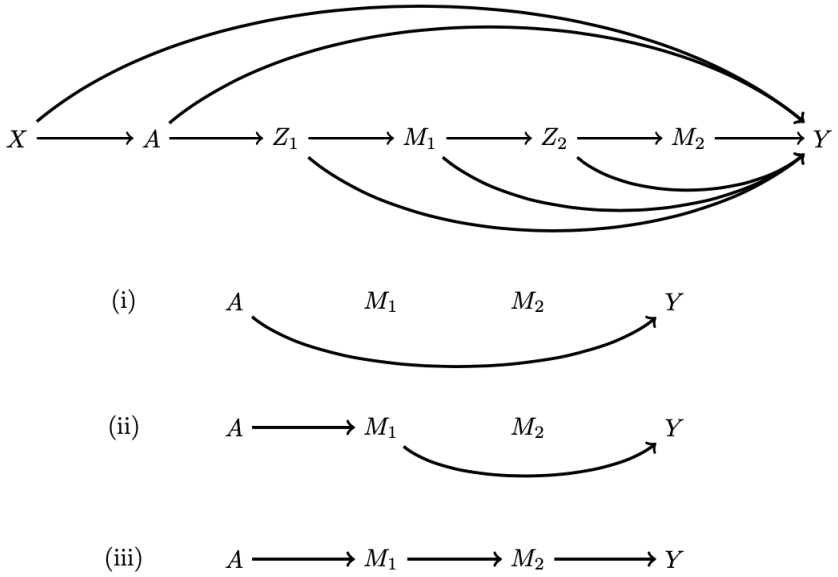

Figure 1: Causal Relationships with Two Monotonic Mediators Shown in a Directed Acyclic Graph (DAG) and the 3 Monotonic Path Specific Effects (MPSEs). A denotes an initial transition of interest, Y, an outcome, and M_1 and M_2 are two causally ordered, monotonic mediators. The set (X, Z_1, Z_2) captures pre-treatment and intermediate confounders.

3 Semiparametric, EIF-Based Estimation

The identification results outlined above suggest that the proposed decomposition can be estimated via several approaches, including outcome-based modeling, models for the treatment and mediators via inverse probability weighting, as well as doubly robust approaches. Parametric procedures are attractive because of their conceptual simplicity and ease of implementation: in Supplementary Material A I show how, under a set of linear models for the outcome and mediators, the θ_k components in Equation 5 can be read off from simple functions of coefficients in these linear models. However when X and $\overline{Z}_K$ are high-dimensional, parametric estimators which require a user-defined specification of the data-generating process may suffer biases resulting from model misspecification. In order

to reduce model dependency, in this section I provide a nonparametric estimation approach which draws on a debiased machine learning (DML) approach. My DML approach is characterized by two components: first, the use of a Neyman orthogonal estimating equation based on the efficient influence function (EIF) for the target parameters, which makes estimates of the parameter "locally robust" to estimates of the nuisance functions; second, the use of a K-fold cross-fitting algorithm (Chernozhukov et al., 2017).

Let $O = (X, A, \overline{Z}_K, \overline{M}_K, Y)$ denote the observed data, and $\mathcal{P}$ a nonparametric model over O wherein all laws satisfy the positivity assumption described in Section 2. Before proceeding, I define the following auxiliary functions, as introduced in Section 2: $\psi_{km_k} \triangleq \mathbb{E}[Y(\overline{1}_k, m_k)]$ and $\phi_k \triangleq \mathbb{E}[M_{k+1}(\overline{1}_{k+1})]$, for all $k \in [K], M_0 \triangleq A$. Using the identification results given in Section 2, ψ_{km_k} can be written in terms of expectations of observed data:

$$\psi_{km_k} = \mathbb{E}_X \mathbb{E}_{Z_1|X,1} \ldots \mathbb{E}_{Z_k|X,\overline{Z}_{k-1},\overline{1}_k} \mathbb{E}[Y|X, \overline{Z}_k, \overline{1}_k, m_k].$$

For each $j \in [k]$, we can thus define $\mu^k_{jm_k}\left(X, \bar{Z}_k\right)$ iteratively as

$$\begin{aligned}
&\mu^k_{km_k}\left(X, \bar{Z}_k\right) \triangleq \mathbb{E}\left[Y \mid X, \bar{Z}_k, \overline{1}_k, m_k\right], \\
&\mu^k_{jm_k}\left(X, \bar{Z}_j\right) \triangleq \mathbb{E}\left[\mu^k_{j+1m_k}\left(X, \bar{Z}_{j+1}\right) \mid X, \bar{Z}_j, \overline{1}_{j+1}\right] \forall j \in [k-1].
\end{aligned}$$

Further, let $\pi_{km_k}\big(X, \overline{Z}_k\big) \triangleq \Pr[M_k = m_k \mid X, \overline{Z}_k, \overline{1}_k] \forall k \in [K]$, and $\pi_{01}\big(X\big) \triangleq \Pr[A = 1|X]$. The efficient influence function (EIF) of ψ_{km_k} is closely related to the EIF for the g-formula, and can be written as

$$\psi_{km_k}(O) = \sum_{j=0}^{k+1} \varphi_j(O),$$

where

$$\varphi_0(O) = \mu^k_{0m_k}(X) - \psi_{km_k}$$
$$\varphi_j(O) = \frac{A}{\pi_{01}(X)}\left(\prod_{l=1}^{j-1}\frac{M_l}{\pi_{l1}(X,\overline{Z}_l)}\right)\left(\mu^k_{jm_k}\left(X,\overline{Z}_j\right) - \mu^k_{j-1m_k}\left(X,\bar{Z}_{j-1}\right)\right), \quad j \in \{1\ldots,k\}$$
$$\varphi_{k+1}(O) = \frac{A}{\pi_{01}(X)}\left(\frac{\mathbb{I}(M_k = m_k)}{\pi_{km_k}(X,\overline{Z}_k)}\prod_{k=1}^{k-1}\frac{M_l}{\pi_{l1}(X,\overline{Z}_l)}\right)\left(Y - \mu^k_{km_k}\left(X,\bar{Z}_k\right)\right).$$

For a proof, see Rotnitzky et al. (2017). The semiparametric efficiency bound for any asymptotically linear estimator of ψ_{km_k} in $\mathcal{P}$ is therefore $\mathbb{E}[\left(\varphi_{km_k}(O)\right)^2]$. The EIF motivates an EIF-based estimator for ψ_{km_k}, obtained by solving the empirical moment condition $\mathbb{P}_n[\varphi_{km_k}(O;\hat{\eta})] = 0$, where $\mathbb{P}_n[\cdot]$ denotes an empirical average, and where $\varphi_{km_k}(O;\hat{\eta})$ denotes the estimated EIF, evaluated using plug-in estimators for the nuisance functions. Specifically,

$$\begin{aligned}\hat{\psi}^{\text{eif}}_{km_k} = \mathbb{P}_n\Bigg[&\frac{A}{\hat{\pi}_{01}(X)}\left(\frac{\mathbb{I}(M_k = m_k)}{\hat{\pi}_{km_k}(X,\overline{Z}_k)}\prod_{k=1}^{k-1}\frac{M_l}{\hat{\pi}_{l1}(X,\overline{Z}_l)}\right)\left(Y - \hat{\mu}^k_{km_k}\left(X,\bar{Z}_k\right)\right)\\ &+\sum_{j=1}^{k}\frac{A}{\hat{\pi}_{01}(X)}\left(\prod_{l=1}^{j-1}\frac{M_l}{\hat{\pi}_{l1}(X,\overline{Z}_l)}\right)\left(\hat{\mu}^k_{jm_k}\left(X,\overline{Z}_j\right) - \hat{\mu}^k_{j-1m_k}\left(X,\bar{Z}_{j-1}\right)\right)\\ &+\hat{\mu}^k_{0m_k}(X)\Bigg].\end{aligned} \tag{9}$$

A similar EIF-based estimator can be used for ϕ_k to estimate the π_k terms in Equation 5. This estimator is based on the following nuisance functions for estimation (see Supplementary Material I for details):

$$\gamma_k\left(X, \bar{Z}_k\right) \triangleq \mathbb{E}\left[M_{k+1} \mid X, \bar{Z}_k, \bar{1}_{k+1}\right],$$
$$\gamma_j\left(X, \bar{Z}_j\right) \triangleq \mathbb{E}\left[\gamma_{j+1}\left(X, \bar{Z}_{j+1}\right) \mid X, \bar{Z}_k, \bar{1}_{j+1}\right] \forall j \in [k-1].$$

Next, following Kennedy (2022, p. 15), let $\mathbb{IF} : \Psi \to L_2(\mathbb{P})$ denote the operator mapping the functionals $\{\Delta_k, \pi_k, \eta_k\} : \mathcal{P} \to \mathbb{R}$, $\forall \in [K]$ to their respective influence functions under the nonparametric model $\mathcal{P}$. Because the (Δ_k, τ_k) components of the decomposition are linear in ψ_{km_k}, by linearity of the EIF, $(\mathbb{IF}(\Delta_k), \mathbb{IF}(\tau_k))$ can be expressed as linear combinations of $\varphi_{km_k}(O)$. In particular, $\mathbb{IF}(\tau_k) = \varphi_{(k+1)1}(O) - \varphi_{k,0}(O)$ and $\mathbb{IF}(\Delta_k) = \varphi_{(k+1)0}(O) - \varphi_{k,0}(O)$. The EIFs of η_k and θ_k, $\forall k \in [K]$ under $\mathcal{P}$ are derived as in Proposition 1:

Proposition 1. *The EIFs of η_k, θ_k $\forall k \in [1, \ldots, K]$ under P are given, respectively, by*

$$\mathbb{IF}(\eta_k) = \mathbb{IF}(\tau_{k-1}) - \mathbb{IF}(\Delta_{k-1}) - \tau_k \mathbb{IF}(\pi_k) - \pi_k \mathbb{IF}(\tau_k),$$
$$\mathbb{IF}(\theta_k) = \mathbb{IF}(\Delta_k) \prod_{j=1}^{k} \pi_j + \Delta_k \sum_{j=1}^{k} \mathbb{IF}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \pi_l + \mathbb{IF}(\eta_k) \prod_{j=1}^{k-1} \pi_j + \eta_k \sum_{j=1}^{k-1} \mathbb{IF}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k-1} \pi_l,$$

for $k \in \{1, \ldots K\}$, with $\theta_0 = \Delta_0$, and where $\mathbb{RIF}(\phi) = \mathbb{IF}(\phi) + \phi$, denotes the recentered EIF of a parameter (about the truth). Their corresponding EIF-based estimators are (see Supplementary Material I for derivations):

$$\hat{\eta}_k^{eif} = \widehat{\mathbb{RIF}}(\tau_{k-1}) - \widehat{\mathbb{RIF}}(\Delta_{k-1}) - \hat{\tau}_k\widehat{\mathbb{RIF}}(\pi_k) - \hat{\pi}_k\widehat{\mathbb{RIF}}(\tau_k) + \hat{\pi}_k\hat{\tau}_k,$$

$$\hat{\theta}_k^{eif} = \widehat{\mathbb{RIF}}(\Delta_k)\prod_{j=1}^{k}\hat{\pi}_j + \hat{\Delta}_k\sum_{j=1}^{k}\widehat{\mathbb{RIF}}(\pi_j)\prod_{\substack{l=1\\l\neq j}}^{k}\hat{\pi}_l + \widehat{\mathbb{RIF}}(\eta_k)\prod_{j=1}^{k-1}\hat{\pi}_j + \hat{\eta}_k\widehat{\mathbb{RIF}}(\pi_j)\prod_{\substack{l=1\\l\neq j}}^{k-1}\hat{\pi}_l$$
$$- k\hat{\Delta}_k\prod_{j=1}^{k}\hat{\pi}_j - (k-1)\hat{\eta}_k\prod_{j=1}^{k-1}\hat{\pi}_j.$$

where $\widehat{\mathbb{RIF}}(\phi) = \widehat{\mathbb{IF}}(\phi) + \phi$, *and* $\widehat{\mathbb{IF}}(\phi)$ *denotes the influence function of a parameter evaluated at estimates of its component nuisance functions (see Supplementary Material I for derivations).*

When machine learning estimators are used to compute the nuisance functions, in order to ensure the convergence rates outlined in Proposition 2 below, one could assume Donsker-type conditions for the nuisance function estimators, which restricts the set of estimators available to use. Alternatively, to expand the class of estimators that can be used for estimating the nuisance functions, sample-splitting can be used. In particular, Chernozhukov et al. (2017) suggest a "cross-fitting" procedure, which comprises the following steps: (1) Randomly split data into J folds: $\{S_1, ...S_J\}$; (2) For each fold S_j, use the remaining $(j-1)$ folds (training sample) to fit a flexible machine-learning model for each of the nuisance functions involved in the estimating equations; (3) For each observation in j (estimation sample), use estimates of the above models to construct a set of estimated RIF functions for $\Delta_k \forall k \in \{0, \ldots, K-1\}$, and for $(\pi_k, \tau_k, \eta_k, \theta_k) \forall k \in [K]$; (4) Compute an estimate of the decomposition components by averaging the estimated RIF functions across all subsamples S_1 through S_J. When all nuisance functions are estimated via data-adaptive methods and cross-fitting, the semiparametric efficiency of θ_k^{rEIF} is given in the following proposition:

Proposition 2. *Semiparametric efficiency. Under Assumption 4, and under suitable regularity conditions (e.g. Chernozhukov et al., 2018), then* $\hat{\theta}_k^{eif}$ *is semiparametric efficient if* $\sum_{j=k}^{k+1}\left[\sum_{l=0}^{j} R_n\big(\hat{\pi}_{l1}\big)R_n\big(\hat{\mu}_{l0}^{j}\big)\right] + \sum_{j=0}^{k-1}\left[R_n\big(\hat{\pi}_{j1}\big)R_n\big(\hat{\mu}_{j0}^{k-1}\big) + R_n\big(\hat{\pi}_{j1}\big)R_n\big(\hat{\mu}_{j1}^{k-1}\big)\right] + \sum_{j=0}^{k}\left[\sum_{l=0}^{j} R_n\big(\hat{\pi}_{l1}\big)R_n\big(\hat{\gamma}_{l}^{j}\big)\right] = o(n^{-1/2})$*, where* $R_n\big(\,\cdot\,\big)$ *denotes a mapping from a nuisance function to its* $L_2(P)$ *convergence rate, and where* $\hat{\mu}_{l0}^{K+1} \triangleq \hat{\mu}_{l1}^{K}$.

To gain some intuition for the result in Proposition 2, we can focus on $\theta_1 = \pi_1\Delta_1 + \eta_1$, i.e., the MPSE through M_1 when $K = 1$. Note that estimation of $\theta_1 = \pi_1\Delta_1 + \eta_1$ requires estimating the following decomposition components: $(\pi_1, \Delta_1, \tau_{0,}, \Delta_0, \tau_1)$. To estimate these components, it suffices to estimate the following quantities: $(\phi_1, \psi_{01}, \psi_{00}, \psi_{10}, \psi_{11})$. In order for $\hat{\theta}_1^{\text{eif}}$ to be semiparametric efficient, we require that the estimators employed for the set $(\phi_1, \psi_{01}, \psi_{00}, \psi_{10}, \psi_{11})$, i.e., $(\hat{\phi}_1^{\text{eif}}, \hat{\psi}_{01}^{\text{eif}}, \hat{\psi}_{00}^{\text{eif}}, \hat{\psi}_{10}^{\text{eif}}, \hat{\psi}_{11}^{\text{eif}})$, are themselves semiparametric efficient. Thus, a sufficient (but not necessary) condition in order for $\hat{\theta}_1^{\text{eif}}$ to obtain the semiparametric efficiency bound is if, for any two nuisance functions involved in $(\hat{\phi}_1^{\text{eif}}, \hat{\psi}_{01}^{\text{eif}}, \hat{\psi}_{00}^{\text{eif}}, \hat{\psi}_{10}^{\text{eif}}, \hat{\psi}_{11}^{\text{eif}})$, the product of their convergence rates is $o(n^{-1/2})$. In this way, $\hat{\theta}_1^{\text{eif}}$ will obtain the semiparametric efficiency bound if all of its constituent nuisance functions converge at a rate faster than $n^{-1/4}$ (although it will also obtain the efficiency bound under a variety of alternative conditions).

When data-adaptive methods are used to estimated the nuisance functions, inference on all components can be conducted via the variance of the empirical analog of the EIF, i.e. $\mathbb{P}_n[\big(\hat{\psi}_{km_k}^{\text{EIF}}\big)^2]/n$. For example, inference on τ_1 can be conducted by estimating $\mathbb{P}_n[\big(\hat{\psi}_{11}^{\text{EIF}} - \hat{\psi}_{10}^{\text{EIF}}\big)^2]/n$.

4 Empirical Analysis

To illustrate my approach empirically, I draw on data from the National Longitudinal Survey of Youth 1997 (NLSY97). I parse out the direct effect of high school graduation on adult earnings and its indirect or continuation effects via (i) college attendance, (ii) college graduation, and (iii) graduate school attendance. My analytic sample comprises $N = 7,305$ respondents.

I construct four types of variables: educational transitions, adult earnings, a set of confounders for the effect of high school graduation on subsequent transitions and earnings, and a single set of intermediate confounders for the effect of college completion on subsequent transitions and earnings. My educational transition variables contain a binary treatment denoting whether a respondent had graduated high school by age 22, and three binary mediators denoting whether the respondent had attended a 4-year college by age 22, whether the respondent had received a BA degree by age 29, and whether the respondent had enrolled in a graduate level program by age 29, respectively. I assume that all individuals who make a given educational transition have made all previous educational transitions. Thus, by construction, my coding strategy disallows for cases which violate the monotonicity assumption.[3] My outcome of interest is logged average annual earnings at ages 32-36, which I define to be the (logged) average of a respondent's self-reported wage, salary income, and business income. Earnings are adjusted for inflation to 2023 dollars using the personal consumption expenditures (PCE) index. After dropping respondents with missing earnings information, I accommodate those with zero earnings by adding a small

[3] Assuming away cases in which an individual makes a particular educational transition without having made *all* previous transitions serves as a reasonable approximation to reality. Among the set of individuals who have non-missing earnings information in the NLSY97 (i.e., those who comprise my analytic sample), 94% of individuals observed to attend graduate school by age 29 also completed a BA by age 29; 93% of respondents who completed a BA by age 29 had attended a 4-year college by age 22 (6% of those who completed a BA by age 29 first attended a 4-year college between ages 23 and 26 inclusive), and 99% of respondents who attended a 4-year college by age 22 had also completed high school.

constant of $\$1,000$ to observed earnings (though in Supplementary Material F, I replicate my main analyses under alternative definitions of earnings).

