
RAPSEM: IDENTIFYING LATENT MEDIATORS WITHOUT SEQUENTIAL IGNORABILITY VIA A RANK-PRESERVING STRUCTURAL EQUATION MODEL

Sofia Morelli
Methods Center
Eberhard Karls Universität Tübingen
72074 Tübingen
sofia.morelli@uni-tuebingen.de

Roberto Faleh
Methods Center
Eberhard Karls Universität Tübingen
72074 Tübingen

Holger Brandt
Methods Center
Eberhard Karls Universität Tübingen
72074 Tübingen

October 2, 2025

ABSTRACT

The identification of latent mediator variables is typically conducted using standard structural equation models (SEMs). When SEM is applied to mediation analysis with a causal interpretation, valid inference relies on the strong assumption of no unmeasured confounding, that is, all relevant covariates must be included in the analysis. This assumption is often violated in empirical applications, leading to biased estimates of direct and indirect effects. We address this limitation by weakening the causal assumptions and proposing a procedure that combines g -estimation with a two-stage method of moments to incorporate latent variables, thereby enabling more robust mediation analysis in settings common to the social sciences. We establish consistency and asymptotic normality of the resulting estimator. Simulation studies demonstrate that the estimator is unbiased across a wide range of settings, robust to violations of its underlying no-effect-modifier assumption, and achieves reasonable power to detect medium to large effects for sample sizes above 500, with power increasing as the strength of treatment–covariate interactions grows.

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

1 Introduction

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 are supposed to help to explain how and why treatments work.

Identifying valid mediators offers several advantages. First, they 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. For instance, if maladaptive beliefs are shown to mediate treatment outcomes,

new CBT protocols can be more precisely tailored to target these thought patterns. 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.

To realize these advantages, it is essential to validate the causal role of potential mediators. Mediator candidates are typically chosen based on theoretical claims regarding their involvement in the treatment mechanism. These claims must be rigorously tested empirically and rejected if unsupported by data. Given the risk of confounding and the inherent complexity of psychological mechanisms, it is crucial to employ robust analytical methods to detect valid mediation effects while minimizing the risk of spurious findings under realistic conditions.

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. Consequently, SEM has become the standard approach for identifying latent mediator variables (MacKinnon et al., 2007).

When SEM is used in mediation analysis to identify causal effects, it relies on the assumption of no unmeasured confounding, often referred to as sequential ignorability. This assumption requires that all relevant covariates influencing both the mediator and the outcome must be included in the model. Many studies apply it either explicitly (e.g., Leite et al., 2021; Sun et al., 2021; Wang et al., 2021) or tacit (e.g., Danner et al., 2015; Goldsmith et al., 2018; Hopwood, 2007; Sim et al., 2022).

In practice, however, the sequential ignorability requirement is rarely achievable (Ten Have et al., 2007). In the example of a study examining whether maladaptive beliefs mediate the effect of CBT on social anxiety, 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. Omitting even a single relevant confounder can severely bias mediation estimates, limiting the validity of the findings (Fritz et al., 2016). Consequently, conclusions may apply only to a narrow, unobserved subgroup, undermining the broader goal of identifying generalizable mechanisms of change.

These limitations, which also apply to standard regression with observed variables, have motivated researchers in fields such as (bio)medicine to develop methods that are robust to omitted variable bias (Vansteelandt & Joffe, 2014). One promising approach, exploiting the structure of randomized trials, is rank-preserving models (RPMs) based on g -estimation (Brandt, 2020; Ten Have et al., 2007; Zheng & Zhou, 2015; Zheng et al., 2015). These models rely on a no-effect-modifier assumption, also known as no-essential-heterogeneity (Heckman et al., 2006). Unlike the no-unmeasured-confounding assumption, this alternative allows for omitted covariates, provided they do not interact with the treatment or the mediator. Empirical studies have shown that RPMs can identify mediation effects even under these relaxed assumptions (Zheng & Zhou, 2015).

Despite their advantages, RPMs have not yet been extended to latent variable models. To overcome this significant limitation for fields like psychology, where measurement error is common, we introduce the Rank-Preserving Structural Equation Model (RAPSEM). This novel framework integrates RPM with factor score regression and interaction corrections (Wall & Amemiya, 2000; Wall & Amemiya, 2003), incorporating latent constructs while maintaining the relaxed no-effect-modifier assumption. RAPSEM thus enables robust mediation analysis even in the presence of measurement error and omitted confounders between the mediator and outcome.

In the following sections, we first define the causal effects of interest and introduce the RAPSEM framework. We then describe the estimation procedure and the assumptions required for consistent and asymptotically normal inference, which we formally establish thereafter. Model performance is evaluated in two simulation studies: the first compares the robustness of RAPSEM to standard SEM under violations of key assumptions, and the second examines the power of RAPSEM across varying scenarios. We conclude with a discussion of the method’s strengths and limitations.

2 Causal Effect Definitions

We model the mediation mechanism illustrated in Figure 1 in a controlled trial with randomized treatment assignment R . 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 et al., 2000), adjusting for measured covariates η_X . We define these effects within the potential outcomes framework (Rubin, 2005), using the latent potential outcome

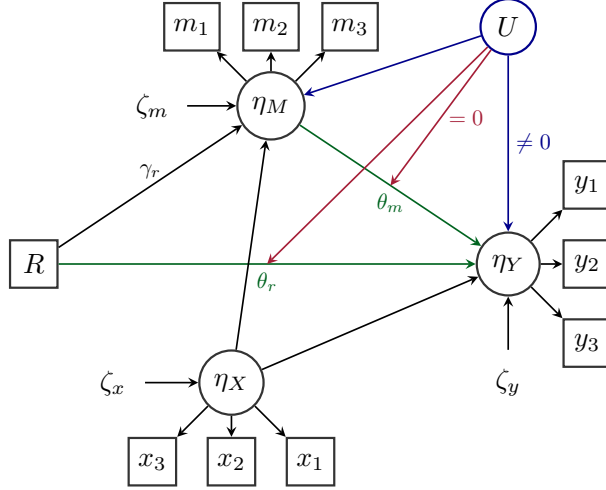


Figure 1: Path diagram of the RAPSEM model estimating the controlled direct effect (θ_r) of treatment R on outcome η_Y and the mediation effect (θ_m) of mediator η_M on outcome η_Y , adjusting for latent covariates η_X . Each latent factor (η_M , η_Y and η_X) is measured by three observed indicators (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-effect-modifier assumption highlighted in red.

η_{Yi}^{rm} for individual i receiving treatment level $R = r$ and experiencing mediator level $\eta_{mi} = m$. We consider a binary treatment that takes values 0 and 1.

The controlled direct effect captures the contrast between potential outcomes under different treatment levels while holding the mediator fixed, corresponding to the parameter θ_r in the path diagram under correct identification:

$$\text{CDE} = \text{E} [\eta_{yi}^{1m} - \eta_{yi}^{0m} \mid \eta_{xi}] = \theta_r. \quad (1)$$

The controlled mediation effect reflects the contrast between different mediator levels, m^1 and m^2 , while holding treatment fixed, corresponding to the parameter θ_m :

$$\text{CME} = \text{E} [\eta_{yi}^{rm^1} - \eta_{yi}^{rm^2} \mid \eta_{xi}] = \theta_m. \quad (2)$$

3 Model Formulation

In RAPSEM, all variables except the treatment R are modeled as latent factors, extending the original RPM to represent latent constructs. The RAPSEM model specification is mathematically equivalent to a standard SEM (Bollen, 1989), comprising a measurement part and a structural part, both of which we introduce below.

3.1 Measurement Model

We consider a measurement model involving three types of latent variables: a mediator η_m measured by a vector of observed indicators $\mathbf{m}$, a set of K latent covariates η_x measured by indicators $\mathbf{x}$, and an outcome variable η_y measured by indicators $\mathbf{y}$. For subject i , the $p \times 1$ vector of all observed indicators is given by

$$\mathbf{z}_i = (\mathbf{m}_i \quad \mathbf{x}_i \quad \mathbf{y}_i)^\top, \quad (3)$$

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

$$\boldsymbol{\eta}_i = (\eta_{m_i} \quad \eta_{x_i} \quad \eta_{y_i})^\top. \quad (4)$$

We define the relationship between the observed indicators and their underlying latent constructs as a linear function:

$$\mathbf{z}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \quad (5)$$

where $\boldsymbol{\tau}$ is the vector of intercepts, $\boldsymbol{\Lambda}$ the factor loading matrix, and $\boldsymbol{\epsilon}_i$ the measurement error.

