This is a "preproof" accepted article for *Psychometrika*.
This version may be subject to change during the production process.
DOI: 10.1017/psy.2026.10153

Psychometrika (YEAR), **xx**:xx 1–37

CAMBRIDGE
UNIVERSITY PRESS

Theory and Methods

RAPSEM: Identifying Latent Mediators Without Sequential Ignorability via a Rank-Preserving Structural Equation Model

Sofia Morelli[1], Roberto Faleh[1] and Prof. Dr. Holger Brandt[1]

[1]Methods Center, Eberhard Karls Universität Tübingen, Tübingen, 72076, Baden-Württemberg, Germany.

Keywords: Latent variable mediation, Structural equation modeling (SEM), Causal effect identification, Sequential ignorability

MSC Codes: *Primary* – 62H25; *Secondary* – 62P15, 62F12, 62-04

Abstract

Standard structural equation models (SEMs) are often used to identify latent mediators. However, valid inference typically relies on the strong, frequently violated Sequential Ignorability assumption. We introduce the Rank-Preserving Structural Equation Model (RAPSEM), which increases robustness through *G*-estimation while maintaining the measurement model's integrity through a two-stage method of moments (2SMM) for factor score corrections. RAPSEM replaces the no unmeasured mediator-outcome confounding with the weaker no unobserved effect modification assumption. By leveraging treatment randomization, RAPSEM achieves identification in a manner equivalent to instrumental variable estimation through structurally emerging instruments. Specifically, identification relies on treatment-covariate interactions that influence the mediator but have no direct effect on the outcome, allowing researchers to utilize natural heterogeneity in treatment response as a testable source of identification. We provide a robustness assessment for the core identifying assumption and establish the consistency and asymptotic normality of the resulting estimator. Simulation studies demonstrate that RAPSEM remains unbiased under unobserved confounding, whereas standard SEM yields biased results. RAPSEM achieves reasonable power for sample sizes above 500, depending on the strength of the structural instruments. The method is implemented in the accompanying `rapsem` R package, and its practical utility is illustrated through an empirical example from educational research.

Mediation analysis is a central topic in applied psychological research, particularly when designing and evaluating intervention studies (Windgassen et al., 2016). A mediator variable represents part of the causal pathway connecting an intervention to an outcome (Holland, 1988). That is, the intervention influences the mediator, which in turn affects the outcome. Positioned temporally and conceptually between intervention and outcome, mediators explain how and why treatments work, which offers several advantages.

First, mediators provide insight into the mechanisms through which interventions exert their effects. For example, cognitive behavioral therapy (CBT) may reduce social anxiety symptoms by first reducing maladaptive beliefs, suggesting that cognitive change is the driving force behind therapeutic improvement (Boden et al., 2012; Castella et al., 2015). Second, understanding which variables serve as mediators can guide the refinement of interventions: if a specific mechanism is identified, treatment protocols can be more precisely tailored to target it. Third, mediators can act as early indicators of treatment success because they precede the outcome in the causal sequence. When such variables are easier or less costly to measure, this allows for more timely and resource-efficient evaluations.

A difficulty in identifying mediators is potential unobserved confounders. While conducting a randomized controlled trial (RCT) protects the treatment-outcome and treatment-mediator pathways, it does not shield the mediator-outcome relationship from confounding. Experimentally isolating this effect

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

would require the direct manipulation and randomization of the mediator itself. However, such an approach is typically challenging or even impossible for psychological constructs. Consequently, most applied studies rely on the assumption of Sequential Ignorability, which presumes the absence of unmeasured mediator-outcome confounding (Imai, Keele, & Yamamoto, 2010). Vo et al. (2020) found that 96% of mediation studies employ traditional methods based on this assumption (for examples see Brandt, 2024).

Sequential ignorability requires that all relevant covariates influencing both the mediator and the outcome are included in the model, a standard that is rarely achievable in practice. Researchers might adjust for variables such as baseline symptom severity, age, gender, or duration of symptoms. Yet other plausible confounders, such as interpersonal sensitivity, peer rejection history, or recent social stressors, are more difficult to assess. Biological markers, such as genetic predispositions and cortisol reactivity, may be prohibitively expensive to obtain, and contextual factors, like population density or seasonal social activity levels, are easily overlooked. Indeed, Vo et al. (2020) found that only 50% of mediation studies adjusted for mediator-outcome confounders, and those that did often relied solely on baseline values of the mediator and outcome.

The failure to account for confounding has serious statistical consequences. Omitting even a single relevant confounder can severely bias mediation estimates, typically resulting in an overestimation of the mediated effect (Fritz et al., 2016). This compromises the validity of the findings and may lead to overly optimistic conclusions about treatment mechanisms.

These limitations have motivated researchers in fields such as biomedicine to develop methods that are robust to omitted variable bias, with *G*-estimation being a particularly promising direction that remains underused due to its complex implementation (Vansteelandt & Joffe, 2014). Ten Have et al. (2007) formulated causal mediation analysis within the potential outcomes framework using *G*-estimation under the rank-preserving (RP) assumption from which the approach in this paper takes its name. Others have built on this approach, developed similar methods in parallel, and pointed out connections to the instrumental variable (IV) framework (Albert, 2008; Dunn & Bentall, 2007; Emsley et al., 2010; Lynch et al., 2008; Small, 2012; Zheng & Zhou, 2015).

Zheng and Zhou (2015) introduced a relaxed version of the rank-preserving assumption that does not require the strict individual-level constancy of the original RPM, while retaining the core identification strategy, which we adopt here. The corresponding identifying assumption is weaker than sequential ignorability (Small, 2012; Zheng & Zhou, 2015): it requires only that unmeasured confounders do not modify the mediator effect or the direct effect of treatment. That is, unmeasured variables are permitted as long as they do not interact with treatment or mediator on the outcome. Identification is achieved by requiring that at least one baseline covariate interact with treatment to predict the mediator but not the outcome. Simulation results further illustrate that this *G*-estimation approach yields approximately unbiased estimates of direct and mediated effects under unmeasured mediator–outcome confounding, whereas standard regression and structural equation approaches show systematic bias under the same conditions and even maintained robustness to unobserved effect modification (Brandt, 2020; Zheng & Zhou, 2015).

Compared to alternative IV-based approaches to mediation, the *G*-estimation framework occupies a distinct position. A rich literature proposes the use of external instruments to identify mediation effects (e.g., Frölich & Huber, 2017). In this context, the exclusion restriction requires that the instrument affects the outcome only through the mediator, with no unobserved path connecting the instrument directly to the outcome. This assumption is not empirically testable and must be justified through domain knowledge.[1] Finding credible external instruments remains a fundamental challenge in psychological research, where the absence of natural experiments and the interconnectedness of psychological constructs make it difficult to identify variables that plausibly satisfy the exclusion restriction (Bullock et al., 2010). The *G*-estimation approach avoids the need for external instruments by leveraging the standard structure of an RCT, specifically the treatment randomization and the baseline covariates typically collected in such trials.

[1] Over-identification tests can provide partial indirect evidence when multiple instruments are available (Angrist et al., 1996).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Despite their advantages, RPMs have not yet been extended to latent variable models. In psychological research, variables of interest are often measured indirectly using self-report questionnaires. Standard regression cannot properly handle such measures, especially when multiple indicators are highly correlated or represent a multidimensional construct (Bollen, 1989). Structural equation modeling (SEM) overcomes these limitations by explicitly modeling latent variables through measurement models. By estimating relationships at the latent level, SEM accounts for measurement error and the shared variance among indicators, thereby increasing statistical power to detect effects.

To overcome this significant limitation for fields like psychology, where measurement error is common, we introduce the Rank-Preserving Structural Equation Model (RAPSEM). This framework integrates the methods developed by Ten Have et al. (2007) and Zheng and Zhou (2015) with factor score regression and interaction corrections (Wall & Amemiya, 2000; Wall & Amemiya, 2003), incorporating latent constructs while maintaining the relaxed identifying assumption. RAPSEM thus enables robust mediation analysis even in the presence of measurement error and omitted confounders between the mediator and outcome.

Thereby, we make the following contributions. We formally extend G-estimation for mediation analysis to the latent variable case, moving beyond observed-variable models to account for measurement error in mediators, covariates, and outcomes. We explicate the identification strategy underlying RAPSEM, detailing the required assumptions, its origins in Structural Nested Mean Models (SNMMs), and its connections to the IV framework. We describe the estimation procedure and its implementation in an `R` package and formally establish consistency and asymptotic normality of the estimator. We show how to access the robustness to effect modification by unobserved confounders within the RAPSEM framework, following Zheng and Zhou (2015). Model performance is then evaluated in two simulation studies: the first compares the robustness of RAPSEM to standard SEM under violations of the underlying causal assumptions, and the second examines the statistical power of RAPSEM across varying scenarios. An empirical example from educational research further illustrates the practical differences between approaches.

1. Setting and Causal Effect Definitions

We model the mediation mechanism illustrated in Figure 1, assuming the data are collected in an RCT in which the treatment R is randomly assigned.[2] The path diagram depicts the mediator, outcome, and covariates using the latent variables η_M, η_Y, and η_X, each measured by three observed indicators (m_1–m_3, y_1–y_3, and x_1–x_3).

The goal is to estimate the controlled direct effect (CDE) of treatment R on the outcome η_Y, and the controlled mediation effect (CME) of the mediator η_M on the outcome η_Y (Pearl, 2001; J. M. Robins & Greenland, 1992; Vanderweele, 2011), adjusting for measured covariates η_X. We define these effects within the potential outcomes framework (Rubin, 1974), using the latent potential outcome $\eta_{Y_i}^{r,\eta_m}$ for individual i receiving treatment $R = r$ when the mediator takes value $\eta_{M_i} = \eta_m$.[3]

The controlled direct effect captures the contrast between potential outcomes under different treatment levels while holding the mediator η_m fixed and conditioning on covariates $\boldsymbol{\eta}_{\mathbf{x}_i}$. Under an additive linear model with a binary treatment $r \in \{0, 1\}$ and without any higher-order structural terms including treatment–mediator interactions, it reduces to the difference between the treated and control group captured by θ_r:

$$\text{CDE} = \text{E}\left[\eta_Y^{1,\eta_m} - \eta_Y^{0,\eta_m} \mid \boldsymbol{\eta}_{\mathbf{X}} = \boldsymbol{\eta}_{\mathbf{x}}\right] = \theta_r. \tag{1.1}$$

The controlled mediation effect captures the contrast between potential outcomes when the mediator is set to $\eta_{m'}$ rather than η_m, while holding the treatment r fixed and conditioning on covariates $\boldsymbol{\eta}_{\mathbf{x}_i}$. For a discrete mediator, it corresponds to the difference between two mediator categories; for a continuous

[2] While an RCT setting is preferred, the framework can be extended to observational data under appropriate covariate adjustment, for example via balancing methods, provided that the Ignorability Assumption (see C1) holds.

[3] Zheng and Zhou (2015) show how the controlled effects relate to natural direct and indirect effects.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

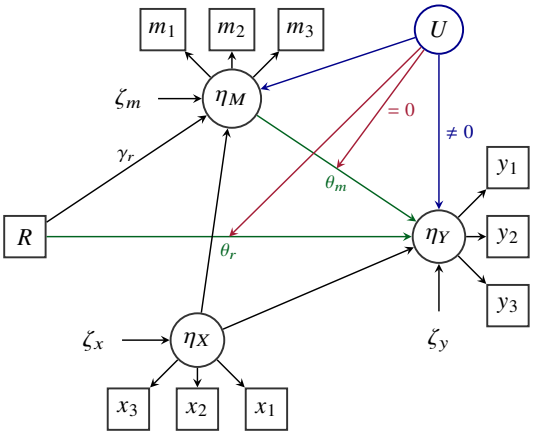


Figure 1: Path diagram of the RAPSEM model estimating the controlled direct effect (θ_r) of treatment R on outcome η_Y and the mediation effect (θ_m) of mediator η_M on outcome η_Y, adjusting for covariates η_X. Each latent factor (η_M, η_Y and η_X) is measured by three observed indicators (m_1–m_3, y_1–y_3, and x_1–x_3, respectively). The model accounts for an unobserved confounder U between mediator and outcome, with the No Unobserved Effect Modification assumption highlighted in red.

mediator, it represents the effect of a finite change in the mediator from η_m to $\eta_{m'}$. In the additive linear case, it is given by:

$$\text{CME} = \text{E}\left[\eta_Y^{r,\eta_{m'}} - \eta_Y^{r,\eta_m} \mid \boldsymbol{\eta}_{\mathbf{X}} = \boldsymbol{\eta}_{\mathbf{x}}\right] = \theta_m(\eta_{m'} - \eta_m). \tag{1.2}$$

2. Model Formulation

RAPSEM comprises a measurement component and a structural component, both detailed below. While mathematically equivalent to classical SEM (Bollen, 1989), the framework diverges significantly in its interpretive logic and the scope of its theoretical claims. This distinction is rooted in the causal nature of RAPSEM, which derives its structural part from the Potential Outcomes Framework and utilizes the causal inference technique of G-estimation. A notation table is provided in Section A.

2.1. Measurement Model

We consider a measurement model involving three types of latent variables: a mediator η_M measured by a vector of observed indicators $\mathbf{m}$,[4] a set of latent covariates $\boldsymbol{\eta}_X$ measured by indicators $\mathbf{x}$, and an outcome variable η_Y measured by indicators $\mathbf{y}$. Let K be the total number of latent variables and P the total number of indicators. For subject i, the $P \times 1$ vector of all observed indicators is given by[5]

$$\mathbf{z}_i = \left(\mathbf{m}_i \;\; \mathbf{x}_i \;\; \mathbf{y}_i\right)^\top, \tag{2.1}$$

and the corresponding $K \times 1$ vector of latent variables is

$$\boldsymbol{\eta}_i = \left(\eta_{M_i} \;\; \boldsymbol{\eta}_{\mathbf{X}_i} \;\; \eta_{Y_i}\right)^\top. \tag{2.2}$$

[4]While multiple latent mediators may be incorporated, such an extension would necessitate a modification of the underlying causal estimands.

[5]Throughout, $^\top$ applied to a row of blocks denotes vertical stacking, i.e. $\left(\mathbf{A} \;\; \mathbf{B}\right)^\top := \begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}$, the blocks themselves not being transposed; applied to a single matrix or vector, $^\top$ denotes the ordinary transpose.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

The measurement model links the observed indicators to their respective latent constructs. In general, our framework accommodates a broad class of latent variable models of the form

$$\begin{aligned} \mathbf{z}_i \mid \boldsymbol{\eta}_i &\sim F\big(\mathrm{E}[\mathbf{z}_i \mid \boldsymbol{\eta}_i], \mathrm{Var}(\mathbf{z}_i \mid \boldsymbol{\eta}_i)\big), \\ \text{with} \quad f\big(\mathrm{E}[\mathbf{z}_i \mid \boldsymbol{\eta}_i]\big) &= \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i. \end{aligned} \tag{2.3}$$

Here, F denotes the conditional distribution of $\mathbf{z}_i$ given the latent variables $\boldsymbol{\eta}_i$, characterized by its conditional mean $\mathrm{E}[\mathbf{z}_i \mid \boldsymbol{\eta}_i]$ and, where applicable, its conditional covariance matrix $\mathrm{Var}(\mathbf{z}_i \mid \boldsymbol{\eta}_i)$. The link function f maps the conditional expectation of the indicators to the latent scale, representing the observed means as a linear function of the latent variables with intercepts $\boldsymbol{\tau}$ and a factor loading matrix $\boldsymbol{\Lambda}$.

The specification of f and F depends on the nature of the observed indicators. For continuous indicators, f is the identity link and F is typically multivariate normal. For dichotomous indicators, f is a logit or probit link, which extends to multinomial or ordered models for categorical or ordinal indicators and can incorporate item-specific parameters, as in item response theory (IRT) models. In these cases, F represents the joint distribution of the resulting discrete outcomes, typically Bernoulli or multinomial.

In what follows, we focus on the linear Gaussian specification, which is well-suited for Likert-type item responses that can be treated as approximately continuous:

$$\mathbf{z}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \tag{2.4}$$

where $\boldsymbol{\epsilon}_i$ represents measurement error with $\mathrm{Var}(\mathbf{z}_i \mid \boldsymbol{\eta}_i) = \mathrm{Var}(\boldsymbol{\epsilon}_i)$. To ensure model identification, we adopt standard constraints by fixing the scale of each latent variable. Specifically, for each latent factor, one corresponding indicator is selected as a reference, with its loading fixed to 1 and its intercept to 0. Under this identification scheme, we partition the intercept vector $\boldsymbol{\tau}$ and loading matrix $\boldsymbol{\Lambda}$ as

$$\boldsymbol{\tau} = \begin{pmatrix} \boldsymbol{\tau}_{\text{free}} & \mathbf{0}_{\mathbf{K}\times\mathbf{1}} \end{pmatrix}^\top, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{\text{free}} & \mathbf{I}_K \end{pmatrix}^\top, \tag{2.5}$$

where $\boldsymbol{\tau}_{\text{free}}$ is a $(P-K)\times 1$ vector of free intercepts, $\boldsymbol{\Lambda}_{\text{free}}$ is a $(P-K)\times K$ matrix of free factor loadings, $\mathbf{0}_{K\times 1}$ is a $K\times 1$ vector of zeros, and $\mathbf{I}_K$ is the $K\times K$ identity matrix.

2.2. Structural Model

In line with (Zheng & Zhou, 2015), the general latent variable form of the potential outcome for person i under treatment r and mediator value η_m can be expressed as a linear combination of the structural component, a nuisance term, and a remaining error term:

$$\eta_{Y_i}^{r,\eta_m} = \underbrace{g(\boldsymbol{\eta}_{\mathbf{X}_i})}_{\text{nuisance term}} + \underbrace{\sum_{s=1}^{S} \theta_s \cdot h_s(r, \eta_M, \boldsymbol{\eta}_{\mathbf{X}_i})}_{\text{structural component}} + \underbrace{\zeta_{Y_i}^{r,\eta_m}}_{\text{error term}}. \tag{2.6}$$

As detailed in Zheng and Zhou (2015), no inherent restrictions need to be imposed on the functional form of $h(\cdot)$, provided that the resulting parameters can be mapped to well-defined causal contrasts under the prevailing identification assumptions. Furthermore, the model may, in principle, include interactions between the treatment and the mediator, as well as interactions between treatment or mediator and covariates, thus explicitly capturing effect heterogeneity. We focus, however, on the additive linear structure represented in Figure 1 and consistent with the original model of Ten Have et al. (2007).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

This choice preserves the interpretability of the CDE and CME by ensuring that θ_r and θ_m function as invariant structural parameters.[6]

The nuisance term $g(\boldsymbol{\eta}_{\mathbf{X}_i}) = \mathrm{E}\left[\eta_{Y_i}^{0,0} \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right]$ represents the covariate-conditional mean of the baseline outcome under no treatment and no mediation. In principle, it may be modeled as an arbitrary function of the covariates, denoted by $g(\boldsymbol{\eta}_{\mathbf{X}_i})$, including nonparametric forms (e.g., using splines in Brandt, 2020; or a deep neural network in Faleh et al., 2026). Here, however, we approximate $g(\cdot)$ as a linear combination of covariates, potentially including polynomial terms and interactions. Such transformations can be accommodated using the factor score corrections proposed by Wall and Amemiya (2000), whereas more general nonlinear specifications are not supported by existing latent factor score correction methods (Hayes & Usami, 2020). Consequently, we model the nuisance term as $g(\boldsymbol{\eta}_{\mathbf{X}_i}) = q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})\,\boldsymbol{\beta}_x$, where $q_{yx}(\cdot)$ denotes the admissible transformations.

