Biometrics 000, 000–000
000 0000

DOI: 000

arXiv:2611.18180v1 [stat.ML] 13 Nov 2026

LaMe: a Joint Latent Mediation Framework for High-Dimensional Systems

Gaël Poux-Médard[1,3,*], Kaniav Kamary[2], Yohann De Castro[3,4], and Vivian Viallon[1]

[1]Nutrition and Metabolism Branch, International Agency for Research on Cancer, WHO, Lyon, France
[2]Institut Camille Jordan, INSA Lyon, Centrale Lyon, Université Lyon 1, CNRS, Villeurbanne, France
[3]Institut Camille Jordan, École Centrale Lyon, CNRS UMR 5208, Lyon, France
[4]Institut Universitaire de France
*email: pouxg@iarc.who.int

Summary: Mediation analysis seeks to understand how an exposure influences an outcome through intermediate variables. However, as increasingly rich datasets become available, hundreds to thousands of potential mediators make these systems difficult to model and hard to interpret. Latent-variable approaches offer a natural way to accommodate these high-dimensional settings. However, existing latent mediation methods face limitations, including overparameterization, biased inference, or limited scalability. We develop LaMe, a joint linear-Gaussian framework that summarizes a high-dimensional exposure-mediator-outcome system through a small number of latent variables. We identify natural direct and indirect effects under an explicit assumption that the manifest measurements fully capture the latent mediators, and characterize conditions under which the latent representation effectively restricts the observed cross-covariance structure. Simulations assess recovery of observed association parameters, natural effects, and manifest features associated with latent mediation pathways across a range of challenging settings. Four real-world applications demonstrate competitive predictive performance with compact latent representations, including single-dimensional representations of biomarker and proteomic measurements. Together, these results establish LaMe as a practical framework for estimating and summarizing multivariate mediation systems under explicit structural and measurement assumptions.

Key words: joint modeling; latent mediators; mediation analysis; probabilistic inference

1. Introduction

Deciphering how complex exposures exert their effects through intermediary processes is a cornerstone of modern epidemiology. In the NHANES data considered here, for example, lifestyle characteristics may directly influence glycohemoglobin levels, but might also affect blood biomarkers which in turn cause these variations. However, the continued emergence of large-scale, high-dimensional datasets poses new challenges to established mediation analysis methods. Analyses conducted on biological data are particularly affected, as molecular profiling can yield hundreds or thousands of correlated molecular measurements, only a subset of which may be relevant to the mediating process linking an exposure to an outcome. Latent variable models have been introduced to extend classical mediation models to these settings (Albert, Geng, and Nelson, 2016; Chén, Crainiceanu, Ogburn, Caffo, Wager, and Lindquist, 2017; Derkach, Pfeiffer, Chen, and Sampson, 2019; Nath, Caffo, Wager, and Lindquist, 2023). This approach brings a number of improvements over non-latent mediation techniques (Zeng, Shao, and Zhou, 2021; Zhu, He, Ma, Hu, Li, and Li, 2024) (e.g. based on OLS, LASSO or ElasticNet): noise decomposition, information compression, interpretability and robustness of the estimates.

However, existing latent mediation approaches face several limitations. Stepwise methods estimate latent scores separately from subsets of the observed data and use the resulting scores in subsequent regressions on remaining observed variables, potentially inferring latent variables that are irrelevant to mediation systems (Chén et al., 2017; Nath et al., 2023). Joint approaches avoid this issue by estimating latent variables along with their relation to observed variables, but may be overparameterized, prone to overfitting, limited in scalability, and typically assume univariate exposures, univariate outcomes, and the conditional independence among manifest variables given the latent mediators whereas biological measurement are generally correlated (Albert et al., 2016; Derkach et al., 2019). To address these limitations, we introduce the **Latent Mediation (LaMe)** framework, which jointly models multivariate exposures, outcomes and high-dimensional manifest variables through a small number of latent mediators without assuming diagonality of noise covariances, while explicitly accounting for direct exposure-outcome effects.

This approach makes three main methodological contributions. First, LaMe is novel in that it infers latent mediators from a joint probabilistic model with multivariate exposures and outcomes, high-dimensional manifest variables, and non-diagonal residual covariance matrices. As a result, it avoids biases associated with estimating latent scores separately from the rest of the system as in stepwise procedures and estimates all its parameters based on all available information rather