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

$$\boldsymbol{\tau} = (\boldsymbol{\tau}_{\text{free}} \quad \mathbf{0}_{k \times 1})^\top, \quad \boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_{\text{free}} \quad \mathbf{I}_k)^\top, \quad (6)$$

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.

3.2 Structural Model

We formulate the structural model for the outcome, consistent with the path diagram in Figure 1, within the potential outcomes framework to emphasize its causal interpretation (as in Ten Have et al. (2007)). Interaction terms involving the treatment or mediator are excluded to preserve the interpretability of the controlled direct and mediation effects defined previously.¹ The outcome for person i is specified as

$$\eta_{y_i}^{rm} = \eta_{y_i}^{00} + \theta_r \cdot r + \theta_m \cdot \eta_m + \zeta_{y_i}^{rm}, \quad (7)$$

where the baseline outcome under no treatment and no mediation, $\eta_{y_i}^{00}$, may in principle be modeled as an arbitrary function of the covariates, denoted as $g(\boldsymbol{\eta}_{x_i})$. This potential outcome model corresponds to the structural equation

$$\eta_{y_i} = g(\boldsymbol{\eta}_{x_i}) + \theta_r \cdot r + \theta_m \cdot \eta_m + \zeta_{y_i},$$

which specifies the latent outcome for each individual under their realized treatment and mediator values.

While $g(\cdot)$ may in theory take any functional form, we model it as a linear combination of covariates because existing latent factor score corrections do not extend to nonlinear specifications (Hayes & Usami, 2020; Wall & Amemiya, 2000). The structural equation can thus be written in matrix notation as

$$\boldsymbol{\eta}_y = \boldsymbol{\Xi}_y \boldsymbol{\theta} + \boldsymbol{\zeta}_y \quad (8)$$

with parameter vector

$$\boldsymbol{\theta} = (\theta_r \quad \theta_m \quad \theta_x)^\top$$

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

$$\boldsymbol{\xi}_{y,i} = (r_i \quad \eta_{m_i} \quad t_{yx}(\boldsymbol{\eta}_{x_i})),$$

where $t_{yx}(\cdot)$ denotes a transformation of the latent covariates. We adopt this formulation for the estimation procedure.

To complete the model, we specify a separate structural equation for the latent mediator η_{m_i} . Here, we include interaction terms between treatment and covariates to account for treatment effect heterogeneity that is relevant for the performance of the RAPSEM. The mediator model can then be written as

$$\boldsymbol{\eta}_m = \boldsymbol{\Xi}_m \boldsymbol{\gamma} + \boldsymbol{\zeta}_m, \quad (9)$$

with parameter vector

$$\boldsymbol{\gamma} = (\gamma_r \quad \gamma_x \quad \gamma_{rx})^\top$$

and the rows of the design matrix given by

$$\boldsymbol{\xi}_{m,i} = (r_i \quad t_{mx}(\boldsymbol{\eta}_{x_i}) \quad r_i \cdot t_{mrx}(\boldsymbol{\eta}_{x_i})),$$

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

In principle, the transformations $t(\cdot)$ may be arbitrary (e.g., splines in Brandt, 2020). However, when involving interactions between latent variables, the factor score approach with corrections from Wall and Amemiya (2000) is restricted to accommodating polynomial terms.

For the Simulation Studies, we use a pair of covariates which gets included without transformation in the outcome and mediator model, such that

$$t_{yx}(\boldsymbol{\eta}_{x_i}) = t_{mx}(\boldsymbol{\eta}_{x_i}) = t_{mrx}(\boldsymbol{\eta}_{x_i}) = (\eta_{x1_i} \quad \eta_{x2_i}).$$

This yields the complete structural model for two covariates:

$$\begin{aligned} \eta_{y_i} &= \theta_r \cdot r_i + \theta_m \cdot \eta_{m_i} + \theta_{x1} \cdot \eta_{x1_i} + \theta_{x2} \cdot \eta_{x2_i} + \zeta_{y_i}, \\ \eta_{m_i} &= \gamma_r \cdot r_i + \gamma_{x1} \cdot \eta_{x1_i} + \gamma_{x2} \cdot \eta_{x2_i} + \gamma_{x1r} \cdot r_i \eta_{x1_i} + \gamma_{x2r} \cdot r_i \eta_{x2_i} + \zeta_{m_i}. \end{aligned} \quad (10)$$

¹Interaction terms between treatment and mediator or between treatment/mediator and covariates explicitly capturing some effect heterogeneity can be included as shown in Zheng and Zhou (2015), but they complicate both the definition and identification of causal effects.

4 Estimation

The 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.

4.1 Notation

Let η_i denote the true underlying factor scores in Equation (4), which are used in Equation (8). We later partition these scores into the outcome factor η_{y_i} and the predictor factors $\eta_{\text{pred},i}$, which include the mediator factor η_{m_i} and the covariate factors η_{x_i} . The theoretical estimator of the factor scores based on the true measurement parameters in Equation (5) is denoted by $\tilde{\eta}_i$, while $\hat{\eta}_i$ refers to the practical estimation of $\tilde{\eta}_i$ obtained by plugging in the estimated measurement parameters.

4.2 First Stage

In the first stage, we adopt the method proposed by Wall and Amemiya (2000) to estimate the latent factor scores η_i . Let

$$\Upsilon_1 = (\tau'_{\text{free}}, (\text{vec } \Lambda_{\text{free}})', (\text{vec } \mathbf{H})')',$$

collect the set of free parameters from the measurement model specified in Equations (5) to (6), where Ψ is the residual covariance matrix and vec denotes the column-wise vectorization of a matrix. Given Υ_1 , the estimator for the latent factor scores of subject i is defined as

$$\tilde{\eta}_i = (-\mathbf{H} \quad \mathbf{I}_k + \mathbf{H}\Lambda_{\text{free}}) \left[\mathbf{z}_i - (\tau_{\text{free}} \quad \mathbf{0}_{k \times 1})^\top \right], \quad (11)$$

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

$$\mathbf{H} = (\mathbf{0}_{k \times (p-k)} \quad \mathbf{I}_k) \Psi (\mathbf{I}_{(p-k)} \quad -\Lambda_{\text{free}}^T)^\top \left[(\mathbf{I}_{(p-k)} \quad -\Lambda_{\text{free}}) \Psi (\mathbf{I}_{(p-k)} \quad -\Lambda_{\text{free}}^T)^\top \right]^{-1}.$$

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

$$\tilde{\eta}_i = \eta_i + \mathbf{e}_i, \quad (12)$$

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

$$\mathbf{e}_i = (-\mathbf{H} \quad \mathbf{I}_k + \mathbf{H}\Lambda_{\text{free}}) \epsilon_i. \quad (13)$$

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

$$\Sigma_{ee} = (-\mathbf{H} \quad \mathbf{I}_k + \mathbf{H}\Lambda_{\text{free}}) \Psi (\mathbf{0}_{(p-k) \times k} \quad \mathbf{I}_k)^\top. \quad (14)$$

The parameters in Υ_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 $\hat{\Upsilon}_1$ into Equations (11) and (14) yields the estimated factor scores $\hat{\eta}_i$ and the associated error covariance structure $\hat{\Sigma}_{ee}$, which serve as inputs for the second stage.

4.3 Second stage

In the second stage, the structural equation in Equation (7) is solved with the g -estimation approach introduced by Ten Have et al. (2007) and generalized by Zheng and Zhou (2015) 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).