The error term $\zeta_{Y_i}^{r,\eta_m}$ may explicitly depend on covariates $\boldsymbol{\eta}_{\mathbf{X}_i}$ as well as on unobserved confounders U, being only subject to $\mathrm{E}\left[\zeta_{Y_i}^{r,\eta_m} \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right] = 0$ for every (r, η_m). We return to this point when discussing the identification of the causal estimands.

Assuming the conditions in the next section hold, the additive linear potential outcome model for a continuous outcome can be represented by the following structural model:

$$\eta_{Y_i} = \theta_r \cdot R_i + \theta_m \cdot \eta_{M_i} + q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})\,\boldsymbol{\beta}_x + \zeta_{Y_i},$$

which corresponds to the model specification supported by RAPSEM. This model specifies the latent outcome for each individual under their realized treatment and mediator values. To facilitate the formulation of the estimation procedure, we adopt matrix notation

$$\boldsymbol{\eta}_Y = \boldsymbol{\Xi}_y \boldsymbol{\alpha} + \boldsymbol{\zeta}_y \tag{2.7}$$

with parameter vector

$$\boldsymbol{\alpha} = \begin{pmatrix} \theta_r & \theta_m & \boldsymbol{\beta}_x \end{pmatrix}^\top$$

and design matrix $\boldsymbol{\Xi}_y$ composed of row vectors containing the predictors of each subject i

$$\boldsymbol{\xi}_{Y_i} = \begin{pmatrix} R_i & \eta_{M_i} & q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i}) \end{pmatrix}.$$

In addition to the outcome model, we specify a separate structural equation for the latent mediator η_{M_i}. In this specification, we include interaction terms between the treatment and covariates to account for treatment effect heterogeneity, which is relevant for the power of RAPSEM as discussed in the assumptions and identification sections and further illustrated in the simulation study. The mediator model can then directly be written in matrix notation as

$$\boldsymbol{\eta}_M = \boldsymbol{\Xi}_m \boldsymbol{\gamma} + \boldsymbol{\zeta}_m, \tag{2.8}$$

with parameter vector

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_r & \boldsymbol{\gamma}_x & \boldsymbol{\gamma}_{xr} \end{pmatrix}^\top$$

and the rows of the design matrix given by

$$\boldsymbol{\xi}_{M_i} = \begin{pmatrix} R_i & q_{mx}(\boldsymbol{\eta}_{\mathbf{X}_i}) & R_i \cdot q_{mrx}(\boldsymbol{\eta}_{\mathbf{X}_i}) \end{pmatrix},$$

[6] As noted by J. M. Robins (2003), the identification of natural effects typically requires a no-interaction assumption to decompose the total effect. While controlled effects remain identifiable in the presence of interaction, this assumption is still necessary to represent them as single, invariant parameters. Including an interaction term $\theta_{rm}(r \cdot \eta_M)$ would transform these effects into functions of the levels of other variables.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

where $q_{mx}(\cdot)$ and $q_{mrx}(\cdot)$ denote transformations of the latent covariates in the main and interaction effects, respectively.

For the simulation study, we introduce a further simplification by considering two covariates that enter both the outcome and mediator models without transformation. Specifically, we set

$$q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i}) = q_{mx}(\boldsymbol{\eta}_{\mathbf{X}_i}) = q_{mrx}(\boldsymbol{\eta}_{\mathbf{X}_i}) = \begin{pmatrix} \eta_{X_{1i}} & \eta_{X_{2i}} \end{pmatrix},$$

which yields the full structural model[7]

$$\begin{aligned} \eta_{Y_i} &= \theta_r R_i + \theta_m \eta_{M_i} + \beta_{x1}\eta_{X_{1i}} + \beta_{x2}\eta_{X_{2i}} + \zeta_{Y_i}, \\ \eta_{M_i} &= \gamma_r R_i + \gamma_{x1}\eta_{X_{1i}} + \gamma_{x2}\eta_{X_{2i}} + \gamma_{x1r} R_i \eta_{X_{1i}} + \gamma_{x2r} R_i \eta_{X_{2i}} + \zeta_{M_i}. \end{aligned} \tag{2.9}$$

3. Assumptions

The assumptions listed in Table 1 are required to identify contrasts in Equations (1.1) to (1.2) from the observed data and to ensure the consistency and asymptotic normality of the effect estimates. We discuss the most relevant assumptions for identification, C1, C2 and S1, in the following and define the remaining ones in Section B.

Table 1: Overview of the assumptions required for identification, consistency, and asymptotic normality, grouped by category.

Category	**Assumption Names**
Causal Assumptions	C1 Ignorability C2 Unobserved Heterogeneity Mean Independence C3 Consistency C4 Positivity C5 Causal Efficacy
Structural Assumption	S1 Covariate-Treatment Interaction
Regularity Assumptions	R1 IID Sampling R2 Correct Model Specification R3 Measurement Model Regularity R4 Measurement Error R5 Factor Score Finite Moments R6 Structural Equation Error

Causal Assumptions

We focus here on the causal assumptions regarding unobserved confounding, while the remaining standard assumptions of the potential outcomes framework are detailed in Section B.

Assumption C1 - Ignorability: Conditional on measured baseline covariates η_x, the treatment assignment is independent of all potential outcomes and potential mediator values:

$$R \perp\!\!\!\perp (\eta_M^r, \eta_Y^{r,\eta_m}) \mid \boldsymbol{\eta}_{\mathbf{X}} \quad \forall\, r, \eta_m.$$

[7]Note that factor score corrections are still required for the weight calculation (see the estimation section).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

This assumption states that, after adjusting for observed covariates, there are no unmeasured confounders between the treatment and the mediator and no unmeasured confounders between the treatment and the outcome. This assumption holds for randomized interventions.

Assumption C2 - Unobserved Heterogeneity Mean Independence: As in Zheng and Zhou (2015), we assume the conditional expectation of the potential outcome error is invariant to the potential treatment and mediator levels. Specifically, there exists a function F such that

$$E\left[\zeta_{Y_i}^{r,\eta_m}(U_i, \boldsymbol{\eta}_{X_i}) \mid R_i, \eta_{M_i}, \boldsymbol{\eta}_{X_i}\right] = F(R_i, \eta_{M_i}, \boldsymbol{\eta}_{X_i}), \quad \forall r, \eta_m, \tag{3.1}$$

where F does not depend on the potential the potential treatment and mediator levels (r, η_m). This implies that the realized treatment and mediator value assignments carry no additional information about unobserved outcome heterogeneity beyond the measured covariates.

Another way to interpret this assumption is through the lens of effect modification by unmeasured confounders. If an unmeasured confounder U modifies the treatment-outcome and/or the mediator-outcome effect, the effects would differ across strata defined by U. Since this would cause the mean of ζ^{rm} to shift with the observed treatment and mediator, the above assumption would be violated. This rules out interaction effects between the unobserved confounder and the mediator in determining the outcome, unless explicitly modeled (see Zheng & Zhou, 2015, for such extensions), as indicated by the red paths in Figure 1. A comparison with the original Rank Preserving and the No Essential Heterogeneity (Heckman et al., 2006) assumption can be found in Section C.

Assumption C2 is central to our framework, as it replaces the commonly invoked Sequential Ignorability assumption. It can be considered weaker for the following reason: In the context of an RCT, Sequential Ignorability reduces to requiring independence between the potential outcomes of the mediator and outcome, conditional on covariates and treatment, which formally demands the error terms of the outcome and mediator equations to be uncorrelated. This rules out any unmeasured mediator-outcome confounding and any effect modification by such confounders. In contrast, C2 permits the error terms of the outcome and mediator equations to be correlated, allowing for unmeasured mediator-outcome confounding. It only rules out that this correlation manifests as a systematic shift in the mean of the outcome error across treatment-mediator conditions, conditional on observed covariates.

As noted in the introduction, no unmeasured confounding represents a strong, often unrealistic requirement. We argue that no effect modification by unobserved confounders may be more plausible in the same settings. For instance, consider genetic predisposition as an unobserved confounder in the studies mentioned in the introduction, where CBT reduces the outcome social anxiety symptoms by first decreasing the mediator maladaptive beliefs. While genetics likely influences both the baseline frequency of maladaptive belief and the physiological threshold for anxiety, it is less likely to act as an effect modifier of the mediator-outcome relationship. That is, while genetics may determine a patient's initial symptom levels (the intercept), it is unlikely to fundamentally alter the psychological mechanism (the slope) through which a reduction in maladaptive beliefs translates into symptom relief. Once conditioned on observed covariates like age and baseline severity, this mechanism is assumed to be a stable process that operates independently of the genetic factors that established the initial symptom levels.

Structural Assumption

Next, we define the critical structural requirement necessary to relax the Sequential Ignorability assumption. Other standard regularity conditions are detailed in Section B.

Assumption S1 - Treatment-Covariate Interaction: There exists at least one observed covariate that interacts with the treatment to significantly affect the mediator (e.g., $r \times \eta_{x_k} \rightarrow \eta_m$), providing sufficient variation to ensure that the weight matrix introduced in the next section has full column rank. This

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

treatment–covariate interaction affects the outcome solely through its effect on the mediator and has no direct influence on the outcome.

4. Identification

In the following, we establish the identification of θ_m and θ_r as causal effects. First, we review the derivation of orthogonality conditions via G-estimation, as developed by Ten Have et al. (2007) and Zheng and Zhou (2015). Then, we examine the formal connections between these conditions and the use of instrumental variables (IVs). Throughout this section, we treat the latent variables as if they were observed and establish identification of $\boldsymbol{\theta} = (\theta_r, \theta_m)$ from the joint distribution of $(R, \eta_M, \eta_Y, \boldsymbol{\eta}_{\mathbf{X}})$. By the measurement part of Assumption R2, the moments of this distribution required for the orthogonality conditions are identified from the observed indicators. Assumption C5 ensures that replacing the latent variables by their measurement-model-based estimates does not alter the interpretation of $\boldsymbol{\theta}$ as causal effects.

4.1. Orthogonality Conditions

Mediation analysis inherently follows a longitudinal structure, involving three temporally ordered measurement occasions following the assessment of baseline covariates: treatment, mediator, and outcome. From this perspective, it is natural to employ generalized causal inference methods, commonly referred to as G-methods, which were developed to address time-varying exposures and confounding in longitudinal settings (J. Robins & Hernan, 2008). While the broader class of G-methods includes the G-formula (or G-computation) and inverse probability weighting, the approach of Ten Have et al. (2007) is based on G-estimation.

G-estimation proceeds by relating the observed outcome η_{Y_i} to the potential outcome $\eta_{Y_i}^{r,\eta_m}$ in Equation (2.6) via a Structural Nested Mean Model (SNMM). This relation is established by sequentially "blipping down" the mediator and treatment effects to remove their respective influences from the observed data:

$$\eta_{Y_i}^{0,0}(\boldsymbol{\theta}) = \eta_{Y_i} - \theta_m \eta_{M_i} - \theta_r R_i.^{8}$$

Under Assumptions C3, C4, and the outcome part of R2, the blipped-down outcome evaluated at the true parameter values $\boldsymbol{\theta}^* = (\theta_r^*, \theta_m^*)$ satisfies

$$\eta_{Y_i}^{0,0}(\boldsymbol{\theta}^*) = g(\boldsymbol{\eta}_{\mathbf{X}_i}) + \zeta_{Y_i}.$$

Subtracting the nuisance term yields

$$\zeta_{Y_i} = \eta_{Y_i}^{0,0}(\boldsymbol{\theta}^*) - g(\boldsymbol{\eta}_{\mathbf{X}_i}).$$

Assumptions C1 and C2 imply that R_i carries no additional information about ζ_{Y_i} beyond what $\boldsymbol{\eta}_{\mathbf{X}_i}$ already provides:

$$\mathrm{E}\left[\zeta_{Y_i} \mid R_i, \boldsymbol{\eta}_{\mathbf{X}_i}\right] = \mathrm{E}\left[\zeta_{Y_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right] = 0.$$

While this resembles a standard ignorability statement, it serves here as the basis for the orthogonality condition

$$\mathrm{Cov}\left(\zeta_{Y_i},\, \mathbf{A}(R_i, \boldsymbol{\eta}_{\mathbf{X}_i}) \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right) = 0,$$

[8] In general $\eta_{Y_i}^{0,0}(\boldsymbol{\theta}) = \eta_{Y_i} - \sum_{s=1}^{S} \theta_s \cdot h_s(r, \eta_M, \boldsymbol{\eta}_{\mathbf{X}_i})$.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

defined for any weighting function $\mathbf{A}$. Since the adjusted residual is mean-centered, this condition translates to the population estimation equation

$$\mathrm{E}\left[\mathbf{A}(R_i, \boldsymbol{\eta}_{\mathbf{X}_i})^\top \left(\eta_{Y_i}^{0,0}(\boldsymbol{\theta}^*) - g(\boldsymbol{\eta}_{\mathbf{X}_i})\right)\right] = 0 \tag{4.1}$$

provided all entries of $\mathbf{A}$ satisfy the centering constraint $\mathrm{E}\left[a_s(R_i, \boldsymbol{\eta}_{\mathbf{X}_i}) \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right] = 0$. The corresponding sample estimating equation is

$$\sum_{i=1}^{N} \mathbf{A}(R_i, \boldsymbol{\eta}_{\mathbf{X}_i})^\top \left\{\eta_{Y_i}^{0,0}(\boldsymbol{\theta}) - \hat{g}(\boldsymbol{\eta}_{\mathbf{X}_i})\right\} = 0. \tag{4.2}$$

where $\hat{g}$ is a working model for the nuisance function g.

Zheng and Zhou (2015) showed that the most efficient weights for the outcome model in Equation (2.6) are given by

$$\mathbf{a}_s = \left(\mathrm{E}\left[h_s(R_i, \eta_{M_i}, \boldsymbol{\eta}_{\mathbf{X}_i}) \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i\right] - \mathrm{E}\left[h_s(R_i, \eta_{M_i}, \boldsymbol{\eta}_{\mathbf{X}_i}) \mid \boldsymbol{\eta}_{\mathbf{X}_i}\right]\right) \cdot \Omega_{\boldsymbol{\eta}_{\mathbf{X}_i}}^{-1}, \tag{4.3}$$

if the covariance of the error term conditioned on the covariates $\Omega_{\boldsymbol{\eta}_{\mathbf{X}_i}}$ does not depend on r and η_m. These weights capture the variation in $h_s(R_i, \eta_{M_i}, \boldsymbol{\eta}_{\mathbf{X}_i})$ that is attributable to the treatment, conditioned on the covariates.

4.2. Instrumental Variable Interpretation

This G-estimation approach aligns closely with the theory of instrumental variables (Angrist et al., 1996), a connection established by Dunn and Bentall (2007). In the classical IV framework, an external variable serves as a valid instrument if it satisfies two core conditions: it must be correlated with the endogenous predictor (relevance) and uncorrelated with the outcome error term (exclusion restriction).

Under the G-estimation framework, the randomized treatment R_i functions as an instrument for the mediator η_{M_i}. Relevance is satisfied via the interaction term required in the first sentence of Assumption S1, while the exclusion restriction is maintained through the conditional independence between the treatment and the potential outcome (Assumption C1) and the structural constraint in the second sentence of Assumption S1. Consequently, the instrument is constructed from the model structure by leveraging covariates that moderate the mediator but not the outcome effect. Notably, the strength of these moderating relationships can be empirically tested, e.g., using the first-stage F-statistic from the regression of the mediator on treatment, covariates, and their interaction.

The orthogonality conditions in Equation (4.2) directly mirror IV moment conditions, with the weighting functions in Equation (4.3) acting as residualized instruments. As demonstrated by Lee (2021) and Lee and Han (2024), using such centered instruments provides the advantage of maintaining consistency even for non-continuous outcomes.

5. Estimation

RAPSEM adopts a limited-information estimation strategy in which latent variable effects are recovered through a two-step procedure. In the first stage, factor scores are estimated based solely on the measurement model, treating it independently from the structural component. In the second stage, these estimated scores are used as observed proxies for the latent constructs in the structural equation. In this last step, we replace ordinary least squares with latent G-estimation. Details on implementation in the R package `rapsem` (Morelli et al., 2025), available at https://github.com/PsychometricsMZ/rapsem., available on GitHub, are provided in Section D. Refer to Section A for a table of all used symbols.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

5.1. Notation

Let $\boldsymbol{\eta}_i$ denote the true underlying factor scores in Equation (2.2), which are used in Equation (2.7). We later partition these scores into the outcome factor η_{Y_i} and the predictor factors $\boldsymbol{\eta}_{\text{pred},i}$, which include the mediator factor η_{M_i} and the covariate factors $\boldsymbol{\eta}_{\mathbf{X}_i}$. The theoretical estimator of the factor scores based on the true measurement parameters in Equation (2.4) is denoted by $\tilde{\boldsymbol{\eta}}_i$, while $\hat{\boldsymbol{\eta}}_i$ refers to the practical estimation of $\tilde{\boldsymbol{\eta}}_i$ obtained by plugging in the estimated measurement parameters.