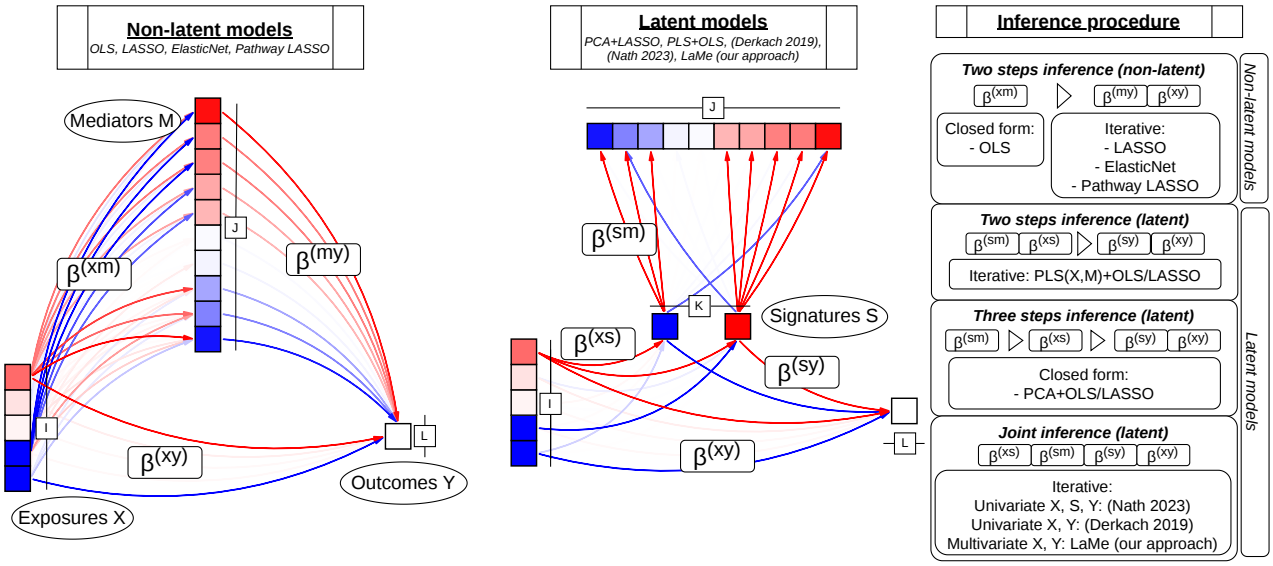


Figure 1. Non-latent (left-panel) and latent (middle-panel) mediation systems, and overview of several associated inference procedures (right-panel) — (left and middle panels) Variables X, M, S and Y are specified in the ovals, structural coefficients matrices β are shown in rounded rectangles, and the system's dimensions are given in white squares. Arrows represent causal relations between variables. For variables with a causal path to the outcome, squares' color represents the total causal effect of a unit intervention on the corresponding variable on the outcome Y under the model specified in Eq. (2)–in this case, the sum of all causal paths leading to Y; in the latent graph, the colors of the manifest-variable squares represent their conditional associations with Y, not causal effects. For both squares and arrows, blue represents a negative relation, red represents a positive relation, and color intensity represents its strength. (right-panel) Common optimization approaches to mediation systems and the order in which they infer parameters. Alt text: three-panel schematic comparing non-latent and latent mediation systems and their inference strategies. Exposures influence outcomes directly and through latent mediators; red and blue arrows indicate positive and negative relationships. The right panel groups methods by two-step, three-step, or fully joint estimation.

than subsets of observed variables. We characterize when the representation imposes a genuine restriction on cross-block relationships, distinguishing dimension reduction from a reparameterization of unrestricted regression. Second, LaMe connects joint probabilistic estimation methods (Albert et al., 2016; Derkach et al., 2019) with representations in which latent mediators are exactly recoverable from the manifest measurements (Chén et al., 2017; Nath et al., 2023; Zhao, Lindquist, and Caffo, 2020). Under this explicit restriction and suitable rank conditions, the aggregate natural direct and indirect effects are statistically identified from the observed data. Throughout this article, identifiability refers to statistical identification from the observed-data distribution, rather than causal identification of natural effects. We also introduce a score to assess which manifest variables are associated with exposures and informative about outcomes through the latent mediators, independently of how the latent variables are represented. Third, we derive closed-form EM updates for two equivalent likelihood factorizations, allowing computational efficiency to be improved when manifest measurements and exposures have substantially different dimensions. Simulations assess parameter recovery, effect recovery under the exact recoverability assumption, and feature ranking; real-world applications assess predictive performance and the compactness of the learned representations. These results position LaMe as a useful methodological option.

2. Background

Let $N > 1$ be the number of observations and, for each observation, let $I \geqslant 1$ be the number of exposures, $J \geqslant 1$ the number of manifest variables, and $L \geqslant 1$ the number of outcomes. Then, let $X \in \mathbb{R}^{N \times I}$ denote the exposure matrix, $M \in \mathbb{R}^{N \times J}$ the manifest variables matrix and $Y \in \mathbb{R}^{N \times L}$ the outcome matrix.