4.3.1 g -estimation

The g -estimation equations for estimating θ_r and θ_m consist of the orthogonality conditions

$$\begin{aligned} \mathbf{a}_r \cdot \boldsymbol{\zeta}_y &= 0, \\ \mathbf{a}_m \cdot \boldsymbol{\zeta}_y &= 0, \end{aligned} \quad (15)$$

which require the residual vector $\boldsymbol{\zeta}_y$ to be orthogonal to weight vectors associated with the treatment and the mediator which must satisfy the conditional mean restriction

$$\mathbb{E}[\mathbf{a}_j(R, \boldsymbol{\eta}_x) \mid \boldsymbol{\eta}_x] = 0 \text{ for } j \in r, m \quad (16)$$

Zheng and Zhou (2015) showed that the most efficient weights are given by

$$\mathbf{a}_j = (\mathbb{E}[\boldsymbol{\xi}_{y,j} \mid \boldsymbol{\eta}_x, R] - \mathbb{E}[\boldsymbol{\xi}_{y,j} \mid \boldsymbol{\eta}_x]) \cdot \Omega_{\boldsymbol{\eta}_x}^{-1}, \quad (17)$$

where $\boldsymbol{\xi}_{y,j}$ refers to the predictor in Ξ_y corresponding to the weight $\mathbf{a}_j$ and $\Omega_{\boldsymbol{\eta}_x}^{-1}$ to the inverse of the residual covariance matrix of the covariates.

For general nonlinear models, an iterative estimation procedure as in Zheng and Zhou (2015) is required. However, when the outcome is modeled as a linear function of transformed baseline covariates $t_{yx}(\boldsymbol{\eta}_{x_i})$, the estimation simplifies considerably. In this case, the estimation can be carried out in a single step, without estimating $\Omega_{\boldsymbol{\eta}_x}^{-1}$.

Under this linear specification of g , the estimation problem reduces to solving the system

$$\mathbf{W}^\top \boldsymbol{\zeta}_y = 0, \quad (18)$$

which combines the orthogonality conditions from Equation (15) and the regression of the transformed covariates $t_{yx}(\boldsymbol{\eta}_{x_i})$ on the outcome η_y . The matrix $\mathbf{W}$ incorporates both the model weights and the transformed covariates, with each row defined as

$$\mathbf{w}_i = (w_{r_i} \quad w_{m_i} \quad t_{yx}(\boldsymbol{\eta}_{x_i})^\top).$$

Using the observed outcome model in Equation (8), we can compute a naive, sample-based estimate of the structural parameter vector as

$$\bar{\boldsymbol{\theta}} = (\mathbf{W}^\top \Xi_y)^{-1} \mathbf{W}^\top \boldsymbol{\eta}_y. \quad (19)$$

This g -estimation approach aligns closely with the theory of instrumental variables (IV), where weights $\mathbf{w}_r$ and $\mathbf{w}_m$ serve as instruments for the potentially endogenous predictors R and η_m . The closed-form solution in Equation (19) corresponds directly to the IV estimator. The orthogonality conditions in Equation (18) mirror the IV moment condition that instruments are uncorrelated with the residuals.

For a binary treatment, the treatment weight is given by

$$\mathbf{w}_r = \mathbf{r} - \mathbb{E}[R] \quad (20)$$

where $\mathbb{E}[R]$ denotes the sample mean of the treatment indicator, effectively centering the treatment variable.

The mediator weight is constructed as the difference in the expected values of the mediator conditional on the covariates under treatment and control

$$\mathbf{w}_m = (\mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 1] - \mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 0]) \cdot \mathbf{w}_r. \quad (21)$$

This difference can be obtained by fitting the mediator model and computing the predicted mediator values under both treatment conditions. Considering for example the specification from Equation (10), we have

$$\begin{aligned} \mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 1] &= \gamma_r + \gamma_{x1}\eta_{x1_i} + \gamma_{x2}\eta_{x2_i} + \gamma_{x1r} \cdot \eta_{x1_i} + \gamma_{x2r} \cdot \eta_{x2_i} \\ \mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 0] &= \gamma_{x1}\eta_{x1_i} + \gamma_{x2}\eta_{x2_i} \end{aligned}$$

so that the difference simplifies to

$$\mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 1] - \mathbb{E}[\eta_M \mid \boldsymbol{\eta}_X, \mathbf{r} = 0] = \gamma_r + \gamma_{x1r} \cdot \eta_{x1_i} + \gamma_{x2r} \cdot \eta_{x2_i}. \quad (22)$$

4.3.2 Factor Score Correction

The structural form of the outcome factor score $\eta_{y,i}$ may depend on polynomial terms of the covariate and mediator scores $\eta_{\text{pred},i}$. Any interactions or higher-order terms involving η_i must be corrected using the moments of the measurement error $\mathbf{e}_i$. We define these moments of $\mathbf{e}_i$ needed for the correction as Υ_2 . Assuming normally distributed measurement errors ϵ_i , they reduce to linear combinations of elements in the estimated covariance matrix $\hat{\Sigma}_{ee}$.

Instead of using the naive sample-based estimator in Equation (19), which ignores measurement error in the factor scores, we estimate the true population parameter

$$\theta = E[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top]^{-1} E[\mathbf{w}_i \eta_{y_i}] \quad (23)$$

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

$$\hat{\theta} = \hat{\mathbf{M}}^{-1} \hat{\mathbf{m}}, \quad (24)$$

where

$$\begin{aligned} \hat{\mathbf{M}} &= \frac{1}{N} \sum_{i=1}^N M(\hat{\eta}_{\text{pred},i}, \hat{\Upsilon}_2), \\ \hat{\mathbf{m}} &= \frac{1}{N} \sum_{i=1}^N m(\hat{\eta}_i, \hat{\Upsilon}_2) \end{aligned} \quad (25)$$

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

$$\begin{aligned} E[M(\tilde{\eta}_{\text{pred},i}, \Upsilon_2) | \eta_i] &= E[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top], \\ E[m(\tilde{\eta}_i, \Upsilon_2) | \eta_i] &= E[\mathbf{w}_i \eta_{y_i}] \end{aligned} \quad (26)$$

where $\tilde{\eta}_i$ denotes the theoretical factor score estimator based on the true measurement parameters.

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

$$(M(\tilde{\eta}_{\text{pred},i}, \Upsilon_2) \quad m(\tilde{\eta}_i, \Upsilon_2)) = \sum_{j=0}^J (-1)^j A_j(\tilde{\eta}_i, \Upsilon_2), \quad (27)$$

where the correction terms are recursively defined as

$$\begin{aligned} A_0(\tilde{\eta}_i, \Upsilon_2) &= \tilde{\mathbf{W}}^\top (\tilde{\Xi}_y \quad \tilde{\eta}_{y,i}), \\ A_j(\tilde{\eta}_i, \Upsilon_2) &= E[A_{j-1}(\tilde{\eta}_i, \Upsilon_2) | \eta_i] - A_{j-1}(\tilde{\eta}_i, \Upsilon_2) \text{ for } j \in \{1, \dots, J\}. \end{aligned} \quad (28)$$

Each A_j term captures contributions of order up to $2(J-j)$ in $\eta_{\text{pred},i}$ and up to order j in the measurement error moments. The term A_0 corresponds to the uncorrected estimates, while A_j for $j > 0$ serves as a correction based on higher-order moments of $\mathbf{e}_i$.

Plugging in the estimates $\hat{\eta}_i$ and $\hat{\Sigma}_{ee}$ from the first stage, we yield $\hat{\theta}$ in Equation (24).

In the concrete model defined in Equation (10), the polynomial order is $J = 1$, so the moment expansion consists of A_0 and A_1 . A_0 provides the uncorrected terms

$$\begin{aligned} A_0(\hat{\eta}_i, \hat{\Upsilon}_{12}) &= (\hat{\mathbf{w}}_r \quad \hat{\mathbf{w}}_m \quad \hat{\eta}_{1x} \quad \hat{\eta}_{2x})^\top (\mathbf{r} \quad \hat{\eta}_m \quad \hat{\eta}_{1x} \quad \hat{\eta}_{2x} \quad \hat{\eta}_{y,i}) \\ &= \begin{pmatrix} \hat{\mathbf{w}}_r^\top \mathbf{r} & \hat{\mathbf{w}}_r^\top \hat{\eta}_m & \hat{\mathbf{w}}_r^\top \hat{\eta}_{1x} & \hat{\mathbf{w}}_r^\top \hat{\eta}_{2x} & \hat{\mathbf{w}}_r^\top \hat{\eta}_{y,i} \\ \hat{\mathbf{w}}_m^\top \mathbf{r} & \hat{\mathbf{w}}_m^\top \hat{\eta}_m & \hat{\mathbf{w}}_m^\top \hat{\eta}_{1x} & \hat{\mathbf{w}}_m^\top \hat{\eta}_{2x} & \hat{\mathbf{w}}_m^\top \hat{\eta}_{y,i} \\ \hat{\eta}_{1x}^\top \mathbf{r} & \hat{\eta}_{1x}^\top \hat{\eta}_m & \hat{\eta}_{1x}^\top \hat{\eta}_{1x} & \hat{\eta}_{1x}^\top \hat{\eta}_{2x} & \hat{\eta}_{1x}^\top \hat{\eta}_{y,i} \\ \hat{\eta}_{2x}^\top \mathbf{r} & \hat{\eta}_{2x}^\top \hat{\eta}_m & \hat{\eta}_{2x}^\top \hat{\eta}_{1x} & \hat{\eta}_{2x}^\top \hat{\eta}_{2x} & \hat{\eta}_{2x}^\top \hat{\eta}_{y,i} \end{pmatrix}. \end{aligned}$$