In an effort to satisfy the sequential ignorability assumption (Assumption 3), I include a large array of covariates in my models. This set of covariates is more expansive than those used in previous, observational studies of returns to education (see in particular Scott-Clayton and Wen, 2019). In particular, in addition to including information on respondent demographics (gender, race, ethnicity, age in 1997), and observed pre-college performance such as overall high school GPA and test score on the Armed Services Vocational Aptitude Battery (ASVAB), I include detailed information on socioeconomic background. Since my proposed decomposition also facilitates the inclusion of a distinct set of observed intermediate confounders for each transition, I include two postsecondary characteristics (Z) to adjust for confounders of the effect of BA completion and graduate school attendance on earnings: field of study and college GPA. To assess the robustness of my main conclusion to forms of unobserved confounding, in Supplementary Material C, I produce a set of "bias-corrected" estimates of the decomposition components under certain assumptions about the nature of the confounding.

A large proportion (just under 50%) of respondents are missing information on covariates X and Z. For my main analyses, I impute missing values on these covariates via multiple imputation to increase efficiency, but in Supplementary Material E, I replicate these analyses restricted to the sample of respondents with complete information. This exercise produces substantively similar results (for covariate means for each of these analytic samples, see Supplementary Material D). After constructing the analytical sample, I apply both the DML estimator described in Section 3 as well as a parametric, regression-with-residuals (RWR) algorithm (described in Supplementary Material A) to implement

the proposed decomposition. For the DML approach, I estimate all nuisance functions, using a super learner composed of the Lasso and random forest and, following Chernozhukov et al. (2017), use five-fold cross-fitting. All weights involved in computing the rEIFs are censored at their $1st$ and $99th$ percentiles. Supplementary Material H gives further details about the particular models required given my assumed data generation process.

Figure 2 shows my estimates of the average total effect (ATE) on log earnings and its direct and continuation components under both the DML and RWR procedures. Both procedures return similar estimates, though deviate in the estimated magnitude of MPSE θ_1, and DML estimates come expectedly with a significantly greater amount of precision. The first column shows that the estimated ATE of attending high school on log earnings under DML (RWR) is 0.67 (0.63), which implies an earnings premium of approximately 96%. The next two columns indicate that the vast majority (69% under DML and 75% under RWR) of the ATE operates directly, i.e. net of college attendance, BA completion and graduate school attendance (MPSE θ_0, $A \to Y$). Specifically, high school graduates who do not proceed to college can be expected to earn on average 0.46 (0.47) log earnings more than high school non-completers under DML (RWR), an earnings premium of 59%.

Under DML, the continuation effects of high school graduation via college attendance without BA completion (MPSE θ_1, $A \to M_1 \to Y$) and via BA completion without graduate school participation (MPSE θ_2, $A \to M_1 \to M_2 \to Y$) both mediate roughly 15% of the ATE, and correspond to an earnings premium of approximately 10%. The RWR estimate of θ_1 is notably lower at 0.03 and is also imprecisely estimated. Under both estimation procedures, the continuation effect via graduate school attendance ($A \to M_1 \to M_2 \to M_3 \to Y$) is very small and fails to reach conventional levels of significance. In sum, the total effect of high school graduation on earnings is determined overwhelmingly by its direct

effect on earnings.

Table 1 shows DML and RWR estimates of the various components (the direct effects (Δ_k), probabilities (π_k) and covariance terms (η_k)) that constitute the continuation effects θ_k. Several points are of note. First, the components in the table offer insights into the economic and educational returns to different educational stages. The direct effects of each educational transition (Δ_k) are highly variable: they are largest for high school graduation and for college completion (both at 0.46 under DML), and lowest for college attendance and graduate school participation (at 0.2 and 0.12, respectively, under DML). Note that the payoff to graduate school attendance could be depressed by the fact that I observe individuals at a maximum age of only 36, if graduate school earnings premia materialize only much later in the life course. The counterfactual continuation probabilities (π_k) also provide insight into barriers in educational participation. In particular, even if an individual were to complete high school (possibly contrary to fact), that individual would have under a 50% chance of continuing to a 4-year college without further intervention to increase individuals' college application, admissions and enrollment rates. Further, even if individuals were to counterfactually both complete high school and attend a 4-year college, only a very small proportion ($\pi_1 \cdot \pi_2 = 0.24$) would be expected to complete their BA degree without further intervention at the college-level.

Second, the fine-grained nature of the MPSE decomposition enables us to trace the continuation effects to their constituent components. In particular, while the direct effect of high school completion is comparable to the direct effect of BA graduation on earnings, suggesting an earnings premium of 59% relative to college attendance without completion, the continuation effect via BA completion that it informs (MPSE θ_2, $A \to M_1 \to M_2 \to Y$) only mediates a small amount of the overall ATE because because θ_2 is approximately

(plus the small value of η_2) equal to Δ_2 scaled by the product $\pi_1 \cdot \pi_2 = 0.24$. In words, despite the relatively large direct effect of BA completion on earnings, given individuals' low counterfactual probability of BA completion, this transition is not an important mediating pathway of the total effect of high school completion on earnings. The result is that college attendance without completion mediates high school graduation's earnings effects as much as BA completion, despite the fact that college attendance without completion yields a much smaller earnings return for high school graduates than BA completion without graduate school attendance does for college enrollees.

One instructive point of comparison for these results are instrumental variable (IV) estimates of returns to years of schooling, typically estimated in the range of 6% to 12% (Angrist and Krueger, 1991, 1992; Kane and Rouse, 1993; Card, 1994; Ashenfelter and Zimmerman, 1997; Angrist and Chen, 2011). While my estimate of the overall return to high school graduation (τ_0) could appear large in this light, several factors could reconcile this difference. First, τ_0 captures the direct *and* continuation effects of high school completion (whereas IV estimates of schooling returns capture schooling's direct effects). Further, τ_0 captures the effect of multiple additional years of schooling (as the high school graduates and high school non-completers that form the comparison group differ by multiple years of schooling), as opposed to a single year's additional return. In fact, we can more directly compare my DML estimate of the direct return to high school graduation (Δ_0) of 0.46 (corresponding to an earnings premium of 58%) using the fact that, in the NLSY97, high school non-completers attained on average 3.7 fewer years of schooling than high school completers. An IV estimate of 12%, for example, would therefore imply an earnings return to 3.7 additional years of approximately 52% - broadly in line with my result. Still, to assess the robustness of the above findings to potential violations of Assumption 3 (Sequential

Ignorability), I implement a sensitivity analysis in Supplementary Material C. Under the stated assumptions about the pattern of unobserved confounding, my primary finding that the ATE of high school graduation is overwhelmingly mediated via its direct effect remains highly robust to unobserved confounding.

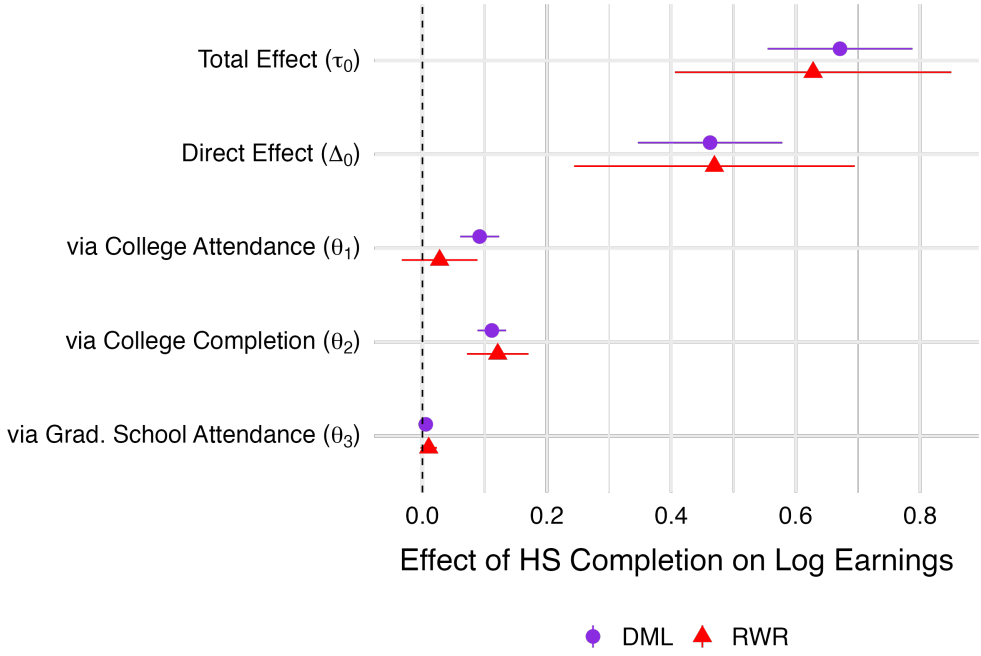

Figure 2: Decomposition of the Average Total Effect (ATE) of High School Graduation on Logged Earnings via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

5 Conclusion

In this article, I have developed a causal mediation framework for analyzing education effects on earnings. First, I have demonstrated that the total effect of any level of education can be decomposed into a direct effect and K mutually exclusive "continuation" effects. All of these effects are identifiable under the assumption of sequential ignorability. Importantly, this property allows for the effect of each educational transition to be confounded by a distinct set of observed covariates - a property which allows for weaker identification

Table 1: Direct Effects (Δ_k), Probabilities (π_k) and Covariance Terms (η_k) Involved in Decomposition via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	Δ_0	Δ_1	Δ_2	Δ_3	π_1	π_2	π_3	η_1	η_2	η_3
DML	0.462	0.200	0.463	0.122	0.427	0.554	0.315	0.006	0.005	-0.016
	(0.059)	(0.034)	(0.046)	(0.029)	(0.009)	(0.015)	(0.022)	(0.007)	(0.007)	(0.009)
RWR	0.469	0.117	0.491	0.160	0.374	0.515	0.219	-0.016	0.071	0.017
	(0.115)	(0.082)	(0.097)	(0.113)	(0.015)	(0.066)	(0.018)	(0.007)	(0.015)	(0.046)

Note: The Δ_k parameters capture the average effect of completing the kth mediator *but no subsequent mediator* on earnings, relative to completing the $k-1$th mediator. For instance, Δ_0 denotes the effect of completing high school (M_1) but not attending college nor, under Assumption 1, completing any subsequent mediators, relative to attending high school but not completing it ($M_0 \equiv A$). The π_k terms capture the average of individuals' counterfactual completion status of the kth mediator under completion of all prior mediators $M_0, \dots M_{k-1}$. For example, π_1 denotes individuals' average counterfactual college attendance, after - possibly contrary to fact - their completion of high school. Finally, the η_k terms refer to the covariance between individuals' own counterfactual completion status of the kth mediator, and their own "gross" effect of completing the kth mediator on earnings. To recall, the "gross" effect of the kth mediator captures the effect of completing that mediator, relative to completing only the $k-1$th mediator, irrespective of whether that effect operates directly (net of subsequent mediators) or via subsequent transitions. For example, η_1 denotes the covariance between each individual's counterfactual college attendance status and their gross effect of college attendance on earnings.

conditions compared with conventional mediation-based decompositions of the ATE (Miles et al., 2017; Zhou, 2022b).

Although my empirical motivation is the estimation of educational returns, the proposed framework applies widely to a range of demographic and organizational settings characterized by "state dependency" between treatment and mediators. This characteristic is particularly salient in demographic phenomena, which often involve sequential transitions over the life course. Certain demographic events are rigid in their monotonicity as a result of their definition. For example, researchers may be interested in discerning the degree to which positive effects of marriage on outcomes such as earnings and life satisfaction are undermined by the negative effects of divorce and separation (and, in turn, their mitigation via re-marriage) (Kenney, 2004; Sweeney and Phillips, 2004). Divorce can be "attained" only by individuals who are already married. Similarly, the effect of parenthood on earnings can be seen as operating directly, through the effect of having a first child net of

subsequent children, as well as operating indirectly through the effects of having multiple children, transitions which are clearly monotonic in nature. A similar perspective may be taken in a criminal justice context: the total effect of early-stage police contact (such as being searched for contraband) on educational and socio-psychological outcomes can be decomposed into path-specific effects via subsequent arrest and incarceration (Weaver and Lerman, 2010; Kirk and Sampson, 2013; Sugie and Turney, 2017).

Finally, although in this paper I have considered a decomposition of the average treatment effect for the case of binary monotonic mediators, as shown in Supplementary Material G, the framework could straightforwardly be extended to accommodate categorical transitions. Given the heterogeneity of higher-educational trajectories in the US, such an extension would prove useful for modeling the relative payoffs to distinct educational pathways.

SUPPLEMENTARY MATERIAL

Supplementary Materials (To Appear Online): This document (PDF) contains: (A) details about a parametric, regression-with-residuals estimation procedure; (B) a simulation study, (C) a sensitivity analysis, (D) further information on sample construction and the data used, (E-F) results with alternative sample restrictions and definitions of earnings; (G) a discussion of extending the decomposition to multivalued, discrete mediators; (H) a description of the EIFs used in the empirical illustration, and (I) proofs and technical details.

References

Jeffrey M Albert and Suchitra Nelson. Generalized causal mediation analysis. *Biometrics*, 67:1028–1038, 2011. ISSN 0006-341X.

Joshua D Angrist and Stacey H Chen. Schooling and the vietnam-era gi bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics*, 3:96–118, 2011. ISSN 1945-7782.

Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106:979–1014, 1991. ISSN 1531-4650.

Joshua D Angrist and Alan B Krueger. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association*, 87:328–336, 1992. ISSN 0162-1459.

Orley Ashenfelter and David J Zimmerman. Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers. *Review of Economics and Statistics*, 79:1–9, 1997. ISSN 0034-6535.

Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. 2005.

Kelli A Bird, Benjamin L Castleman, Jeffrey T Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger. Nudging at scale: Experimental evidence from fafsa completion campaigns. *Journal of Economic Behavior Organization*, 183:105–128, 2021. ISSN 0167-2681.

Sandra E Black, Jeffrey T Denning, Lisa J Dettling, Sarena Goodman, and Lesley J Turner. Taking it to the limit: Effects of increased student loan availability on attainment, earn-

ings, and financial well-being. *American Economic Review*, 113:3357–3400, 2023. ISSN 0002-8282.

Zachary Bleemer. Affirmative action, mismatch, and economic mobility after california's proposition 209. *The Quarterly Journal of Economics*, 137:115–160, 2022. ISSN 0033-5533.

Hugo Bodory, Martin Huber, and Lukáš Lafférs. Evaluating (weighted) dynamic treatment effects by double machine learning. *The Econometrics Journal*, 25:628–648, 2022. ISSN 1368-4221.

Jennie E Brand and Yu Xie. Who benefits most from college? evidence for negative selection in heterogeneous economic returns to higher education. *American sociological review*, 75:273–302, 2010. ISSN 0003-1224.

David Card. Earnings, schooling, and ability revisited. 1994.

David Card. *The causal effect of education on earnings*, volume 3, pages 1801–1863. Elsevier, 1999. ISBN 1573-4463.

David Card. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69:1127–1160, 2001. ISSN 0012-9682.

Pedro Carneiro, James J Heckman, and Edward J Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101:2754–2781, 2011. ISSN 0002-8282.

Benjamin L Castleman, Denise Deutschlander, and Gabrielle Lohner. Pushing college advising forward: Experimental evidence on intensive advising and college success. *EdWorkingPapers. com*, 2020.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107:261–265, 2017. ISSN 0002-8282.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

Raj Chetty, David J Deming, and John N Friedman. Diversifying society's leaders? the causal effects of admission to highly selective private colleges, 2023.

Christina Ciocca-Eller. What makes a quality college? re-examining the equalizing potential of higher education in the united states. 2023. ISSN 0002-9602.

Sarah R Cohodes and Joshua S Goodman. Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6:251–285, 2014. ISSN 1945-7782.

Rhian M Daniel, Bianca L De Stavola, Simon N Cousens, and Stijn Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71:1–14, 2015. ISSN 0006-341X.

Susan Dynarski, C J Libassi, Katherine Michelmore, and Stephanie Owen. Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111:1721–1756, 2021. ISSN 0002-8282.

Christina Ciocca Eller and Thomas A DiPrete. The paradox of persistence: Explaining the black-white gap in bachelor's degree completion. *American Sociological Review*, 83: 1171–1214, 2018. ISSN 0003-1224.

Helmut Farbmacher, Martin Huber, Lukáš Lafférs, Henrika Langen, and Martin Spindler. Causal mediation analysis with double machine learning. *The Econometrics Journal*, 25: 277–300, 2022. ISSN 1368-4221.

Joshua Goodman, Michael Hurwitz, and Jonathan Smith. Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35:829–867, 2017. ISSN 0734-306X.

James J Heckman, John Eric Humphries, and Gregory Veramendi. Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126:S197–S246, 2018. ISSN 0022-3808.

Michael Hout. Social and economic returns to college education in the united states. *Annual review of sociology*, 38:379–400, 2012.

Michael Hurwitz and Jessica Howell. Estimating causal impacts of school counselors with regression discontinuity designs. *Journal of Counseling Development*, 92:316–327, 2014. ISSN 0748-9633.

Thomas J Kane and Cecilia E Rouse. Labor market returns to two-and four-year colleges: is a credit a credit and do degrees matter? 1993.

Edward H Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022.