5.2. First Stage

In the first stage, we adopt the method proposed by Wall and Amemiya (2000) to estimate the latent factor scores $\boldsymbol{\eta}_i$. We collect the free CFA parameters of the measurement model specified in Equations (2.4) to (2.5) in

$$\boldsymbol{\Upsilon}_1 = \left(\boldsymbol{\tau}_{\text{free}} \;\; \text{vec}\, \boldsymbol{\Lambda}_{\text{free}} \;\; \text{vech}\, \text{Var}(\boldsymbol{\epsilon}_i)\right) \tag{5.1}$$

where vec denotes the column-wise vectorization of a matrix and vech the vectorization of the non-redundant elements of a symmetric matrix. Given $\boldsymbol{\Upsilon}_1$, the estimator for the latent factor scores of subject i is defined as

$$\tilde{\boldsymbol{\eta}}_i = \left(-\mathbf{H} \;\; \mathbf{I}_K + \mathbf{H}\boldsymbol{\Lambda}_{\text{free}}\right) \left[\mathbf{z}_i - \left(\boldsymbol{\tau}_{\text{free}} \;\; \mathbf{0}_{K\times 1}\right)^\top\right], \tag{5.2}$$

where $\mathbf{z}_i$ denotes the vector of observed indicators, and the matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \left(\mathbf{0}_{K\times(P-K)} \;\; \mathbf{I}_K\right) \text{Var}(\boldsymbol{\epsilon}_i) \left(\mathbf{I}_{(P-K)} \;\; -\boldsymbol{\Lambda}_{\text{free}}^\top\right)^\top \left[\left(\mathbf{I}_{(P-K)} \;\; -\boldsymbol{\Lambda}_{\text{free}}\right) \text{Var}(\boldsymbol{\epsilon}_i) \left(\mathbf{I}_{(P-K)} \;\; -\boldsymbol{\Lambda}_{\text{free}}^\top\right)^\top\right]^{-1}.$$

To account for the propagation of measurement error, we treat the factor score estimator $\tilde{\boldsymbol{\eta}}_i$ as a noisy proxy of the true latent variables $\boldsymbol{\eta}_i$, assuming the relationship

$$\tilde{\boldsymbol{\eta}}_i = \boldsymbol{\eta}_i + \mathbf{e}_i, \tag{5.3}$$

where the estimation error $\mathbf{e}_i$ is linearly related to the measurement error $\boldsymbol{\epsilon}_i$ via

$$\mathbf{e}_i = \left(-\mathbf{H} \;\; \mathbf{I}_K + \mathbf{H}\boldsymbol{\Lambda}_{\text{free}}\right) \boldsymbol{\epsilon}_i. \tag{5.4}$$

For use in the errors-in-variables estimation in the subsequent stage, we also need the second moment of $\mathbf{e}_i$, i.e., its covariance matrix

$$\boldsymbol{\Sigma}_{ee} = \left(-\mathbf{H} \;\; \mathbf{I}_K + \mathbf{H}\boldsymbol{\Lambda}_{\text{free}}\right) \text{Var}(\boldsymbol{\epsilon}_i) \left(\mathbf{0}_{(P-K)\times K} \;\; \mathbf{I}_K\right)^\top. \tag{5.5}$$

The parameters in $\boldsymbol{\Upsilon}_1$ are estimated using standard confirmatory factor analysis (CFA), which minimizes the discrepancy between the model-implied and observed covariance matrices (Bollen, 2002). Substituting the CFA estimates $\hat{\boldsymbol{\Upsilon}}_1$ into Equations (5.2) and (5.5) yields the estimated factor scores $\hat{\boldsymbol{\eta}}_i$ and the associated error covariance structure $\hat{\boldsymbol{\Sigma}}_{ee}$, which serve as inputs for the second stage.

5.3. Second stage

In the second stage, the structural equation in Equation (2.6) is solved with the G-estimation approach introduced in the identification section with an additional two-stage method-of-moments (2SMM) correction for latent variable interactions, as proposed by Wall and Amemiya (2000). In our implementation, we adopt their modified 2SMM estimator and incorporate an additional ridge-inspired variance term to improve numerical stability (Hoerl & Kennard, 2000).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

5.3.1. Solving the G-Estimation Equation

In the identification section, we derived the orthogonality condition in Equation (4.2). Here, we show how to solve it for the observed outcome model in Equation (2.7). We begin by specifying the optimal weights introduced in Equation (4.3).

For a binary treatment that is randomized independently of the covariates, the treatment weight simplifies to the centered treatment indicator:[9]

$$a_{r,i} = R_i - \mathrm{E}[R]. \tag{5.6}$$

The mediator weight in this setting is given by the treatment-induced shift in the conditional expectation of the mediator:

$$a_{m,i} = \left(\mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 1\right] - \mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 0\right]\right) \cdot a_{r,i}. \tag{5.7}$$

To construct this weight, we require the difference in the expected value of the mediator conditional on the covariates under treatment and control. This quantity can be obtained by fitting the mediator model in Equation (2.8) and computing the predicted mediator values under both treatment conditions. Considering for example the specification from Equation (2.9), we have

$$\begin{aligned}
\mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 1\right] &= \gamma_r + \gamma_{x1}\eta_{X_{1i}} + \gamma_{x2}\eta_{X_{2i}} + \gamma_{x1r}\eta_{X_{1i}} + \gamma_{x2r}\eta_{X_{2i}}, \\
\mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 0\right] &= \gamma_{x1}\eta_{X_{1i}} + \gamma_{x2}\eta_{X_{2i}},
\end{aligned}$$

so that the difference simplifies to

$$\mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 1\right] - \mathrm{E}\left[\eta_{M_i} \mid \boldsymbol{\eta}_{\mathbf{X}_i}, R_i = 0\right] = \gamma_r + \gamma_{x1r}\eta_{X_{1i}} + \gamma_{x2r}\eta_{X_{2i}}. \tag{5.8}$$

In addition, we introduce pseudo-weights for the covariate terms, given by the transformed baseline covariates $q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})$ themselves. Since the nuisance term is modeled as a linear combination of these transformed covariates, we can define an extended weight matrix $\mathbf{W}$ that incorporates both the structural weights and the pseudo-weights.[10] Each row of $\mathbf{W}$ is defined as

$$\mathbf{w}_i = \left(a_{r,i} \;\; a_{m,i} \;\; q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})\right).$$

The estimating equation then becomes

$$\mathbf{W}^\top \boldsymbol{\zeta}_Y = 0, \tag{5.9}$$

which can be solved in a single step by replacing expectations with their sample analogues and solving for the structural parameters. This yields the closed-form, sample-based estimator

$$\bar{\boldsymbol{\alpha}} = (\mathbf{W}^\top \boldsymbol{\Xi}_y)^{-1} \mathbf{W}^\top \boldsymbol{\eta}_Y, \tag{5.10}$$

which corresponds to a linear IV estimator, with $\mathbf{W}$ serving as the matrix of residual (structural) instruments.

5.3.2. Factor Score Correction

The structural form of the outcome factor score η_{Y_i} may depend on polynomial terms of the covariate and mediator scores $\boldsymbol{\eta}_{\text{pred},i}$. Any interactions or higher-order terms involving $\boldsymbol{\eta}_i$ must be corrected using the

[9] In the general formulation, the optimal weight additionally involves the covariance matrix $\Omega_{\boldsymbol{\eta}_{\mathbf{X}_i}}$, which accounts for heteroskedasticity of the outcome error across covariate values. For simplicity, we set it to the identity matrix, yielding a consistent, though not necessarily fully efficient, estimator.

[10] For a non-linear nuisance term, defining a common weight function is not possible; instead, an iterative estimation procedure is required, alternating between the estimation of the nuisance term and the weights for the structural component.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

moments of the measurement error $\mathbf{e}_i$. We define these moments of $\mathbf{e}_i$ needed for the correction as $\boldsymbol{\Upsilon}_2$. Assuming normally distributed measurement errors $\boldsymbol{\epsilon}_i$, they reduce to linear combinations of elements in the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{ee}$.

Instead of using the naive sample-based estimator $\bar{\boldsymbol{\alpha}}$ in Equation (5.10), which ignores measurement error in the factor scores, we estimate the true population parameter

$$\boldsymbol{\alpha} = \mathrm{E}[\mathbf{w}_i^\top \boldsymbol{\xi}_{Y_i}]^{-1}\, \mathrm{E}[\mathbf{w}_i^\top \eta_{Y_i}] \tag{5.11}$$

with the two stage method-of-moments estimator from Wall and Amemiya (2000)

$$\hat{\boldsymbol{\alpha}} = \hat{\mathbf{M}}^{-1}\hat{\mathbf{m}}, \tag{5.12}$$

where

$$\begin{aligned} \hat{\mathbf{M}} &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{M}(\hat{\boldsymbol{\eta}}_{\text{pred},i}, \hat{\boldsymbol{\Upsilon}}_2), \\ \hat{\mathbf{m}} &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2) \end{aligned} \tag{5.13}$$

are computed from the estimated factor scores $\hat{\boldsymbol{\eta}}_i$ and the estimated measurement error moments $\hat{\boldsymbol{\Upsilon}}_2$. The moment functions $\mathbf{M}(\cdot)$ and $\mathbf{m}(\cdot)$ are defined such that their conditional expectations given the true factor scores $\boldsymbol{\eta}_i$ and the treatment R_i satisfy

$$\begin{aligned} \mathrm{E}\big[\mathbf{M}(\tilde{\boldsymbol{\eta}}_{\text{pred},i}, \boldsymbol{\Upsilon}_2) \mid \boldsymbol{\eta}_i, R_i\big] &= \mathbf{w}_i^\top \boldsymbol{\xi}_{Y_i}, \\ \mathrm{E}\big[\mathbf{m}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2) \mid \boldsymbol{\eta}_i, R_i\big] &= \mathbf{w}_i^\top \eta_{Y_i}, \end{aligned} \tag{5.14}$$

where $\tilde{\boldsymbol{\eta}}_i$ denotes the theoretical factor score estimator based on the true measurement parameters, so that, by iterated expectations, their means equal the population moments in Equation (5.11).

Let J denote the highest order in which any component of $\boldsymbol{\eta}_{\text{pred},i}$ appears in $\mathbf{W}$ or $\boldsymbol{\Xi}_y$. Then $\mathbf{M}$ and $\mathbf{m}$ are constructed via the expansion

$$\Big(\mathbf{M}(\tilde{\boldsymbol{\eta}}_{\text{pred},i}, \boldsymbol{\Upsilon}_2)\ \ \mathbf{m}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2)\Big) = \sum_{j=0}^{J} (-1)^j D_j(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2), \tag{5.15}$$

where the correction terms are recursively defined as

$$\begin{aligned} D_0(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2) &= \tilde{\mathbf{w}}_i^\top \begin{pmatrix} \tilde{\boldsymbol{\xi}}_{Y_i} & \tilde{\eta}_{Y_i} \end{pmatrix}, \\ D_j(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2) &= \mathrm{E}\left[D_{j-1}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2) \mid \boldsymbol{\eta}_i, R_i\right] - D_{j-1}(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2) \quad \text{for } j \in \{1, \ldots, J\}. \end{aligned} \tag{5.16}$$

Each D_j term captures contributions of order up to $2(J-j)$ in $\boldsymbol{\eta}_{\text{pred},i}$ and up to order j in the measurement error moments. The term D_0 corresponds to the uncorrected estimates, while D_j for $j > 0$ serves as a correction based on higher-order moments of $\mathbf{e}_i$. The specification for the model in Equation (2.9) with $J = 1$ is detailed in Section E. Plugging in the estimates $\hat{\boldsymbol{\eta}}_i$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ from the first stage, we obtain $\hat{\boldsymbol{\alpha}}$ in Equation (5.12).

5.3.3. Modifications for Numerical Stability

To improve efficiency and ensure that $\hat{\mathbf{M}}$ keep the corrected moment matrix well conditioned, particularly in small-sample settings, we adopt the modification of the 2SMM estimator in Equation (5.13) proposed

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

by Wall and Amemiya (2000). We separate the uncorrected estimates

$$\left(\hat{\mathbf{M}}_1 \;\; \hat{\mathbf{m}}_1\right) = \frac{1}{N}\sum_{i=1}^{N} D_0(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2),$$

from the correction terms

$$\left(\hat{\mathbf{M}}_2 \;\; \hat{\mathbf{m}}_2\right) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}(-1)^j D_j(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2).$$

Then, we define

$$\mathbf{R}_1 = \begin{pmatrix} \hat{\mathbf{M}}_1 & \hat{\mathbf{m}}_1 \\ \hat{\mathbf{m}}_1^\top & \frac{1}{N}\sum_{i=1}^{N}\hat{\eta}_{Y_i}^2 \end{pmatrix} \text{ and } \mathbf{R}_2 = \begin{pmatrix} -\hat{\mathbf{M}}_2 & -\hat{\mathbf{m}}_2 \\ -\hat{\mathbf{m}}_2^\top & \hat{\boldsymbol{\Sigma}}_{ee,k_y k_y} \end{pmatrix},$$

where $\hat{\boldsymbol{\Sigma}}_{ee,k_y k_y}$ denotes the estimated error variance of the outcome factor score, which is the last of the k latent variables in Equation (2.2). Let $\hat{\lambda}$ denote the largest eigenvalue of

$$\mathbf{R}_1^{-1}\mathbf{R}_2.$$

Then, the modified estimator is given by

$$(\check{\mathbf{M}}, \check{\mathbf{m}}) = \begin{cases} \left(\hat{\mathbf{M}}_1, \hat{\mathbf{m}}_1\right) + \left(1 - \frac{c}{N}\right)\left(\hat{\mathbf{M}}_2, \hat{\mathbf{m}}_2\right), & \text{if } \frac{1}{\hat{\lambda}} \geq 1 + \frac{1}{N}, \\ \left(\hat{\mathbf{M}}_1, \hat{\mathbf{m}}_1\right) + \left(\frac{1}{\hat{\lambda}} - \frac{1}{N} - \frac{c}{N}\right)\left(\hat{\mathbf{M}}_2, \hat{\mathbf{m}}_2\right), & \text{otherwise,} \end{cases}$$

where $c \in [0, \dim(\boldsymbol{\alpha}) + 5]$ is an empirically chosen tuning parameter (Wall & Amemiya, 2000).

While this adjustment bounds the measurement-error correction in finite samples, the resulting matrix may still be poorly conditioned, particularly in small samples or when the moments are highly collinear (e.g., if the effect modification required by Assumptions S1 is weak). To improve numerical stability when inverting the matrix, we therefore introduce an additional regularization term by adding a small variance v to the diagonal of $\check{\mathbf{M}}$, analogous to a ridge penalty in linear regression. The final estimator then becomes

$$\check{\boldsymbol{\alpha}} = (\check{\mathbf{M}} + v\mathbf{I})^{-1}\check{\mathbf{m}}. \tag{5.17}$$

6. Asymptotic Properties

We first consider the measurement model parameters, collected in $\hat{\boldsymbol{\Upsilon}} = (\hat{\boldsymbol{\Upsilon}}_1, \hat{\boldsymbol{\Upsilon}}_2)$.

Theorem 6.1. *Under Assumptions R1, R2, R3, and R4 the Z-estimator* $\hat{\boldsymbol{\Upsilon}}$ *is* $\sqrt{N}$*-consistent and satisfies*

$$\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Delta}_i + o_p\left(N^{-1/2}\right),$$

where $\{\boldsymbol{\Delta}_i\}$ *are i.i.d. with* $\mathrm{E}[\boldsymbol{\Delta}_i] = 0$ *and finite variance.*

Next, we regard the behavior of the factor score estimator $\hat{\boldsymbol{\eta}}_i$ relative to $\tilde{\boldsymbol{\eta}}_i$ defined in Equation (5.2).

Theorem 6.2. *Under Theorem 6.1,*

$$\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = \hat{\mathbf{B}}_i\,(\hat{\boldsymbol{\Upsilon}}_1 - \boldsymbol{\Upsilon}_1),$$

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

where

$$\mathbf{B}_i = \begin{pmatrix} \mathbf{H} & \mathbf{H}(\mathbf{q}_i^\top \otimes \mathbf{I}_{p-k}) & -(\mathbf{p}_i^\top \otimes \mathbf{I}_k) \end{pmatrix}$$

with

$$\mathbf{q}_i = \begin{pmatrix} 0 & \mathbf{I}_k \end{pmatrix} \mathbf{z}_i \quad \textit{and} \quad \mathbf{p}_i = \begin{pmatrix} \mathbf{I}_{p-k} & -\boldsymbol{\Lambda}_{\text{free}} \end{pmatrix} \left[\mathbf{z}_i - \begin{pmatrix} \boldsymbol{\tau}_{\text{free}} & \mathbf{0}_{k\times 1} \end{pmatrix}^\top \right].$$

Consequently, $\hat{\boldsymbol{\eta}}_i$ is $\sqrt{N}$-consistent.

Finally, we can formulate the properties of the structural parameter estimator $\hat{\boldsymbol{\alpha}}$ defined in Equation (5.12).

Theorem 6.3. *Under Assumption S1 and R1–R6, $\hat{\boldsymbol{\alpha}}$ is consistent and asymptotically normally distributed with asymptotic variance*

$$\mathbf{G}^{-1}\mathbf{S}\mathbf{G}^{-\top},$$

where

$$\mathbf{G} = \mathrm{E}\left[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top\right],$$

$$\mathbf{S} = \mathrm{Var}[\mathbf{d}_i], \qquad \mathbf{d}_i = \mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}) + \overline{\mathbf{C}}\,\boldsymbol{\Delta}_i,$$

$$\mathbf{l}(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}) = \mathbf{m}(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2) - \mathbf{M}(\boldsymbol{\eta}_{\text{pred},i}, \boldsymbol{\Upsilon}_2)\,\boldsymbol{\alpha},$$

$$\overline{\mathbf{C}} = \mathrm{E}\left[\left(\frac{\partial \mathbf{l}}{\partial \boldsymbol{\eta}_i^\top}\bigg|_{\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}} \mathbf{B}_i \quad \frac{\partial \mathbf{l}}{\partial \boldsymbol{\Upsilon}_2^\top}\bigg|_{\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}} \right)\right].$$

All proofs are provided in Section F. Theorem 6.3 also holds for the modified 2SMM estimator $\check{\mathbf{M}}^{-1}\check{\mathbf{m}}$ (Wall & Amemiya, 2000). Our adapted version in Equation (5.17), however, introduces a deliberate bias towards zero, whereby the reduction in variance increases overall efficiency in line with the standard rationale of regularization. Specifically, the bias derivation based on Singular Value Decomposition (SVD) shows that the ridge penalty v shrinks the coordinates of the true coefficient along the principal components of the data (Hastie, 2020). For a standardized principal component with unit variance, this bias simplifies to:

$$\mathrm{Bias}(\check{\alpha}_j) = \frac{v}{1+v}\alpha_j^*.$$

7. Robustness Assessment

Causal assumptions cannot be fully tested as they concern unobserved quantities. However, we can assess the vulnerability to violations of Assumption C2, following the ideas in Zheng and Zhou (2015).[11]

Let U be an unobserved confounder that interacts with the latent mediator η_M, and let this unobserved interaction, $\eta_M U$, affect the outcome with effect size δ. If this interaction is present but omitted from the estimation, the corresponding asymptotic bias of the estimator $\check{\boldsymbol{\alpha}}$ defined in Equation (5.17) is given by

$$\mathrm{Bias}_U(\check{\boldsymbol{\alpha}}) = \delta\,(\mathbf{M}')^{-1}\,\mathrm{E}\left[\mathbf{w}_i^\top \eta_{M_i} U_i\right] \tag{7.1}$$

with $\mathbf{M}' = \mathrm{E}[\mathbf{w}_i^\top \boldsymbol{\xi}_{Y_i}] + v\mathbf{I}$. The second term $\mathrm{E}[\mathbf{w}_i^\top \eta_{M_i} U_i]$ captures the dependence between the weights and the interaction term. Its ℓ-th element can be decomposed as

$$\mathrm{E}[w_{i\ell}\,\eta_{M_i} U_i] = \mathrm{Cov}(w_{i\ell}, \eta_{M_i} U_i) + \mathrm{E}[w_{i\ell}]\,\mathrm{E}[\eta_{M_i} U_i].$$

[11] For sensitivity analyses regarding Sequential Ignorability, see, for example, Imai, Keele, and Yamamoto (2010) and VanderWeele (2010).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

In our setting, this simplifies to the covariance $\mathrm{Cov}[w_{i\ell}, \eta_{M_i} U_i]$. This simplification arises because the elements of $\mathbf{W}$ are centered by construction: the structural weights $a_{r,i}$ and $a_{m,i}$ must fulfill the centering condition to serve as residualized instruments, and we additionally center the covariate terms $q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})$. Consequently, $\mathrm{E}[w_{i\ell}] = 0$ for all ℓ, which causes the second term of the decomposition to vanish.