The correction term $A_1(\hat{\eta}_i, \hat{\Sigma}_{ee})$ involves expectations over the measurement error variance and covariances. Under the latent factor ordering assumed in Equation (4), and using the form of $\hat{\mathbf{w}}_m$ from Equations (21) to (22), it takes the form

$$A_1(\hat{\eta}_i, \hat{\Upsilon}_{12}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\gamma}_{x2r} \hat{\Sigma}_{ee,31}^+ \hat{\mathbf{w}}_r^\top & \hat{\gamma}_{x2r} \hat{\Sigma}_{ee,32}^+ \hat{\mathbf{w}}_r^\top & \hat{\gamma}_{x2r} \hat{\Sigma}_{ee,33}^+ \hat{\mathbf{w}}_r^\top & \hat{\gamma}_{x2r} \hat{\Sigma}_{ee,34}^+ \hat{\mathbf{w}}_r^\top \\ 0 & \hat{\Sigma}_{ee,21} & \hat{\Sigma}_{ee,22} & \hat{\Sigma}_{ee,23} & \hat{\Sigma}_{ee,24} \\ 0 & \hat{\Sigma}_{ee,31} & \hat{\Sigma}_{ee,32} & \hat{\Sigma}_{ee,33} & \hat{\Sigma}_{ee,34} \end{pmatrix}.$$

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

4.3.3 Modifications for Numerical Stability

To improve efficiency and ensure that M remains positive definite, particularly in small-sample settings, we adopt the modification of the 2SMM estimator in Equation (25) proposed by Wall and Amemiya (2000). We separate the uncorrected estimates

$$(\hat{\mathbf{M}}_1 \quad \hat{\mathbf{m}}_1) = \frac{1}{N} \sum_{i=1}^N A_0(\hat{\eta}_i, \hat{\Upsilon}_{12}),$$

from the correction terms

$$(\hat{\mathbf{M}}_2 \quad \hat{\mathbf{m}}_2) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J (-1)^j A_j(\hat{\eta}_i, \hat{\Upsilon}_{12}).$$

Then, we define

$$\mathbf{R}_1 = \begin{pmatrix} \hat{\mathbf{M}}_1 & \hat{\mathbf{m}}_1 \\ \hat{\mathbf{m}}_1^\top & \frac{1}{N} \sum_{i=1}^N \hat{\eta}_{y,i}^2 \end{pmatrix} \text{ and } \mathbf{R}_2 = \begin{pmatrix} -\hat{\mathbf{M}}_2 & -\hat{\mathbf{m}}_2 \\ -\hat{\mathbf{m}}_2^\top & \hat{\Sigma}_{ee,11} \end{pmatrix}.$$

Let $\hat{\lambda}$ denote the largest eigenvalue of

$$\mathbf{R}_1^{-1/2} \mathbf{R}_2 \mathbf{R}_1^{-1/2}.$$

Then, the modified estimator is given by

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

where $\tau \in [0, J + 5]$ is an empirically chosen tuning parameter (Wall & Amemiya, 2000).

To further stabilize estimation, we allow a small variance v to be added to the diagonal of $\check{\mathbf{M}}$, analogous to applying a ridge penalty in linear regression. The final estimator becomes

$$\check{\boldsymbol{\theta}} = (\check{\mathbf{M}} + v\mathbf{I})^{-1} \check{\mathbf{m}}. \quad (29)$$

5 Assumptions

The following assumptions (see Table 1 for an overview) are required to ensure the identifiability of the causal effect parameters θ_r and θ_m and the consistency and asymptotic normality of the estimates.

Causal and Identification Assumptions

We first outline the assumptions that define the underlying causal structure and provide the conditions under which the parameters θ_r and θ_m can be identified using the latent g -estimation framework.

Assumption C1 - Consistency: The realized outcome corresponding to a given treatment and mediator assignment equals the potential outcome under those values

$$\eta_y(r, \eta_m) = \eta_y^{rm}.$$

This assumption ensures a unique mapping from treatment and mediator values to potential outcomes.

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

Category	Assumption Names
Causal and Identification Assumptions	C1 Consistency
	C2 Positivity
	C3 SUTVA
	C4 Treatment Randomization
	C5 No Effect Modification
Statistical and Regularity Assumptions	S1 Full Rank of Moment Matrix
	S2 Model Specification
	S3 IID Sampling
	S4 Finite Moments
	S5 Measurement Error
	S6 Structural Equation Error

Assumption C2 - Positivity: For all levels of baseline covariates η_x with positive probability, the probability of receiving each treatment level is strictly between 0 and 1:

$$0 < P(r \mid \eta_x) < 1.$$

Similarly, for mediator values, we require

$$0 < P(m \mid r, \eta_x) < 1$$

for all relevant (r, m, η_x) , ensuring that causal effects are estimable across the covariate space.

Assumption C3 - Stable Unit Treatment Value Assumption (SUTVA): There is a single well-defined version of each treatment, and no interference exists between units; that is, one unit’s treatment does not affect another unit’s outcome.

Assumption C4 - Randomization: Conditional on measured baseline covariates η_x , the treatment assignment is independent of all potential outcomes and potential mediator values:

$$r \perp (\eta_m(r), \eta_y(r, m)) \mid \eta_x.$$

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 always holds for randomized interventions.

Assumption C5 - No Effect Modification: The causal effects of the treatment and mediator are homogeneous across both observed covariates η_x and unmeasured confounders u :

$$\theta_r(\eta_x, u) = \theta_r, \quad \theta_m(\eta_x, u) = \theta_m.$$

Equivalently, the latent outcome model satisfies

$$\eta_Y = \theta_r R + \theta_m \eta_M + f(\eta_X, u) + \zeta_Y,$$

where $f(\eta_x, u)$ is an arbitrary function capturing the joint influence of η_x and u on the baseline level of the outcome, and ζ_y is an independent error term. Crucially, this formulation excludes any interaction terms involving R or η_M . That is, neither θ_r nor θ_m varies with η_x or u , unless explicitly modeled (see Zheng & Zhou, 2015, for such extensions).

While Assumptions C1–C4 are standard in causal inference, Assumption C5 is central to our framework because it is weaker than the commonly invoked sequential ignorability assumption. Sequential ignorability requires that the mediator be independent of potential outcomes, conditional on treatment and covariates. In contrast, the no-effect-modification assumption does not impose this independence. Instead, it allows for unmeasured confounding between the mediator and the outcome but rules out effect heterogeneity arising from interactions between treatment and either covariates or the mediator, as illustrated in Figure 1. Violations of both assumptions are examined in Simulation Study 1.

The no-effect-modification assumption is often plausible in intervention studies because standardized treatments are designed to produce consistent effects across participants. For instance, blood pressure-lowering medications typically reduce pressure in a similar manner within patient subgroups stratified by known covariates, such as age or baseline severity. Similarly, structured psychological interventions, such as a cognitive-behavioral therapy (CBT) protocol for anxiety (Flückiger, 2014), are designed to provide comparable benefits to individuals following the same protocol-driven procedures. In these contexts, the relative ordering of treatment effects is generally preserved, making no-effect-modification a reasonable working assumption.

In contrast, as noted in the introduction, sequential ignorability is frequently unrealistic. Unobserved factors, including diet or stress in blood pressure studies and baseline motivation or social support in CBT interventions, may confound the relationships between mediators and outcomes.

Statistical and Regularity Assumptions

Consistent and asymptotically normal estimation of the parameters requires the modeling assumptions and conditions detailed in Section A. Here, we highlight the most crucial condition for g -estimation.