Catherine Kenney. Cohabiting couple, filing jointly? resource pooling and us poverty policies. *Family Relations*, 53:237–247, 2004. ISSN 0197-6664.

David S Kirk and Robert J Sampson. Juvenile arrest and collateral educational damage in the transition to adulthood. *Sociology of education*, 86:36–62, 2013. ISSN 0038-0407.

Greg Lewis and Vasilis Syrgkanis. Double/debiased machine learning for dynamic treatment effects via g-estimation. *arXiv preprint arXiv:2002.07285*, 2020.

Sheng-Hsuan Lin and Tyler VanderWeele. Interventional approach for path-specific effects. *Journal of Causal Inference*, 5:20150027, 2017. ISSN 2193-3685.

Robert D Mare. Social background and school continuation decisions. *Journal of the American Statistical Association*, 75:295–305, 1980. ISSN 0162-1459.

Caleb H Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J Tchetgen Tchetgen. Quantifying an adherence path-specific effect of antiretroviral therapy in the nigeria pepfar program. *Journal of the American Statistical Association*, 112:1443–1452, 2017. ISSN 0162-1459.

Caleb H Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J Tchetgen Tchetgen. On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika*, 107:159–172, 2020. ISSN 0006-3444.

Jack Mountjoy. Community colleges and upward mobility. *American Economic Review*, 112:2580–2630, 2022. ISSN 0002-8282.

Jack Mountjoy and Brent R Hickman. The returns to college (s): Relative value-added and match effects in higher education, 2021.

Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3, 2009. ISSN 19357516. doi: 10.1214/09-SS057.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical modelling*, 7:1393–1512, 1986. ISSN 0270-0255.

Andrea Rotnitzky, James Robins, and Lucia Babino. On the multiply robust estimation of the mean of the g-functional. *arXiv preprint arXiv:1705.08582*, 2017.

Judith Scott-Clayton and Qiao Wen. Estimating returns to college attainment: Comparing survey and state administrative data–based estimates. *Evaluation Review*, 43:266–306, 2019. ISSN 0193-841X.

Jonathan Smith, Joshua Goodman, and Michael Hurwitz. The economic impact of access to public four-year colleges, 2020.

Thomas D Snyder, Cristobal de Brey, and Sally A Dillow. Digest of education statistics 2014, 50th edition. nces 2016-006, 2016.

Johan Steen, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt. Flexible mediation analysis with multiple mediators. *American journal of epidemiology*, 186:184–193, 2017. ISSN 0002-9262.

Naomi F Sugie and Kristin Turney. Beyond incarceration: Criminal justice contact and mental health. *American Sociological Review*, 82:719–743, 2017. ISSN 0003-1224.

Zachary Sullivan, Benjamin L Castleman, and Eric Bettinger. College advising at a national scale: Experimental evidence from the collegepoint initiative. 2019.

Megan M Sweeney and Julie A Phillips. Understanding racial differences in marital disruption: Recent trends and explanations. *Journal of Marriage and Family*, 66:639–650, 2004. ISSN 0022-2445.

Lesley J Turner and Oded Gurantz. Experimental estimates of college coaching on post-secondary re-enrollment, 2024.

Tyler VanderWeele and Stijn Vansteelandt. Mediation analysis with multiple mediators. *Epidemiologic methods*, 2:95–115, 2014. ISSN 2194-9263.

Stijn Vansteelandt and Rhian M Daniel. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, 28:258–265, 2017. ISSN 1044-3983.

Davide Viviano and Jelena Bradic. Dynamic covariate balancing: estimating treatment effects over time. *arXiv preprint arXiv:2103.01280*, 2021.

Vesla M Weaver and Amy E Lerman. Political consequences of the carceral state. *American Political Science Review*, 104:817–833, 2010. ISSN 1537-5943.

Xiang Zhou. Attendance, completion, and heterogeneous returns to college: A causal mediation approach. *Sociological Methods Research*, page 00491241221113876, 2022a. ISSN 0049-1241.

Xiang Zhou. Semiparametric estimation for causal mediation analysis with multiple causally ordered mediators. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84:794–821, 2022b. ISSN 1369-7412.

Xiang Zhou and Guanghui Pan. Higher education and the black-white earnings gap. *American Sociological Review*, 88:154–188, 2023. ISSN 0003-1224.

Seth D Zimmerman. The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32:711–754, 2014. ISSN 0734-306X.

Supplemental Materials (to appear online)

A Parametric, regression-with-residuals (RWR) estimation

In this section, I propose a linear regression-with-residuals (RWR) approach for the MPSE decomposition. The approach relies on two steps. The first involves residualizing pre-treatment confounders with respect to their marginal means, and intermediate on causally prior confounders, ie., $X^{\perp} \triangleq X - \mathbb{E}[X]$, and $Z_k^{\perp} \triangleq M_{k-1}\big[Z_k - \mathbb{E}[Z_k \mid X, \overline{Z}_{k-1}, M_{k-1} = 1]\big]$ for all $k \in [K]$, $M_0 \triangleq A$. For now, we are agnostic about the functional form used for $\mathbb{E}[Z_k \mid X, \overline{Z}_{k-1}, M_{k-1} = 1]$. The second step involves fitting three sets of models. The first is simply a model for the outcome given pre-treatment covariates and the treatment, namely,

$$\mathbb{E}[Y \mid X, A] = \lambda_0 + \lambda_1 A + \alpha_1^T X^{\perp} + \alpha_2^T A X^{\perp}; \tag{10}$$

The second is a set of models for the outcome given covariates, the treatment and M_k for all $k \in [K]$, i.e.,

$$\begin{aligned}\mathbb{E}[Y|X, \overline{Z}_k, A, \overline{M}_k] =& \beta_{k,0} + c_{k,0}A + \sum_{j=1}^{k} \beta_{k,j} M_j + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} A X^{\perp} + \sum_{j=1}^{k-1} \eta_{k,j}^T M_j X^{\perp} \\ &+ \sum_{j=1}^{k} \gamma_{k,j}^T Z_j^{\perp} + \sum_{j=1}^{k-1} M_j \sum_{l=1}^{j} \xi_{k,k,l}^{\top} Z_l^{\perp},\end{aligned} \tag{11}$$

while the third is a set of models for each mediator given covariates, the treatment, conditional on the treatment and all prior mediators, i.e., for all $k \in [K-1]$,

$$\mathbb{E}[M_{k+1} \mid X, \overline{Z}_k, \overline{1}_{k+1}] = \theta_{k,0} + \delta_{k,1}^T X^{\perp} + \sum_{j=1}^{k} \delta_{k,j+1}^T Z_j^{\perp}. \tag{12}$$

These models differ from conventional linear regression in that (i) pre-treatment variables are centered around their marginal means, and (ii) post-treatment confounders $Z_k \forall k \in \{1, \dots K\}$ are centered around their conditional means given all antecedent variables. Under Assumptions 2-4 in the main text, and assuming that the outcome and mediators are linear in pre- and post-treatment confounders, the treatment, and prior mediators, and that all necessary interaction terms have been accounted for, then the ATE τ_0 can be obtained from the linear model $\mathbb{E}[Y \mid X, A]$ as λ_1, and coefficients from the models $\mathbb{E}[Y \mid X, A, \overline{Z}_k, \overline{M}_k]$ and $\mathbb{E}[M_{k+1} \mid X, A, \overline{Z}_k, \overline{M}_k]$ yield estimates of the components of the decomposition as follows:

$$\begin{aligned}
\tau_k &= \mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)] = \beta_{k,k}, \forall k \in \{1, \dots, K\}, \\
\Delta_k &= \mathbb{E}[Y(\overline{1}_{k+1}, \underline{0}_{k+2}) - Y(\overline{1}_k, \underline{0}_{k+1})] = \beta_{k+1,k-1}, \forall k \in \{0, \dots, K-1\}, \\
\pi_{k+1} &= \mathbb{E}[M_{k+1}(\overline{1}_{k+1})] = \theta_{k,0}, \forall k \in \{0, \dots, K-1\}.
\end{aligned}$$

I state the RWR estimation procedure formally in the following algorithm:

Algorithm 1. *RWR.*

1. For each of the baseline confounders, compute $\hat{X}^{\perp} = X - \mathbb{P}_n[X]$, where $\mathbb{P}_n[\cdot]$ denotes empirical average.

2. Fit $\hat{\mathbb{E}}[Y \mid X, A]$ using the linear specification shown above; an estimate of τ_0 is given by $\hat{\lambda_1}$.

3. For each set of post-treatment confounders Z_k, $k \in \{1, \dots, K\}$, compute $Z_k^{\perp} = M_{k-1}\big[Z_k - \mathbb{E}[Z_k \mid X, \overline{Z}_{k-1}, M_{k-1} = 1]\big]$ where an overbar denotes a vector of vari-

ables such that $\overline{Z}_k = (Z_1, \ldots, Z_k)$, by fitting a regression of Z_k on X and $\overline{Z}_{k-1}$ among units with $M_{k-1} = 1$ and then calculating the residuals.

4. For each $k \in \{1, \ldots K\}$:

   (a) compute least squares estimates of equations 11 and 12, using estimates of $X^{\perp}$ and $Z_k^{\perp}$.

   (b) compute $\hat{\tau}_k^{\text{RWR}} = \hat{\beta}_{k,k}$, $\hat{\Delta}_{k-1}^{\text{RWR}} = \hat{\beta}_{k,k-1}$, and $\hat{\pi}_k^{\text{RWR}} = \hat{\theta}_{k-1,0}$.

5. Compute the decomposition using $\hat{\tau}_k$, $\hat{\Delta}_k$ and $\hat{\pi}_{k+1}$, and estimating the covariance terms as $\hat{\eta}_k^{\text{RWR}} = \hat{\beta}_{k-1,k} - \hat{\beta}_{k,k-1} - \hat{\beta}_{k,1}\hat{\theta}_{k-1,k}$, and the continuation effects as $\hat{\theta}_k^{\text{RWR}} = (\Pi_{j=1}^{k} \hat{\pi}_j^{\text{RWR}})\hat{\Delta}_k^{\text{RWR}} + (\Pi_{j=1}^{k-1} \hat{\pi}_j^{\text{RWR}})\hat{\eta}_k^{\text{RWR}}$.

Standard errors and confidence intervals can then be obtained via the non-parametric bootstrap, or by using their asymptotic analytic variance. Specifically, let $\hat{\theta}_k^* \triangleq (\hat{\beta}_{k,0}, \hat{\beta}_{k,1}, \theta_{k,0})$ denote a set of parameters. Under the above models, I have that $\hat{\theta}^* = \{\hat{\lambda}_1, \theta_1^*, \ldots \theta_K^*\}$ solves $\mathbb{P}_n[g(O; \hat{\theta}^*)] = 0$, where $g(O; \theta^*)$ is the set of stacked moment conditions with solution $\hat{\theta}^*$. Under standard regularity conditions (Newey and McFadden, 1994), under correct specification of Models 9-11 wherein all residualized quantities are estimated via linear models, the set $\hat{\theta}^*$ is consistent and asymptotically normal, such that $\sqrt{n}\big(\hat{\theta}^* - \theta^*\big)$ converges to a mean-zero normal distribution with finite variance $V = G^{-1}\Omega(G^{-1})^{\top}$, where $\Omega = \mathbb{E}[g(O; \theta^*)g(O; \theta^*)^{\top}]$, and where $G = \mathbb{E}[\frac{\partial g(O;\theta^*)}{\partial \theta^{\top}}]$. It follows by a simple application of the Delta Method that the set $\hat{\gamma_k^*} \triangleq (\hat{\tau}_k^{\text{RWR}}, \hat{\Delta}_{k-1}^{\text{RWR}}, \hat{\pi}_k^{\text{RWR}}, \hat{\eta}_k^{\text{RWR}}, \hat{\theta}_k^{\text{RWR}}) \forall k \in [K]$ is also consistent and asymptotically normal.

B A simulation study

In this section, I evaluate the finite-sample performance of my two estimation procedures via a simulation experiment. Specifically, I compare how the DML estimator proposed in Section 3 (as well as the parametric, RWR estimator described in Appendix A) perform under different degrees of misspecification. Without loss of generality, I focus on the single-mediator setting, and focus on the path-specific effect $A \to M \to Y$. I generate simulations of observed data $O = (X_1, X_2, A, Z, M, Y)$ as follows:

$$
\begin{aligned}
U_1, U_2, U_3, U_4 &\sim \text{MVN}\big(0_4, I_4\big) \\
X_1 &\sim \text{N}\big((U_1, U_2, U_3, U_4)\beta_{X_1}, 1\big) \\
X_2 &\sim \text{N}\big((U_1, U_2, U_3, U_4)\beta_{X_2}, 1\big) \\
A &\sim \text{Bern}\big(\text{logit}^{-1}\big[(1, X)\beta_A\big]\big) \\
Z|A = 1 &\sim \text{Bern}\big(\text{logit}^{-1}\big[(1, X)\beta_Z\big]\big) \\
M|A = 1 \sim \text{Bern}\big(\text{logit}^{-1}\big[(1, X, Z)\beta_M\big]\big) \quad & M|A = 0 = 0 \\
Y &\sim \text{N}\big((1, A, X, AZ, AM)\beta_Y, 1\big).
\end{aligned}
$$

The coefficients $(\beta_{X_1}, \beta_{X_2}, \beta_Y)$ are drawn from a $\text{Unif}\big(-1, 1\big)$, while the coefficient β_A is drawn from a $\text{Unif}\big(-0.5, 0.5\big)$ distribution. In order to test how the DML and RWR methods perform when the relevant models are misspecified, I also construct transformations of the observed covariates (X_i^*, Z_i^*) as follows, employing a similar setup to Kang and Schafer (2007):

$$
\begin{aligned}
X_1^* &= (\exp(X_1/2) - 1)^2 \\
X_1^* &= X_2/(1 + \exp(X_2)) + 10 \\
Z^*|A = 1 &= (X_1 \cdot Z/25 + 0.6)^3
\end{aligned}
$$

For each simulated dataset, I construct two estimates of the path-specific effect θ_1 $(A \to$

$M \to Y$) by estimating the parameter set $(\tau_0, \Delta_0, \Delta_1, \pi_1)$ via the RWR and DML procedures described in Section 3. Standard errors for the coverage rates are computed via the estimated variance of the estimated EIFs for the DML approach, and via the nonparametric bootstrap with 1000 replications for the RWR procedure. For the DML estimator, for each component involved in the decomposition, I construct a Neyman-orthogonal "signal" using its EIF. The recentered EIFs for each component are shown below:

$$
\begin{aligned}
M^*(1) &= \gamma_1(X) + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(M - \gamma_1(X)), \\
Y^*(a) &= \mu_0(X,a) + \frac{\mathbb{I}(A=a)}{\pi_0(X,a)}(Y - \mu_0(X,a)), \text{ for } a \in \{0,1\} \\
Y^*(1,m_1) &= \nu_1(X,m) + \frac{\mathbb{I}(A=1)\mathbb{I}(M=m)}{\pi_0(X,1)\pi_1(X,m_1)}(Y - \mu_1(X,Z,m)) \\
&\quad + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(\mu_1(X,Z,m) - \nu_1(X,m)), \text{ for } m \in \{0,1\}
\end{aligned}
$$

where

$$
\begin{aligned}
\pi_0(X,a) &\triangleq \Pr[A = a \mid X] \\
\pi_1(X,m_1) &\triangleq \Pr[M = m \mid X, A = 1] \\
\gamma_1(X) &\triangleq \mathbb{E}[M \mid X, A = 1] \\
\mu_0(X,a) &\triangleq \mathbb{E}[Y|X, A = a] \\
\mu_1(X,Z,m) &\triangleq \mathbb{E}[Y|X, A = 1, Z, M = m] \\
\nu_1(X,m) &\triangleq \mathbb{E}[\mu_1(X,Z,m)|X, A = 1].
\end{aligned}
$$

I run 1000 replications of this DGP and compute the average bias, the root mean square error (RMSE), and the coverage of nominal 95% confidence intervals for sample sizes of 250, 500, and 1000 and using either the "correctly specified" covariates (X_1, X_2, Z) and the "incorrectly specified", transformed versions (X_1^*, X_2^*, Z^*). I calculated the true value of θ_1 by recovering the true values of the parameter set $(\tau_0, \Delta_0, \Delta_1, \pi_1)$ in each Monte Carlo

simulation.

Figure 1 presents the results of this simulation experiment. Under correctly specified models, the DML and RWR estimators perform similarly, with each displaying low bias, low RMSE and close to nominal coverage at all sample sizes. Under incorrectly specified models, however, the DML and RWR approaches diverge in performance. Whereas the DML estimator under an incorrect feature space performs similarly to when it is used on the correct feature space, the RWR estimator performs much more poorly, displaying a large amount of bias that in fact grows with the sample size, a large RMSE, and coverage rates that are not close to nominal. In short, when the models are correctly specified, both the parametric and semiparametric approaches perform well; the strong performance of a semiparametric approach compared with a parametric estimation strategy becomes clear under model misspecification.