Bias emerges if this covariance is non-zero. The plausibility of such a higher-order nonlinear dependence is generally impossible to evaluate empirically. Quantifying this bias would require specifying the exact functional form of the dependency between the weights, the mediator, and the unobserved confounder. This requirement precludes the use of standard sensitivity analysis techniques, such as those by Imai, Keele, and Yamamoto (2010), which rely on bounded parameters (e.g., the correlation between residuals $\rho \in [-1, 1]$) to provide a finite, interpretable space for simulation. In contrast, our bias term depends on a higher-order interaction involving $\mathbf{u}$ that lacks such natural bounds, as the scale of $\mathbf{u}$ and the functional form of its interaction with $\mathbf{w}_j$ and $\boldsymbol{\eta}_M$ are entirely unconstrained. Consequently, unlike traditional sensitivity analysis, which explores a known range of possible violations, the magnitude of this third-order dependence remains practically unquantifiable without making assumptions about the joint distribution of U that are rarely justifiable.[12]

Instead, we focus on the inverse matrix $(\check{\mathbf{M}}')^{-1}$ in Equation (7.1), which is a calculable and informative quantity. This matrix serves as a scaling factor that determines the extent to which an unobserved interaction is amplified and distributed across the parameter vector. Specifically, the diagonal elements reflect the direct vulnerability of each coefficient to the omitted moderation, whereas the off-diagonal elements indicate the degree of bias transmission between covariates.

The influence of $\check{\mathbf{M}}'$ is fundamentally tied to the stability of its inversion: as the matrix approaches singularity, the entries of the inverse grow significantly, boosting the impact of unobserved effect modification. This allows us to draw several links to our structural specifications. First, Assumption S1 ensures that $\check{\mathbf{M}}'$ has full rank and is thus invertible, a condition that fails if the mediator weight becomes collinear with the treatment. weight. For example, if $\gamma_{x1r} = \gamma_{x2r} = 0$ in Equation (5.8), the mediator weight reduces to $\gamma_r a_{r,i}$, creating collinearity with the treatment weight, leaving the G-estimation equation without a unique solution. Consequently, a stronger influence of the covariate-treatment interactions on the mediator not only ensures identification but also serves as a structural defense against bias by improving the conditioning of the system. Second, the numerical stability modifications mitigate the risk of extreme bias in settings of weak identification by bounding the elements of the inverse matrix, although the ridge term cannot restore identification.

To quantify the sensitivity due to structural instability, Zheng and Zhou (2015) suggest considering the condition number, which for RAPSEM is given by:

$$\kappa(\check{\mathbf{M}}') = \|\check{\mathbf{M}}'\| \cdot \|\check{\mathbf{M}}'^{-1}\| = \frac{\sigma_{\max}(\check{\mathbf{M}}')}{\sigma_{\min}(\check{\mathbf{M}}')} \tag{7.2}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ represent the maximum and minimum singular values, respectively. For improved interpretability, we propose the reciprocal condition number, $1/\kappa \in [0, 1]$, as a robustness index: A value near 1 indicates a well-conditioned system where the potential influence of unobserved effect modification is naturally attenuated. Conversely, a value approaching zero signals that the identification is fragile, warning that the causal estimates are highly sensitive to even minor violations of Assumption C2.

8. Simulation Studies

We evaluate the proposed method through two simulation studies. The first compares the response of standard SEM and RAPSEM to unobserved confounding, while the second investigates RAPSEM's

[12] Zheng and Zhou (2015) bound the bias by the maximum or total violation across interaction terms. However, the coefficients of these bounds again depend on the joint distribution of the unobserved confounder and are computable only in simple models.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

statistical power. In the latter, we focus on the structural requirements of Assumption S1, specifically exploring the relationship between the covariate-treatment interaction strength and the robustness coefficient. The entire code is available on GitHub.

8.1. Data Generation

We generated data based on Equation (2.9) with an additional unobserved confounding variable u_i

$$\begin{aligned}\eta_{y_i} &= 0.224 \cdot r_i + \theta_m \cdot \eta_{m_i} + 0.240 \cdot \eta_{x1_i} + 0.240 \cdot \eta_{x2_i} + \delta_u \cdot u_i + \delta_{um} \cdot u_i \eta_{m_i} + \zeta_{y_i}, \\ \eta_{m_i} &= 0.316 \cdot r_i + 0.277 \cdot \eta_{x1_i} + 0.277 \cdot \eta_{x2_i} + \gamma_{x1r} \cdot r_i \eta_{x1_i} + \gamma_{x2r} \cdot r_i \eta_{x2_i} + \delta_u \cdot u_i + \zeta_{m_i},\end{aligned} \tag{8.1}$$

where the parameters θ_r, $\boldsymbol{\theta}_x$, γ_r, $\boldsymbol{\gamma}_x$ were fixed to standardized effect sizes corresponding to R^2 of 5%, 5%, 15% and 20%, respectively, across all simulations. When varying the other effect sizes, we employed Cohen's d values of 0.2, 0.5, 0.65, and 0.8. These represent small, medium, and large effects (Cohen, 1977) and correspond to R^2 values of 1%, 6%, 10%, and 14%, respectively. For each parameter combination, we generated 100 datasets, applying the same seeds across conditions to ensure comparability.

The treatment R was sampled from a discrete uniform distribution on $\{-1, 1\}$.[13] Covariates η_{x_1} and η_{x_2} were drawn from a multivariate standard normal distribution with correlation $\rho = 0.3$, while the confounder U followed an independent standard normal distribution. Residuals ζ_m and ζ_y were sampled from $N(0, \psi)$, where the residual variance $\psi = 1 - R^2$ was determined by population parameters to maintain a theoretical unit variance for η_m and η_y.

Each latent variable was measured by three indicators with parameters randomized per dataset to enhance generalizability. Item-specific reliabilities rel_j were sampled from $U(\text{rel} - 0.1, \text{rel} + 0.1)$, where rel denotes the reliability level defined by the simulation condition. Factor loadings were set to $\lambda_j = \sqrt{\text{rel}_j}$ with corresponding residual variances of $1-$
rel_j, and item intercepts ν_j were drawn from $U(-1, 1)$.

8.2. Simulation Study 1: Robustness against Unobserved Confoundedness and Effect Modification

We compared three estimation methods to evaluate their performance in the presence of unobserved confounding and effect modification:

1. **RAPSEM**: Our proposed framework utilizing latent G-estimation implemented as R package `rapsem`
2. **Uncorrected**: A two-stage latent regression approach using the predictor matrix $\boldsymbol{\Xi}_y$ in place of the weight matrix $\mathbf{W}$ in Equation (5.10). This represents factor score regression with the same correction for measurement error but without G-estimation corrections.
3. **Lavaan**: A standard `lavaan` implementation utilizing a two-group model to capture treatment-covariate interactions, with factor loadings and intercepts constrained to equality across groups.

Data conditions

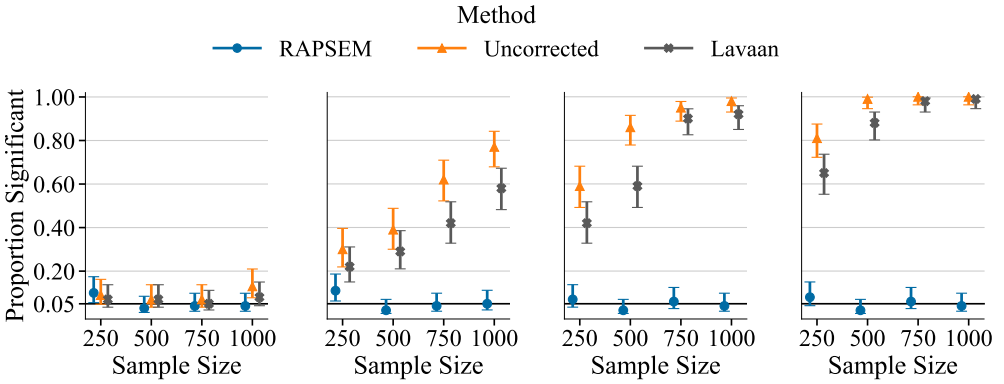
The confounding effect sizes δ_u and δ_{um} were varied across different sample sizes N according to the conditions in Table 2. When varying the interaction effect δ_{um}, we held the main effect δ_u at its maximum (0.374) to respect the hierarchical principle that interactions manifest alongside their corresponding main effects. To evaluate the type I error rate, the mediator-outcome effect θ_m was fixed at zero. Reliability rel was set to 0.9 and interaction parameters $\boldsymbol{\gamma}_{\mathbf{x}xr}$ were fixed at 0.232, representing a substantial treatment-covariate interaction ($R^2 = 14\%$) to ensure high statistical power of RAPSEM.

[13] Note that the CDE thus corresponds to $\frac{1}{2}\theta_r$.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Table 2: Varied parameters in Study 1: unobserved confounding effect size (δ_u), unobserved-confounder-mediator interaction effect size (δ_{um}), and sample size (N).

Variable	**Levels**			
δ_u	0.100	0.245	0.316	0.374
δ_{um}	0.100	0.245	0.316	0.374
N	250	500	750	1000



(a) $\delta_u = 0.100$ (b) $\delta_u = 0.245$ (c) $\delta_u = 0.316$ (d) $\delta_u = 0.374$

Figure 2: False positive rates of $\check{\theta}_m$ with Wilson score 95% confidence intervals under varying confounding levels and sample size.

Results

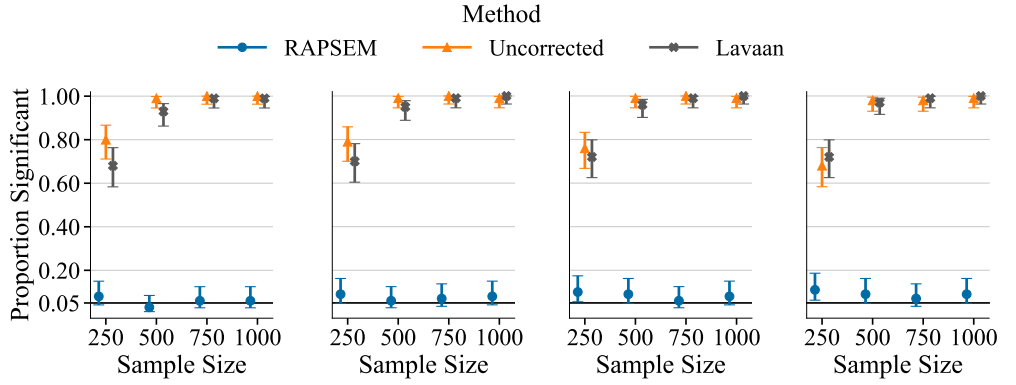
False positive rates (Figure 2) were severely inflated for standard SEM approaches (Uncorrected and Lavaan) in the presence of unobserved confounding. In moderate to high confounding conditions ($\delta_u \geq 0.316$), the Type I rates exceeded 80% and approached 100% as the sample size increased. In contrast, RAPSEM consistently maintained the false positive error rate around the nominal 5% level across all settings.

Effect modification (δ_{um}) had minimal influence on the type one error rates (Figure 3). For standard SEM, these rates were, interestingly, slightly reduced; for RAPSEM, they only marginally increased, such that the false positive rate stayed close to 5%. This robustness to the presence of unobserved effect modification provides further evidence of RAPSEM's validity, consistent with findings in Brandt (2020).

RAPSEM estimates of θ_m remained centered around the true value of zero, whereas standard SEM estimates shifted upward with increasing δ_u, indicating substantial positive bias (Figure 7 in Section I). As expected, this bias is not reflected in typical SEM model fit indices, which remained within acceptable ranges despite the lack of sequential ignorability (Figure 8 in Section I). RAPSEM's sampling distributions also reveal increased variance in small samples, reflecting an inherent trade-off between robustness and small-sample efficiency, with the gap narrowing at $N = 1000$.

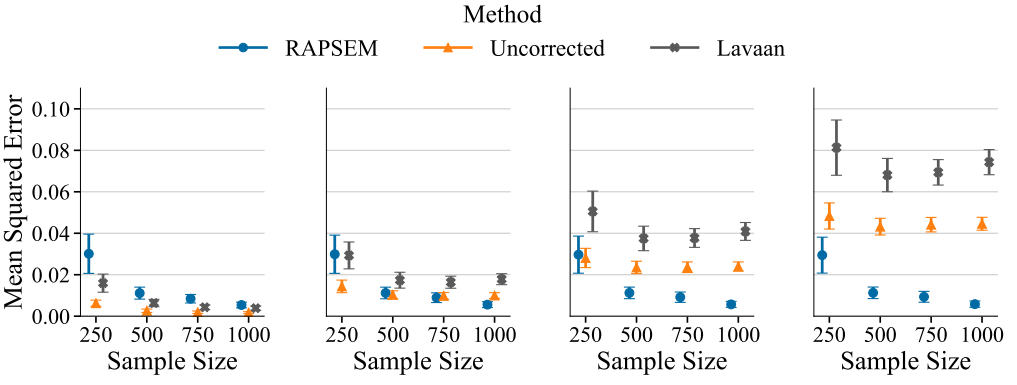
These two properties jointly determine MSE behavior (Figure 4). In the absence of confounding, RAPSEM's higher variance dominates, yielding a larger MSE than standard SEM at small sample sizes and reaching parity only at $N = 1000$. In all confounding scenarios, however, the bias term causes standard SEM's MSE to grow significantly with sample size, eventually exceeding that of RAPSEM.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.



(a) $\delta_{um} = 0.100$ (b) $\delta_{um} = 0.245$ (c) $\delta_{um} = 0.316$ (d) $\delta_{um} = 0.374$

Figure 3: False positive rates of $\check{\theta}_m$ with Wilson score 95% confidence intervals under varying confounding levels and sample size.



(a) $\delta_u = 0.100$ (b) $\delta_u = 0.245$ (c) $\delta_u = 0.316$ (d) $\delta_u = 0.374$

Figure 4: Mean squared error of $\check{\theta}_m$ with Normal approximation 95% confidence intervals under varying levels of confounding and sample size.

8.3. Simulation Study 2: Statistical Power

Building on the findings of Study 1, we further investigated the efficiency–robustness trade-off by identifying the conditions under which RAPSEM achieves adequate statistical power. Additionally, we evaluated the framework's robustness to violations of C2 using the condition number κ, as introduced in the sensitivity analysis section.

Data Conditions

We identified four primary factors that determine the power and sensitivity of RAPSEM: the mediator effect size $\check{\theta}_m$, the strength of the covariate–treatment interaction $\boldsymbol{\gamma}_{xr}$, the reliability of the indicators rel, and the sample size N. These parameters were systematically varied according to the levels specified in Table 3, without unobserved confounding ($\delta_u = 0$). In Table 4, we present the theoretical F-statistics for

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

the joint test of the interaction terms calculated using the nested model comparison formula:

$$F = \frac{\Delta R^2_{int}/m}{(1 - R^2_{total})/(N - k - 1)} \tag{8.2}$$

where ΔR^2_{int} is the variance uniquely explained by the $m = 2$ interactions, $R^2_{total} = 35\% + \Delta R^2_{int}$ is the total variance explained by the full model (including a 15% treatment effect and 20% covariate main effects), N is the sample size, and $k = 5$ represents the total number of predictors. The robustness $1/\kappa$ was calculated in additional runs with rel = 0.9 and $\theta_m = 0.0$ to mirror the first study, as the mediation effect size does not influence this coefficient. δ_u and δ_{um} were set to zero.

Table 3: Varied parameters in Study 2: levels of mediator effect size $\breve{\theta}_m$, covariate–treatment interaction effect size $\boldsymbol{\gamma}_{xr}$, indicator reliability rel, and sample size N.

Variable	**Levels**			
θ_m	0.100	0.243	0.316	0.371
$\boldsymbol{\gamma}_{xr}$	0.062	0.152	0.196	0.232
rel	0.5	0.7	0.9	
N	250	500	750	1000

Table 4: Theoretical F-statistics for the joint test of covariate–treatment interactions across conditions in Study 2.

Sample Size (N)	**Interaction Effect Size** ($\boldsymbol{\gamma}_{xr}$)			
	0.062	0.152	0.196	0.232
250	1.9	12.4	22.2	33.5
500	3.9	25.1	44.9	67.8
750	5.8	37.8	67.6	102.1
1000	7.8	50.5	90.4	136.4

Results

The power to detect a non-zero mediation effect (Figure 5) increased systematically with sample size (N), measurement reliability (rel), and the strength of the covariate–treatment interaction (γ_{xr}). Overall, the findings suggest that large-scale studies are essential. A sample size of $N = 250$ was insufficient for detecting even large effects, while $N = 500$ serves as a critical threshold where success becomes highly contingent on the remaining parameters.

Regarding mediation effect sizes, power remained insufficient for small effects ($\breve{\theta}_m = 0.100$) across all scenarios, failing to reach the 80% threshold even at $N = 1000$. While power naturally improves as $\breve{\theta}_m$ increases, reaching an acceptable level for medium-to-large effects depends on the quality of the measurement and the interaction strength.