Assumption S1 - Full Rank of Moment Matrix: The matrix $\mathbf{W}^\top \Xi_y$ must be of full rank to ensure invertibility. In particular, this requires at least one treatment–covariate interaction effect on the mediator (e.g., $r \times \eta_{x_k} \rightarrow \eta_m$).

If this condition fails, the mediator weight becomes collinear with the treatment. As shown in Equation (22), when $\gamma_{x1r} = \gamma_{x2r} = 0$, only the constant term γ_r remains, and the g -estimation equation has no unique solution. We investigate the consequences of such violations in Simulation Study 2.

6 Asymptotic Properties

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

Theorem 1. *Let $\hat{\Upsilon}$ solve the estimation equation*

$$\frac{1}{N} \sum_{i=1}^N \psi(\mathbf{z}_i, \hat{\Upsilon}) = \mathbf{0},$$

and let Υ be the unique solution to $E[\psi(\mathbf{z}_i, \Upsilon)] = \mathbf{0}$, with nonsingular Jacobian

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

Under Assumptions C2–S6,

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

where $\{\Delta_i\}$ are i.i.d. with $E[\Delta_i] = 0$ and finite variance. Hence, $\hat{\Upsilon}$ is $\sqrt{N}$ -consistent.

Next, we regard the behavior of the factor score estimator $\hat{\eta}_i$ relative to $\tilde{\eta}_i$ defined in Equation (11).

Theorem 2. *Under Theorem 1,*

$$\hat{\eta}_i - \tilde{\eta}_i = \hat{\mathbf{B}}_i (\hat{\Upsilon}_1 - \Upsilon_1),$$

where

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

with

$$\mathbf{q}_i = (0 \quad \mathbf{I}_k) \mathbf{z}_i \quad \text{and} \quad \mathbf{p}_i = (\mathbf{I}_{p-k} \quad -\Lambda_{\text{free}}) \left[\mathbf{z}_i - (\boldsymbol{\tau}_{\text{free}} \quad \mathbf{0}_{k \times 1})^\top \right].$$

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

Finally, we can formulate the properties of the structural parameter estimator $\hat{\boldsymbol{\theta}}$ defined in Equation (24).

Theorem 3. *Under Assumptions S1–S6, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed with asymptotic variance*

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

where

$$\begin{aligned} \mathbf{G} &= E\left[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top\right], \\ \mathbf{S} &= \text{Var}[\mathbf{d}_i], \quad \mathbf{d}_i = \mathbf{l}(\tilde{\eta}_i, \Upsilon_2, \boldsymbol{\theta}) + \overline{\mathbf{C}} \Delta_i, \\ \mathbf{l}(\eta_i, \Upsilon_2, \boldsymbol{\theta}) &= \mathbf{m}(\eta_i, \Upsilon_2) - \mathbf{M}(\eta_{\text{pred},i}, \Upsilon_2) \boldsymbol{\theta}, \\ \overline{\mathbf{C}} &= E\left[\left(\frac{\partial \mathbf{l}}{\partial \eta_i} \Big|_{\tilde{\eta}_i, \Upsilon_2, \boldsymbol{\theta}} \mathbf{B}_i \quad \frac{\partial \mathbf{l}}{\partial \Upsilon_2} \Big|_{\tilde{\eta}_i, \Upsilon_2, \boldsymbol{\theta}}\right)\right]. \end{aligned}$$

Under Assumptions C1–C5, the parameters θ_r and θ_m in Equation (8) are globally identifiable and correspond to the causal effects of the treatment and mediator as defined in Equations (1) to (2), respectively.

All proofs are provided in the Appendix.

Theorem 3 also holds for the modified 2SMM estimator $\tilde{\theta} = \tilde{M}^{-1}\tilde{m}$ (Wall & Amemiya, 2000). Our adapted version in Equation (29), however, introduces bias towards zero. This bias–variance trade-off is deliberate, as the reduction in variance increases efficiency in line with the standard rationale of regularization. In practice, we choose v sufficiently small so that effect size estimates remain essentially unaffected.

7 Implementation

We provide an implementation of the latent g -estimation procedure in the R package `rapsem`, available at <https://github.com/PsychometricsMZ/RAPSEM>. The package allows estimation of structural models as specified in Equation (5), Equation (8), and Equation (9). While 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 A_0 and A_1), but the framework can be extended to higher orders by incorporating additional correction terms $A_{j>1}$.

Estimation is carried out via the function `est_med`, 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. For the measurement model, we impose the constraints specified in Equation (6), estimate factor intercepts, loadings, and residual variances with `lavaan`, and compute factor scores according to Equation (11). Structural parameters are then estimated using the modified 2SMM regularized g -estimator in Equation (29), with default settings $\tau = 5$ and $v = 10^{-4}$. Because the ridge penalty introduces bias, variances were estimated via bootstrapping rather than the analytic asymptotic variance formula in Theorem 3, using a default of 100 bootstrap samples.

8 Simulation Studies

In this section, we present two Simulation Studies. First, we assess the sensitivity of RAPSEM to violations of the no-effect-modifier assumption and compare its robustness to a standard SEM when the no-unmeasured-confounder assumption is violated. Then, we examine the power of RAPSEM under different conditions.

8.1 Data Generation

We generated data based on Equation (10) with an additional confounding variable u_i

$$\begin{aligned}\eta_{y_i} &= 0.125 \cdot r_i + \theta_m \cdot \eta_{m_i} + 0.226 \cdot \eta_{x1_i} + 0.226 \cdot \eta_{x2_i} + \delta_u \cdot u_i + \delta_{ur} \cdot u_i r_i + \zeta_{y_i}, \\ \eta_{m_i} &= 0.3 \cdot r_i + 0.3 \cdot \eta_{x1_i} + 0.3 \cdot \eta_{x2_i} + \gamma_{x1r} \cdot r_i \eta_{x1_i} + \gamma_{x2r} \cdot r_i \eta_{x2_i} + \delta_u \cdot u_i + \zeta_{m_i},\end{aligned}\tag{30}$$

where the parameters $\theta_r, \theta_x, \gamma_r, \gamma_x$ were fixed across all simulations with values taken from Brandt (2020).

The treatment r was randomly sampled with replacement, taking values -1 and 1 . The covariates η_{1x} and η_{2x} were drawn from a common multivariate standard normal distribution with correlation $\rho = 0.2$. The confounder u was drawn from a standard normal distribution independent of the covariates. The residual variables ζ_m and ζ_y were each normally distributed with a variance such that η_m and η_y had a variance of one.

For each latent variable, we generated three indicator variables with residual variances and intercepts sampled from specified distributions. This procedure was applied separately for each dataset and each item to enable a more comprehensive evaluation of model performance, rather than relying on identical values. Item-specific reliabilities κ_j were drawn from a uniform distribution $U(\kappa - 0.1, \kappa + 0.1)$, where the respective reliability condition determined κ . The residual variance $\frac{1}{\kappa_j} - 1$ was then added to the generated latent variable. Consequently, the standardized factor loadings varied across items. In addition, item intercepts were randomly drawn from a uniform distribution $U(-1, 1)$.