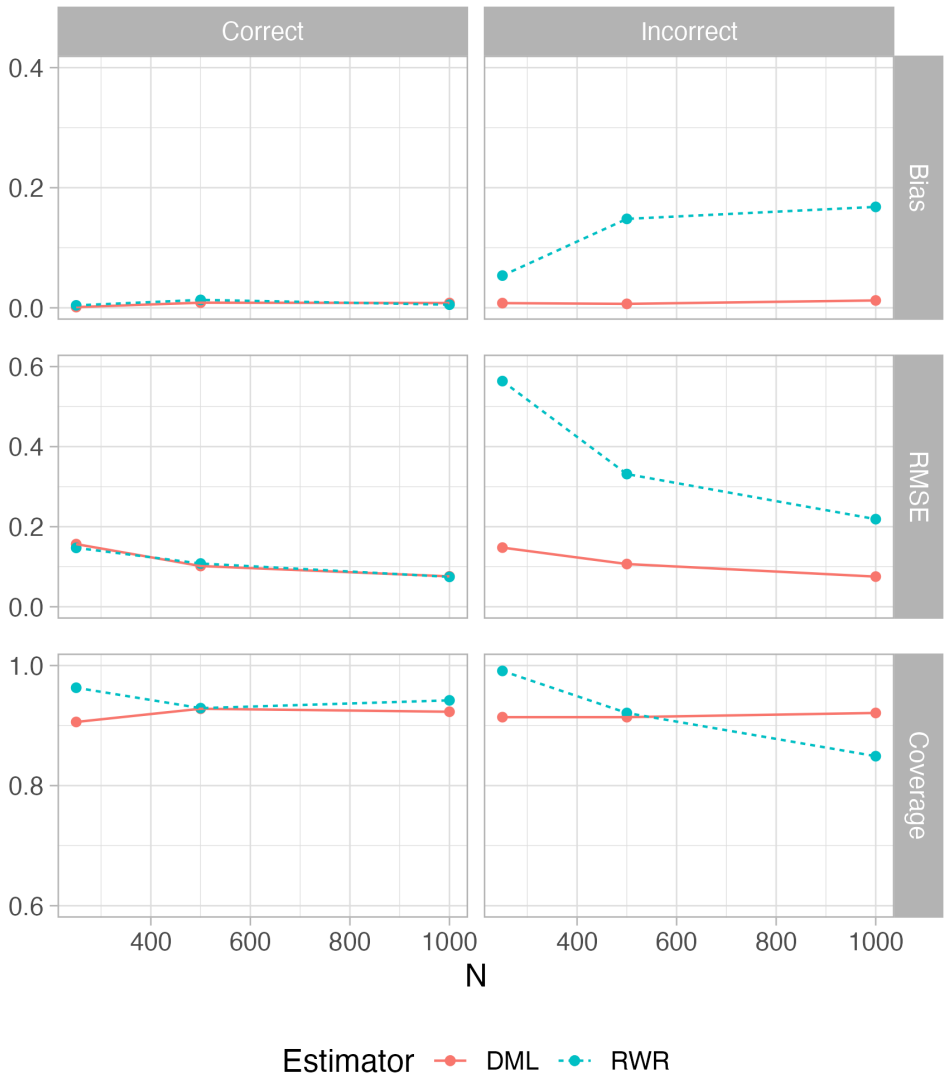

Figure 1: Bias, RMSE, and coverage of DML and RWR estimators for $n = 250, 500, 1000$. The left panel shows the performance of the DML and RWR estimators when the correct feature matrix is supplied to the estimators; the right panel shows the performance of the two estimators when an incorrect feature matrix is supplied to the estimators, as described in the main text.

C Sensitivity analysis

How do my estimates tally with previous findings on the labor market returns to education? While previous work does not estimate quantities analogous to the direct and indirect effects of interest (i.e., the θ_k terms), some prior educational returns estimates are closely related to the net effect (τ_k) terms that inform the total, direct and indirect components of the decomposition. My estimate of the net effect of 4-year college enrollment (τ_1) is large, at 54%, but not implausibly so. While Zimmerman (2014) and Smith et al. (2020) recover college earnings returns at around 20by age 30 (exploiting admissions discontinuities in the Florida and Georgia state university systems), both of these studies estimate the earnings premium from attending a less selective 4-year college rather than a community college, for the marginally qualified university attendee. By contrast, τ_1 captures the effect of 4-year college enrollment compared with community college *and* no college enrollment, pooling across the less selective colleges examined in Zimmerman (2014) and Smith et al. (2020) as well as more selective colleges which could have greater earnings effects. Moreover, since τ_1 represents an effect averaged over all individuals, it reflects college return among a broader population than the marginal college-goers in previous studies, for whom economic theories of "positive selection" suggest the effect of college attendance is lowest (Willis and Rosen, 1979; Carneiro et al., 2011).[6] One additional point of comparison are instrumental variable (IV) estimates of returns to years of schooling, which are typically estimated in the range of 6 to 12 (Angrist

[6]My estimate of Δ_1 (the direct effect of 4-year college attendance on earnings) further tallies with a similar quantity estimated by Scott-Clayton and Wen (2019). On the intensive margin of employment (i.e. dropping respondents with zero observed earnings), the authors estimate a return to college attendance without degree completion of 0.21. While, theoretically, one might expect my estimate - which corresponds to the extensive employment margin (including respondents with zero observed earnings - to be larger, the fact that it is slightly smaller could reflect several factors, including the richer array of pre-college controls I use in my models, model mis-specification resulting from linearities imposed in prior work and, perhaps most importantly, collider-stratification biases induced by conditioning on BA completion in Scott-Clayton and Wen (2019)'s models (biases that are likely reduced by the inclusion of time-varying controls).

and Krueger, 1991, 1992; Kane and Rouse, 1993; Card, 1994; Ashenfelter and Zimmerman, 1997; Angrist and Chen, 2011) - although one estimate puts the return for women at 18.5% (Butcher and Case, 1994). While my estimate of the overall return to high school graduation (τ_0) could appear large in this light, several factors could explain this discrepancy. In particular, IV estimates of schooling returns capture direct effects of schooling, whereas τ_0 captures both the direct effect of schooling as well as its continuation value via subsequent education. Moreover, the ATE τ_0 captures the effect of multiple years of schooling: the high school graduates and high school non-completers that form the comparison groups differ by multiple years of schooling: in the NLSY97, high school non-completers attained on average 3.7 fewer years of schooling than high school completers.

Of course, an alternative explanation is that my estimates are upwardly biased by a large degree of unobserved confounding. While the sequential ignorability assumption facilitates identification of educational effect pathways under a weaker set of conditions than might be typically invoked in mediation settings, it is still strong and fundamentally unverifiable. To assess potential bias of the estimated MPSEs due to unobserved confounders not picked up in my covariate set $(X, \overline{Z}_K)$, I propose a sensitivity analysis for each of the MPSEs.

Assume first that we have a binary unobserved confounder, U, for the treatment-outcome relationship. Assuming that $\alpha_0 = \mathbb{E}[Y|x, a, U = 1] - \mathbb{E}[Y|x, a, U = 0]$ does not depend on x or a, and further that $\beta_0 = \Pr[U = 1|x, A = 1] - \Pr[U = 1|x, A = 0]$ does depend on x, for $\tau_0 = \mathbb{E}[Y(1) - Y(0)] \triangleq$ ATE, then bias$(\tau_0) = \alpha\beta$ (VanderWeele and Arah, 2011),

Next, consider an unobserved binary confounder, U_k that affects both M_k and Y for any $k \in \{1, \ldots, K\}$. Then, under a weaker instantiation of Assumption 3 (Sequential Ignorability), i.e.,

$$Y(\overline{1}_k, m_k) \perp\!\!\!\perp (A, \overline{M}_k)|X, A, U_k, \overline{Z}_k, \overline{M}_{k-1} \forall k \in [K], \tag{13}$$

which states that potential outcomes under an arbitrary transition sequence are independent of observed treatment and mediator values conditional on observed confounders $(X, \overline{Z}_k)$

and unobserved confounders U_k. Under the following set of assumptions: (Assumption A_k) $\alpha_k = \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 0]$ does not depend on $(x, \overline{z}_k, \overline{1}_k, m_k)$, and (Assumption B_k), $\beta_k = \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, m_k] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k])$ does not depend on $(x, \overline{z}_k)$, I can show that, for any $k \in \{1, \dots K\}$,

$$\text{bias}(\tau_k) = \alpha_k \beta_k,$$

and, further, that

$$\text{bias}(\Delta_{k-1}) = -\alpha_k \beta_k \pi_k,$$

where $\pi_k = \int_x \int_{\overline{z}_k} \Pr[M_k = 1|x, \overline{z}_k, \overline{1}_k] \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})] dP(x)$, and is estimable from observed using the estimation strategies described previously. A contour plot showing bias-adjusted estimates of Δ_{k^*} and τ_k then enables assessment of how strong the unobserved confounder would need to be to reduce estimates of the direct and gross effects to zero. I illustrate these techniques in my empirical illustration below.

In order to assess the robustness of my empirical findings to potential violations of Assumption 3 (Sequential Ignorability), I implement my sensitivity analysis as discussed in Section C.

Figure 2 below displays a set of contour plots, which capture the bias-corrected estimates of the (Δ_k, τ_k) terms under varying degrees of confounding (that is, under different values of α_k and β_k). For example, the level set marked "0" corresponds to values of (α_k, β_k) required in order for the unobserved confounder to fully "explain away" estimates (Δ_k, τ_k) (i.e., to reduce their true values to zero). Importantly, each row corresponds to a different set of (α_k, β_k) terms for a given U, such that the top row corresponds to (α_0, β_0), while the second row corresponds to (α_1, β_1), and so on.

For simplicity, I consider U to be an unmeasured binary confounder that is (marginally) positively associated with each transition $A, M_1, \dots M_3$ as well as with adult earnings Y.

To benchmark the hypothetical behavior of U, for each plot, I also display the values of (α_k, β_k) that would correspond to a U that behaved similarly to a given confounder that I do observe in the data: an indicator for whether an individual's test score on the ASVAB is above the median. In each plot, I mark this point and label it "Ability". For each plot, I also mark the point on the zero contour that corresponds to $\alpha_k = \beta_k$ (i.e, the point at which the unobserved confounder's associations with the treatment and with the outcome are equal, and reduce the true value of the parameter to zero).

I focus on estimates of τ_0 and Δ_0 to assess how robust my primary conclusion - that the ATE of high school completion is overwhelmingly mediated by high school's direct effect on earnings - is to unobserved confounding. Bias-adjusted estimates of the ATE τ_0 are presented in the top row of Figure 2. Since U is assumed to be positively associated with both A and Y, τ_0 is overestimated and suffers from a bias of $\alpha_0\beta_0$. My estimate of τ_0 at 0.67 is nevertheless quite robust: if U had similar effects to ability, the effect would be reduced by 0.06 log points, to 0.61, still implying a high earnings premium to high school completion overall in excess of 84%.

How do my estimates of the direct effect of high school completion Δ_0 (and, in particular, about the proportion of the total effect that is direct) fare under unobserved confounding? The second row of Figure 2 considers bias-adjusted estimates of $\Delta_0 = \theta_0$ under different values of (α_1, β_1), which correspond to the effects of an unobserved confounder U (marginally) positively associated with both M_1 and with Y. As described above, in this scenario, Δ_0 is affected by a bias of $-\alpha_1\beta_1\pi_1$. Importantly, even if the unobserved confounder U is *marginally* positively associated with high school graduation (A), the conditional association between U and A may be zero or even negative since M_1 is a collider of A and U. In the case that the conditional association between U and A is negative, $-\alpha_1\beta_1\pi_1$ would be positive, implying an overestimation of the direct effect θ_0. On the plot, I show estimates of U if it behaved similarly to the ability variable. Indeed, despite the fact that ability is marginally positively associated with high school completion (top row of Figure 2), its conditional association -

conditional on college attendance - is depressed to zero. Thus, it would take an extreme form of confounding for θ_0 to be largely different from its estimated value of .46. In this way, my primary finding that the ATE of high school graduation is overwhelmingly mediated via its direct effect remains highly robust to patterns of unobserved confounding, under my set of simplifying assumptions.

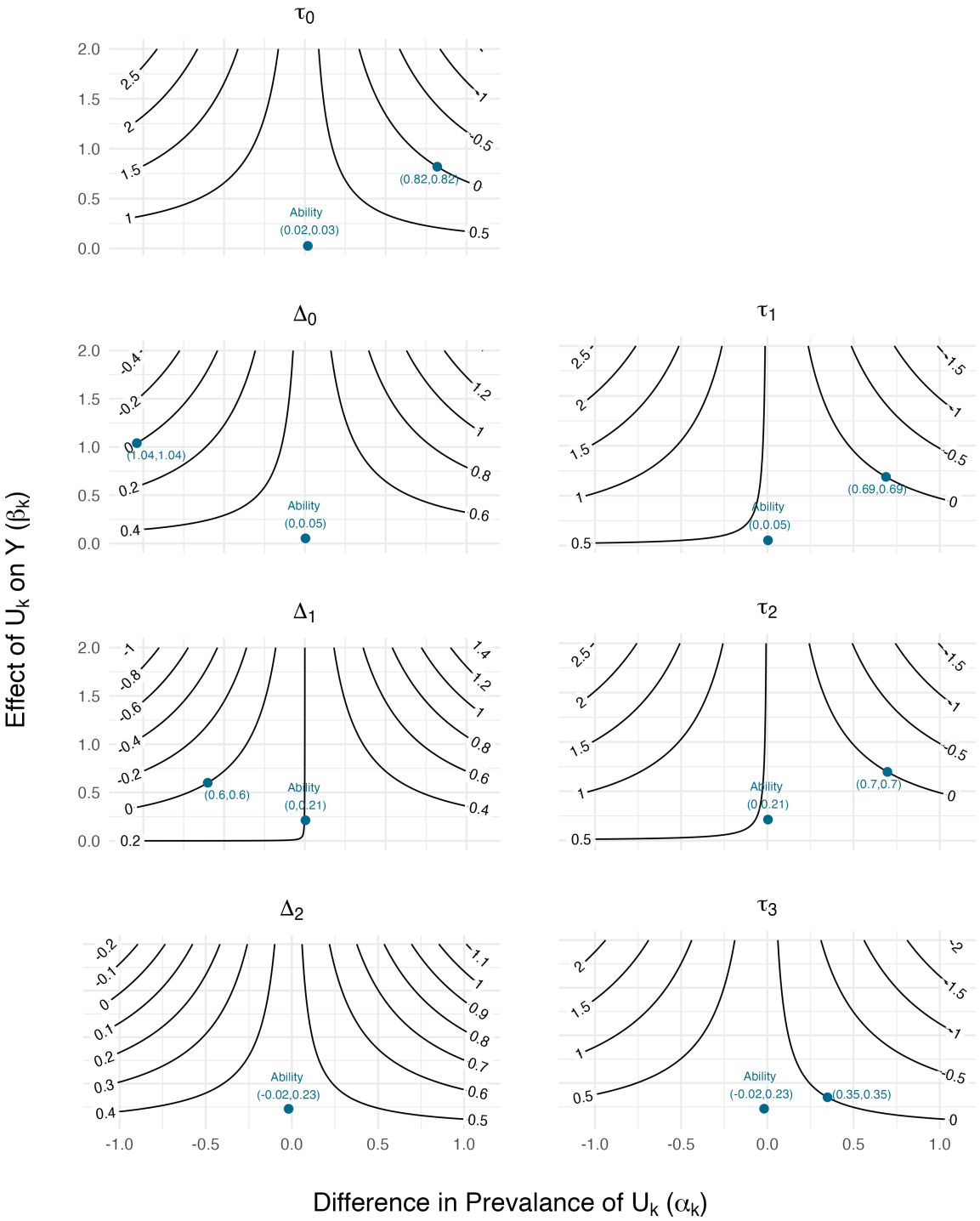

Figure 2: Sensitivity Analysis for the "gross effect" (τ_k) and "direct effect" (Δ_k) terms in decomposition. Each row corresponds to a different set of (α_k, β_k) terms, where $\alpha_k = \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 0]$ parameterizes the effect of U_k on Y, and $\beta_k = \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, m_k] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k])$ parameterizes the effect of M_k on U_k. Each row corresponds to a different set of (α_k, β_k) terms. For example, the top row corresponds to (α_0, β_0), while the second row corresponds to (α_1, β_1), and so on.

D Further details on variable construction and education groups

Variable construction

In an effort to satisfy the sequential ignorability assumption (Assumption 3), I include a large array of covariates in my models for the effects of completing educational transitions on labor market outcomes. Figure 3 summarizes my assumed data-generating process for the empirical example. In addition to including information on respondent demographics (gender, race, ethnicity, age at 1997), and observed pre-college performance such as overall high school GPA and test score on the Armed Services Vocational Aptitude Battery (ASVAB), I include detailed information on socioeconomic background (parental education, parental income, parental asset, co-residence with both biological parents, presence of a paternal figure, rural residence, southern residence), an index of substance use, an index of delinquency, whether the respondent had any children by age 18), and peer and school-level characteristics (measures of peers' college expectations and behaviors). Both parental income and parental asset variables are transformed to 2023 dollars.

Since my proposed decomposition also facilitates the inclusion of a distinct set of observed intermediate confounders for each transition to adjust for selection processes that may confound the causal effects of each transition on earnings (i.e., the $A - Y$ and $M_k - Y$ relationships, for $k \in \{1, \dots K\}$, I include two postsecondary characteristics (Z) to adjust for confounders of the effect of BA completion and graduate school attendance on earnings, namely, field of study, and college GPA. Specifically, I use college self-reported major field of study, drawing on the NLSY survey instrument asking respondents about their choice of major in each month in which they were enrolled in college, and using a dummy variable to denote whether whether a respondent majored in a STEM or non-STEM field by age 29. Finally, college GPA is measured using the respondent's cumulative GPA from the

Post-Secondary Transcript Study. I treat two of the Z_k sets as empty (namely, Z_1 and Z_3), assuming that the effects of the first mediator (college attendance, M_1) on subsequent transitions and adult earnings are unconfounded given background characteristics (X), and that the effects of the third mediator (graduate school attendance, M_3) are unconfounded given background characteristics (X) and postsecondary characteristics (Z). [7]

How convincingly do I satisfy the sequential ignorability assumption? Despite the inclusion of a comprehensive set of background covariates in my models, it is possible that observed variables do not perfectly proxy for all important confounders jointly affecting education and earnings. In particular, researchers often argue that important variables, such as students' innate ability, ambition, and detailed forms of socioeconomic advantage, confound observational estimates of educational returns (e.g. Carneiro et al., 2011). While some research suggests that observational estimates of earnings returns may well capture actual returns to education and that the degree of observational bias may be rather small (Card, 1999), it is of course impossible to quantify the true extent of the bias in the estimates I produce. The sensitivity analysis described above provides a step towards this goal.