2.1 *Non-latent models*

In their simplest form, linear mediation models are specified by the following set of structural causal equations (see Fig. 1-left):

$$\begin{cases} X = \varepsilon_x \\ M = X\beta^{(xm)} + \varepsilon_m \\ Y = X\beta^{(xy)} + M\beta^{(my)} + \varepsilon_y \end{cases} \tag{1}$$

The random vectors $\varepsilon_{x,n}$, $\varepsilon_{m,n}$ and $\varepsilon_{y,n}$ corresponding to the n^{th} observation are usually assumed to be mean-zero Gaussians and mutually independent. Formally, for each observation n, $\varepsilon_{x,n} \sim \mathcal{N}(0, \Sigma_x)$, $\varepsilon_{m,n} \sim \mathcal{N}(0, \Sigma_{m|x})$ and $\varepsilon_{y,n} \sim \mathcal{N}(0, \Sigma_{y|x,m})$; these terms account for noise and/or exogenous variables. Coefficient matrices are $\beta^{(xm)} \in \mathbb{R}^{I\times J}, \beta^{(xy)} \in \mathbb{R}^{I\times L}, \beta^{(my)} \in \mathbb{R}^{J\times L}$ and noise covariances are $\Sigma_{m|x} \in \mathbb{R}^{J\times J}, \Sigma_{y|x,m} \in \mathbb{R}^{L\times L}$. We consider that all variables are centered. As shown in Web Appendix 6, the inclusion of confounders C between X, M and Y is straightforward and is therefore omitted here for clarity.

According to recent surveys (Rijnhart, Lamp, Valente, MacKinnon, Twisk, and Heymans, 2021; Zeng et al., 2021; Zhu et al., 2024), most high-dimensional mediation analyses published in the literature rely on non-latent linear mediation models, under which the parameters of Eq. (1) are estimated using sequential regression procedures combined with penalization. Classical mediation analysis is based on ordinary least squares (OLS) estimation of $\beta^{(xm)}$, $\beta^{(my)}$ and $\beta^{(xy)}$, while extensions to high-dimensional settings incorporate regularization to select the relevant mediators or improve estimation or prediction accuracy. Typical examples include the LASSO, Ridge and ElasticNet regressions. Another example is the Pathway LASSO model (Zhao and Luo, 2022), which introduces a nonlinear penalty targeting entire indirect paths rather than individual coefficients. These approaches allow for straightforward coefficient estimation, quickly converge to global optima, and provide readily interpretable results. On the other hand, they heavily rely on hyperparameter tuning and may yield unstable and selection-inconsistent estimates in the presence of strongly correlated mediators or high-dimensional regimes, where mediator selection can become effectively arbitrary (Wang, Lengerich, Aragam, and Xing, 2018; Ye, Zhu, and Coffman, 2021).

2.2 *Latent models*

2.2.1 *Latent mediators.*

Structural equations. To overcome some of these limitations, latent models have been suggested, which assume the existence of unobserved variables, denoted by $S \in \mathbb{R}^{N\times K}$, where K is the number of latent dimensions. Following the latent mediation formulations of (Albert et al., 2016; Derkach et al., 2019), LaMe is specified by the following structural equations (see Fig. 1-middle):

$$\begin{cases} X = \varepsilon_x \\ S = X\beta^{(xs)} + \varepsilon_s \\ M = S\beta^{(sm)} + \varepsilon_m \\ Y = X\beta^{(xy)} + S\beta^{(sy)} + \varepsilon_y \end{cases} \tag{2}$$

The coefficient matrices of the model are $\beta^{(xs)} \in \mathbb{R}^{I\times K}$, $\beta^{(sm)} \in \mathbb{R}^{K\times J}$, $\beta^{(xy)} \in \mathbb{R}^{I\times L}$, $\beta^{(sy)} \in \mathbb{R}^{K\times L}$. The noise vectors $\varepsilon_{x,n}, \varepsilon_{s,n}, \varepsilon_{m,n}$ and $\varepsilon_{y,n}$ corresponding to the n^{th} observation are assumed to be mean-zero Gaussians and mutually independent. Formally, for each observation n, $\varepsilon_{x,n} \sim \mathcal{N}(0, \Sigma_x)$, $\varepsilon_{s,n} \sim \mathcal{N}(0, \Sigma_{s|x})$, $\varepsilon_{m,n} \sim \mathcal{N}(0, \Sigma_{m|s})$ and $\varepsilon_{y,n} \sim \mathcal{N}(0, \Sigma_{y|x,s})$, for some noise covariance matrices $\Sigma_x \in \mathbb{R}^{I\times I}$, $\Sigma_{m|s} \in \mathbb{R}^{J\times J}$, $\Sigma_{s|x} \in \mathbb{R}^{K\times K}$, $\Sigma_{y|x,s} \in \mathbb{R}^{L\times L}$.

To simplify the exposition, we omit observed confounders (say C) from the subsequent derivations. As shown in Web Appendix 6, the inclusion of confounders between X, S and Y is straightforward and requires only replacing the exposure variable with a composite variable $\xi := [X, C]$, and appropriately partitioning the estimated coefficients corresponding to X and C.