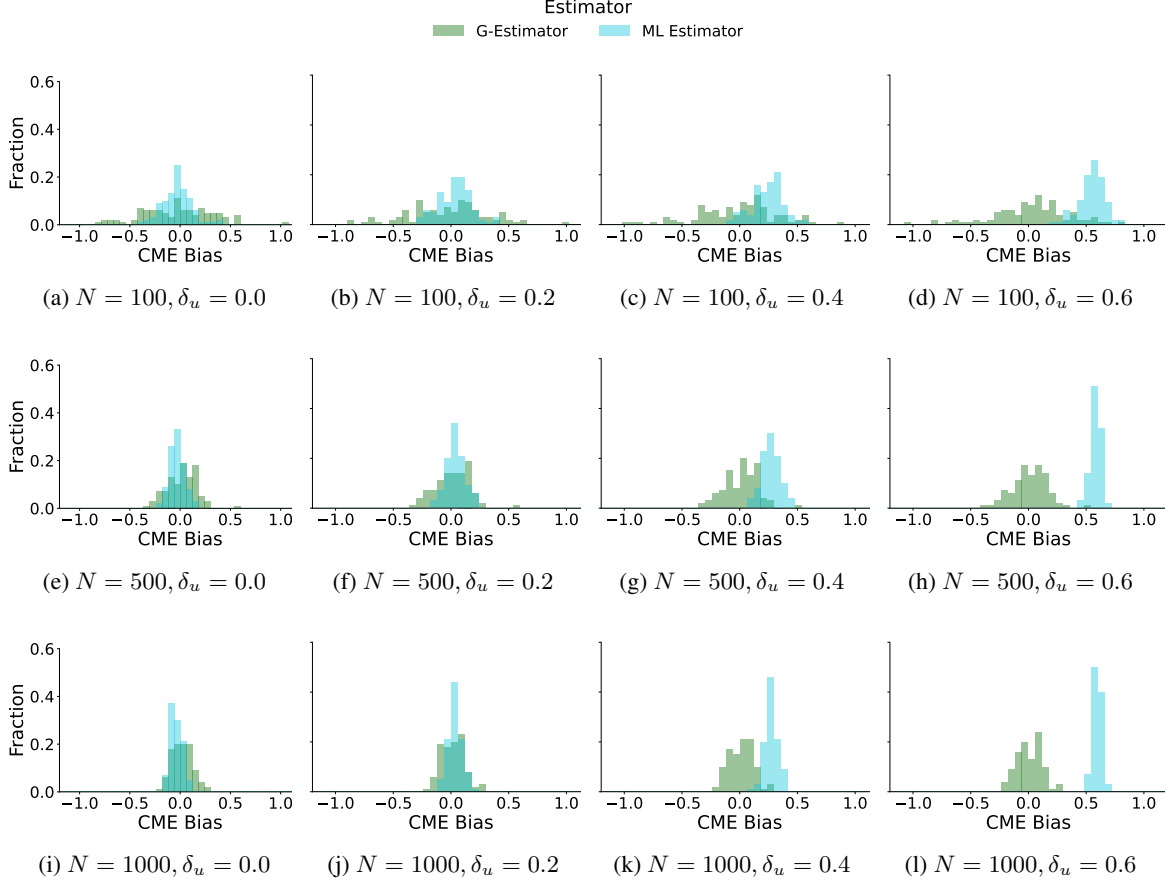
We used 100 data sets per data condition. To facilitate direct comparability and minimize variability due to random sampling, the same data sets were applied across conditions by fixing the random seed. The code is also available on GitHub.

8.2 Study 1

To test for robustness, data were generated under violations of (a) the no-unmeasured confounder assumption and (b) the no-effect modifier assumption. The parameters γ_{x1r} and γ_{x2r} were fixed at 0.204, and the reliability κ was chosen at 0.75, representing a setting with high statistical power and thus high potential for spurious effects. The parameter

Table 2: Data conditions for Simulation Study 1. For δ_u , and δ_{ur} , the explained variance is indicated in parentheses.

Confounding effect size δ_u	0.0 (0%)	0.2 (5%)	0.4 (10%)	0.6 (15%)	
Modification effect size δ_{ur}	0.0 (0%)	0.3 (5%)	0.6 (10%)	0.9 (15%)	
Sample size N	100	250	500	750	1000

Figure 2: Histograms of CME bias under varying confounding levels and sample sizes. Each panel shows the relative frequency of bias values for the ML Estimator and G-Estimator. Rows correspond to selected sample sizes ($N = 100, 500, 1000$), while columns correspond to different confounding levels ($\delta_u = 0.0, 0.2, 0.4, 0.6$).

θ_m was set to zero to enable the assessment of the type I error rate. For comparison, a standard SEM was estimated alongside the RAPSEM.

Data conditions The data conditions are summarized in Table 2. Sample sizes were varied across $N \in \{100, 250, 500, 750, 1000\}$.

Violations of the no-unmeasured-confounder assumption were introduced by setting $\delta_u \in \{0.2, 0.4, 0.6\}$, explaining 5%, 10%, and 15% of the variance. A baseline condition with $\delta_u = 0$ was also included.

Similarly, the no-effect-modifier assumption was manipulated by either setting $\delta_u = 0$ (assumption holds) or by violating it with $\delta_{ur} \in \{0.3, 0.6, 0.9\}$, which corresponded to 5%, 10%, and 15% explained variance.

Results Violations of sequential ignorability in SEM induced substantial positive bias (Figure 2). Under moderate to strong confounding ($\delta_u = 0.4, 0.6$), bias distributions were markedly shifted away from zero, highlighting SEM's sensitivity to modest violations of sequential ignorability. By contrast, RAPSEM's bias distributions remained

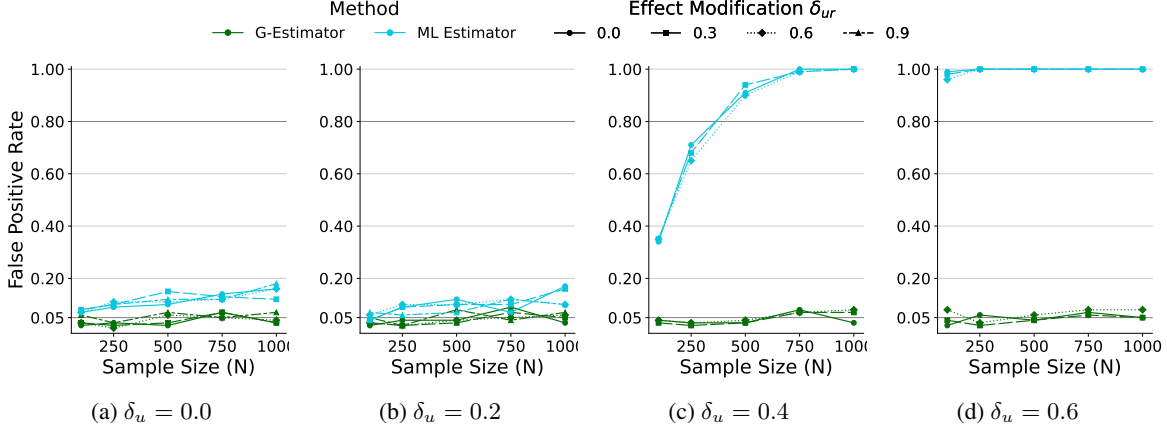


Figure 3: False positive rates under varying levels of confounding. Each panel shows the false positive rate as a function of sample size for a fixed confounding level ($\delta_u = 0.0, 0.2, 0.4, 0.6$). Colors distinguish estimation methods (ML Estimator vs. G-Estimator), while line styles and markers correspond to different effect modification levels ($\delta_{ur} = 0.0, 0.3, 0.6, 0.9$).

consistently centered around zero. At small sample sizes, RAPSEM exhibited greater variability than SEM, reflecting a trade-off of small-sample efficiency for robustness. With increasing sample size ($N = 1000$), this variance difference diminished, though RAPSEM’s variability remained slightly higher.

Type I error rates (Figure 3) were severely inflated under SEM, approaching 100% in the presence of moderate to strong confounding, with inflation worsening as sample size increased. RAPSEM, in contrast, consistently controlled Type I error across all settings, remaining below the nominal 5% level and never exceeding 10%.

Effect modification (δ_{ur}) had minimal influence on the performance of either method, as shown by the overlapping colored lines (Figure 3). For SEM, this insensitivity is expected given its modeling assumptions. For RAPSEM, the robustness in the presence of effect modification provides further evidence of its validity, consistent with findings in Brandt (2020).

8.3 Study 2

We identified four key factors that determine the power of RAPSEM: the effect size θ_m , the sample size N , the reliability of the indicators κ , and the strength of the covariate–treatment interaction γ_{xr} . These parameters were systematically varied to assess the conditions under which the method achieves adequate power. Meanwhile, we fixed the effect of the confounder variable at $\delta_u = 0.4$, thereby violating the no-unmeasured confounder assumption, while setting $\delta_{ur} = 0$ to ensure that the no-effect-modifier assumption was satisfied.

Data conditions The data conditions are summarized in Table 3. Sample sizes were varied across $N \in \{250, 500, 750, 1000\}$.

We considered two values for the conditional main effect (CME) θ_m : 0.29, representing a medium effect with 5% explained variance, and 0.41, representing a large effect with 10% explained variance.

Indicator reliabilities were set to $\kappa = 0.4, 0.5, 0.667$ and 0.8, corresponding to residual variances of 1.5, 1, 0.5 and 0.25, respectively.