I note that my assumption of ignorability of M_3 without conditioning on intermediate variables Z_3 is perhaps the strongest assumption I make. For example, many individuals take time off to work before enrolling in graduate school, and labor market experience and earnings gained in the interim period between college completion and graduate school enrollment may confound the latter variable's effects on earnings. Nevertheless, including a measure of labor market characteristics for this period is difficult because some respondents enroll directly in graduate school after BA completion, such that pre-graduate school earnings variables would be undefined for these individuals.

[7] To be clear, assuming that Z_1 and Z_3 are empty is not to say thatM_1 and M_3 are marginally unconfounded; rather, it means that the set of covariates that confound the effects of M_1 is assumed to be the same as those that confound the effects of A, and that the set of covariates that confound the effects of M_3 are assumed to be the same as those that confound the effects of M_2. This assumption is in part data-driven, given the few variables observed chronologically post high school graduation and pre college attendance.

Table 1 shows conditional means of respondent attributes X and Z for the full (imputed) and restricted (non-imputed) samples, showing first the mean among the full population of high school goers, and progressively restricting the sample from (i) high school (HS) non-completers, to (ii) HS graduates, to (iii) college attendees and, finally, to (iv) BA completers. Imputed and non-imputed means - shown without and with brackets, respectively - are highly similar across variables. As I progressively restrict the sample to those who attained higher educational levels, variables capturing components of socioeconomic advantage (such as parental income, parental education and household net worth) increase monotonically in value. Background covariates measuring aspects of the school environment (such as peers' college expectations - which is an indicator for whether over 90 of a respondent's peers expected to go to college) behave similarly. I also see that students who progress to higher educational levels have higher levels of pre-college ability: HS non-completers have on average an ASVAB Percentile score of 22.3, compared with only in excess of 70 among BA completers. Similarly, college-goers average high school GPA is approximately .5 higher than high school graduates overall (regardless of whether or not they proceed to college). Nevertheless, the association between high school GPA and attainment declines at higher educational levels: BA completers have only on average a .11 higher a high school GPA than the pooled group of college goers, irrespective of their BA completion status. At this stage, college GPA appears to matter more: college goers overall have on average a college GPA of 2.77, while BA completers' average college GPA is 3.07.

Educational groups: raw mean earnings

Table 2 (column 2) presents the proportion of individuals who have attained each level of education constructed above. By age 22, a small, but not insignificant, proportion of individuals who enroll in high school do not complete their studies (13), and by this same age, just over 40% of individuals have attended a 4-year college. By age 29, 29 of individuals have attained a Bachelor's degree or higher. These estimates of high school completion and

BA completion align closely both with those reported in previous studies that employ the NLSY97 (e.g. Scott-Clayton and Wen (2019), as well as with those reported in the Current Population Survey (CPS). Table 2 (columns 3-4) also shows mean log earnings by educational group (column 3), alongside the estimated gap between these means and mean log earnings among high school non-completers (column 4). High school dropouts earn an average of 9.07 log earnings, while groups with higher levels of attainment earn successively more than high school dropouts, though at a decreasing rate. High school graduates earn on average 1.11 log earnings more than high school non-completers, implying an earnings premium in excess of 200% ($\exp(1.11) - 1$), while college goers earn on average 1.5 log earnings more than high school non-completers (or 0.6 log earnings more than high school graduates). At the highest end, graduate school goers earn on average 10.88 log earnings. These educational premia are extremely high, since they reflect both the causal effect of a given educational level as well as the effects of individual, geographic and family factors correlated both with attainment and with adult earnings. To net out these patterns of selection, we need to turn to estimates of the MPSE decomposition, as well as its constituent components.

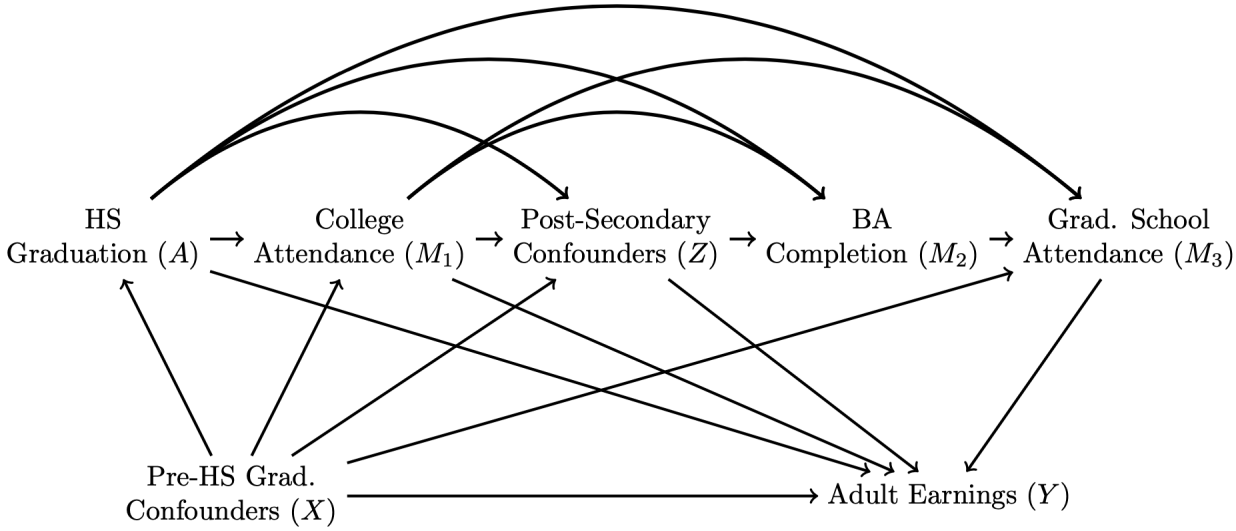

Figure 3: DAG showing the hypothesized causal relationships between high school completion A and adult earnings Y via mediators M_1, M_2 and M_3.

Table 1: Conditional means in background and college-level attributes by sample type (imputed dataset and non-imputed dataset).

	Full Population	HS Non-Completers	HS Graduates	College Goers	BA Completers
Female	0.50 (0.50)	0.44 (0.47)	0.51 (0.51)	0.55 (0.56)	0.57 (0.58)
Black	0.17 (0.16)	0.25 (0.24)	0.16 (0.15)	0.13 (0.10)	0.10 (0.09)
Hispanic	0.14 (0.12)	0.20 (0.18)	0.13 (0.12)	0.09 (0.08)	0.08 (0.07)
Parental Income	80,235 (79,243)	43,308 (43,863)	85,898 (83,865)	110,240 (109,636)	118,026 (114,329)
Parental Education	12.85 (12.84)	11.18 (11.24)	13.11 (13.04)	14.06 (14.13)	14.38 (14.34)
Household Net Worth	175,513 (176,097)	61,997 (62,318)	192,920 (190,959)	272,606 (284,029)	306,177 (307,101)
Lived w Biological Parents	0.52 (0.51)	0.30 (0.28)	0.55 (0.55)	0.67 (0.69)	0.72 (0.73)
Father Figure Present	0.75 (0.75)	0.61 (0.61)	0.77 (0.76)	0.84 (0.83)	0.86 (0.86)
Lived in Rural Area	0.28 (0.30)	0.25 (0.27)	0.28 (0.31)	0.28 (0.32)	0.28 (0.31)
Lived in South	0.36 (0.35)	0.44 (0.42)	0.35 (0.34)	0.33 (0.30)	0.31 (0.29)
Children by 18	0.07 (0.07)	0.21 (0.24)	0.05 (0.05)	0.01 (0.01)	0.01 (0.01)
Substance Abuse Score	1.09 (1.10)	1.32 (1.37)	1.06 (1.07)	0.86 (0.82)	0.82 (0.77)
Delinquency Score	1.38 (1.39)	2.07 (2.13)	1.27 (1.29)	0.91 (0.88)	0.81 (0.79)
Peers' College Expectations	0.56 (0.56)	0.40 (0.38)	0.59 (0.58)	0.69 (0.69)	0.72 (0.71)
Property Stolen at School	0.24 (0.23)	0.28 (0.29)	0.23 (0.22)	0.21 (0.20)	0.20 (0.19)
Threatened at School	0.22 (0.23)	0.30 (0.33)	0.20 (0.21)	0.14 (0.13)	0.13 (0.11)
In a Fight at School	0.16 (0.16)	0.32 (0.34)	0.14 (0.13)	0.07 (0.07)	0.06 (0.06)
ASVAB Percentile	48.36 (48.52)	22.27 (21.77)	52.37 (52.02)	67.22 (69.66)	70.66 (71.82)
High School GPA	2.84 (2.81)	2.12 (2.14)	2.95 (2.90)	3.30 (3.33)	3.41 (3.43)
Stem Major				0.17 (0.19)	0.18 (0.19)
College GPA				2.78 (2.86)	3.07 (3.12)
Earnings ($)	46,505 (45,151)	21,597 (21,566)	46,505 (45,151)	64,984 (64,595)	72,374 (70,212)
Log (Earnings + c)	10.03 (10.06)	9.07 (9.14)	10.03 (10.06)	10.57 (10.65)	10.73 (10.78)

Note: Numbers denote means for the imputed sample (non-imputed sample). Means are adjusted for multiple imputation via Rubin's (1987) method, and all statistics are calculated using NLSY97 sampling weights.

Table 2: Means of observed log earnings by educational participation, and earnings gaps (versus high school non-completers).

Group	Population Proportion	Log Earnings	Gap (vs HS Non-Completers)
HS Non-Completers	0.13	9.07 (0.05)	
HS Graduates	0.87	10.18 (0.02)	1.11 (0.05)
College Goers	0.41	10.57 (0.03)	1.5 (0.05)
BA Completers	0.29	10.73 (0.03)	1.66 (0.06)
Grad. School Goers	0.09	10.88 (0.05)	1.81 (0.07)

Note: The category "High School Non-Completers" captures all individuals who attended high school but did not obtain a high school diploma; "High School Graduates" refers to those individuals who graduated high school, regardless of their subsequent educational experiences (i.e., whether or not they proceeded to college); "College Goers" refers to individuals who attended a 4-year college, irrespective of whether they completed their degree; "BA Completers" denotes individuals who completed a Bachelor's degree, while "Grad. School Goers" captures individuals who participated in a graduate-level degree program. A small constant of $\$1,000$ is added to observed earnings before taking the log. All statistics are computed with a monotonicity assumption imposed on the observed data (i.e. such that all individuals who complete a given educational level are coded as having completed all prior levels). All statistics are calculated using NLSY97 sampling weights, and standard errors are in parentheses.

E Results without imputation of missing covariates

In the main text, I report estimates of the MPSE decomposition for the ATE of high school attendance on logged annual earnings for the full sample of NLSY97 respondents with non-missing educational information and non-missing earnings ($N = 7,305$). A very large number (approximately 50%) of these respondents are missing information on one or more of the covariates (X, Z) used in the models in order to identify the decomposition components, due especially to the large amount of missingness on variables such as parental assets (25% of the sample), and the Armed Services Vocational Aptitude Battery (ASVAB) test (20% of the sample). To assess the sensitivity of my primary conclusions to the use of multiple imputation as opposed to dropping observations with missing values, I replicate my DML and RWR estimates on a non-imputed analytic sample ($N = 3,735$). Figure 4 below shows the results of this exercise. For both estimation procedures, results under multiple imputation and non-imputation are highly similar. As is to be expected, imputation reduces standard errors significantly, especially for the parametric RWR procedure. Further, the greatest variability between imputed and non-imputed results come from effects pertaining to high school completion, perhaps because patterns of missingness are correlated with educational attainment. Despite this, because the total effect τ_0 and direct effect θ_0 are similarly attenuated in the imputed sample, the overall conclusion about the importance of the direct effect in explaining the ATE remains unaffected.

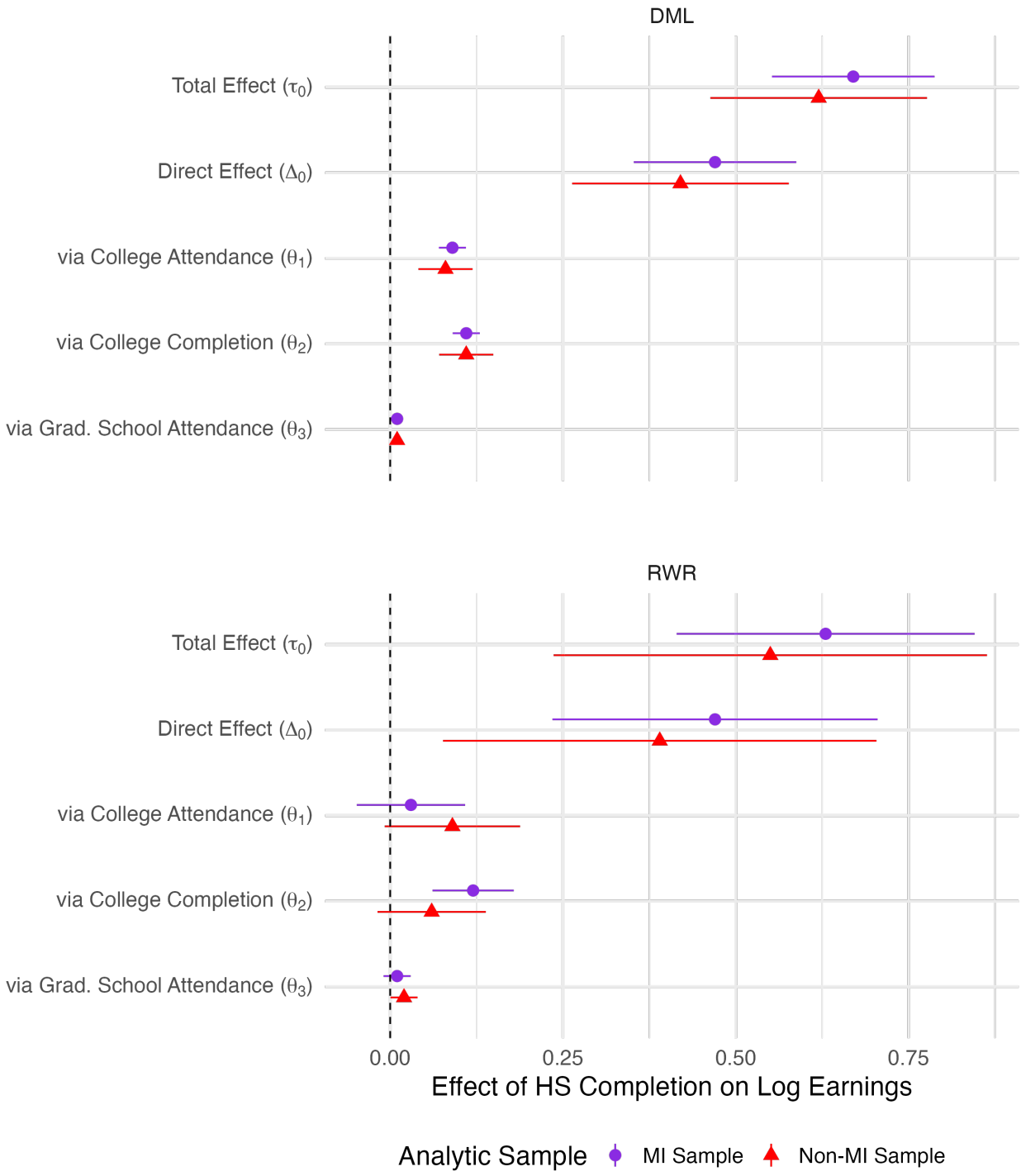

Figure 4: Decomposition of the Average Total Effect (ATE) of High School Graduation on Logged Earnings Under Multiple Imputation (MI) and Under Dropping Observations with Missing (X, Z) Values. Results with multiple imputation (purple lines) are reproduced from the main text ($N = 7,305$); results without multiple imputation (red lines) employ a sample restricted to respondents with observed values for all covariates used ($N = 3,735$).

F Results under alternative definitions of earnings

In the main text, I report estimates of the MPSE decomposition components for the ATE of high school attendance on logged annual earnings. Logged annual earnings in the main text are defined as the log of observed annual earnings plus a small constant of $\$1,000$, in order to accommodate respondents with zero observed annual earnings. In order to assess the sensitivity of the reported results to the choice of this constant, I replicate the main analyses under alternative definitions of earnings. Figure 5 reports estimates of the direct and indirect effects under a series of different constants c added to pre-logged annual earnings, for $c \in \{10, 100, 1000\}$, while Figure 6 shows estimates of these direct and indirect effects for observed annual earnings in dollar values. Beginning with Figure 5, I see that, while for the indirect effects θ_1, θ_2, and θ_3, both DML and RWR estimates are quite consistent under these different constants, estimates of the total effect τ_0 as well as the direct effect Δ_0 are quite sensitive to the choice of constant. Specifically, lower constant values correspond with large increases in the DML estimate of τ_0 from 0.67 ($c = 1000$) to 1.20 ($c = 10$), and of Δ_0 from 0.47 ($c = 1000$) to 0.89 ($c = 10$). This is because individuals with less than a high school degree are more likely than their higher-educated counterparts to have zero or low earnings, making their logged earnings rather sensitive to the choice of constant. Nevertheless, because the total effect τ_0 and direct effect θ_0 are similarly affected by the change in constant value, the importance of the direct effect in explaining the ATE of high school completion is reinforced. In particular, the proportion of the total effect that is direct is estimated to be 70%, 74% and 72% under each of $c = 10, 100, 1000$. Turning next to Figure 5, under DML, estimates of high school completion increases earnings in expectation by roughly $\$16,500$, corresponding to an earnings return of approximately 53 relative to a baseline of $\$31,300$ without high school graduation ($\mathbb{E}[Y(0)]$) . Almost half ($\frac{7824}{16454.} \cdot 100 = 47.5\%$) of the total effect is estimated to operate directly, with 29% and 22% mediated via college attendance and college completion, respectively.