Low reliability (rel = 0.5) significantly hinders the detection of moderate effects, often keeping power well below acceptable thresholds. While the differences between reliability levels of 0.7 and 0.9 are less pronounced, higher reliability consistently yields marginal gains in power.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Ultimately, the covariate–treatment interaction emerges as the most critical determinant of model performance apart from sample size. The strongest interaction levels consistently yield the highest power. Conversely, a small interaction effect ($\gamma_{xr} = 0.062$) fails to reach 80% power in any tested scenario.

This interaction is not merely a matter of statistical power but a fundamental requirement for model identification. We could not include a zero interaction scenario in our simulations because, without this effect, the weight matrix becomes collinear and the inverse fails to converge, as detailed in the sensitivity analysis section.

This structural dependency is empirically demonstrated by the robustness coefficient $1/\kappa$ (Figure 6). While $1/\kappa$ remains approximately constant across sample sizes, it increases sharply with the interaction effect size, rising from approximately 0.006 for a small interaction to 0.08 for a large interaction. By a standard rule of thumb (Cheney & Kincaid, 2007), these values correspond to a loss of about two digits and one digit of precision in the calculation of $\check{\boldsymbol{\alpha}}$, respectively, with the latter indicating greater structural robustness. This behavior confirms that stable and reliable RAPSEM estimation depends fundamentally on the interaction effect size γ_{xr}.

9. Empirical Example

We further illustrate the differences between our proposed method and the two alternative approaches investigated in the simulation study through an empirical example from educational psychology. We use publicly available data by Russ et al. (2026), who investigated the effects of various self-study strategies (generative learning activities) on cognitive learning outcomes in an RCT.

We emphasize the illustrative nature of this analysis. The modeling approach was adapted to facilitate a methodological comparison rather than to provide new evidence for or against the specific psychological relationships studied. Consequently, our modeling choices deviate substantially from those of the original authors. While their specifications were based on domain expertise and prior literature, we adapted the model specifically to demonstrate how diverging results can emerge in real-world applications when using different statistical frameworks.

Data were collected in a classroom setting involving 30 seventh- and eighth-grade classes from 11 secondary schools in Southwest Germany, with a resulting sample size of $N = 590$ students. Contrary to the original study, which utilized manifest z-scores in a multilevel model, we employ SEM with a measurement model but without a hierarchical structure as implemented in RAPSEM. This involves omitting the nested structure of the data, specifically the clustering within schools and classes, a choice that leads to results differing from the original findings. This departure is useful for our illustration because, while correlated random intercepts in a multilevel model can account for some unobserved confounding, our approach is designed to address such confounding through the structural model itself. By removing the hierarchical layers, we can more clearly highlight the impact of unobserved confounding and demonstrate how the different methods studied in this paper respond to it. Consistent with the original study, missing data were handled via multiple imputation by chained equations (MICE) under the assumption of data missing at random (MAR), and all results are pooled across 20 imputed datasets.

While the original experiment included four treatment conditions, we restricted our analysis to two: the control condition, in which students restudied the lecture content, and the treatment condition, in which they explained the content to a fictitious, non-present peer via voice messages. This results in a binary but unbalanced treatment variable, with only 150 students in the control group. This imbalance is noteworthy because it reduces the overall statistical power to detect a mediation effect compared to a balanced design of equal total size.

The outcome, conceptual knowledge, was measured using 15 binary items. To facilitate factor score estimation, we aggregated these items into three parcels, with item assignment randomized for each imputation. We selected task interest as the mediator, which was identified as a key predictor of conceptual

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

knowledge in the original study, modeling it as a continuous latent variable measured by two items on a 4-point Likert scale.

While the original study collected numerous control variables, we selected three specific covariates for this illustration: students' interest in physics, conscientiousness in physics, and perceived teacher support (each measured with four items all on a 4-point Likert scale). These specific variables were selected because they exhibited significant individual and joint interaction effects with the treatment on the mediator (see Table 5). We additionally included perceived cognitive activation (six items combined to three parcels) as a confounder, given its significant main effects on both the mediator and the outcome. Notably, while the original study's pre-test cognitive assessment would be a primary confounder to include, we deliberately omitted it to accentuate the influence of the other factors in our methodological comparison.

Table 5: Testing individual and joint treatment-covariate interaction effects on the mediator: reported are the interaction coefficients: interaction coefficient γ_{xr}, partial F-statistic, and p-value for the comparison between models with and without interaction terms.

Covariate	**Effect Size**	F**-Statistic**	p**-value**
interest in physics	-0.231	6.6	0.011
conscientiousness in physics	-0.220	5.4	0.020
perceived teacher support	-0.290	6.2	0.013
All Combined	—	5.1	0.002

We compared two model specifications:

- **Full Model M1**: Included all the aforementioned variables.
- **Reduced Model M2**: Omitted the confounder perceived cognitive activation, artificially treating it as an unobserved confounder. This reflects a scenario with unobserved confounding where a critical dimension of the learning environment might be overlooked during study design.

9.1. Results

The empirical results (Table 6) highlight the distinct trade-offs between RAPSEM and standard SEM (Uncorrected and Lavaan).

Table 6: Comparison of mediation effects ($\breve{\theta}_m$) across estimation methods. Values represent point estimates and 95% confidence intervals ($\breve{\theta} \pm 1.96\, \breve{\sigma}_{\text{pooled}}$) aggregated across imputations.

Model	**Estimation Method**		
	G-Estimator	**Uncorrected**	**Lavaan**
M1 (Full)	−0.06 [−0.75, 0.62]	0.06 [−0.10, 0.22]	0.14 [0.00, 0.28]
M2 (Reduced)	−0.08 [−0.71, 0.55]	0.09 [−0.01, 0.19]	0.17 [0.05, 0.28]

The most critical finding concerns comparative robustness. For the standard approaches, omitting confounders produces a shift in the $\breve{\theta}_m$ point estimate, indicating susceptibility to omitted variable bias: the estimates of the standard factor score regression and the `lavaan` group model increase from 0.06 to 0.09 and from 0.14 to 0.17, respectively, i.e., by about half of their standard errors. RAPSEM estimates,

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

by contrast, remain stable relative to their uncertainty, changing from −0.06 to −0.08, i.e., by less than 0.1 standard errors.

The substantially wider confidence intervals for RAPSEM reflect greater variance relative to the standard models, consistent with our simulation findings. Given that the effective sample size is $N_{\text{eff}} = (4 \cdot N_{\text{treat}} \cdot N_{\text{control}})/N = 447$ and McDonald's ω ranges from 0.63 to 0.86 (Table 9 in Section J), measurement quality is acceptable. The excess variance is therefore primarily attributable to weak instruments. The detection of multiple interacting covariates in a dataset not originally designed to capture effect modification suggests that treatment-covariate interactions may be a realistic expectation rather than an edge case. Nevertheless, the combined F-statistic of 5.13 falls below the threshold of $F > 10$ suggested by Stock et al. (2002). Indeed, comparing with Simulation Study 2, this value is only marginally above the theoretical one for $N = 500$ with the smallest interaction effect size ($R^2 = 1\%$) in Table 4. This places the application in the low-power regime illustrated by the dark-blue markers in Figure 5, which fail to reach the conventional power threshold of 0.8. Accordingly, also the small estimated robustness coefficient ($1/\kappa = 0.007$) closely tracks our theoretical expectations (see Figure 6).

Table 7: $\breve{\theta}_m$ estimates under artificially increased instrument strength: The parameter b represents the artificial boost applied to the interaction between treatment and task interest within the mediator equation, F_{interact} denotes the F-statistic for all interaction terms, and $1/\kappa$ is the robustness parameter..

b	F_{interact}	$\hat{\theta}_m$ [95 % CI]	$\hat{\kappa}$	$1/\hat{\kappa}$
−0.2	11.3	−0.03 [−0.38, 0.31]	52.9	0.019
−0.4	21.0	−0.02 [−0.28, 0.23]	34.7	0.029
−0.6	34.2	−0.02 [−0.22, 0.18]	28.2	0.035

To assess how results might change if the requirements for instrument strength were fully met, we artificially increased the influence of the interaction between treatment and task interest by adding the term $b \cdot (\text{treat} \times \text{task interest})$ to the original mediator. As the magnitude of this interaction increased (Table 7), the confidence intervals for RAPSEM shrank substantially.[14], from a width of 0.69 at $F_{\text{interact}} = 11.3$ to 0.40 at $F_{\text{interact}} = 34.2$, while the point estimates remained close to zero. While $1/\kappa$ increases alongside instrument strength, it remains numerically small. Even with large interactions, it does not approach the theoretical upper bound of 1.

This exercise confirms that the wide intervals observed in our empirical example are not a fundamental flaw of the RAPSEM algorithm itself, but rather a direct consequence of the low-power regime inherent in the original data. When supported by a sufficiently strong instrument, RAPSEM approaches the precision of standard SEM while maintaining its unique advantage of robustness against omitted variable bias.

Taken together, these results empirically demonstrate that RAPSEM requires larger samples and stronger interactions to achieve high precision, but offers estimates that are robust to the omission of observed confounders. In educational research settings where variables such as cognitive activation or personality traits frequently go unmeasured, this robustness represents a meaningful safeguard against bias that standard SEM does not provide.

10. Discussion

This article introduces a latent variable model for mediation analysis that is robust to unobserved confounding. The proposed approach builds on the G-estimation framework within a rank-preserving

[14]While the absolute effect size also increases in this scenario, this is a direct consequence of modifying the mediator's distribution rather than a lack of estimator stability.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

model of Ten Have et al. (2007) and extends the general formulation of Zheng and Zhou (2015) by incorporating a two-stage method of moments for polynomial structural equation models (Wall & Amemiya, 2000). This integration enables the identification of mediation effects of latent variables under the No Unobserved Effect Modification assumption, rather than the stronger Sequential Ignorability assumption.

By relying on structural assumptions to identify the causal effects, the *G*-estimation approach is semi-parametric. While no specific distributional form is assumed for the error terms, the framework requires linear models for the outcome and mediator, subject to Assumption S1. Consequently, it does not provide the general, nonparametric identification established by Pearl (2009, 2012) via the Mediation Formula or by Imai, Keele, and Tingley (2010), which can be seen as a trade-off to avoid Sequential Ignorability. Indeed, if Sequential Ignorability is actually satisfied, adopting our framework is suboptimal, as the additional constraints result in a loss of statistical power and less efficient estimates compared to methods that fully exploit those stronger ignorability conditions.

We establish the consistency and asymptotic normality of the resulting estimator and evaluate its performance in simulation studies. Our results show that while standard methods exhibit substantial bias under unobserved confounding, RAPSEM remains unbiased. It achieves acceptable power for medium-to-large effects at sample sizes of $N > 500$, provided there are significant interaction effects between covariates and treatment on the mediator. These results align with previous findings (e.g., Brandt, 2020; Zheng et al., 2015). Here, we found that an increased reliability of the indicator variables will also improve power.

Given these results, our findings suggest that RAPSEM offers a promising framework for robustly identifying mediation effects in the presence of unobserved confounding, particularly in large-scale intervention studies. The need for relatively large sample sizes to achieve well-powered estimation is a limitation, but it also reflects the broader challenges of causal identification. Without strong assumptions such as Sequential Ignorability, mediation effects are difficult to identify, particularly in small-sample settings or when indicator variables have low reliability. In this sense, the reliance on larger datasets highlights an important practical consideration: findings from small studies may provide only limited evidence about mediation effects.

Beyond sample size, a critical point for the use of RAPSEM is the requirement for effect modification of the mediator to serve as a structurally emerging instrument, a condition that can be challenging to meet. However, this requirement should be seen in the light of the trade-off it provides. Sequential Ignorability requires the exhaustive identification and precise measurement of all confounders to avoid bias. In contrast, our approach only requires finding at least one significant mediator-covariate interaction. Importantly, the interaction effect does not necessarily need to be causal itself, and the instrument strength can be accumulated from different covariates.

Furthermore, a significant advantage of this framework is that the strength and relevance of these variables can be empirically tested using standard diagnostic statistics for instruments. This provides a more transparent and falsifiable alternative to the inherently untestable assumption of sequential ignorability, which can lead to significant hidden bias if violated. While an interaction was present in our empirical example, the effect was small, resulting in a weak instrument. This is unsurprising given that the study was not originally designed with this requirement in mind. Future studies can directly target this by proactively identifying and collecting data on variables hypothesized to interact with the treatment.

On the methodological side, the RAPSEM framework already provides the theoretical basis for implementations involving multiple treatments, treatment–mediator interactions, and non-linear measurement models, e.g., capturing dichotomous mediators or covariates. In contrast, adapting the framework for dichotomous outcomes requires more than a simple change to the measurement model, as the causal estimands and the identification strategy must be specifically adapted for non-linear transformations. Such a generalization (e.g., along the lines of Dukes & Vansteelandt, 2018) represents a promising direction for future research. Further expansions may include higher-order polynomials and interaction terms among

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

latent variables; for instance, introducing a nonlinear specification for baseline effects can increase power when linearity assumptions are violated (Brandt, 2020; Faleh et al., 2026). These enhancements would broaden the method's applicability and strengthen its ability to capture complex causal structures.

Ultimately, as actual randomized trials in this context remain scarce, extending this approach to address confounding within observational data is a promising direction. In a similar vein, the extension of the RPM formulation to latent multilevel or longitudinal variable models can open new opportunities to broaden the flexibility of the approach. Such extensions, however, will require a very thorough evaluation and adaptation of the underlying (causal) assumptions.

Acknowledgments. We thank Samuel Merk for his insights on our method and for connecting us with Heike Russ and Andreas Lachner. We extend our gratitude to Heike Russ and Andreas Lachner for sharing their high-quality open-access data and reviewing the manuscript. We also thank Stijn Vansteelandt and Yves Rosseel for their valuable feedback and suggestions.

Funding Statement. This research was funded by the German Research Foundation (DFG) under grant BR 5175/2-1. The authors also acknowledge support from the state of Baden-Württemberg through bwHPC and from the DFG through grant INST 35/1597-1 FUGG.

Competing Interests. The authors declare no competing interests.

Data Availability Statement. Replication data and code can be found at https://github.com/PsychometricsMZ/rapsem.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

AI Usage. Adhering to the ethical guidelines for generative AI in academic research by Porsdam Mann et al. (2024), the authors acknowledge the use of Anthropic 4, Gemini 3, and Opus 5 in this work. These tools were employed to enhance the readability of the manuscript and to support code validation and testing, as well as plot creation. Every author has contributed substantially to the research, has verified the accuracy of the content, and accepts full responsibility for the integrity of the final work.

Author Contributions. Conceptualization: S.M., H.B. Methodology: S.M., R.F. Software, data curation, formal analysis, investigation, resources, visualization, writing original draft, project administration: S.M. Editing, validation: H.B., R.F. Funding acquisition, supervision: H.B. All authors approved the final submitted draft.

References

Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in Medicine*, *27*(8), 1282–1304. https://doi.org/https://doi.org/10.1002/sim.3016

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Boden, M. T., John, O. P., Goldin, P. R., Werner, K., Heimberg, R. G., & Gross, J. J. (2012). The role of maladaptive beliefs in cognitive-behavioral therapy: Evidence from social anxiety disorder. *Behaviour Research and Therapy*, *50*(5), 287–291. https://doi.org/https://doi.org/10.1016/j.brat.2012.02.007

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*(Volume 53, 2002), 605–634. https://doi.org/https://doi.org/10.1146/annurev.psych.53.100901.135239

Brandt, H. (2020). A more efficient causal mediator model without the no-unmeasured-confounder assumption. *Multivariate Behavioral Research*, *55*(4), 531–552. https://doi.org/10.1080/00273171.2019.1656051

Brandt, H. (2024). Causal definitions versus casual estimation: Reply to valente et al. (2022). *Psychological Methods*, *29*(3), 589–602. https://doi.org/10.1037/met0000544

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism?(don't expect an easy answer). *Journal of personality and social psychology*, *98*(4), 550.

Castella, K. D., Goldin, P., Jazaieri, H., Heimberg, R. G., Dweck, C. S., & Gross, J. J. (2015). Emotion beliefs and cognitive behavioural therapy for social anxiety disorder [PMID: 25380179]. *Cognitive Behaviour Therapy*, *44*(2), 128–141. https://doi.org/10.1080/16506073.2014.974665

Cheney, E. W., & Kincaid, D. R. (2007). Numerical mathematics and computing. Cengage Learning.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, rev*. Lawrence Erlbaum Associates, Inc.

Dukes, O., & Vansteelandt, S. (2018). A note on g-estimation of causal risk ratios. *American Journal of Epidemiology*, *187*(5), 1079–1084. https://doi.org/10.1093/aje/kwx347

Dunn, G., & Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine*, *26*(26), 4719–4745. https://doi.org/https://doi.org/10.1002/sim.2891

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions [PMID: 19608601]. *Statistical Methods in Medical Research*, *19*(3), 237–270. https://doi.org/10.1177/0962280209105014

Faleh, R., Morelli, S., & Brandt, H. (2026). *Representation learning for semiparametric causal mediation analysis under no essential heterogeneity* [Manuscript in preparation].

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model [PMID: 27739903]. *Multivariate Behavioral Research*, *51*(5), 681–697. https://doi.org/10.1080/00273171.2016.1224154

Frölich, M., & Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(5), 1645–1666. https://doi.org/10.1111/rssb.12232

Hastie, T. (2020). Ridge regularization: An essential concept in data science [PMID: 36033922]. *Technometrics*, *62*(4), 426–433. https://doi.org/10.1080/00401706.2020.1791959

Hayes, T., & Usami, S. (2020). Factor score regression in the presence of correlated unique factors [PMID: 31933491]. *Educational and Psychological Measurement*, *80*(1), 5–40. https://doi.org/10.1177/0013164419854492

Heckman, J. J., Urzúa, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. https://hdl.handle.net/10419/34082

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *42*(1), 80–86. https://doi.org/10.1080/00401706.2000.10485983

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, *1988*(1), i–50.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, *15*(4), 309.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, *25*(1), 51–71. https://doi.org/10.1214/10-STS321

Lee, M.-j. (2021). Instrument residual estimator for any response variable with endogenous binary treatment. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(3), 612–635. https://doi.org/10.1111/rssb.12442

Lee, M.-j., & Han, C. (2024). Ordinary least squares and instrumental-variables estimators for any outcome and heterogeneity. *The Stata Journal*, *24*(1), 72–92. https://doi.org/10.1177/1536867X241233645

Lynch, K. G., Cary, M., Gallop, R., & Ten Have, T. R. (2008). Causal mediation analyses for randomized trials. *Health Services and Outcomes Research Methodology*, *8*(2), 57–76.