Latent interaction effects γ_{x1r} and γ_{x2r} were set to 0.102, 0.145, 0.176, or 0.204, corresponding to 2.5%, 5%, 7.5%, and 10% explained variance for the two interactions combined ($r \times \eta_{x1}$ and $r \times \eta_{x2}$). This range reflects small (1.25% each) to large (5% each) interaction effects, consistent with empirical evidence (e.g., Jaccard et al., 1990).

Results The power to detect a non-zero CME (Figure 4) increased systematically with sample size, measurement reliability (κ), and the covariate–treatment interaction effect (γ_{xr}).

A desired rate of 80% power could be achieved only for the larger CME effect size of $\theta_m = 0.41$ (corresponding to 10% explained variance) under favorable conditions—namely, $N \geq 500$, medium-to-large interaction effects, and at

Table 3: Data conditions for Simulation Study 2. For γ_{x1r} , γ_{x2r} , and θ_m , the explained variance is indicated in parentheses.

CME effect size θ_m	0.29 (5%)	0.41 (10%)		
Reliability κ	0.4	0.5	0.667	0.8
Interaction effect size	0.102	0.145	0.176	0.204
$\gamma_{x1r} = \gamma_{x2r}$	(1.25%)	(2.5%)	(3.75%)	(5%)
Sample size N	250	500	750	1000

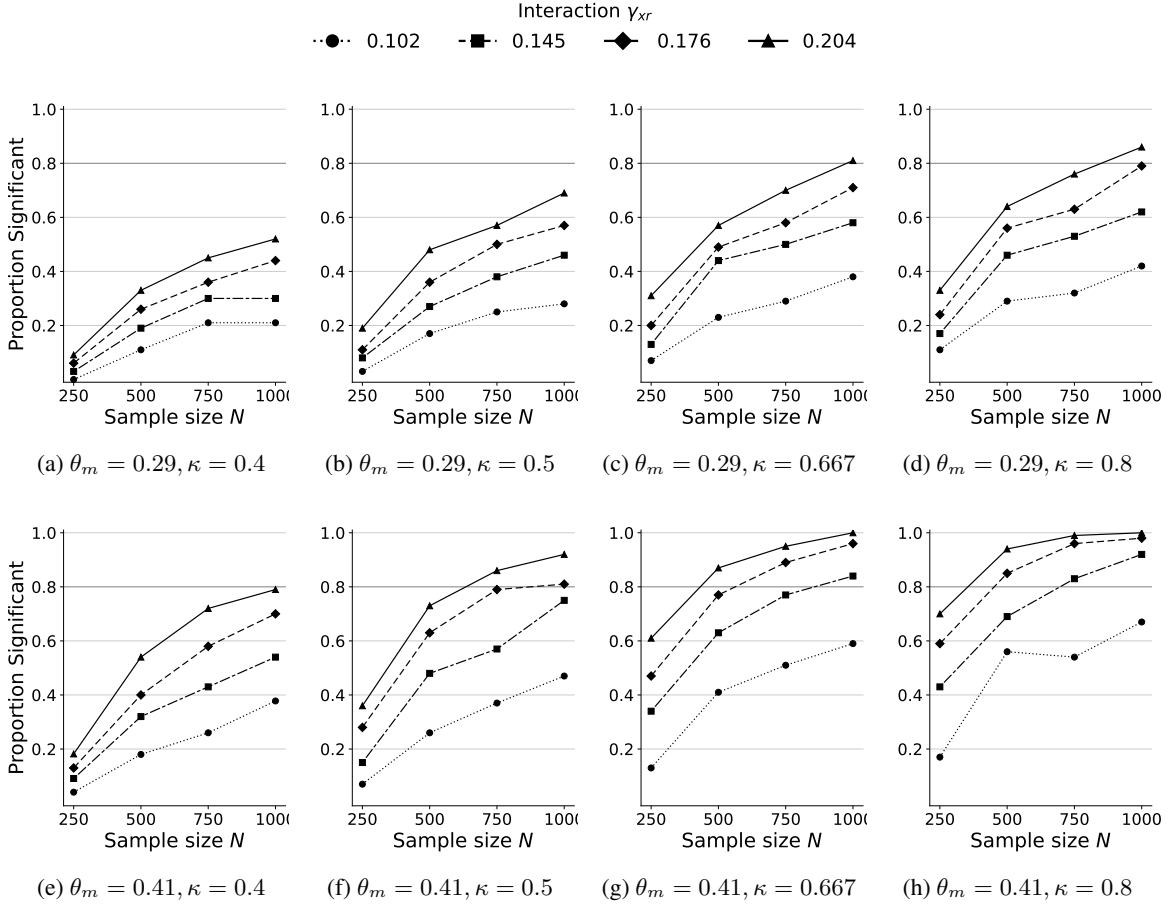


Figure 4: Power of detecting CME θ_m under varying reliability and effect sizes. Each panel shows the rate of significant values for different covariate-treatment interaction effect sizes ($\gamma_{xr} = 0.102, 0.145, 0.176, 0.204$). Rows correspond to CME effect size ($\theta_m = 0.29, 0.41$), while columns correspond to increasing reliabilities ($\kappa = 0.4, 0.5, 0.667, 0.8$).

least moderate reliability. For the smaller CME effect size ($\theta_m = 0.29$), power remained well below 80% except in the most optimistic scenarios (large γ_{xr} , high reliability, and $N = 1000$).

These results underscore the joint importance of sufficient reliability and strong covariate–treatment interactions in enabling adequate power to detect CME effects, even when sample sizes are moderately large.

9 Discussion

This article introduced 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 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 weaker effect-modification assumption, rather than the stronger sequential ignorability assumption.

We establish both consistency and asymptotic normality of the resulting estimator, and further implement a regularized version to ensure numerical stability. Simulation studies demonstrate that the proposed estimation method yields unbiased estimates across a range of conditions, exhibits robustness to violations of the no-effect modifier assumption, and achieves reasonable power to detect medium to large effects when sample sizes exceed $N = 500$. Moreover, statistical power is affected by interaction effects between covariates and treatment, and increases with the magnitude of these interactions, which is in line with previous findings with the RPM (e.g., Zheng et al., 2015). Here, we found that also an increased reliability of the indicator variables will improve power.

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 a 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.

Notably, the proposed approach holds strong potential for robustly identifying mediation effects in the presence of unobserved confounding, particularly in large-scale intervention studies. However, actual randomized trials in this context are scarce, making it even more important to extend the approach to address confounding in the context of selected instead of randomized treatments.

Future work could extend the framework by incorporating multiple treatments and treatment–mediator interactions within the structural model, as well as accommodating non-linear and dichotomous measurement models. Introducing a nonlinear specification for baseline effects may further increase power when linearity assumptions are violated (Brandt, 2020). Furthermore, higher-order polynomials and interaction terms among latent variables can be readily integrated. These enhancements would broaden the method’s applicability and strengthen its ability to capture complex causal structures.

The extension of the RPM formulation to latent variable models opens new opportunities to broaden the flexibility of the approach, for example, for longitudinal data (via latent growth curve models), or intensive longitudinal data (via DSEM). Extensions though will need a very thorough evaluation and adaptation of the underlying (causal) assumptions.

Acknowledgement

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.

References

- 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>
- Castella, K. D., Goldin, P., Jazaieri, H., Heimberg, R. G., Dweck, C. S., & and, J. J. G. (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>
- Danner, D., Hagemann, D., & Fiedler, K. (2015). Mediation analysis with structural equation models: Combining theory, design, and statistics. *European Journal of Social Psychology*, 45, 460–481. <https://doi.org/10.1002/ejsp.2106>
- Flückiger, C. (2014). The adherence/resource priming paradigm—a randomised clinical trial conducting a bonafide psychotherapy protocol for generalised anxiety disorder. *BMC psychiatry*, 14, 1–8.
- 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>
- Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M., & Pickles, A. (2018). Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. *Psychological Methods*, 23(2), 191–207. <https://doi.org/10.1037/met0000154>
- 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., & Vytlačil, 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.
- Hopwood, C. J. (2007). Moderation and mediation in structural equation modeling: Applications for early intervention research. *Journal of Early Intervention*, 29(3), 262–272. <https://doi.org/10.1177/105381510702900305>
- Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage publications.
- Leite, W. L., Shen, Z., Marcoulides, K., Fisk, C. L., & Harring, J. R. (2021). Using ant colony optimization for sensitivity analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(1), 47–56. <https://doi.org/10.1080/10705511.2021.1881786>
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol.*, 58(1), 593–614.
- Pearl, J., et al. (2000). Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2), 3.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association*, 100(469), 322–331.
- Sim, M., Kim, S.-Y., & Suh, Y. (2022). Sample size requirements for simple and complex mediation models. *Educational and Psychological Measurement*, 82(1), 76–106. <https://doi.org/10.1177/00131644211003261>
- Sun, R., Zhou, X., & Song, X. (2021). Bayesian causal mediation analysis with latent mediators and survival outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 778–790. <https://doi.org/10.1080/10705511.2020.1863154>
- 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.
- 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>
- 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.
- Wang, W., Xu, J., Schwartz, J., Baccarelli, A., & Liu, Z. (2021). Causal mediation analysis with latent subgroups. *Statistics in Medicine*, 40(25), 5628–5641. <https://doi.org/10.1002/sim.9144>
- 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.