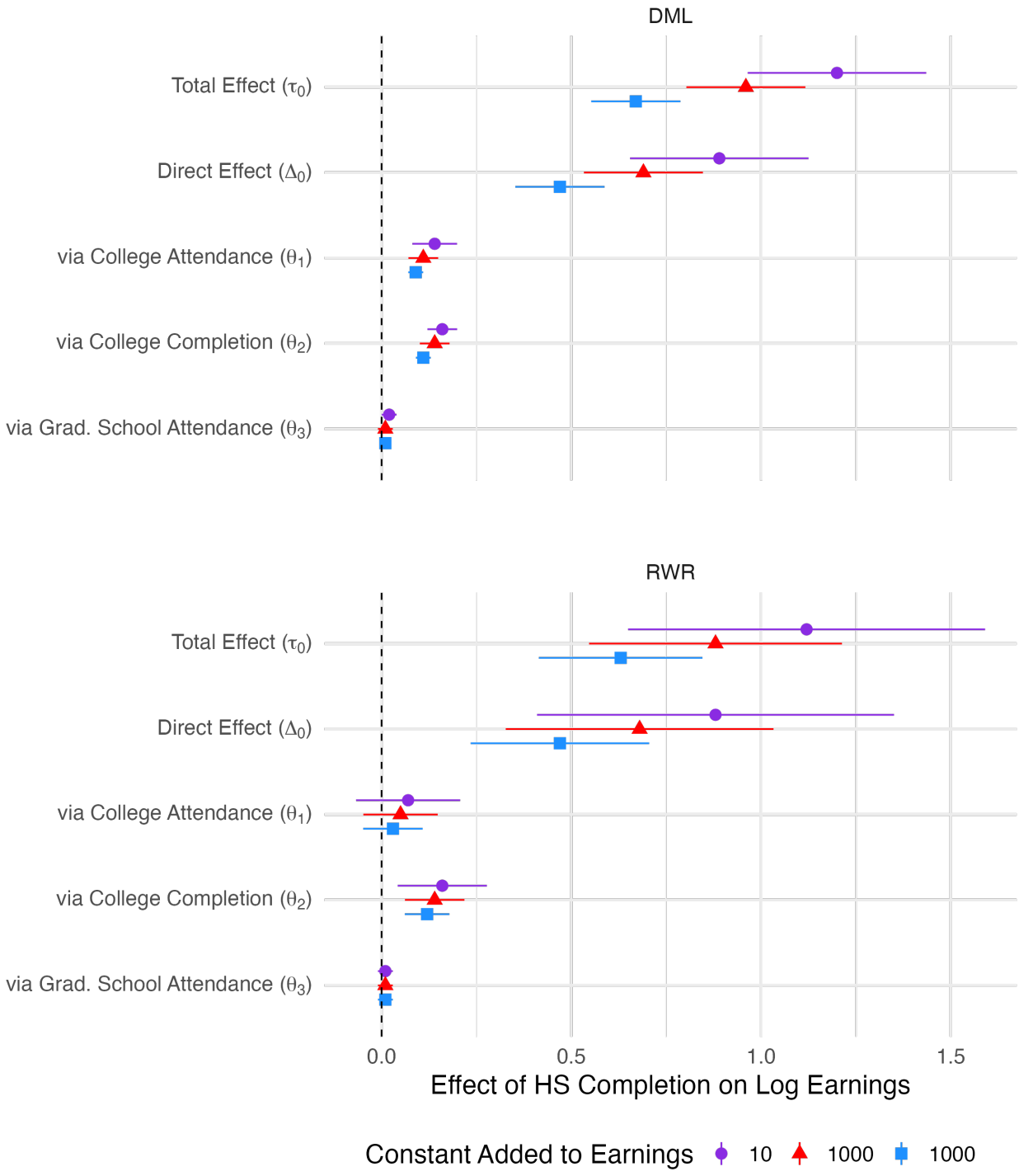

Figure 5: Decomposition of the Average Total Effect (ATE) of High School Graduation on Logged Earnings Under Alternative Definitions of Earnings. The figure shows estimates of the total effect (τ_0) as well as the indirect effects $\Delta_0, \Delta_1, \ldots\ldots, \Delta_K$ when constants of 10, 100 and 1000, respectively, are added to raw annual earnings (in dollar amounts) before taking the log.

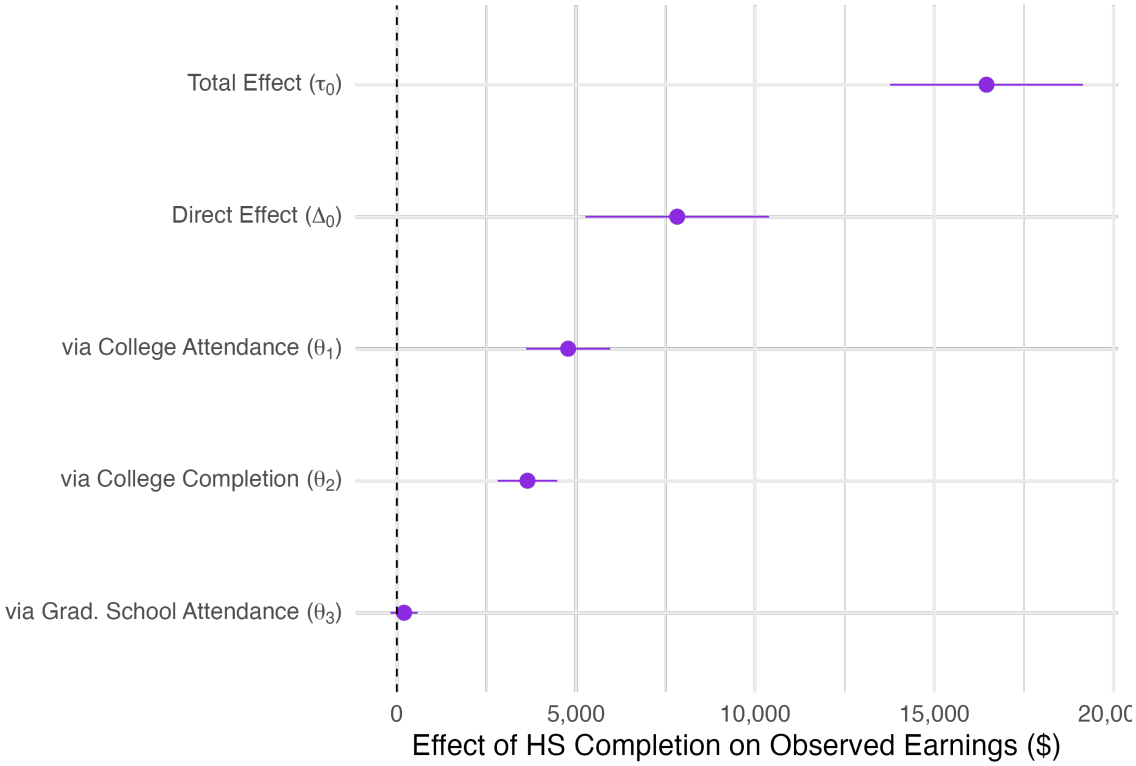

Figure 6: Decomposition of the Average Total Effect (ATE) of High School Graduation on Logged Earnings with Different Definitions of Earnings. The figure shows estimates of the total effect (τ_0) as well as the indirect effects $\Delta_0, \Delta_1, \ldots\ldots, \Delta_K$ when constants of 10, 100 and 1000 are added to raw annual earnings (in dollar amounts) before taking the log.

G Extension to multivalued, discrete mediators

The main text considers a decomposition of the ATE in the case of binary monotonic mediators (i.e., educational transitions), but the framework can straightforwardly be extended to accommodate categorical transitions. Such a decomposition is especially appealing when we consider the variegated and complex trajectories that individuals take through the US postsecondary system, that dichotomizing transitions invariably misses. In particular, in the early 2010s, under 40% of high school graduates immediately enrolled in a four-year college, with around 30% immediately attending a two-year college. Approximately one third of immediate two-year college attendees will then progress to a four-year college; indeed, nearly 50% of BA recipients previously attended a public two-year college in their educational careers. To illustrate the MPSE decomposition for multivalued, discrete mediators in the case of a single mediator, consider high school graduation A, and immediate college enrollment $M \in \{\text{none}, \text{2yr}, \text{4yr}\}$ denotes whether an individual did not pursue postsecondary education, or instead attended a 2-year or 4-year college. Assumption 1 is similarly assumed to hold in this context, such that $M(0) = 0$ (high school non-completers cannot pursue *any* form of postsecondary education). Under this assumption, I obtain the following multivariate decomposition:

$$
\begin{aligned}
&\text{ATE}\mathbb{E}[Y(1) - Y(0)] \\
&\quad = \mathbb{E}[Y(1,0) - Y(0)] \\
&\quad + \mathbb{P}[M(1) = \text{2yr}]\mathbb{E}[Y(1,\text{2yr}) - Y(1,0)] + \text{cov}\big[\mathbb{I}[M_1(1) = \text{2yr}], Y(1,\text{2yr}) - Y(1,0)\big] \qquad (14) \\
&\quad + \mathbb{P}[M(1) = \text{4yr}]\mathbb{E}[Y(1,\text{4yr}) - Y(1,0)] + \text{cov}\big[\mathbb{I}[M_1(1) = \text{4yr}], Y(1,\text{4yr}) - Y(1,0)\big]. \qquad (15)
\end{aligned}
$$

Each of these quantities can be identified under Assumptions 2-4 above. Further, Equations 14 and 15 can be further decomposed into the direct and gross effects of each of these transitions in a way analogous to Equation 4 (for example, the gross effect of 2-year attendance can be decomposed into its direct effects, net of subsequent transfer to a 4-year college, and its indirect effect via 4-year transfer, which itself can be further decomposed into direct and gross effects).

H Description of EIFs used in empirical illustration

For each component involved in the MPSE, I construct a Neyman-orthogonal "signal" using its EIF, whose exact form depends on whether each set of intermediate confounders is empty or not. Figure 3 in the main text shows a potential data-generating process for the direct and indirect (continuation) effects of high school graduation on adult earnings, via three transitions: college attendance (M_1), BA completion (M_2), and graduate school attendance (M_3). I assume that a set of pre-college characteristics serve as confounders for the $A - (M_1, M_2, M_3, Y)$ relationships, and that a set of post-secondary confounders Z confound the $M_2 - (M_3, Y)$ relationships.

Under these assumptions for the various sets of confounders, my MPSE decomposition implies that, in the case of the four transitions (one treatment and three mediators), it suffices to estimate the following three sets of parameters: (i) four direct effects $\Delta_k, k \in [0 \ldots, 3]$, where $\Delta_3 = \tau_3$, (ii) four gross effects $\tau_k, k \in [0 \ldots, 3]$, where $\tau_0 = \text{ATE}$, (iii) three mediator terms, $\pi_k, k \in [1 \ldots, 3]$. All components in the three-mediator decomposition can then be estimated as functions of these parameters. For each of these target parameters, I construct a Neyman-orthogonal signal using its efficient influence function. Because of my assumed data-generating process, which maintains that there is only a single set of intermediate confounders (as opposed to a separate set of confounders for each mediator), the EIF for each estimand involved in the decomposition simplifies somewhat. Specifically, the recentered EIFs for each component in the decomposition are shown below:

$$
\begin{aligned}
M_1^*(1) &= \gamma_1(X) + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(M_1 - \gamma_1(X)), \\
M_2^*(1,1) &= \gamma_2(X) + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(M_2 - \gamma_2(X)), \\
M_3^*(1,1,1) &= \mathbb{E}[\gamma_3(X,Z)|X, A=1, M_1=1]
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(\gamma_3(X,Z) - \mathbb{E}[\gamma_3(X,Z)|X, A=1, M_1=1]) \\
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=1)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,1)}(M_3 - \gamma_3(X,Z)), \\
Y^*(a) &= \mu_0(X,a) + \frac{\mathbb{I}(A=a)}{\pi_0(X,a)}(Y - \mu_0(X,a)), \text{ for } a \in \{0,1\} \\
Y^*(1,m_1) &= \mu_1(X,m_1) + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=m_1)}{\pi_0(X,1)\pi_1(X,m_1)}(Y - \mu_1(X,m_1)), \text{ for } m_1 \in \{0,1\} \\
Y^*(1,1,m_2) &= \mathbb{E}[\mu_2(X,Z,m_2)|X, A=1, M_1=1] \\
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(\mu_2(X,Z,m_2) - \mathbb{E}[\mu_2(X,Z,m_2)|X, A=1, M_1=1]) \\
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=m_2)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,m_2)}(Y - \mu_2(X,Z,m_2)), \text{ for } m_2 \in \{0,1\} \\
Y^*(1,1,1,m_3) =& \mathbb{E}[\mu_3(X,Z,m_3)|X, A=1, M_1=1] \\
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(\mu_3(X,Z,m_3) - \mathbb{E}[\mu_3(X,Z,m_3)|X, A=1, M_1=1]) \\
&+ \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=1)\mathbb{I}(M_3=m_3)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,1)\pi_3(X,Z,m_3)}(Y - \mu_3(X,Z,m_3)) \text{ for } m_3 \in \{0,1\},
\end{aligned}
$$

where

$$\pi_0(X, a) \triangleq \Pr[A = a \mid X]$$

$$\pi_1(X, m_1) \triangleq \Pr[M_1 = m_1 \mid X, A = 1]$$

$$\pi_2(X, Z, m_2) \triangleq \Pr[M_2 = m_2 \mid X, A = 1, M_1 = 1, Z]$$

$$\pi_3(X, Z, m_3) \triangleq \Pr[M_3 = m_3 \mid X, A = 1, M_1 = 1, Z, M_2 = 1]$$

$$\gamma_1(X) \triangleq \mathbb{E}[M_1 \mid X, A = 1]$$

$$\gamma_2(X) \triangleq \mathbb{E}[M_2 \mid X, A = 1, M_1 = 1]$$

$$\gamma_3(X, Z) \triangleq \mathbb{E}[M_3 \mid X, A = 1, M_1 = 1, Z, M_2 = 1]$$

$$\mu_0(X, a) \triangleq \mathbb{E}[Y|X, A = a]$$

$$\mu_1(X, m_1) \triangleq \mathbb{E}[Y|X, A = 1, M_1 = m_1]$$

$$\mu_2(X, Z, m_2) \triangleq \mathbb{E}[Y|X, A = 1, M_1 = 1, Z, M_2 = m_2]$$

$$\mu_3(X, Z, m_3) \triangleq \mathbb{E}[Y|X, A = 1, M_1 = 1, Z, M_2 = 1, M_3 = m_3].$$

I Proofs and technical details

I.1 EIFs for η_k and θ_k terms (Proposition 1)

Under Assumptions 2-4, the covariance component η_k is identified as $\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k\tau_k$. Following (Kennedy, 2022, , p. 15), I let $\mathbb{IF} : \Psi \to L_2(\mathbb{P})$ denote the operator mapping the functionals $\{\Delta_k, \pi_k, \eta_k\} : \mathcal{P} \to \mathbb{R}$, $\forall \in [K]$ to their respective influence functions under the nonparametric model $\mathcal{P}$. First, by linearity of the EIF, $\mathbb{IF}(\eta_k)$ is given by

$$\mathbb{IF}(\eta_k) = \mathbb{IF}(\tau_{k-1}) - \mathbb{IF}(\Delta_{k-1}) - \mathbb{IF}(\pi_k\tau_k).$$

Since $\mathbb{IF}(\pi_k\tau_k)$ can be written as follows $\mathbb{IF}(\pi_k\tau_k) = \tau_k\mathbb{IF}(\pi_k) + \pi_k\mathbb{IF}(\tau_k)$, $\mathbb{IF}(\eta_k)$ can be written as

$$\begin{aligned}\mathbb{IF}(\eta_k) &= \mathbb{IF}(\tau_{k-1}) - \mathbb{IF}(\Delta_{k-1}) - \big(\tau_k\mathbb{IF}(\pi_k) + \pi_k\mathbb{IF}(\tau_k)\big) \\ &= \mathbb{IF}(\tau_{k-1}) - \mathbb{IF}(\Delta_{k-1}) - \tau_k\mathbb{IF}(\pi_k) - \pi_k\mathbb{IF}(\tau_k).\end{aligned}$$

Noticing that I can rewrite this expression as

$$\begin{aligned}\mathbb{IF}(\eta_k) &= \mathbb{RIF}(\tau_{k-1}) - \tau_{k-1} - \mathbb{RIF}(\Delta_{k-1}) + \Delta_{k-1} - \tau_k\mathbb{RIF}(\pi_k) + \tau_k\pi_k - \pi_k\mathbb{RIF}(\tau_k) + \pi_k\tau_k \\ &= \mathbb{RIF}(\tau_{k-1}) - \mathbb{RIF}(\Delta_{k-1}) - \tau_k\mathbb{RIF}(\pi_k) - \pi_k\mathbb{RIF}(\tau_k) + \pi_k\tau_k - \eta_k,\end{aligned}$$

where $\mathbb{RIF}(\phi) = \mathbb{IF}(\phi) + \phi$, I can obtain the corresponding EIF-based estimator for η_k can obtained by solving the empirical moment condition obtained by setting the average of the above equation equal to 0, and plugging in the set of estimated nuisance functions.

$$\hat{\eta}_k^{\text{eif}} = \hat{\mathbb{RIF}}(\tau_{k-1}) - \hat{\mathbb{RIF}}(\Delta_{k-1}) - \tau_k \hat{\mathbb{RIF}}(\pi_k) - \pi_k \hat{\mathbb{RIF}}(\tau_k) + \pi_k \tau_k,$$

where $\hat{\mathbb{RIF}}(\phi) = \hat{\mathbb{IF}}(\phi) + \phi$, and $\hat{\mathbb{IF}}(\phi)$ denotes the influence function of a parameter evaluated at estimates of its component nuisance functions.