Natural effects. Under sequential ignorability, namely $\{Y(x,s), S(x')\} \perp X|C = c$ and $Y(x,s) \perp S(x')|X = x', C = c$ where $Y(x,s)$ and $S(x')$ denote the counterfactual outcome and mediator under the corresponding interventions, the natural direct and indirect effects associated with a one-unit increase in the exposures are given by (Albert et al., 2016; Derkach et al., 2019):

$$NDE = \beta^{(xy)} \qquad ; \qquad NIE = \beta^{(xs)}\beta^{(sy)}, \tag{3}$$

2.2.2 *Stepwise approaches.*

Stepwise approaches to latent mediation generally define S as a deterministic linear combination of manifest variables. These methods typically derive latent scores from an initial model, for instance from Principal Component Analysis (PCA), Partial Least Squares (PLS) or their variants; then, these scores enter as observed mediators in the subsequent estimation of a non-latent mediation model (Cai, Wang, and Zhu, 2023; Hu, DeSimone, and Wang, 2024; Kawano, Fujisawa, Takada, and Shiroishi, 2018; Zhao et al., 2020).

However, most existing stepwise approaches do not use all available data to derive their latent scores. For instance, approaches based on PCA optimize S to explain the variance of M, and approaches based on PLS optimize S to explain the covariance between X and M, or M and Y, regardless of whether the resulting scores best characterize mediation between X

and Y. Conversely, a correctly specified model accounting for the joint variance and covariance of (X, M, Y) does not require latent scores to be selected using only a subset of the observed relationships.

2.2.3 *Joint modeling.* A joint approach that maximizes a likelihood corresponding to the full structural model may therefore be preferable.

The latent approaches of (Chén et al., 2017; Nath et al., 2023) both consider an univariate exposure and outcome, and both estimate the first latent direction and the associated path coefficients from a joint objective based on all observed data. In (Chén et al., 2017), the model extracts subsequent directions through approximate deflation while keeping earlier directions fixed. Consequently, the retained directions generally differ from those estimated by maximizing the joint likelihood of a full multivariate latent model; in particular, the sequential objective does not include cross-score residual covariances. Finally, the authors remark that inference may become cumbersome when $J \gg 1$ without resorting to additional data augmentation techniques. The model introduced in (Nath et al., 2023) extends this approach beyond linear relationships to model S as an arbitrary deterministic function of M. However, this extension is limited to a single latent variable. Moreover, score selection and the estimated direct–indirect decomposition are sensitive to the scale of Y, as shown in Web Appendix 9. Finally, standardization of the univariate score is imposed algorithmically rather than incorporated into the objective function, complicating the analytical characterization of the resulting estimator.

Joint probabilistic formulations, on the other hand, explicitly account for invariance to latent rescaling. In (Albert et al., 2016), the authors formulate a joint probabilistic latent model with a single latent mediator. The approach was later extended by (Derkach et al., 2019) to handle multivariate latent mediators. In that work, the authors assume a univariate exposure and outcome, and constrain the conditional noise covariance of manifest variables given latent mediators $\Sigma_{m|s}$ to be diagonal. In this setting, however, we show in Web Appendix 10 that the non-regularized and correctly specified model from (Derkach et al., 2019) is equivalent to Factor Analysis of M followed by linear regressions on the resulting S when $K > 1$. Additionally, the diagonal constraint on $\Sigma_{m|s}$ imposes conditional independence of manifest variables given the latent mediators, which is not necessarily true when M represents biomarkers. Finally, Condition 1 in (Derkach et al., 2019) is required for identification of the effects, but the proposed model does not guarantee that this condition holds.

3. Methods

We develop LaMe to extend existing approaches by jointly modeling all observed variables, allowing non-diagonal noise covariance matrices and multivariate exposures, outcomes and latent information bottleneck. Under the exact recoverability assumption for S, the framework also permits identification of natural effects.

We first detail the exact recoverability and information bottleneck assumptions made by our approach, then we develop a general inference procedure for an auxiliary model that does not assume exact recoverability, and we finally show in Proposition 1 that exact recoverability can be recovered through a likelihood-equivalent reparametrization of the model post-estimation. This proposition further establishes that the parametric representations of natural direct and indirect effects are statistically identified under this condition.

3.1 *Assumptions*

3.1.1 *Exact recoverability.* Following score-based mediation approaches, which define latent mediators as linear combinations of manifest variables (Chén et al., 2017; Nath et al., 2023; Zhao et al., 2020), we assume the same kind of deterministic relationship holds within our probabilistic framework. Specifically, we will work under what we call the exact recoverability assumption, under which $S = \mathbb{E}[S|M] := M\beta^{(ms)}$. We emphasize that this equality does not imply that the measurements M are causes of S; they remain the manifest variables causally influenced by S, with S causally influenced by X. This requirement is equivalent to:

$$\begin{aligned} \Sigma_{m|s}\beta^{(ms)} &= 0 \\ \text{where } \beta^{(ms)} &= \underbrace{((\beta^{(sm)})^T\Sigma_s\beta^{(sm)} + \Sigma_{m|s})^{-1}}_{\Sigma_m^{-1}}(\beta^{(sm)})^T \underbrace{(\Sigma_{s|x} + (\beta^{(xs)})^T\Sigma_x\beta^{(xs)})}_{\Sigma_s} \end{aligned} \tag{4}$$