A Additional Assumptions

Assumption S2 - Model Specification: The linear measurement model in Assumption Equation (5) holds, and the structural dependence among factor scores is correctly specified by a polynomial model that is linear in its parameters. The structural model in Equation (8) is a special case of this specification. In addition, the mediator model in Equation (9) is assumed to be correctly specified.

Assumption S3 - IID Sampling of Latent Variables and Observed Data: The observations $\{(r_i, \eta_{m_i}, \eta_{x_i}, \eta_{y_i})\}_{i=1}^n$ are iid draws from the joint distribution P .

Assumption S4 - Finite Moments: The factor scores $\eta_{\text{pred},i}$ satisfy finite moment conditions of sufficiently high order:

$$\mathbb{E}[|\eta_{\text{pred},i}|^{4J-2}] < \infty$$

Assumption S5 - Measurement Error: The measurement errors ϵ_i are iid, independent of the latent variables η_i , have zero mean, and satisfy

$$\mathbb{E}[|\epsilon_i|^{4J}] < \infty.^2$$

Assumption S6 - Structural Equation Error: The structural equation errors ζ_i comprising ζ_{y_i} and ζ_{m_i} are iid, independent of η_i , have zero mean, and finite variance:

$$\text{Var}[\zeta_i] < \infty.$$

B Proof of Theorem 1

Proof. Under standard M-estimation conditions—including existence and uniqueness of the population root, continuity and differentiability of the estimating function, a non-singular Jacobian, the uniform law of large numbers, and finite variance to satisfy a central limit theorem—we can linearize the sample estimating equation around Υ using a first-order Taylor expansion:

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

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

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

Defining

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

we obtain the desired expansion. By definition, $\mathbb{E}[\psi(\mathbf{z}_i, \Upsilon)] = 0$, so $\mathbb{E}[\Delta_i] = 0$. Since $\psi(\mathbf{z}_i, \Upsilon)$ is polynomial in $\eta_{\text{pred},i}$ and linear in (ϵ_i, ζ_i) , Assumptions S4–S6 ensure $\text{Var}(\Delta_i) < \infty$. Hence,

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

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

C Proof of Theorem 2

Proof. From the definition of the latent factor score estimator in Equation (11), we can express

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

with

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

Expanding this difference and regrouping by parameters yields

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

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

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 2. By Theorem 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. □

D Proof of Theorem 3

A Consistency

Proof. Under Assumptions S3 and S4, we define the population moments

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

The g -estimation equation,

$$\mathbb{E}[\mathbf{w}_i \zeta_{y_i}] = \mathbf{0}, \text{ with } \zeta_{y_i} = \eta_{y_i} - \boldsymbol{\xi}_{y_i}^\top \boldsymbol{\theta},$$

implies

$$\mathbb{E}[\mathbf{w}_i \eta_{y_i}] = \mathbb{E}[\mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top] \boldsymbol{\theta} \Leftrightarrow \mathbf{h} = \mathbf{G} \boldsymbol{\theta}.$$

By Assumption S1, $\mathbf{G}$ is invertible, so the population parameter is

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

From (25) and Theorems 1–2, together with the Law of Large Numbers, the empirical moments satisfy

$$\begin{pmatrix} \hat{M} \\ \hat{m} \end{pmatrix} = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} M(\hat{\boldsymbol{\eta}}_{\text{pred},i}, \hat{\boldsymbol{\Upsilon}}_2) \\ m(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2) \end{pmatrix} \xrightarrow{p} \mathbb{E} \begin{pmatrix} M(\tilde{\boldsymbol{\eta}}_{\text{pred}}, \boldsymbol{\Upsilon}_2) \\ m(\tilde{\boldsymbol{\eta}}, \boldsymbol{\Upsilon}_2) \end{pmatrix}.$$

Using Equation (26) and applying the law of iterated expectations, we have

$$\mathbb{E} \begin{pmatrix} M(\tilde{\boldsymbol{\eta}}_{\text{pred}}, \boldsymbol{\Upsilon}_2) \\ m(\tilde{\boldsymbol{\eta}}, \boldsymbol{\Upsilon}_2) \end{pmatrix} = \mathbb{E} \begin{pmatrix} \mathbb{E}[M(\tilde{\boldsymbol{\eta}}_{\text{pred},i}, \boldsymbol{\Upsilon}_2) \mid \boldsymbol{\eta}_i] \\ \mathbb{E}[m(\boldsymbol{\eta}_i, \boldsymbol{\Upsilon}_2) \mid \boldsymbol{\eta}_i] \end{pmatrix} = \mathbb{E} \begin{pmatrix} \mathbf{w}_i \boldsymbol{\xi}_{y_i}^\top \\ \mathbf{w}_i \eta_{y_i} \end{pmatrix} = \begin{pmatrix} \mathbf{G} \\ \mathbf{h} \end{pmatrix}.$$

Finally, the estimator defined in (24) can be shown to be consistent: As $n \rightarrow \infty$,

$$\hat{\boldsymbol{\theta}} = \hat{M}^{-1} \hat{m} \xrightarrow{p} \mathbf{G}^{-1} \mathbf{h} = \boldsymbol{\theta}. \quad \text{□}$$

B Asymptotic normality

Proof. To show asymptotic normality, we rewrite

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{N}(\hat{M}^{-1} \hat{m} - \boldsymbol{\theta}) = \hat{M}^{-1} \sqrt{N}(\hat{m} - \hat{M} \boldsymbol{\theta}).$$

With the definition

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

the second term becomes

$$\hat{m} - \hat{M} \boldsymbol{\theta} = \frac{1}{N} \sum_i \mathbf{l}(\hat{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\theta}).$$

A first-order Taylor expansion of the moment function $\mathbf{l}$ around $(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\theta}_0)$ yields

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

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

$$\mathbf{l}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\Upsilon}}_2, \boldsymbol{\theta}_0) = \mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\theta}_0) + \overline{\mathbf{C}} \boldsymbol{\Delta}_i + o_p(n^{-1/2}),$$

where

$$\bar{\mathbf{C}} = \mathbb{E} \left[\begin{pmatrix} \frac{\partial \mathbf{l}}{\partial \boldsymbol{\eta}_i^\top} \mathbf{B}_i & \frac{\partial \mathbf{l}}{\partial \boldsymbol{\Upsilon}_2^\top} \end{pmatrix} \right].$$

By the central limit theorem for i.i.d. variables (Assumption S3) and the finite-moment conditions (Assumptions S4, S6), we obtain

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

with

$$\mathbf{S} = \text{Var} [\mathbf{l}(\tilde{\boldsymbol{\eta}}_i, \boldsymbol{\Upsilon}_2, \boldsymbol{\theta}_0) + \bar{\mathbf{C}} \boldsymbol{\Delta}_i].$$

Finally, using Assumption S1 and applying the delta method gives

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

□

C Global Identifiability

Proof. Under Assumptions C1–C5 and S1–S2, the instruments are exogenous and relevant, guaranteeing that the structural parameters θ_r and θ_m can be uniquely recovered despite any unmeasured confounding (Assumption C5 ensures constant effects). Combined with the full-rank condition and correct model specification (Assumptions S1–S2), this implies a one-to-one mapping from $\boldsymbol{\theta}$ to the population moments defined by the instruments. Hence, the parameters are globally identifiable and correspond to the causal effects in Equations (1) and (2). □