Morelli, S., Faleh, R., & Brandt, H. (2025). *Rapsem: Latent g-estimation for structural equation models* [R package version 1.0]. https://github.com/sofiamorelli/rapsem

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Uncertainty in artificial intelligence: Proceedings of the seventeenth conference* (pp. 411–420). Morgan Kaufmann.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. https://doi.org/10.1214/09-SS057

Pearl, J. (2012). The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prevention Science*, *13*(4), 426–436. https://doi.org/10.1007/s11121-011-0270-1

Porsdam Mann, S., Vazirani, A. A., Aboy, M., Earp, B. D., Minssen, T., Cohen, I. G., & Savulescu, J. (2024). Guidelines for ethical use and acknowledgement of large language models in academic writing. *Nature Machine Intelligence*, *6*(11), 1272–1274. https://doi.org/10.1038/s42256-024-00922-7

Robins, J., & Hernan, M. (2008, August). Estimation of the causal effects of time-varying exposure. In *Longitudinal data analysis* (pp. 553–599). Chapman; Hall/CRC.

Robins, J. M. (2003). Semantics of causal dag models and the identification of direct and indirect effects. In P. J. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems*. Oxford University Press. https://doi.org/10.1093/oso/9780198510550.003.0007

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*(2), 143–155. https://journals.lww.com/epidem/fulltext/1992/03000/identifiability_and_exchangeability_for_direct_and.13.aspx

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Russ, H., Sibley, L., Flegr, S., Kuhn, J., Hoogerheide, V., Scheiter, K., & Lachner, A. (2026). Combining non-interactive teaching and drawing fosters conceptual knowledge but not monitoring accuracy from guided inquiry in science learning. *Journal of Educational Psychology*, *118*(3), 345–367. https://doi.org/10.1037/edu0000971

Small, D. S. (2012). Mediation analysis without sequential ignorability: Using baseline covariates interacted with random assignment as instrumental variables. https://arxiv.org/abs/1109.1070

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529. https://doi.org/10.1198/073500102288618658

Ten Have, T. R., Joffe, M. M., Lynch, K. G., Brown, G. K., Maisto, S. A., & Beck, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics*, *63*, 926–934.

van der Vaart, A. W. (1998). M–and z-estimators. In *Asymptotic statistics* (pp. 41–84). Cambridge University Press.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*(4), 540–551. https://doi.org/10.1097/EDE.0b013e3181df191c

Vanderweele, T. J. (2011). Controlled direct and mediated effects: Definition, identification and bounds. *Scandinavian Journal of Statistics*, *38*(3), 551–563. https://doi.org/https://doi.org/10.1111/j.1467-9469.2010.00722.x

VanderWeele, T. J., & Vansteelandt, S. (2022). A statistical test to reject the structural interpretation of a latent factor model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *84*(5), 2032–2054. https://doi.org/10.1111/rssb.12555

Vansteelandt, S., & Joffe, M. (2014). Structural Nested Models and G-estimation: The Partially Realized Promise. *Statistical Science*, *29*(4), 707–731. https://doi.org/10.1214/14-STS493

Vo, T.-T., Superchi, C., Boutron, I., & Vansteelandt, S. (2020). The conduct and reporting of mediation analysis in recently published randomized controlled trials: Results from a methodological systematic review. *Journal of Clinical Epidemiology*, *117*, 78–88. https://doi.org/https://doi.org/10.1016/j.jclinepi.2019.10.001

Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the Statistical American Association*, *95*, 929–940. https://doi.org/10.1080/01621459.2000.10474283

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*(1), 47–63.

Windgassen, S., Goldsmith, K., Moss-Morris, R., & Chalder, T. (2016). Establishing how psychological therapies work: The importance of mediation analysis [PMID: 26732531]. *Journal of Mental Health*, *25*(2), 93–99. https://doi.org/10.3109/09638237.2015.1124400

Zheng, C., & Zhou, X.-H. (2015). Causal mediation analysis in the multilevel intervention and multicomponent mediator case. *Journal of the Royal Statistical Society (Series B)*, *77*, 581–615. https://doi.org/10.1111/rssb.12082

Zheng, C., Atkins, D. C., Zhou, X.-H., & Rhew, I. C. (2015). Causal models for mediation analysis: An introduction to structural mean models. *Multivariate Behavioral Research*, *50*(6), 614–631.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

A. Notation

Table 8: Notation used throughout the manuscript.

Symbol	Meaning	Defined in
Observed data		
N	Sample size (indexed by i)	—
R_i	Randomized binary treatment	—
$\mathbf{m}_i, \mathbf{x}_i, \mathbf{y}_i$	Indicators of mediator, covariates, outcome	(2.1)
$\mathbf{z}_i$	Vector of all indicators, $(\mathbf{m}_i, \mathbf{x}_i, \mathbf{y}_i)^\top$	(2.1)
P	Number of observed indicators, $\dim(\mathbf{z}_i)$	(2.1)
U_i	Unmeasured mediator-outcome confounder	Fig. 1
Latent variables		
$\eta_{M_i}, \eta_{Y_i}, \boldsymbol{\eta}_{\mathbf{X}_i}$	Latent mediator, outcome and covariates	(2.2)
$\boldsymbol{\eta}_i$	Vector of stacked latent variables$(\eta_{M_i} \;\; \boldsymbol{\eta}_{\mathbf{X}_i} \;\; \eta_{Y_i})^\top$	(2.2)
K	Number of latent variables, $\dim(\boldsymbol{\eta}_i)$	(2.2)
$\boldsymbol{\eta}_{\text{pred},i}$	Latent outcome predictors, i.e. $(\eta_{M_i} \;\; \boldsymbol{\eta}_{\mathbf{X}_i})^\top$	(5.13)
J	Highest order in which a component of $\boldsymbol{\eta}_{\text{pred},i}$ enters $\boldsymbol{\Xi}_y$	(5.15)
$\tilde{\boldsymbol{\eta}}_i$	Factor score at the *true* measurement parameters	(5.2)
$\hat{\boldsymbol{\eta}}_i$	Factor score at the *estimated* measurement parameters	(5.2)
$\mathbf{H}$	Weight matrix in the factor-score estimator	—
$\mathbf{e}_i$	Estimation error, $\tilde{\boldsymbol{\eta}}_i = \boldsymbol{\eta}_i + \mathbf{e}_i$	(5.3)
$\boldsymbol{\Sigma}_{ee}$	Variance of the estimation error with elements $\Sigma_{ee,\cdot\cdot}$	(5.5)
Measurement model		
$\boldsymbol{\epsilon}_i$	Measurement error with covariance matrix $\text{Var}(\boldsymbol{\epsilon}_i)$	(2.4)
$\boldsymbol{\tau}, \boldsymbol{\tau}_{\text{free}}$	Indicator intercepts; free subvector, $(P-K)\times 1$	(2.5)
$\boldsymbol{\Lambda}, \boldsymbol{\Lambda}_{\text{free}}$	Loading matrix; free submatrix, $(P-K)\times k$	(2.5)
$\boldsymbol{\Upsilon}_1$	Free CFA parameters	(5.1)
$\boldsymbol{\Upsilon}_2$	Moments of $\mathbf{e}_i$ required by the 2SMM correction	—
$\boldsymbol{\Upsilon}$	$(\boldsymbol{\Upsilon}_1 \;\; \boldsymbol{\Upsilon}_2)^\top$	Thm. 6.1
$\boldsymbol{\Delta}_i$	Influence function of $\hat{\boldsymbol{\Upsilon}}$	Thm. 6.1
$\psi_{\boldsymbol{\Upsilon}_1}(\mathbf{z}_i)$	Z-estimating function for $\boldsymbol{\Upsilon}$	—
$\mathbf{J}$	Jacobian $\text{E}[\partial\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)/\partial\boldsymbol{\Upsilon}^\top]$	Ass. R3
$\mathbf{B}_i, \mathbf{q}_i, \mathbf{p}_i$	Linearisation of $\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i$ in the estimation error	Thm. 6.2
Structural model		
r	Fixed level of the treatment	(2.6)
η_m	Fixed level of the latent mediator	(2.6)
$\eta_{Y_i}^{r,\eta_m}$	Potential outcome under $R_i = r$ and η_{M_i} set to η_m	(2.6)
$\theta_s,\; h_s(\cdot)$	Structural parameters and basis functions	(2.6)
S	Number of structural terms h_s	(2.6)

continued on next page

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Symbol	Meaning	Defined in
θ_r	Controlled direct effect of treatment	(1.1)
θ_m	Controlled mediation effect (per unit of η_m)	(1.2)
$\boldsymbol{\beta}_x$	Covariate coefficients in the outcome equation	(2.7)
q_{yx}, q_{mx}, q_{mrx}	Admissible polynomial transformations of $\boldsymbol{\eta}_{\mathbf{X}_i}$	(2.9)
$g(\boldsymbol{\eta}_{\mathbf{X}_i})$	Nuisance term modeled as $q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i})\,\boldsymbol{\beta}_x$	(2.6)
$\boldsymbol{\xi}_{Y_i}$	Outcome regressors $(R_i \;\; \eta_{M_i} \;\; q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i}))$	(2.7)
$\boldsymbol{\Xi}_y$	$N \times \dim(\boldsymbol{\alpha})$ design matrices with rows $\boldsymbol{\xi}_{Y_i}$	(2.7)
$\boldsymbol{\alpha}$	outcome structural parameters $(\theta_r \;\; \theta_m \;\; \boldsymbol{\beta}_x)^\top$	(2.7)
ζ_{Y_i}	Structural errors of outcome equation	(2.7)
$\boldsymbol{\xi}_{M_i}$	Mediator regressors $(R_i \;\; q_{mx}(\boldsymbol{\eta}_{\mathbf{X}_i}) \;\; R_i\, q_{mrx}(\boldsymbol{\eta}_{\mathbf{X}_i}))$	(2.8)
$\boldsymbol{\Xi}_m$	$N \times \dim(\boldsymbol{\gamma})$ design matrices with rows $\boldsymbol{\xi}_{M_i}$	(2.8)
$\boldsymbol{\gamma}$	$(\gamma_r \;\; \boldsymbol{\gamma}_x \;\; \boldsymbol{\gamma}_{xr})^\top$, mediator-equation parameters	(2.8)
ζ_{M_i}	Structural errors of mediator equation	(2.8)
G-estimation weights		
$\mathbf{A}(R_i, \boldsymbol{\eta}_{\mathbf{X}_i})$	Centred weight (instrument) matrix, $\mathrm{E}[a_j \mid \boldsymbol{\eta}_{\mathbf{X}_i}] = 0$	(4.2)
$a_{r,i}$	Treatment weight	(5.7)
$a_{m,i}$	Mediator weight	(5.7)
$\mathbf{W}$	Full weight matrix with rows $(a_{r,i} \;\; a_{m,i} \;\; q_{yx}(\boldsymbol{\eta}_{\mathbf{X}_i}))$	(5.9)
$\boldsymbol{\Omega}(\boldsymbol{\eta}_{\mathbf{X}})$	Conditional error term covariance $\mathrm{Var}(\zeta_{Y_i}^{r,\eta_m} \mid \boldsymbol{\eta}_{\mathbf{X}_i})$	(4.3)
$\bar{\boldsymbol{\alpha}}$	Naive IV/G-estimator without factor score correction	(5.10)
Moment functions and estimators		
$D_j(\cdot)$	jth 2SMM correction term	(5.16)
$\mathbf{M}(\cdot), \mathbf{m}(\cdot)$	Per-unit corrected moment functions	(5.14)
$\hat{\mathbf{M}}, \hat{\mathbf{m}}$	Sample averages of $\mathbf{M}, \mathbf{m}$ at $(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2)$	(5.13)
$\hat{\mathbf{M}}_1, \hat{\mathbf{m}}_1$	Uncorrected (D_0) part	mod. 2SMM
$\hat{\mathbf{M}}_2, \hat{\mathbf{m}}_2$	Correction part, $\sum_{j\geq 1}(-1)^j D_j$	mod. 2SMM
$\mathbf{R}_1, \mathbf{R}_2$	Blocks of the Fuller-type positive-definiteness modification	mod. 2SMM
$\hat{\lambda}$	Largest eigenvalue of $\mathbf{R}_1^{-1}\mathbf{R}_2$	mod. 2SMM
$\hat{\boldsymbol{\alpha}}$	2SMM estimator $\hat{\mathbf{M}}^{-1}\hat{\mathbf{m}}$	(5.12)
c	Fuller tuning constant $\in [0, \dim(\boldsymbol{\alpha}) + 5]$	mod. 2SMM
ν	Ridge regularisation constant	(5.17)
$\check{\boldsymbol{\alpha}}$	Final estimator $(\check{\mathbf{M}} + \nu\mathbf{I})^{-1}\check{\mathbf{m}}$	(5.17)
$\mathbf{G}, \mathbf{h}$	$\mathrm{E}[\mathbf{w}_i \boldsymbol{\xi}_{y,i}^\top], \mathrm{E}[\mathbf{w}_i \eta_{y_i}]$	Thm. 6.3
$\mathbf{l}(\cdot)$	$\mathbf{m}(\cdot) - \mathbf{M}(\cdot)\boldsymbol{\alpha}$, the moment residual	Thm. 6.3
$\overline{\mathbf{C}}$	Expected derivative matrix linking $\mathbf{l}$ to $\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}$	Thm. 6.3
$\mathbf{d}_i, \mathbf{S}$	$\mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}) + \overline{\mathbf{C}}\boldsymbol{\Delta}_i$ and $\mathrm{Var}(\mathbf{d}_i)$	Thm. 6.3

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

B. Additional Assumptions

Assumption C3 - Consistency: The realized mediator and outcome corresponding to a given treatment and mediator assignment equals the potential outcome under those values

$$\begin{aligned} \eta_{M_i} &= \eta_{M_i}^{r} && \text{whenever } R_i = r, \\ \eta_{Y_i} &= \eta_{Y_i}^{r,\eta_m} && \text{whenever } R_i = r \text{ and } \eta_{M_i} = \eta_m. \end{aligned}$$

This assumption ensures a unique mapping from treatment and mediator values to potential outcomes. In practice, consistency is typically justified under the Stable Unit Treatment Value Assumption (SUTVA), which comprises two conditions: first, that each treatment corresponds to a single well-defined version; and second, that there is no interference between units, meaning that one unit's treatment does not affect another unit's outcome.

Assumption C4 - Positivity: For all $\boldsymbol{\eta}_x$ in the support of $\boldsymbol{\eta}_{\mathbf{X}}$, the probability of receiving each treatment level is strictly between 0 and 1:

$$0 < P(R = r \mid \boldsymbol{\eta}_{\mathbf{X}} = \boldsymbol{\eta}_x) < 1 \quad \forall r.$$

In an RCT, this holds by design, since treatment assignment is independent of the covariates. For the mediator, positivity is required at the levels contrasted by the estimands (e.g. η_m and $\eta_{m'}$ for the CME in Equation (1.2)):

- If the mediator is discrete:

$$P(\eta_M = m \mid R = r, \boldsymbol{\eta}_{\mathbf{X}} = \boldsymbol{\eta}_x) > 0 \quad \forall r, \eta_m.$$

- If the mediator is continuous:

$$f_{\eta_M \mid R, \boldsymbol{\eta}_{\mathbf{X}}}(m \mid r, \boldsymbol{\eta}_x) > 0 \quad \forall r, \eta_m,$$

  where $f_{\eta_M \mid R, \boldsymbol{\eta}_{\mathbf{X}}}$ denotes the conditional density of the mediator.

Assumption C5 - Causal Efficacy: We assume that the latent factor model admits a structural interpretation in the sense of VanderWeele and Vansteelandt (2022). That is, the latent variables rather than their indicators are causally efficacious. Specifically, for each latent variable η with indicator set $\mathbf{z}_\eta$, (i) the indicators $\mathbf{z}_\eta$ have no causal effect on any subsequent variable, and (ii) they are affected by any other variable only through η.

Property (ii) implies that the measurement parameters $(\boldsymbol{\tau}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ are invariant across treatment arms. Since the factor score estimation error in Equation (5.3) is a fixed linear transformation of $\boldsymbol{\epsilon}$, it inherits this independence, so that the instruments remain valid once the 2SMM correction accounts for the measurement error.

Assumption R1 - IID Sampling of Observed Data: The observations $\{(r_i, \mathbf{m}_i, \mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n}$ are independent and identically distributed (iid) draws from the joint distribution P. Note that this iid assumption pertains only to independence across units and places no restrictions on within-unit dependence. In particular, latent variables such as η_{m_i} and η_{y_i} may be arbitrarily dependent, allowing for unobserved confounding between the mediator and the outcome.

Assumption R2 - Correct Model Specification: The measurement model in Equation (2.3) and the structural dependence among latent variables in Equation (2.6) are correctly specified as polynomial models that are linear in their parameters.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

Correct specification of the function $g(\boldsymbol{\eta}_{\mathbf{X}_i})$ and of the mediator model in Equation (2.8) is not required for identification of the controlled effects. Misspecifications may affect the efficiency of estimation, but do not invalidate Equation (4.2).

.

Assumption R3 - Measurement Model Regularity: The observed indicators $\mathbf{z}$ satisfy

$$\mathrm{E}\left[|\mathbf{z}_i|^2\right] < \infty.$$

The true parameter $\boldsymbol{\Upsilon}$ lies in the interior of a compact set $\mathcal{U}$, the estimating function $\boldsymbol{\Upsilon} \mapsto \psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)$ is continuous in $\boldsymbol{\Upsilon}$ almost surely, and the Jacobian

$$\mathbf{J} := \mathrm{E}\left[\partial \psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)/\partial \boldsymbol{\Upsilon}^\top\right],$$

is nonsingular at $\boldsymbol{\Upsilon}$.

Assumption R4 - Measurement Error: The measurement errors $\boldsymbol{\epsilon}_i$ are iid, independent of the latent variables $\boldsymbol{\eta}_i$, have zero mean, and satisfy

$$\mathrm{E}[|\boldsymbol{\epsilon}_i|^{4J}] < \infty.^{15}$$

Assumption R5 - Factor Score Finite Moments: The factor scores $\boldsymbol{\eta}_{\text{pred},i}$ satisfy finite moment conditions of sufficiently high order:

$$\mathrm{E}[|\boldsymbol{\eta}_{\text{pred},i}|^{4J-2}] < \infty$$

Assumption R6 - Structural Equation Error: The structural equation errors $\boldsymbol{\zeta}_i$ comprising $\boldsymbol{\zeta}_{Y_i}$ and $\boldsymbol{\zeta}_{M_i}$ are iid, independent of $\boldsymbol{\epsilon}_i$, hence of $\mathbf{e}_i$, have zero mean, and finite variance:

$$\mathrm{Var}[\boldsymbol{\zeta}_i] < \infty.$$

C. More details on C2

C2 is implied by what econometricians call No Essential Heterogeneity (Heckman et al., 2006). Extending the original single-treatment notion to a joint treatment-mediator setting, it states: individuals may have idiosyncratic responses depending on the treatment-mediator condition, but their selection into that condition cannot be mean-informative about these gains conditional on covariates.