We point out that the exact-recoverability approximation becomes increasingly accurate as the number of manifest variables J grows, making it well suited to high-dimensional settings. To see this, consider the difference between the unconstrained latent variable S in Eq. 2 and its exactly recoverable counterpart $S^* = M\beta^{(ms)}$. Assuming that additional manifest variables provide non-vanishing, non-redundant information about S, the measurement precision $\beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T$ grows linearly with J. Then, we show in Web Appendix 7 that $\mathbb{E}[||S^* - S||^2] = \mathcal{O}(J^{-1})$, meaning that $S \overset{J\to\infty}{\to} S^*$. Thus, as the number of informative manifest variables increases, the latent variable S of the unconstrained model in Eq. 2 asymptotically converges to the latent mediator that is exactly recoverable from M. Although exact recoverability is an idealization for finite J, it therefore becomes a reasonable approximation when sufficiently many informative manifest variables are available, supporting its use in high-dimensional settings.

We show in Proposition 1 below that the joint model specified in Eq (2) can always be formulated to satisfy exact recoverability.

3.1.2 *Information bottleneck.* Introducing low-dimensional latent variables addresses several limitations of non-latent least-squares mediation models. When the number of observations is finite, sampling noise can introduce spurious association patterns; by imposing a low-rank structure on the cross-covariance between X and M, and between M and Y, latent models alleviate this by forcing the model to focus on the most relevant covariation patterns, typically preventing overfitting the noise and improving robustness and generalization of structural parameters $\beta^{(xs)}, \beta^{(sm)}, \beta^{(xy)}, \beta^{(sy)}$. In this sense, the latent representation acts as an information bottleneck between observed variables (Chechik, Globerson, Tishby, and Weiss, 2005). Likewise, information bottlenecks can suppress weaker associations between observed variables, allowing the system to be approximated by a simplified representation that only captures its most defining patterns of covariation, which is a recurring objective of dimension-reduction approaches to mediation analysis.

Under the assumptions from Eq. (2) (linear relations, full noise covariances and mutually independent noises), a necessary and sufficient condition for an information bottleneck can be formulated as (see Web Appendix 8 for details):

$$K < \text{rank}\begin{pmatrix}\text{Cov}(X,M)\\ \text{Cov}(Y,M|X)\end{pmatrix} \qquad \text{where rank}\begin{pmatrix}\text{Cov}(X,M)\\ \text{Cov}(Y,M|X)\end{pmatrix} \leqslant \min(I+L, J), \tag{5}$$

where $\text{Cov}(\cdot,\cdot)$ denotes an observed covariance. Conversely, $K \geqslant \min(J, I+L)$ is a sufficient condition ensuring that every possible cross-covariance matrix can be represented exactly. Then, the observable regression coefficients between X, M and Y estimated from the latent model in Eq. (2) are a reparametrization of those estimated by a non-latent OLS model, as shown in Web Appendix 8, limiting the interest of the approach. Note that assuming a diagonal conditional noise covariance $\Sigma_{m|s}$ does not affect the condition above. In practice, the latent dimension K can be selected via cross-validation or information criteria such as AIC or BIC to balance model complexity and goodness-of-fit.

3.2 *LaMe model*

3.2.1 *Likelihood.* Starting from Eq. (2), $\theta = \{\beta^{(xs)}, \beta^{(sm)}, \beta^{(xy)}, \beta^{(sy)}, \Sigma_x, \Sigma_{s|x}, \Sigma_{m|s}, \Sigma_{y|x,s}\}$ denote the full set of model parameters, and $p(\theta)$ its prior distribution. Assuming that mean-zero Gaussian noises are mutually independent in Eq. (2), a natural factorization of the joint distribution is

$$p(X,S,M,Y,\theta) = p(X|\theta)p(S|X,\theta)p(M|S,\theta)p(Y|X,S,\theta)p(\theta) \tag{6}$$

Recall that all the distributions above are Gaussian, with parameters specified as, for example, $p(S|X,\theta) = \mathcal{N}(S; X\beta^{(xs)}, \Sigma_{s|x})$. In the following derivations, we assume a flat prior $p(\theta)$ on all parameters, so that $p(X,S,M,Y,\theta) \propto p(X,S,M,Y|\theta)$. Other priors that regularize the parameters, together with an extension to log-normal variables, are discussed in Web Appendix 13.