Turning next to the influence functions for the continuation effects θ_k, $k \in \{1, \dots K\}$, following the same logic as the above, I can write the EIF of θ_k, $\mathbb{IF}(\theta_k)$, as

$$\mathbb{IF}(\theta_k) = \mathbb{IF}(\Delta_k) \prod_{j=1}^{k} \pi_j + \Delta_k \sum_{j=1}^{k} \mathbb{IF}(\pi_j) \prod_{l:l\neq j}^{k} \pi_l + \mathbb{IF}(\eta_k) \prod_{j=1}^{k-1} \pi_j + \eta_k \sum_{j=1}^{k-1} \mathbb{IF}(\pi_j) \prod_{l:l\neq j}^{k-1} \pi_l.$$

Rewriting this expression as

$$\begin{aligned}\mathbb{IF}(\theta_k) = {} & \mathbb{RIF}(\Delta_k) \prod_{j=1}^{k} \pi_j + \Delta_k \sum_{j=1}^{k} \mathbb{RIF}(\pi_j) \prod_{l:l\neq j}^{k} \pi_l + \mathbb{RIF}(\eta_k) \prod_{j=1}^{k-1} \pi_j + \eta_k \sum_{j=1}^{k-1} \mathbb{RIF}(\pi_j) \prod_{l:l\neq j}^{k-1} \pi_l \\ & - k\Delta_k \prod_{j=1}^{k} \pi_j - (k-1)\eta_k \prod_{j=1}^{k-1} \pi_j - \theta_k,\end{aligned}$$

I obtain the corresponding EIF-based estimator for θ_k can obtained by solving the empirical moment condition obtained by setting the average of the above equation equal to 0, and plugging in the set of estimated nuisance functions, as:

$$\begin{aligned}\hat{\theta}_k^{\text{eif}} = {} & \widehat{\mathbb{RIF}}(\Delta_k) \prod_{j=1}^{k} \hat{\pi}_j + \hat{\Delta}_k \sum_{j=1}^{k} \hat{\mathbb{RIF}}(\pi_j) \prod_{l:l\neq j}^{k} \hat{\pi}_l + \widehat{\mathbb{RIF}}(\eta_k) \prod_{j=1}^{k-1} \hat{\pi}_j + \hat{\eta}_k \widehat{\mathbb{RIF}}(\pi_j) \prod_{l:l\neq j}^{k-1} \hat{\pi}_l \\ & - k\hat{\Delta}_k \prod_{j=1}^{k} \hat{\pi}_j - (k-1)\hat{\eta}_k \prod_{j=1}^{k-1} \hat{\pi}_j.\end{aligned}$$

I.2 Semiparametric efficiency (Proposition 2)

In this section, I establish the conditions required for the semiparametric efficiency of all terms featured in the decomposition. Before proceeding, I establish some notational preliminaries. Let $||g|| = (\int g^\top g dP)^{1/2}$ denote the $L_2(P)$ norm, and let $R_n(\,\cdot\,)$ denote a mapping from a nuisance function to its $L_2(P)$ convergence rate. Let $\hat{\varphi}^{\text{EIF}}_{km_k} = \mathbb{P}_n[m(O;\hat{\eta})]$, where $m(O;\hat{\eta})$ is the quantity inside $\mathbb{P}_n[\cdot]$ in equation 9, and $\hat{\eta} = (\hat{\pi}_0, \ldots, \hat{\pi}_K, \hat{\mu}_0, \ldots, \hat{\mu}_K)$. I have that

$$\begin{aligned}\hat{\varphi}^{\text{EIF}}_{km_k} - \varphi_{km_k} &= \mathbb{P}_n[m(O;\hat{\eta})] - P[m(O;\eta)] \\ &= \mathbb{P}_n[m(O;\eta)] + \underbrace{P[m(O;\hat{\eta}) - m(O;\eta)]}_{\triangleq R_2(\hat{\eta})} + (\mathbb{P}_n - P)\left[m(O;\hat{\eta}) - m(O;\eta)\right], \quad (16)\end{aligned}$$

where $Pg = \int g dP$ denotes the expectation of function g at the truth. The first term in equation 16 is a sample average, and can be analyzed with the central limit theorem. It has an asymptotic variance of $\mathbb{E}\left[(\varphi_{km_k}(O;\eta))^2\right]$. The last term is an empirical process term that will be $o_p(n^{-1/2})$ if either the nuisance functions fall in a Donsker class or if cross-fitting is used to induce independence between $\hat{\eta}$ and O. Thus,$\hat{\theta}^{\text{EIF}}$ will be asymptotically normal and semiparametric efficient if $R_2(\hat{\eta})$ is $o(n^{-1/2})$. To analyze this term, I first note that

$$
\begin{aligned}
P[m(O;\eta)] &= P\Bigg[\frac{A}{\pi_{01}(X)}\left(\frac{\mathbb{I}(M_k=m_k)}{\pi_{km_k}(X,\overline{Z}_k)}\prod_{j=1}^{k-1}\frac{M_j}{\pi_{j1}(X,\overline{Z}_j)}\right)\left(Y-\mu^k_{km_k}\left(X,\bar{Z}_k\right)\right) \\
&\quad+\sum_{j=1}^{k}\frac{A}{\pi_{01}(X)}\left(\prod_{l=1}^{j-1}\frac{M_l}{\pi_{l1}(X,\overline{Z}_l)}\right)\left(\mu^k_{jm_k}\left(X,\overline{Z}_k\right)-\mu^k_{j-1m_k}\left(X,\bar{Z}_{j-1}\right)\right)+\mu_0(X)\Bigg] \\
&= P\Bigg[\frac{A}{\pi_{01}(X)}\left(\frac{\mathbb{I}(M_k=m_k)}{\pi_{km_k}(X,\overline{Z}_k)}\prod_{j=1}^{k-1}\frac{M_j}{\pi_{j1}(X,\overline{Z}_j)}\right)\cdot \\
&\quad\left(\underbrace{\mathbb{E}[Y-\mu^k_{km_k}\left(X,\bar{Z}_k\right)|X,\bar{Z}_k,\overline{1}_k,m_k]}_{=0}\right) \\
&\quad+\sum_{j=1}^{k}\frac{A}{\pi_{01}(X)}\left(\prod_{l=1}^{j-1}\frac{M_l}{\pi_{l1}(X,\overline{Z}_l)}\right)\cdot \\
&\quad\left(\underbrace{\mathbb{E}[\mu^k_{jm_k}\left(X,\overline{Z}_j\right)-\mu_{(j-1)m_j}\left(X,\bar{Z}_{j-1}\right)|X,\bar{Z}_{j-1},\overline{1}_j]}_{=0}\right)+\mu_0(X)\Bigg] \\
&= P\Big[\mu_0(X)\Big].
\end{aligned}
$$

Plugging this result into $R_2(\hat{\eta})$, I have that

$$
\begin{aligned}
R_2(\hat{\eta}) &= P[m(O;\hat{\eta})-m(O;\eta)] \\
&= P\Bigg[\sum_{j=0}^{k-1}\frac{A}{\hat{\pi}_{01}(X)}\left(\prod_{l=1}^{j-1}\frac{M_j}{\hat{\pi}_{l1}(X,\bar{Z}_l)}\right)\cdot \\
&\quad\left(\hat{\pi}_{j1}(X,\overline{Z}_j)-\pi_{j1}(X,\overline{Z}_j)\right)\left(\hat{\mu}^k_{jm_k}(X,\bar{Z}_j)-\mu^k_{jm_k}(X,\bar{Z}_j)\right) \\
&\quad+\frac{A}{\hat{\pi}_{01}(X)}\left(\prod_{l=1}^{k-1}\frac{M_j}{\hat{\pi}_{l1}(X,\bar{Z}_l)}\right)\cdot \\
&\quad\left(\hat{\pi}_{km_k}(X,\overline{Z}_j)-\pi_{km_k}(X,\overline{Z}_j)\right)\left(\hat{\mu}^k_{km_k}(X,\bar{Z}_k)-\mu^k_{km_k}(X,\bar{Z}_k)\right)\Bigg] \\
&= \sum_{j=0}^{k}O_p\big(||\hat{\pi}_{j1}(X,\overline{Z}_j)-\pi_{j1}(X,\overline{Z}_j)||\cdot||\hat{\mu}^k_{jm_k}(X,\bar{Z}_k)-\mu^k_{jm_k}(X,\bar{Z}_k)||\big),
\end{aligned}
$$

where the last equality results from the positivity assumption that $\hat{\pi}_{k1}(X,\bar{Z}_k)$ is bounded

away from zero, for all $k \in [K]$, and from the Cauchy-Schwartz inequality. Then, assuming that the empirical process term is of order $o_p(n^{-1/2})$, I can write $\hat{\varphi}^{\text{EIF}}_{km_k} - \varphi_{km_k}$ as

$$\begin{aligned}\hat{\psi}^{\text{EIF}}_{km_k} - \psi_{km_k} = \mathbb{P}_n[m(O;\eta) - \psi_{km_k}] \\ + \sum_{j=0}^{k} O_p\big(||\hat{\pi}_{j1}\big(X, \overline{Z}_j\big) - \pi_{j1}\big(X, \overline{Z}_j\big)||\big) \cdot O_p\big(||\hat{\mu}^k_{jm_k}\big(X, \bar{Z}_k\big) - \mu^k_{jm_k}\big(X, \bar{Z}_j\big)||\big) \\ + o_p(n^{-1/2}).\end{aligned}$$

Thus, letting $R_n(k, m_k) \triangleq \sum_{j=0}^{k} R_n\big(\hat{\pi}_j\big) R_n\big(\hat{\mu}^k_{km_k}\big)$, $\hat{\varphi}^{\text{rEIF}}_{km_k}$ is consistent if $R_n(k, m_k) = o(1)$ and it is semiparametric efficient if $R_n(k, m_k) = o(n^{-1/2})$. Clearly, then, $\hat{\Delta}^{\text{rEIF}}_k$ is consistent if $\sum_{j=k}^{k+1} R_n(j, 0) = o(1)$ and it is semiparametric efficient if $\sum_{j=k}^{k+1} R_n(j, 0) = o(n^{-1/2})$. Similarly, $\hat{\tau}^{\text{rEIF}}_k$ is consistent if $\sum_{j=0}^{1} R_n(k, j) = o(1)$ and it is semiparametric efficient if $\sum_{j=0}^{1} R_n(k, j) = o(n^{-1/2})$.

Turning next to to $\phi_k \triangleq \mathbb{E}[M_{k+1}(\overline{1}_{k+1})]$, for all $k \in [K-1]$, I can similarly define $\gamma_k\left(X, \bar{Z}_k\right)$ iteratively as

$$\begin{aligned}\gamma_k\left(X, \bar{Z}_k\right) &\triangleq \mathbb{E}\left[M_{k+1} \mid X, \bar{Z}_k, \overline{1}_{k+1}\right] \\ \gamma_j\left(X, \bar{Z}_j\right) &\triangleq \mathbb{E}\left[\gamma_{j+1}\left(X, \bar{Z}_{j+1}\right) \mid X, \bar{Z}_j, \overline{1}_{j+1}\right] \forall j \in [k-1].\end{aligned}$$

Similarly to the previous case, the EIF of ϕ_k is equal to

$$\begin{aligned}\varphi_k(O) = \frac{A}{\pi_{01}\big(X\big)} \left(\prod_{l=1}^{k} \frac{M_j}{\pi_{l1}\big(X, \overline{Z}_l\big)}\right) \left(Y - \gamma_k\left(X, \bar{Z}_k\right)\right). \\ + \sum_{j=0}^{k} \frac{A}{\pi_{01}\big(X\big)} \left(\prod_{l=1}^{j-1} \frac{M_l}{\pi_{l1}\big(X, \overline{Z}_l\big)}\right) \left(\gamma_j\left(X, \overline{Z}_j\right) - \gamma_{j-1}\left(X, \bar{Z}_{j-1}\right)\right) \\ + \gamma_0(X) - \phi_k.\end{aligned}$$

Following similar arguments to the above, we have that

$$\begin{aligned}\hat{\phi}_k^{\text{EIF}} - \phi_k = \mathbb{P}_n[m_2(O;\eta) - \phi_k] \\ + \sum_{j=0}^{k} O_p\big(||\hat{\pi}_{j1}\big(X, \overline{Z}_j\big) - \pi_{j1}\big(X, \overline{Z}_j\big)||\big) \cdot O_p\big(||\hat{\gamma}_j\big(X, \bar{Z}_j\big) - \gamma_j\big(X, \bar{Z}_j\big)||\big) \\ + o_p(n^{-1/2}),\end{aligned}$$

where $m_2(O;\hat{\eta}) = \varphi_k + \phi_k$. Thus, letting $R_n(k,\gamma) \triangleq \sum_{j=0}^{k} R_n\big(\hat{\pi}_j\big) R_n\big(\hat{\gamma}_j\big)$, $\hat{\phi}_k^{\text{EIF}}$ is consistent if $R_n(k,\gamma) = o(1)$ and it is semiparametric efficient if $R_n(k,\gamma) = o(n^{-1/2})$. This result implies that if all nuisance functions are consistently estimated and converge at faster than $n^{1/4}$ rates, then $\hat{\phi}_k^{\text{EIF}}$ is semiparametric efficient. I first establish the following lemma:

Lemma 1. *Let X_n and Y_n denote two convergent sequences, where $X_n = O_p(n^{-1/2})$ and $Y_n = o_p(n^{-1/2})$. Then, (a) $X_n Y_n = o_p(n^{-1/2})$, and (b) $X_n X_n = o_p(n^{-1/2})$.*

Proof. (a) $X_n = O_p(n^{-1/2}) = n^{-1/2} O_p(1) = o_p(1)$. Thus, $X_n Y_n = o_p(1) o_p(n^{-1/2}) = o_p(n^{-1/2})$. (b) $X_n X_n = O_p(n^{-1/2}) O_p(n^{-1/2}) = O_p(n^{-1}) = n^{-1/2} O_p(n^{-1/2}) = n^{-1/2} o_p(1)$ (by (a)). Thus, $X_n X_n = o_p(n^{-1/2})$. □

Using this lemma, I establish rate conditions for the semiparametric efficiency of $\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k$. I can analyze the asymptotic behavior of $\hat{\eta}_k = \hat{\tau}_{k-1} - \hat{\Delta}_{k-1} - \hat{\pi}_k \hat{\tau}_k$ via a distributional expansion of each plug-in estimator:

$$\begin{aligned}\hat{\eta}_k &= \hat{\tau}_{k-1} - \hat{\Delta}_{k-1} - \hat{\pi}_k \hat{\tau}_k \\ &= (\tau_{k-1} + \mathbb{P}_n[\tau_{k-1}^{\text{EIF}}] + \tau_{k-1}^{\text{EP}} + \tau_{k-1}^{\text{R2}}] - (\Delta_{k-1} + \mathbb{P}_n[\Delta_{k-1}^{\text{EIF}}] + \Delta_{k-1}^{\text{EP}} + \Delta_{k-1}^{\text{R2}}] \\ &\quad - [\pi_k + \mathbb{P}_n[\pi_k^{\text{EIF}}] + \pi_k^{\text{EP}} + \pi_k^{\text{R2}}][\tau_k + \mathbb{P}_n[\tau_k^{\text{EIF}}] + \tau_k^{\text{EP}} + \tau_k^{\text{R2}}] \\ &= (\tau_{k-1} + \mathbb{P}_n[\tau_{k-1}^{\text{EIF}}] + o_p(n^{-1/2}) + \tau_{k-1}^{\text{R2}}] - (\Delta_{k-1} + \mathbb{P}_n[\Delta_{k-1}^{\text{EIF}}] + o_p(n^{-1/2}) + \Delta_{k-1}^{\text{R2}}]\end{aligned}$$

$$
\begin{aligned}
&- [\pi_k + \mathbb{P}_n[\pi_k^{\text{EIF}}] + o_p(n^{-1/2}) + \pi_k^{\text{R2}}][\tau_k + \tau_k^{\text{EIF}} + o_p(n^{-1/2}) + \tau_k^{\text{R2}}] \\
&= [\tau_{k-1} - \Delta_{k-1} - \pi_k\tau_k] + \mathbb{P}_n[(\tau_{k-1} - \Delta_{k-1} - \pi_k\tau_k)^{\text{EIF}}] \\
&\quad + \tau_{k-1}^{\text{R2}} + \Delta_{k-1}^{\text{R2}} + \pi_k^{\text{R2}} + \tau_k^{\text{R2}} + O_p(n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1/2})o_p(n^{-1/2}) + o_p(n^{-1}) + o_p(n^{-1/2}) \\
&= [\tau_{k-1} - \Delta_{k-1} - \pi_k\tau_k] + \mathbb{P}_n[(\tau_{k-1} - \Delta_{k-1} - \pi_k\tau_k)^{\text{EIF}}] \\
&\quad + \tau_{k-1}^{\text{R2}} + \Delta_{k-1}^{\text{R2}} + \pi_k^{\text{R2}} + \tau_k^{\text{R2}} + o_p(n^{-1/2}),
\end{aligned}
$$

where the penultimate equality follows from Proposition 1, and the final equality follows from Lemma 1.

Thus, for any $k \in \{1, \ldots, K\}$, $\hat{\eta}_k = \hat{\tau}_{k-1} - \hat{\Delta}_{k-1} - \hat{\pi}_k\hat{\tau}_k$ is semiparametric efficient if $\tau_{k-1}^{\text{R2}} + \Delta_{k-1}^{\text{R2}} + \pi_k^{\text{R2}} + \tau_k^{\text{R2}} = o_p(n^{-1/2})$.