A sufficient condition for both is full homogeneity of the treatment and mediator effects across both observed covariates $\boldsymbol{\eta}_{\mathbf{X}}$ and unmeasured confounders U:

$$\theta_r(\boldsymbol{\eta}_{\mathbf{X}_i}, U_i) = \theta_r, \quad \theta_m(\boldsymbol{\eta}_{\mathbf{X}_i}, U_i) = \theta_m.$$

This homogeneity is, in turn, implied by the additive Rank Preserving assumption, the original approach of Ten Have et al., 2007 is built on. It links potential outcomes through a common shift:

$$\eta_{Y_i}^{r',\eta_{m'}} = \eta_{Y_i}^{r,\eta_m} + \theta_r\,(r' - r) + \theta_m\,(\eta_{m'} - \eta_m),$$

[15] In our implementation, we additionally assume $\boldsymbol{\epsilon}_i$ to be normally distributed, allowing the 2SMM correction terms to be derived directly from $\boldsymbol{\Sigma}_{ee}$. An alternative approach, based on OLS-estimated higher-order error moments and not requiring distributional assumptions on $\boldsymbol{\epsilon}_i$, but yielding identical asymptotic properties, is described in Wall and Amemiya (2000).

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

such that if individual i has a better outcome than j under one treatment-mediator condition, they will also have a better outcome under any other condition:[16]

$$\eta_{Y_i}^{r,\eta_m} > \eta_{Y_j}^{r,\eta_m} \iff \eta_{Y_i}^{r',\eta_{m'}} > \eta_{Y_j}^{r',\eta_{m'}} \quad \forall i,j \forall (r,\eta_m),(r',\eta_{m'}).$$

The formulation C2 is weakest, permitting heterogeneous effects as long as they are not systematically related to treatment-mediator assignment.

D. Implementation

The R package `rapsem` allows estimation of structural models as specified in Equation (2.7) and Equation (2.8). While directly observed variables may be transformed arbitrarily, latent variables can only enter as polynomial terms. The current version supports latent variable terms up to dimension $J = 1$ (i.e., computation of D_0 and D_1, but the framework can be extended to higher orders by incorporating additional correction terms $D_{j>1}$. Furthermore, the current implementation is based on the measurement model in Equation (2.4), treating it as a confirmatory factor analysis (CFA) subject to the constraints in Equation (2.5). However, it can be extended to accommodate alternative factor specifications and modeling techniques, such as Item Response Theory (IRT) for discrete indicators.

Estimation is carried out via the function `med_est`, which takes as input the observed data and a `lavaan` model specifying the structural equation model, and returns results from both the standard regression approach and the G-estimation approach, each using factor score corrections. Factor intercepts, loadings, and residual variances are estimated by maximum likelihood with `lavaan`. By default, the measurement model of each latent variable is fitted separately (which requires at least three indicators per latent variable), since the factor scores do not depend on the factor covariances and misspecification in one measurement block cannot affect the others. Subsequently, the factor scores are computed according to Equation (5.2). Structural parameters are then estimated using the modified 2SMM regularized G-estimator in Equation (5.17), with default settings $c = 5$ and $v = 10^{-4}$ (sufficiently small for the effect size estimates to remain essentially unaffected).

Since the analytic asymptotic variance in Theorem 6.3 does not account for the finite-sample modification and the ridge penalty, uncertainty is quantified by the nonparametric bootstrap (default: 100 bootstrap samples): point estimates, standard errors, and confidence intervals are the mean, standard deviation, and percentile intervals of the bootstrap replicates.

E. Factor Score Corrections for Simulation Study Model

In the concrete model defined in Equation (2.9), the polynomial order is $J = 1$, so the moment expansion consists of D_0 and D_1. D_0 provides the uncorrected terms

$$\begin{aligned} D_0(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2) &= \begin{pmatrix} a_{r,i} & \hat{a}_{m,i} & \hat{\eta}_{X_{1i}} & \hat{\eta}_{X_{2i}} \end{pmatrix}^\top \begin{pmatrix} R_i & \hat{\eta}_{M_i} & \hat{\eta}_{X_{1i}} & \hat{\eta}_{X_{2i}} & \hat{\eta}_{Y_i} \end{pmatrix} \\ &= \begin{pmatrix} a_{r,i}R_i & a_{r,i}\hat{\eta}_{M_i} & a_{r,i}\hat{\eta}_{X_{1i}} & a_{r,i}\hat{\eta}_{X_{2i}} & a_{r,i}\hat{\eta}_{Y_i} \\ \hat{a}_{m,i}R_i & \hat{a}_{m,i}\hat{\eta}_{M_i} & \hat{a}_{m,i}\hat{\eta}_{X_{1i}} & \hat{a}_{m,i}\hat{\eta}_{X_{2i}} & \hat{a}_{m,i}\hat{\eta}_{Y_i} \\ \hat{\eta}_{X_{1i}}R_i & \hat{\eta}_{X_{1i}}\hat{\eta}_{M_i} & \hat{\eta}_{X_{1i}}^2 & \hat{\eta}_{X_{1i}}\hat{\eta}_{X_{2i}} & \hat{\eta}_{X_{1i}}\hat{\eta}_{Y_i} \\ \hat{\eta}_{X_{2i}}R_i & \hat{\eta}_{X_{2i}}\hat{\eta}_{M_i} & \hat{\eta}_{X_{2i}}\hat{\eta}_{X_{1i}} & \hat{\eta}_{X_{2i}}^2 & \hat{\eta}_{X_{2i}}\hat{\eta}_{Y_i} \end{pmatrix}. \end{aligned}$$

[16] The additive form is a special case of general rank preservation, $\eta_{Y_i}^{r',\eta_{m'}} = \varphi(\eta_{Y_i}^{r,\eta_m})$ for a strictly increasing function φ with $\varphi' > 0$. The general form also preserves ranks but does not imply homogeneity, since the individual effect $\varphi(y) - y$ may vary with the outcome level and hence with U_i.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

The correction term $D_1(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2)$ involves expectations over the measurement error variance and covariances. Under the latent factor ordering assumed in Equation (2.2), in which the indices $1, \ldots, 4$ of $\hat{\boldsymbol{\Sigma}}_{ee}$ refer to $\eta_M, \eta_{X_1}, \eta_{X_2}$, and η_Y, and using the form of $\hat{a}_{m,i}$ from Equations (5.7) to (5.8), it takes the form

$$D_1(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\gamma}_{x1r}\hat{\boldsymbol{\Sigma}}_{ee,21} + \hat{\gamma}_{x2r}\hat{\boldsymbol{\Sigma}}_{ee,31})\hat{a}_{r,i} & \hat{\gamma}_{x1r}\hat{\boldsymbol{\Sigma}}_{ee,22} + \hat{\gamma}_{x2r}\hat{\boldsymbol{\Sigma}}_{ee,32})\hat{a}_{r,i} & \hat{\gamma}_{x1r}\hat{\boldsymbol{\Sigma}}_{ee,23} + \hat{\gamma}_{x2r}\hat{\boldsymbol{\Sigma}}_{ee,33})\hat{a}_{r,i} & \hat{\gamma}_{x1r}\hat{\boldsymbol{\Sigma}}_{ee,24} + \hat{\gamma}_{x2r}\hat{\boldsymbol{\Sigma}}_{ee,34})\hat{a}_{r,i} \\ 0 & \hat{\boldsymbol{\Sigma}}_{ee,21} & \hat{\boldsymbol{\Sigma}}_{ee,22} & \hat{\boldsymbol{\Sigma}}_{ee,23} & \hat{\boldsymbol{\Sigma}}_{ee,24} \\ 0 & \hat{\boldsymbol{\Sigma}}_{ee,31} & \hat{\boldsymbol{\Sigma}}_{ee,32} & \hat{\boldsymbol{\Sigma}}_{ee,33} & \hat{\boldsymbol{\Sigma}}_{ee,34} \end{pmatrix}.$$

This correction ensures that second-order bias due to measurement error is appropriately removed from the moment conditions when estimating the structural model parameters.

F. Proof of Theorem 6.1

F.1. Consistency

Proof. By Assumptions R2, $\boldsymbol{\Upsilon}$ is the unique solution of $\Psi(\boldsymbol{\Upsilon}) := \mathrm{E}[\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)] = \mathbf{0}$. Let $\hat{\boldsymbol{\Upsilon}}$ solve $\Psi_N(\boldsymbol{\Upsilon}) := \frac{1}{N}\sum_{i=1}^{N} \psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i) = \mathbf{0}$. Since $\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)$ is polynomial in $\mathbf{z}_i$ with coefficients that are continuous in $\boldsymbol{\Upsilon}$ (Assumption R2), and $\mathcal{U}$ is compact (Assumption R3), there exist finite constants $C > 0$ and integer $q \geq 1$ such that

$$\sup_{\boldsymbol{\Upsilon}\in\mathcal{U}} \|\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)\| \leq C\big(1 + \|\mathbf{z}_i\|^q\big) \quad \text{a.s.}$$

Since Assumptions R3 and R4 jointly ensure $\mathrm{E}[\|\mathbf{z}_i\|^q] < \infty$ for the relevant q (since $\mathbf{z}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$, moments of $\|\mathbf{z}_i\|^q$ are controlled by those of $\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i$, and the estimating equations of Wall and Amemiya (2000) satisfy $q \leq 4J - 2$), the envelope condition

$$\mathrm{E}\Big[\sup_{\boldsymbol{\Upsilon}\in\mathcal{U}} \|\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)\|\Big] < \infty$$

holds, so $\{\psi_{\boldsymbol{\Upsilon}}(\cdot) : \boldsymbol{\Upsilon} \in \mathcal{U}\}$ is Glivenko–Cantelli. The uniform law of large numbers (Assumption R1) thus ensures uniform convergence:

$$\sup_{\boldsymbol{\Upsilon}\in\mathcal{U}} \|\Psi_N(\boldsymbol{\Upsilon}) - \Psi(\boldsymbol{\Upsilon})\| \xrightarrow{p} 0.$$

Almost sure continuity of $\psi_{\boldsymbol{\Upsilon}}(\mathbf{z}_i)$ in $\boldsymbol{\Upsilon}$ (Assumption R3), together with the envelope condition established above, allows an application of the dominated convergence theorem, which implies that $\Psi(\boldsymbol{\Upsilon}') = \mathrm{E}[\psi_{\boldsymbol{\Upsilon}'}(\mathbf{z}_i)]$ is continuous in $\boldsymbol{\Upsilon}'$ on $\mathcal{U}$. Since $\boldsymbol{\Upsilon}$ is the unique zero of Ψ (Assumption R2) and $\mathcal{U}$ is compact (Assumption R3), the set $\{\boldsymbol{\Upsilon}' : \|\boldsymbol{\Upsilon}' - \boldsymbol{\Upsilon}\| \geq \varepsilon\}$ is compact and contains no zero of Ψ for any $\varepsilon > 0$. A continuous function on a compact set that is everywhere nonzero attains a strictly positive infimum, so

$$\inf_{\|\boldsymbol{\Upsilon}'-\boldsymbol{\Upsilon}\|>\varepsilon} \|\Psi(\boldsymbol{\Upsilon}')\| > 0 \quad \text{for every } \varepsilon > 0.$$

Since $\Psi_N(\hat{\boldsymbol{\Upsilon}}) = \mathbf{0}$ by definition of $\hat{\boldsymbol{\Upsilon}}$, uniform convergence of Ψ_N to Ψ and the separation condition above together satisfy the conditions of van der Vaart (1998, Theorem 5.9), giving

$$\hat{\boldsymbol{\Upsilon}} \xrightarrow{p} \boldsymbol{\Upsilon}.$$

□

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

F.2. Asymptotic normality.

Proof. Since $\hat{\boldsymbol{\Upsilon}} \xrightarrow{p} \boldsymbol{\Upsilon}$, we can linearize the sample estimating equation around $\boldsymbol{\Upsilon}$ using a first-order Taylor expansion:

$$\frac{1}{N}\sum_{i=1}^{N}\psi_i(\hat{\boldsymbol{\Upsilon}}) = \frac{1}{N}\sum_{i=1}^{N}\psi_i(\boldsymbol{\Upsilon}) + \mathbf{J}(\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}) + o_p\big(\|\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}\|\big),$$

where $\mathbf{J}$ is nonsingular by assumption. Since the left-hand side equals zero by definition of $\hat{\boldsymbol{\Upsilon}}$, rearranging yields

$$\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon} = -\mathbf{J}^{-1}\frac{1}{N}\sum_{i=1}^{N}\psi_i(\boldsymbol{\Upsilon}) + o_p(N^{-1/2}).$$

Defining

$$\boldsymbol{\Delta}_i := -\mathbf{J}^{-1}\psi_i(\boldsymbol{\Upsilon}),$$

we obtain the desired expansion. By definition, $\mathrm{E}[\psi(\mathbf{z}_i, \boldsymbol{\Upsilon})] = 0$, so $\mathrm{E}[\boldsymbol{\Delta}_i] = 0$. Since $\psi(\mathbf{z}_i, \boldsymbol{\Upsilon})$ is polynomial in $\boldsymbol{\eta}_{\text{pred},i}$ and linear in $(\boldsymbol{\epsilon}_i, \boldsymbol{\zeta}_i)$, Assumption R3 ensures $\mathrm{Var}(\boldsymbol{\Delta}_i) < \infty$. Hence,

$$\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon} = O_p(N^{-1/2}),$$

establishing $\sqrt{N}$-consistency. □

G. Proof of Theorem 6.2

Proof. From the definition of the latent factor score estimator in Equation (5.2), we can express

$$\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = \hat{\mathbf{H}} f(\hat{\boldsymbol{\tau}}_{\text{free}}, \hat{\boldsymbol{\Lambda}}_{\text{free}}) - \mathbf{H} f(\boldsymbol{\tau}_{\text{free}}, \boldsymbol{\Lambda}_{\text{free}}),$$

with

$$f(\boldsymbol{\tau}_{\text{free}}, \boldsymbol{\Lambda}_{\text{free}}) = -\mathbf{z}_i + \boldsymbol{\tau}_{\text{free}} + \boldsymbol{\Lambda}_{\text{free}}\mathbf{z}_i - \boldsymbol{\Lambda}_{\text{free}}\mathbf{0}_{k\times 1}.$$

Expanding this difference and regrouping by parameters yields

$$\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = \mathbf{H}(\hat{\boldsymbol{\tau}}_{\text{free}} - \boldsymbol{\tau}_{\text{free}}) + \mathbf{H}(\mathbf{q}_i^\top \otimes \mathbf{I}_{p-k})\,\mathrm{vec}(\hat{\boldsymbol{\Lambda}}_{\text{free}} - \boldsymbol{\Lambda}_{\text{free}}) - (\mathbf{p}_i^\top \otimes \mathbf{I}_k)\,\mathrm{vec}(\hat{\mathbf{H}} - \mathbf{H}),$$

which can be written compactly as

$$\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = \mathbf{B}_i\,(\hat{\boldsymbol{\Upsilon}}_1 - \boldsymbol{\Upsilon}_1),$$

with $\mathbf{B}_i$, $\mathbf{q}_i$, and $\mathbf{p}_i$ defined in Theorem 6.2. By Theorem 6.1, $\hat{\boldsymbol{\Upsilon}}_1 - \boldsymbol{\Upsilon}_1 = O_p(N^{-1/2})$, which implies

$$\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = O_p(N^{-1/2}),$$

establishing $\sqrt{N}$-consistency. □

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

H. Proof of Theorem 6.3

H.1. Consistency

Proof. The identification argument provides us with

$$\mathrm{E}[\mathbf{w}_i \zeta_{y_i}] = \mathbf{0},$$

where

$$\zeta_{y_i} = \eta_{y_i} - \boldsymbol{\xi}_{y_i}^\top \boldsymbol{\alpha}.$$

By inserting this error term into the above equation, under Assumptions R1 and R5, we obtain

$$\mathrm{E}[\mathbf{w}_i \eta_{y_i}] = \mathrm{E}[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top] \boldsymbol{\alpha}.$$

We define the population moments

$$\mathbf{G} = \mathrm{E}[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top], \quad \mathbf{h} = \mathrm{E}[\mathbf{w}_i \eta_{y_i}].$$

and thus write

$$\mathbf{h} = \mathbf{G}\boldsymbol{\alpha}.$$

By Assumption S1, $\mathbf{G}$ is invertible. Hence, the equation

$$\mathbf{h} - \mathbf{G}\boldsymbol{\alpha} = \mathbf{0}$$

has a unique solution, given by

$$\boldsymbol{\alpha} = \mathbf{G}^{-1} - \mathbf{h}.$$

To show that $\hat{\mathbf{M}} \xrightarrow{p} \mathbf{G}$ and $\hat{\mathbf{m}} \xrightarrow{p} \mathbf{h}$, we decompose each empirical moment into two terms:,

$$\hat{\mathbf{M}} - \mathbf{G} = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \left[M(\hat{\boldsymbol{\eta}}_{\mathrm{pred},i}, \hat{\boldsymbol{\Upsilon}}_2) - M(\tilde{\boldsymbol{\eta}}_{\mathrm{pred},i}, \boldsymbol{\Upsilon}_2) \right]}_{\text{(I)}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N} M(\tilde{\boldsymbol{\eta}}_{\mathrm{pred},i}, \boldsymbol{\Upsilon}_2) - \mathbf{G}}_{\text{(II)}},$$

and analogously for $\hat{\mathbf{m}} - \mathbf{h}$. Term (II) involves i.i.d. summands, since $M(\tilde{\boldsymbol{\eta}}_{\mathrm{pred},i}, \boldsymbol{\Upsilon}_2)$ evaluated at the true parameters depends only on $\mathbf{z}_i$, which are i.i.d. by Assumption R1. The standard LLN therefore gives Term (II) = $o_p(1)$. For Term (I), a first-order Taylor expansion of M around $(\tilde{\boldsymbol{\eta}}_{\mathrm{pred},i}, \boldsymbol{\Upsilon}_2)$ yields

$$M(\hat{\boldsymbol{\eta}}_{\mathrm{pred},i}, \hat{\boldsymbol{\Upsilon}}_2) - M(\tilde{\boldsymbol{\eta}}_{\mathrm{pred},i}, \boldsymbol{\Upsilon}_2) = \left.\frac{\partial M}{\partial \boldsymbol{\Upsilon}^\top}\right|_{\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2} (\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon}) + o_p(N^{-1/2}).$$