The latent nature of S prevents direct optimization of (6). We can, however, compute the incomplete likelihood by marginalizing the complete likelihood over all possible values of S. Using the complete likelihood expression (6), a standard Gaussian integration formula and grouping terms together, we find

$$p(X,M,Y|\theta) = \int_S \underbrace{e^{-\frac{1}{2}\langle S^TS,A\rangle + \langle S,B\rangle + C}}_{p(S,X,M,Y|\theta)} dS = e^{\frac{1}{2}\langle B^TB, A^{-1}\rangle} e^C (2\pi)^{\frac{NK}{2}} |A|^{-\frac{N}{2}} \tag{7}$$

$$\text{where } A := \Sigma_{s|x}^{-1} + \beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T + \beta^{(sy)}\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T$$

$$\text{and } B := X\beta^{(xs)}\Sigma_{s|x}^{-1} + M\Sigma_{m|s}^{-1}(\beta^{(sm)})^T + (Y - X\beta^{(xy)})\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T.$$

Above, $|A|$ denotes the determinant of A and $\langle A,B\rangle := \sum_{i,j} A_{ij}B_{ij}$ is the Frobenius product. The expression of the normalization term C is provided in Web Appendix 12.

The expectation-maximization algorithm then first computes the expectation of latent variables given the current parameter estimates $\theta^{(t)}$ (E-step), and then derives the new optimal parameter set $\theta^{(t+1)}$ given the expectation of latent variables computed in the previous step (M-step). This procedure is iterated until convergence.

3.2.2 *E-step.* Given the parameter estimates at the tth iteration, $\theta^{(t)}$, the E-step computes the conditional distribution of the latent variables S and evaluates expectations of functions of S with respect to this distribution.

We show in Web Appendix 11 that, for any candidate distribution $q_{\theta^{(t)}}(S)$, the observed log-likelihood is lower-bounded as:

$$\log p(X,M,Y|\theta) \geqslant \int_S q_{\theta^{(t)}}(S) \log \frac{p(S,X,M,Y|\theta)}{q_{\theta^{(t)}}(S)} dS \tag{8}$$

Then, it can be shown that this lower-bound is equal to $\log p(X,M,Y|\theta^{(t)})$, and therefore maximal, when:

$$q_{\theta^{(t)}}(S) = p(S|X,M,Y,\theta^{(t)}) = \frac{p(S,X,M,Y|\theta^{(t)})}{p(X,M,Y|\theta^{(t)})} \tag{9}$$

A full derivation of this result can be found in Web Appendix 11.

We finally use $q_{\theta^{(t)}}(S)$ to compute the moments of S under the current parameter estimates before proceeding to the M-step, recalling that N is the number of observations:

$$\overline{S}^{(t)} = BA^{-1} \quad ; \quad \overline{S^TS}^{(t)} = NA^{-1} + A^{-1}B^TBA^{-1} \tag{10}$$

3.2.3 *M-step.* The M-step then computes the parameter set $\theta^{(t+1)}$ that maximizes the lower-bound in Eq. (8) with $q_{\theta^{(t)}}(S)$ from Eq. (9). Taking $\beta^{(sm)}$ as an illustration:

$$\partial_{\beta^{(sm)}} \int_S q_{\theta^{(t)}}(S) \log p(S, X, M, Y|\theta) dS = -\overline{S^T S}^{(t)} \beta^{(sm)} \Sigma_{m|s}^{-1} + (\overline{S}^{(t)})^T M \Sigma_{m|s}^{-1} = 0 \tag{11}$$
$$\Leftrightarrow \beta^{(sm)} = (\overline{S^T S}^{(t)})^{-1} (\overline{S}^{(t)})^T M$$

where $\overline{S}^{(t)}$ and $\overline{S^T S}^{(t)}$ are obtained from Eq. (10). Detailed derivations for other parameters are provided in Web Appendix 12.

The EM algorithm alternates between the E-step, where $\overline{S}^{(t)}$ and $\overline{S^T S}^{(t)}$ are computed given the current parameter estimates $\theta^{(t)}$, and the M-step, where the parameter estimates are updated to $\theta^{(t+1)}$ given $\overline{S}^{(t)}$ and $\overline{S^T S}^{(t)}$, until convergence of the incomplete likelihood. The EM algorithm guarantees a non-decreasing likelihood and convergence to a stationary point (Neal and Hinton, 1998). Note that although we consider all parameters θ to be random variables, the EM optimization algorithm yields point estimates rather than their distribution.

At this point, S may not be exactly recoverable from M as assumed in Eq. (4). In the next section, we show how to satisfy this assumption through a likelihood-equivalent reparametrization of the EM solution.

3.3 *Identifiability*

The parametric representation of the observed moments recalled in Web Appendix 8 shows that the observed likelihood in Eq. (7) is invariant to some transformations of the latent space; the exact recoverability property limits these to invertible linear transformations, which makes natural effects identifiable.

PROPOSITION 1 (Statistical identification of the natural effects): The natural direct and indirect effect through S as defined in Eq. (3), namely $\beta^{(xy)}$ and $\beta^{(xs)}\beta^{(sy)}$, are uniquely identified from the joint distribution of (X, M, Y) under the exact recoverability assumption $S = \mathbb{E}[S|M]$ when the bottleneck condition from Eq. (5) is satisfied.