For the continuation terms $(\theta_k = (\Pi_{j=1}^{k}\pi_j)\Delta_k + (\Pi_{j=1}^{k-1}\pi_j)\eta_k)$, I proceed by induction. Let $\hat{\Delta}_* = \Delta_* + \mathbb{P}_n[\Delta_*^{\text{EIF}}] + \Delta_*^{\text{EP}} + \Delta_*^{\text{R2}}$ and $\hat{\eta}_* = \eta_* + \mathbb{P}_n[\eta_*^{\text{EIF}}] + \eta_*^{\text{EP}} + \eta_*^{\text{R2}}$ be asymptotically linear, where $* \in \{1, \ldots K\}$. For $k = 1$, I can asymptotically expand $\hat{\pi}_1\hat{\Delta}_*$ as

$$
\begin{aligned}
\hat{\pi}_1\hat{\Delta}_* &= (\pi_1 + \mathbb{P}_n[\pi_1^{\text{EIF}}] + \pi_1^{\text{EP}} + \pi_1^{\text{R2}})(\Delta_* + \mathbb{P}_n[\Delta_*^{\text{EIF}}] + \Delta_*^{\text{EP}} + \Delta_*^{\text{R2}}) \\
&= (\pi_1 + \mathbb{P}_n[\pi_1^{\text{EIF}}] + o_p(n^{-1/2}) + \pi_1^{\text{R2}})(\Delta_* + \mathbb{P}_n[\Delta_*^{\text{EIF}}] + o_p(n^{-1/2}) + \Delta_*^{\text{R2}}) \\
&= \pi_1\Delta_* + \mathbb{P}_n[\pi_1^{\text{EIF}}\Delta_* + \Delta_*^{\text{EIF}}\pi_1] + \pi_1^{\text{R2}} + \Delta_*^{\text{R2}} + o_p(n^{-1/2}) \quad \text{(by Lemma S??)} \\
&= \sum_{j=1}^{k}\pi_j\Delta_* + \mathbb{P}_n[(\sum_{j=1}^{k}\pi_j\Delta_*)^{\text{EIF}}] + \Delta_*^{\text{R2}} + \sum_{j=1}^{k}\pi_j^{\text{R2}} + o_p(n^{-1/2}) \quad \text{(by Proposition 1)},
\end{aligned}
$$

and expand $(\Pi_{j=1}^{k-1}\pi_j)\eta_*$, similarly, as

$$
\begin{aligned}
\sum_{j=1}^{k-1}\hat{\pi}_j\hat{\eta}_* &= \eta_* + \mathbb{P}_n[\eta_*{}^{\text{EIF}}] + \tau_{*-1}^{\text{R2}} + \Delta_{*-1}^{\text{R2}} + \pi_*^{\text{R2}} + \tau_*^{\text{R2}} + o_p(n^{-1/2}) \\
&= \sum_{j=1}^{k-1}\pi_j\eta_* + \mathbb{P}_n[(\sum_{j=1}^{k-1}\pi_j\eta_*)^{\text{EIF}}]
\end{aligned}
$$

$$+ \tau^{\text{R2}}_{*-1} + \Delta^{\text{R2}}_{*-1} + \pi^{\text{R2}}_{*} + \tau^{\text{R2}}_{*} + \sum_{j \in \{1...,k^*-1\}: j \neq *} \pi^{\text{R2}}_{j} + o_p(n^{-1/2}),$$

following a similar logic to above. Now, assume that, for $k^* \in \{1, \ldots, K\}$, $(\Pi_{j=1}^{k^*} \hat{\pi}_j)\hat{\Delta}_* = \Delta_k \prod_{j=1}^{k^*} \pi_j + \mathbb{P}_n[(\Delta_* \prod_{j=1}^{k^*} \pi_j)^{\text{EIF}}] + \Delta^{\text{R2}}_* + \sum_{j=1}^{k^*} \pi^{\text{R2}}_j + o_p(n^{-1/2})$ and, further, that $(\Pi_{j=1}^{k^*-1} \hat{\pi}_j)\hat{\eta}_* = \Pi_{j=1}^{k^*-1} \pi_j \eta_* + \mathbb{P}_n[(\Pi_{j=1}^{k^*-1} \pi_j \eta_*)^{\text{EIF}}] + \tau^{\text{R2}}_{*-1} + \Delta^{\text{R2}}_{*-1} + \pi^{\text{R2}}_{*} + \tau^{\text{R2}}_{*} + \sum_{j \in \{1...,k^*-1\}: j \neq *} \pi^{\text{R2}}_j + o_p(n^{-1/2})$. Then, by induction, I have that

$$\begin{aligned}
(\Pi_{j=1}^{k^*+1} \hat{\pi})\hat{\Delta}_* &= \left[(\pi_{k^*+1} + \mathbb{P}_n[\pi^{\text{EIF}}_{k^*+1}] + o_p(n^{-1/2}) + \pi^{\text{R2}}_{k^*+1})\right] \\
&\quad \left[\Delta_* \prod_{j=1}^{k^*} \pi_j + \mathbb{P}_n[(\Delta_* \sum_{j=1}^{k^*} \pi_j)^{\text{EIF}}] + \Delta^{\text{R2}}_* + \sum_{j=1}^{k^*} \pi^{\text{R2}}_j + o_p(n^{-1/2})\right] \\
&= \Delta_* \prod_{j=1}^{k^*+1} \pi_j + \mathbb{P}_n[(\Delta_* \prod_{j=1}^{k^*+1} \pi_j)^{\text{EIF}}] + \Delta^{\text{R2}}_* \\
&\quad + \sum_{j=1}^{k^*+1} \pi^{\text{R2}}_j + O_p(n^{-1/2}) O_p(n^{-1/2}) \\
&\quad + O_p(n^{-1/2}) o_p(n^{-1/2}) + o_p(n^{-1}) + o_p(n^{-1/2}) \\
&= \Delta_* \sum_{j=1}^{k^*+1} \pi_j + \mathbb{P}_n[(\Delta_* \prod_{j=1}^{k^*+1} \pi_j)^{\text{EIF}}] + \Delta^{\text{R2}}_* + \sum_{j=1}^{k^*+1} \pi^{\text{R2}}_j + o_p(n^{-1/2}),
\end{aligned}$$

and that

$$\begin{aligned}
(\Pi_{j=1}^{k^*} \hat{\pi}_j)\hat{\Delta}_* &= \left[(\pi_{k^*+1} + \mathbb{P}_n[\pi^{\text{EIF}}_{k^*+1}] + o_p(n^{-1/2}) + \pi^{\text{R2}}_{k^*+1})\right] \\
&\quad \left[\Pi_{j=1}^{k^*-1} \pi_j \eta_* + \mathbb{P}_n[(\Pi_{j=1}^{k^*-1} \pi_j \eta_*)^{\text{EIF}}] + \tau^{\text{R2}}_{*-1} + \Delta^{\text{R2}}_{*-1} + \pi^{\text{R2}}_{*} + \tau^{\text{R2}}_{*} + \sum_{j \in \{1...,k^*-1\}: j \neq *} \pi_j + o_p(n^{-1/2})\right] \\
&= \Pi_{j=1}^{k^*} \pi_j \eta_* + \mathbb{P}_n[(\Pi_{j=1}^{k^*} \pi_j \eta_*)^{\text{EIF}}] + \tau^{\text{R2}}_{*-1} + \Delta^{\text{R2}}_{*-1} + \pi^{\text{R2}}_{*} + \tau^{\text{R2}}_{*} + \sum_{j \in \{1...,k^*\}: j \neq *} \pi^{\text{R2}}_j + o_p(n^{-1/2})
\end{aligned}$$

It follows that, for any $k \in \{1, \ldots, K\}$,

$$(\Pi_{j=1}^{k}\hat{\pi}_j)\hat{\Delta}_k = \Delta_k \sum_{j=1}^{k}\pi_j + \Pi_{j=1}^{k-1}\pi_j\eta_k + \mathbb{P}_n[(\Delta_k \sum_{j=1}^{k}\pi_j + \Pi_{j=1}^{k-1}\pi_j\eta_k)^{\mathrm{EIF}}] + \sum_{j=k-1}^{k}\Delta_j^{\mathrm{R2}} + \sum_{j=k-1}^{k}\tau_j^{\mathrm{R2}} + \sum_{j=1}^{k}\pi_j^{\mathrm{R2}}.$$

Thus, $\hat{\theta}_k = (\Pi_{j=1}^{k}\hat{\pi}_j)\hat{\Delta}_k + (\Pi_{j=1}^{k-1}\hat{\pi}_j)\hat{\eta}_k$ is semiparametric efficient if $\sum_{j=k-1}^{k}\Delta_j^{\mathrm{R2}} + \sum_{j=k-1}^{k}\tau_j^{\mathrm{R2}} + \sum_{j=1}^{k}\pi_j^{\mathrm{R2}} = o(n^{-1/2})$. Proposition 2 then follows immediately by recognizing the rate conditions required for each of the constituent functionals of $(\pi_k, \Delta_k, \tau_k)$ to be semiparametric efficient.

J Derivation of RWR procedures

For simplicity, throughout the following I let $Z_0 = X$ and $M_0 = A$. I assume the following linear specification of the outcome model:

$$\mathbb{E}[Y \mid \overline{Z}_k, A, \overline{M}_k] = \beta_{k,0} + c_{k,0}A + \sum_{j=1}^{k} \beta_{k,j} M_j + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} A X^{\perp} + \sum_{j=1}^{k-1} \eta_{k,j}^{T} M_j X^{\perp} + \sum_{j=1}^{k} \gamma_{k,j}^{T} Z_j^{\perp} + \sum_{j=1}^{k-1} M_j \sum_{l=1}^{j} \xi_{k,k,l}^{\top} Z_l^{\perp}, \tag{17}$$

where $Z_k^{\perp} = Z_k - \mathbb{E}[Z_k | \overline{Z}_{k-1}, M_{k-1} = 1_{k-1}]$, $\forall k \in [0, \ldots, K]$. In the following derivations, I use the fact that, $\forall k \in \{1, \ldots, K\}$,

$$\begin{aligned} &\int z_k^{\perp} dP(z_k | \overline{z}_{k-1}, m_{k-1} = 1) \\ &= \mathbb{E}[Z_k - \mathbb{E}[Z_k | \overline{z}_{k-1}, m_{k-1} = 1] | \overline{z}_{k-1}, m_{k-1} = 1] \\ &= 0. \end{aligned}$$

Letting $X = Z_0$, the above also implies that $\int z_0^{\perp} dP(z_0) = \mathbb{E}[Z_0 - \mathbb{E}[Z_0]] = 0$. Under sequential ignorability and assuming linearity of the outcome with respect to all antecedent variables, I have that

$$\begin{aligned} \Delta_{k-1} = &\int \mathbb{E}[Y | \overline{M}_{k-1} = \overline{1}_{k-1}, \overline{z}_k, M_k = 0] \prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1) \\ &- \int \mathbb{E}[Y | \overline{M}_{k-2} = \overline{1}_{k-2}, \overline{z}_k, M_{k-1} = 0] \prod_{j=1}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1) \end{aligned}$$

$$= \int \Big[\beta_{k,0} + c_{k,0} + \sum_{j=1}^{k-1} \beta_{k,j} + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} X^{\perp} + \sum_{j=1}^{k-2} \eta_{k,j}^{T} X^{\perp} + \sum_{j=1}^{k} \gamma_{k,j}^{T} Z_j^{\perp} + \sum_{j=1}^{k-2} \sum_{l=1}^{j} \xi_{k,k,l}^{\top} Z_l^{\perp})$$

$$- (\beta_{k,0} + c_{k,0} + \sum_{j=1}^{k-2} \beta_{k,j} M_j + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} A X^{\perp} + \sum_{j=1}^{k-2} \eta_{k,j}^{T} X^{\perp} + \sum_{j=1}^{k} \gamma_{k,j}^{T} Z_j^{\perp} + \sum_{j=1}^{k-3} \sum_{l=1}^{j} \xi_{k,k,l}^{\top} Z_l^{\perp})\Big]$$

$$\prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1)$$

$$= \beta_{k,k-1}.$$

Further, for $\tau_k \forall k \in \{1, \ldots, K\}$ I have that

$$\tau_k = \int \mathbb{E}[Y | A = 1, \overline{M}_k = \overline{1}_k, \overline{z}_k] \prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1)$$

$$- \int \mathbb{E}[Y | A = 1, \overline{M}_{k-1} = \overline{1}_{k-1}, \overline{z}_k, M_k = 0] \prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1)$$

$$= \int \Big[(\beta_{k,0} + c_{k,0} + \sum_{j=1}^{k} \beta_{k,j} + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} X^{\perp} + \sum_{j=1}^{k-1} \eta_{k,j}^{T} X^{\perp} + \sum_{j=1}^{k} \gamma_{k,j}^{T} Z_j^{\perp} + \sum_{j=1}^{k-1} \sum_{l=1}^{j} \xi_{k,j,l}^{\top} Z_l^{\perp})$$

$$- (\beta_{k,0} + c_{k,0} + \sum_{j=1}^{k-1} \beta_{k,j} + \eta_{k,1}^{\top} X^{\perp} + c_{k,1} X^{\perp} + \sum_{j=1}^{k-2} \eta_{k,j}^{T} X^{\perp} + \sum_{j=1}^{k} \gamma_{k,j}^{T} Z_j^{\perp} + \sum_{j=1}^{k-2} \sum_{l=1}^{j} \xi_{k,j,l}^{\top} Z_l^{\perp})\Big]$$

$$\prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1)$$

$$= \beta_{k,k}.$$

Finally, I assume that

$$\mathbb{E}[M_{k+1} \mid A = 1, \overline{Z}_k, \overline{M}_k = \overline{1}_k] = \theta_0 + \sum_{k=0}^{k} \delta_{k+1}^{T} Z_k^{\perp} \quad .$$

Then:

$$\mathbb{E}[M(\overline{1}_{k+1})] = \theta_0 + \int \left[\sum_{k=0}^{k} \delta_{k+1}^T Z_k^{\perp} \prod_{j=0}^{k} dP(z_j | \overline{z}_{j-1}, m_{j-1} = 1) \right]$$

$$= \theta_0.$$

K Further details about simulation study

L Derivation of bias formulae for sensitivity analysis

In this section, I derive the bias formulae for the set (τ_k, Δ_k) for all $k \in [K]$, where K denotes the number of mediators considered in the decomposition, under a sequence of simplifying assumptions. Assume first that I have a binary unobserved confounder, U, for the treatment-outcome relationship. Assuming that $\alpha_0 = \mathbb{E}[Y|x, a, U = 1] - \mathbb{E}[Y|x, a, U = 0]$ does not depend on x or a, and further that $\beta_0 = \Pr[U = 1|x, A = 1] - \Pr[U = 1|x, A = 0]$ does depend on x, for $\tau_0 = \mathbb{E}[Y(1) - Y(0)] \triangleq$ ATE, I then have that $\text{bias}(\tau_0) = \alpha\beta$ (VanderWeele and Arah, 2011).

Next, consider an unobserved confounder, U_k that affects both M_k and Y for any $k \in \{1, \ldots, K\}$. Then, under a weaker iteration of Assumption 3 (Sequential Ignorability), i.e., $Y(\overline{1}_k, m_k) \perp\!\!\!\perp (A, \overline{M}_k)|X, A, U_k, \overline{Z}_k, \overline{M}_{k-1} \forall k \in [K]$,$\mathbb{E}[Y(\overline{1}_k, m_k)]$ is identified as

$$\mathbb{E}[Y(\overline{1}_k, m_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, u_k] \big[dP(u_k|x, \overline{z}_k, \overline{1}_{k-1}) \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x).$$

By contrast, under Assumption 3, my estimator of $\mathbb{E}[Y(\overline{1}_k, m_k)]$, $\tilde{\mathbb{E}}[Y(\overline{1}_k, m_k)]$, converges to

$$\tilde{\mathbb{E}}[Y(\overline{1}_k, m_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, u_k] \big[dP(u_k|x, \overline{z}_k, \overline{1}_k) \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x).$$

I invoke the following three assumptions: (Assumption A_k) $\alpha_k = \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 0]$ does not depend on $(x, \overline{z}_k, \overline{1}_k, m_k)$; (Assumption B_k) $\beta_k = \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, m_k] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k]\big)$ does not depend on $(x, \overline{z}_k)$; Assumption (C)

U_k is binary. Taking the difference between the quantities in the above two equations thus gives that, for any $m_k \in \{0,1\}$, for any $k \in [K]$, I have that

$$\text{bias}(\tilde{\mathbb{E}}[Y(\overline{1}_k, m_k)]) = \int \big(\mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, m_k, U_k = 0]\big) \cdot \big(\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, m_k] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k]\big) \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x). \tag{18}$$

Consider first $\text{bias}(\Delta_{k-1}) = \text{bias}(\tilde{\mathbb{E}}[Y(\overline{1}_k, 0) - Y(\overline{1}_{k-1}, 0)])$. Under mediator monotonicity (Assumption 1), I immediately have that $\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, m_k] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k]\big)$, and thus that $\text{bias}(\Delta_{k-1}) = \text{bias}(\tilde{\mathbb{E}}[Y(\overline{1}_k, 0)])$, which can be written as

$$\begin{aligned}
\text{bias}(\tilde{\mathbb{E}}[Y(\overline{1}_k, 0)]) &= \int \big(\mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0, U_k = 0]\big) \\
&\quad \big(\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, 0] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k]\big) \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x) \\
&= \int \big(\mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0, U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0, U_k = 0]\big) \cdot \\
&\quad \big(\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, 0] - \big(\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_{k+1}]\Pr[M_k = 1|x, \overline{z}_k, \overline{1}_k] \\
&\quad + \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, 0] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, 0]\Pr[M_k = 1|x, \overline{z}_k, \overline{1}_k]\big)\big) \\
&\quad \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x) \\
&= -\int \big(\mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0U_k = 1] - \mathbb{E}[Y|x, \overline{z}_k, \overline{1}_k, 0_k, U_k = 0]\big) \cdot \\
&\quad \big(\big(\Pr[U_k = 1|x, \overline{z}_k, \overline{1}_{k+1}] - \Pr[U_k = 1|x, \overline{z}_k, \overline{1}_k, 0]\big)\Pr[M_k = 1|x, \overline{z}_k, \overline{1}_k]\big) \\
&\quad \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x).
\end{aligned}$$

Next, applying assumptions A_k and B_k, I can write

$$\text{bias}(\tilde{\mathbb{E}}[Y(\overline{1}_{k+1},0)]) = -\alpha_k\beta_k \int_x \int_{\overline{z}_k} \Pr[M_k = 1|x, \overline{z}_k, \overline{1}_k] \prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \overline{1}_{j-1})\big] dP(x).$$

Second, to compute $\text{bias}(\tau_k) = \text{bias}(\mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)])$ for any $k \in \{1, \ldots K\}$, beginning with Equation 18 and applying assumptions A_k and B_k once again, I have that

$$\text{bias}(\tau_k) = \alpha_k\beta_k.$$