Averaging over i, applying the standard LLN to the i.i.d. gradient terms, and using $\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon} = o_p(1)$ from Theorem 6.1 gives Term (I) = $o_p(1)$. Combining both terms,

$$\begin{pmatrix} \hat{\mathbf{M}} \\ \hat{\mathbf{m}} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \mathbf{G} \\ \mathbf{h} \end{pmatrix}.$$

Finally, the estimator defined in (5.12) is consistent: as $N \to \infty$,

$$\hat{\boldsymbol{\alpha}} = \hat{\mathbf{M}}^{-1}\hat{\mathbf{m}} \xrightarrow{p} \mathbf{G}^{-1}\mathbf{h} = \boldsymbol{\alpha}.$$

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

□

H.2. Asymptotic normality

Proof. To show asymptotic normality, we rewrite

$$\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \sqrt{N}(\hat{\mathbf{M}}^{-1}\hat{\mathbf{m}} - \boldsymbol{\alpha}) = \hat{\mathbf{M}}^{-1}\sqrt{N}(\hat{\mathbf{m}} - \hat{\mathbf{M}}\boldsymbol{\alpha}),$$

With the definition

$$\mathbf{l}(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}) = \mathbf{m}(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2) - \mathbf{M}(\boldsymbol{\eta}_{\text{pred},i}, \boldsymbol{\Upsilon}_2)\,\boldsymbol{\alpha},$$

the second term becomes

$$\hat{\mathbf{m}} - \hat{\mathbf{M}}\boldsymbol{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{l}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2, \boldsymbol{\alpha}).$$

A first-order Taylor expansion of the moment function $\mathbf{l}$ around $(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha})$ yields

$$\mathbf{l}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2, \boldsymbol{\alpha}_0) = \mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}_0) + \left.\frac{\partial \mathbf{l}}{\partial \boldsymbol{\eta}_i^\top}\right|_{\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}_0} (\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i) + \left.\frac{\partial \mathbf{l}}{\partial \boldsymbol{\Upsilon}_2^\top}\right|_{\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}_0} (\hat{\boldsymbol{\Upsilon}}_2 - \boldsymbol{\Upsilon}_2) + o_p(N^{-1/2}).$$

Substituting the first-stage expansion $\hat{\boldsymbol{\eta}}_i - \tilde{\boldsymbol{\eta}}_i = \mathbf{B}_i(\hat{\boldsymbol{\Upsilon}} - \boldsymbol{\Upsilon})$ gives

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{l}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2, \boldsymbol{\alpha}_0) = \frac{1}{N}\sum_{i=1}^{N} \left\{\mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}_0) + \overline{\mathbf{C}}\,\boldsymbol{\Delta}_i\right\} + o_p(N^{-1/2}),$$

where

$$\overline{\mathbf{C}} = \mathrm{E}\left[\left(\tfrac{\partial \mathbf{l}}{\partial \boldsymbol{\eta}_i^\top}\mathbf{B}_i \;\; \tfrac{\partial \mathbf{l}}{\partial \boldsymbol{\Upsilon}_2^\top}\right)\right].$$

By the central limit theorem for i.i.d. variables (Assumption R1) and the finite-moment conditions (Assumptions R5, R6), we obtain

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{l}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2, \boldsymbol{\alpha}_0)\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}),$$

with

$$\mathbf{S} = \mathrm{Var}\left[\mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\alpha}_0) + \overline{\mathbf{C}}\,\boldsymbol{\Delta}_i\right].$$

Finally, using Assumption S1 and applying the delta method gives

$$\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{G}^{-1}\mathbf{S}\mathbf{G}^{-\top}).$$

□

I. Additional Plots for Simulation Study 1

Figure 7 displays the sampling distributions of $\hat{\theta}_m$ across the setting of Simulation Study 1. This illustrates the bias and variance patterns discussed in the main text: RAPSEM distributions are centered at zero throughout, while standard SEM distributions shift progressively rightward with increasing unobserved

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

confounding effect size δ_u. The wider spread of RAPSEM's distributions at small sample sizes is also clearly visible, narrowing relative to standard SEM as the sample size N increases.

Figure 8 shows the fit indices for the lavaan model in Simulation Study 1. For all model runs, they remain within the common thresholds for good model fit (below 0.06 for RMSEA, below 0.08 for SRMR, and above 95% for CFA and TLI). Notably, model fit hardly changes for different levels of confounding.

J. Reliabilities in the Empirical Example

To assess the internal consistency of the latent factors, Table 9 reports the pooled McDonald's ω reliability coefficients along with their standard errors.

Table 9: Pooled McDonald's ω Reliability Coefficients for the full Model M1.

Factor	$\hat{\omega}$	**SE**
conceptual knowledge (outcome)	0.65	0.04
task interest (mediator)	0.86	0.01
interest in physics	0.84	0.01
conscientiousness in physics	0.75	0.02
perceived teacher support	0.63	0.04
perceived cognitive activation	0.79	0.03

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

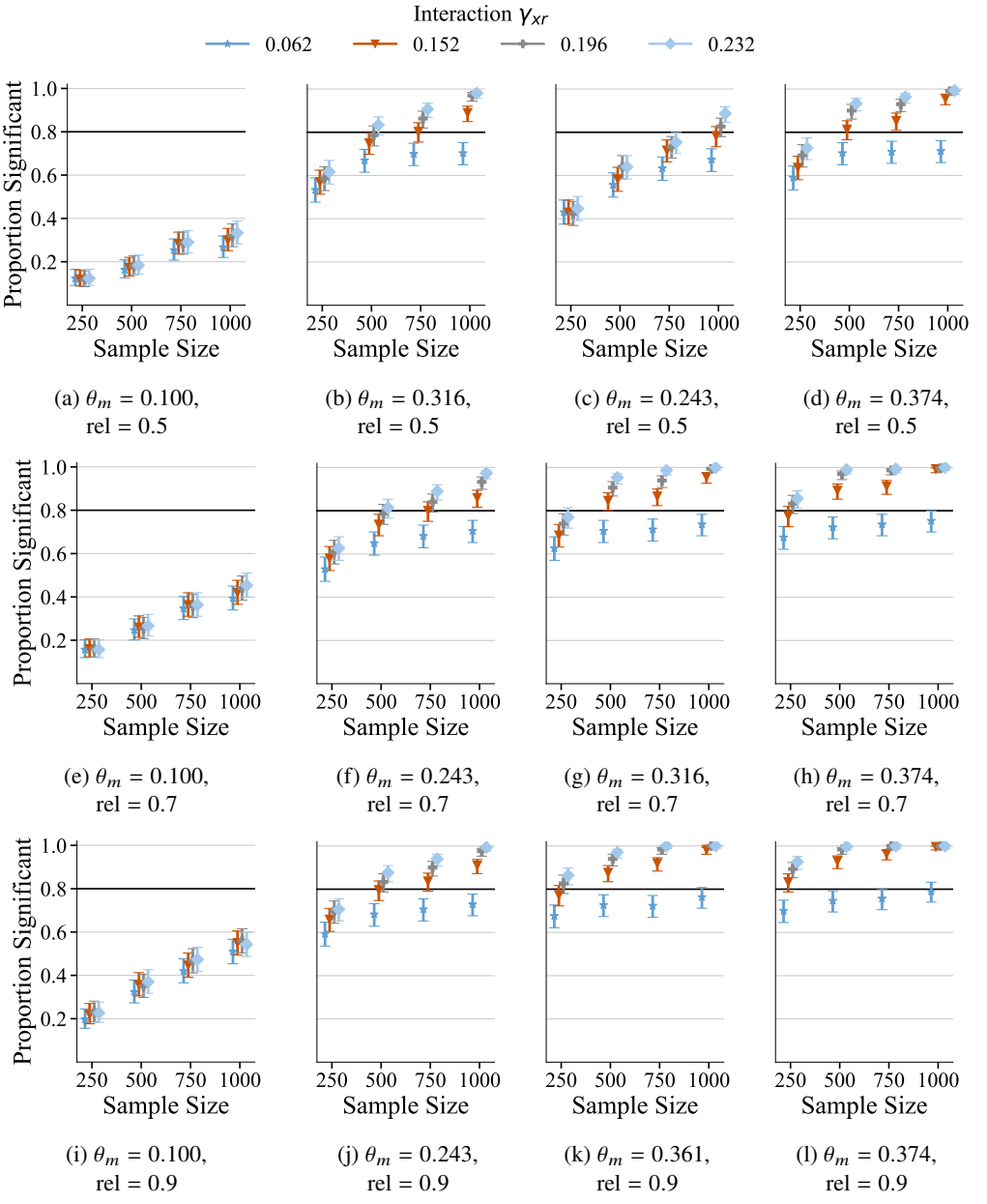


Figure 5: Power of θ_m with Wilson score 95% confidence intervals for varying effect sizes, reliabilities, and covariate-treatment interaction effect sizes.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.

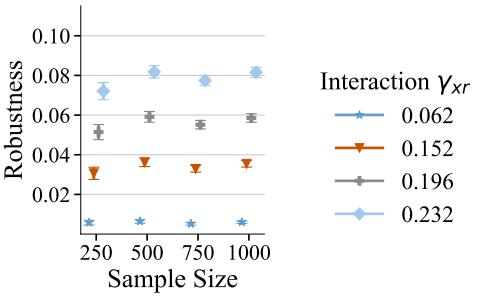
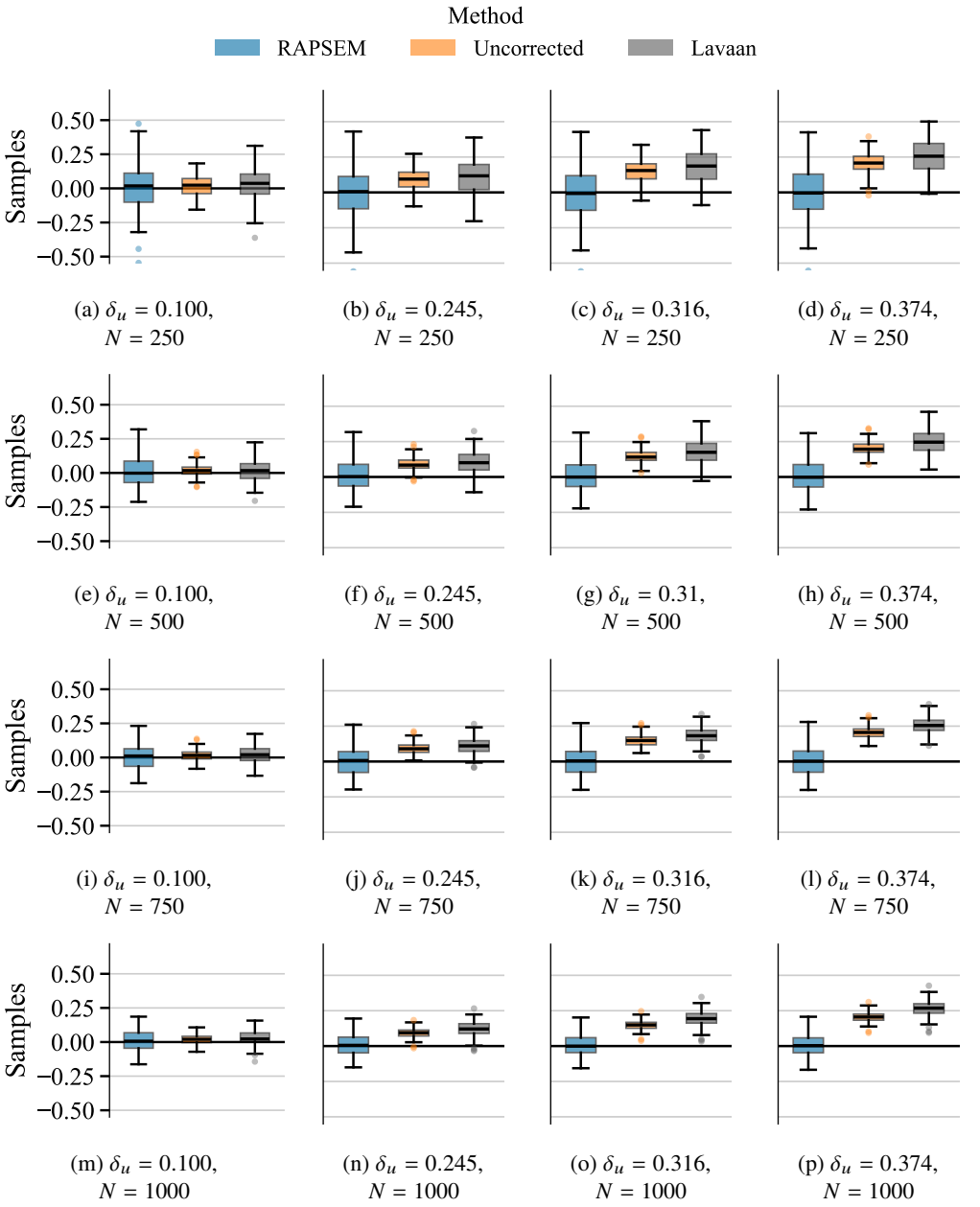


Figure 6: Robustness coefficient $1/\kappa$ with Delta-method approximated 95% confidence intervals for varying covariate-treatment interaction effect sizes.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.



(a) $\delta_u = 0.100$, $N = 250$ (b) $\delta_u = 0.245$, $N = 250$ (c) $\delta_u = 0.316$, $N = 250$ (d) $\delta_u = 0.374$, $N = 250$

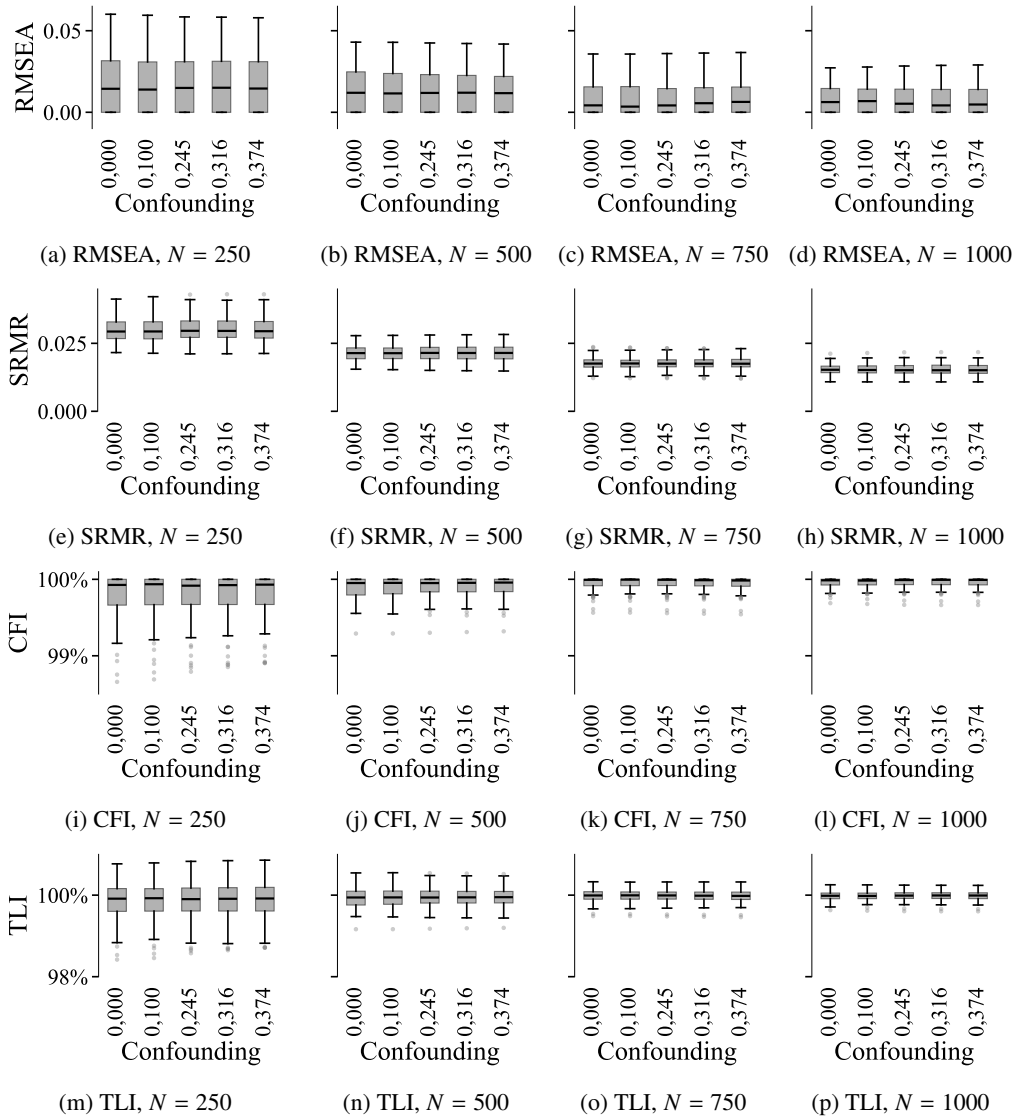
(e) $\delta_u = 0.100$, $N = 500$ (f) $\delta_u = 0.245$, $N = 500$ (g) $\delta_u = 0.31$, $N = 500$ (h) $\delta_u = 0.374$, $N = 500$

(i) $\delta_u = 0.100$, $N = 750$ (j) $\delta_u = 0.245$, $N = 750$ (k) $\delta_u = 0.316$, $N = 750$ (l) $\delta_u = 0.374$, $N = 750$

(m) $\delta_u = 0.100$, $N = 1000$ (n) $\delta_u = 0.245$, $N = 1000$ (o) $\delta_u = 0.316$, $N = 1000$ (p) $\delta_u = 0.374$, $N = 1000$

Figure 7: Sampling distributions of θ_m under varying confounding levels and sample sizes.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.



(a) RMSEA, $N = 250$ (b) RMSEA, $N = 500$ (c) RMSEA, $N = 750$ (d) RMSEA, $N = 1000$

(e) SRMR, $N = 250$ (f) SRMR, $N = 500$ (g) SRMR, $N = 750$ (h) SRMR, $N = 1000$

(i) CFI, $N = 250$ (j) CFI, $N = 500$ (k) CFI, $N = 750$ (l) CFI, $N = 1000$

(m) TLI, $N = 250$ (n) TLI, $N = 500$ (o) TLI, $N = 750$ (p) TLI, $N = 1000$

Figure 8: Lavaan model fit indices across simulation runs under varying confounding levels and sample size.

Downloaded from https://www.cambridge.org/core. 16 Nov 2026 at 08:45:16, subject to the Cambridge Core terms of use.