Proof. [(Proposition 1)]

Likelihood-equivalent reparametrizations. Let $H_M \in \mathbb{R}^{K\times K}$ and $H_Y \in \mathbb{R}^{K\times K}$ denote invertible matrices. Then, we define the elements of a set of transformed parameters denoted by θ^* expressed in terms of θ, H_M and H_Y as:

$$\begin{aligned}
\beta_*^{(xs)} &= \beta^{(xs)} H_M \quad ; \quad \Sigma_{s|x}^* = H_M^T \Sigma_{s|x} H_Y \\
\beta_*^{(sm)} &= H_M^{-1} \beta^{(sm)} \quad ; \quad \Sigma_{m|s}^* = \Sigma_{m|s} + (\beta^{(sm)})^T \Sigma_{s|x} \beta^{(sm)} - (\beta_*^{(sm)})^T \Sigma_{s|x}^* \beta_*^{(sm)} \\
\beta_*^{(sy)} &= H_Y^{-1} \beta^{(sy)} \quad ; \quad \Sigma_{y|x,s}^* = \Sigma_{y|x,s} + (\beta^{(sy)})^T \Sigma_{s|x} \beta^{(sy)} - (\beta_*^{(sy)})^T \Sigma_{s|x}^* \beta_*^{(sy)} \\
\beta_*^{(xy)} &= \beta^{(xy)} + \beta^{(xs)}\beta^{(sy)} - \beta_*^{(xs)}\beta_*^{(sy)} \quad ; \quad \Sigma_x^* = \Sigma_x
\end{aligned} \tag{12}$$

Finally, in the absence of additional assumptions, the relation $P(X, M, Y|\theta) = P(X, M, Y|\theta^*)$ holds. We demonstrate in Web Appendix 14 that no other likelihood-equivalent reparametrizations exist when the bottleneck condition from Eq. (5) is satisfied.

Parametrizations with exact recoverability. From the definition in Eq. (4) and noting that $\Sigma_m^* = \Sigma_m$ (see Web Appendix 15), we derive an expression of $H_M H_Y^{-1}$ such that they define a reparametrization that satisfies the exact recoverability assumption, meaning $\Sigma_{m|s}^* \Sigma_m^{-1} (\beta_*^{(sm)})^T \Sigma_s^* = 0$, as:

$$H_M H_Y^{-1} = \left((\beta^{(sm)} \Sigma_m^{-1} (\beta^{(sm)})^T)^{-1} - (\beta^{(xs)})^T \Sigma_x \beta^{(xs)} \right)^{-1} \Sigma_{s|x}$$

Note that the relation above is defined as long as $\beta^{(sm)} \Sigma_m^{-1} (\beta^{(sm)})^T$ is non-singular, which is necessarily true when the bottleneck condition from Eq. (5) is satisfied. We demonstrate in Web Appendix 15 that the expression above also guarantees the positive semi-definiteness of all transformed noise covariances $\Sigma_{s|x}^*$, $\Sigma_{m|s}^*$ and $\Sigma_{y|x,s}^*$. Therefore, any choice of an invertible H_M, along with the corresponding H_Y from the equation above, guarantees the assumption from Eq. (4) is satisfied.

Invariance of natural effects under exact recoverability. Note that $\beta^{(sm)} \Sigma_m^{-1} (\beta^{(sm)})^T = \Sigma_s^{-1}$ (see Web Appendix 15) when the assumption in Eq. (4) is verified. Injecting it in the expression above yields $H_M H_Y^{-1} = I$. Identification of the natural effects coefficients under the exact recoverability assumption immediately follows:

$$\beta_*^{(xy)} = \beta^{(xy)} + \beta^{(xs)}(I - H_M H_Y^{-1})\beta^{(sy)} = \beta^{(xy)}; \quad \beta_*^{(xs)}\beta_*^{(sy)} = \beta^{(xs)} H_M H_Y^{-1} \beta^{(sy)} = \beta^{(xs)}\beta^{(sy)}$$

We point out that assumptions other than exact recoverability may also make the NDE and NIE identifiable, although they may not provide similar convergence guarantees when $J \to \infty$; for instance, alternative choices may allow S to be a noisy function of M rather than deterministic. This question lies outside the scope of the present article and is left for future work.

3.4 *Alternative factorization*

The causal model specified in Eq. (2) implies two conditional independence relationships, $M \perp X|S$ and $Y \perp M|X,S$, which allow the joint distribution of (X,S,M,Y) to be factorized equivalently in two ways, denoted **LaMe-vX** and **LaMe-vM** hereafter:

$$\begin{aligned} p(X,S,M,Y|\theta) &= p(X|\theta)p(S|X,\theta)p(M|S,\theta)p(Y|X,S,\theta) \quad \text{(LaMe-vX)} \\ &= p(M|\theta)p(S|M,\theta)p(X|S,\theta)p(Y|X,S,\theta) \quad \text{(LaMe-vM)} \end{aligned} \tag{13}$$

Both factorizations describe the same family of joint distributions and can therefore be used interchangeably for parameter estimation, with parameters from either representation recoverable from the other. Note that this reformulation is not necessarily equivalent in the presence of regularization, orthogonality, or structural constraints –for instance, a diagonal conditional noise covariance in (Derkach et al., 2019) prevents such equivalence. However, it holds for any reparametrization based on Eq. (12), including those assuming exact recoverability.

A first remark is that alternative factorizations may not correspond to any relevant causal system. For instance, the coefficient $\beta^{(sx)}$ in LaMe-vM has no causal interpretation in a mediation analysis setting; however, as shown in Web Appendix 16, it can be used to retrieve the relevant causal path coefficient $\beta^{(xs)}$. Second, structural parameters differ depending on the factorization. Third, optimizing the joint model using one factorization instead of the other can make the EM algorithm orders of magnitude faster and the estimation more accurate when $J \gg I$: using the LaMe-vM factorization, $\Sigma_{x|s}$ and $\Sigma_{s|m}$ must be updated ($\mathcal{O}(NI(K+I)+NK(K+J))$) at each iteration whereas Σ_m needs to be updated ($\mathcal{O}(NJ^2)$) only once. We numerically support this result in Web Appendix 19.

The full table of the update equations for both models is available in Web Appendix 12.

4. Results

For both the LaMe-vX and LaMe-vM factorizations (see Eq. (13)), we stop the EM loop once the relative change in the log-likelihood remains below 10^{-5} for ten consecutive iterations. This procedure is repeated 50 times with different random initializations, and the run with the highest likelihood is selected. We consider both unregularized models and models with an ℓ_1 penalty on the regression coefficients linking S and M.

4.1 *Baselines*

We evaluate LaMe against **OLS**, OLS with ℓ_1 regularizer (**LASSO**), OLS with ℓ_1 and ℓ_2 regularizers (**ElasticNet**), PCA on the manifest variables M followed by OLS or LASSO on the resulting (X,S,Y) latent mediation system (**PCA+OLS** and **PCA+LASSO**), PLS on $X \to M$ followed by OLS on $[X,S] \to Y$ (brackets denote column-wise concatenation) (**PLS(X,M)+OLS**), PLS on $X \to [M,Y]$ followed by OLS on $X \to Y_{res}$ (**PLS(X,M,Y)+OLS**), OLS on the direct path $X \to Y$ followed by PLS on the indirect path $X \to [M, Y_{res}]$ (**OLS+PLS(X,M,Y)**) (where Y_{res} are the residuals of Y that are not explained by the model used in the first step), an over-parametrized version of LaMe-vX (**OVP-LaMe**), and an adaptation of the latent mediation model described in (Nath et al., 2023) (Joint Alternate Optimization - Gaussian case **JAOG**). A discussion of our adaptation of (Nath et al., 2023) (originally designed for univariate latent mediators) and why we consider OVP-LaMe rather than the model of (Derkach et al., 2019) is provided in Web Appendix 18.

All experiments were run on an Intel Core Ultra 7 165U CPU; a scalability comparison is provided in Web Appendix 19.

4.2 *Synthetic data*

4.2.1 *Data generation.* Synthetic variables X, M, Y are generated according to the linear-Gaussian structural equations specified by Eq. (2) when data are assumed to stem from a latent model, and according to Eq. (1) when we assume the absence of a shared latent structure. The dimensions I, J, L and possibly K_{true} are specified for each experiment below. Noise covariances are rescaled so the following signal-to-noise ratios (SNR) $SNR(S|X), SNR(M|S), SNR(Y|X,S)$ (in the latent case) and $SNR(M|X), SNR(Y|X,M)$ (in the non-latent case) attain a given value (e.g. $SNR(M|S) = \frac{\text{Var}(S\beta^{(sm)})}{\text{Var}(\mathcal{N}(0,\Sigma_{m|s}))} = \frac{25}{100}$ means 20% of the variations in M are explained by the signal from S, and 80% are due to random noise; see Eq. (2) and Web Appendix 17).

4.2.2 *Recovering underlying parameters.* To compare LaMe against baselines that do not assume the existence of latent mediators (e.g. OLS and OVP-LaMe), we first evaluate the capacity of each model to recover the true marginal association between observed variables X and M ($\beta^{(xm)} := \nabla_X \mathbb{E}[M|X]$) and between M and Y ($\beta^{(my)} := \nabla_M \mathbb{E}[Y|M]$). Not only do these coefficients provide a common basis for comparing models, they are also involved in the computation of the natural effects under the exact recoverability assumption (i.e. $NIE = \beta^{(xs)} H_M H_Y^{-1} \beta^{(sy)} = \beta^{(xm)}\beta^{(my)}$ where $H_M H_Y^{-1}$ are derived from Proposition 1). A detailed assessment of LaMe's capacity to retrieve causal natural effects is developed in the next subsection. We compare the models' estimates with the corresponding parameters used for data generation, $\beta^{(xm)}_{true}$ and $\beta^{(my)}_{true}$, using the mean absolute error across all coefficients as the evaluation metric.

We present results under a challenging configuration with $I = 100$, $J = 100$, $L = 1$, signal-to-noise ratios $SNR(M|S) =$

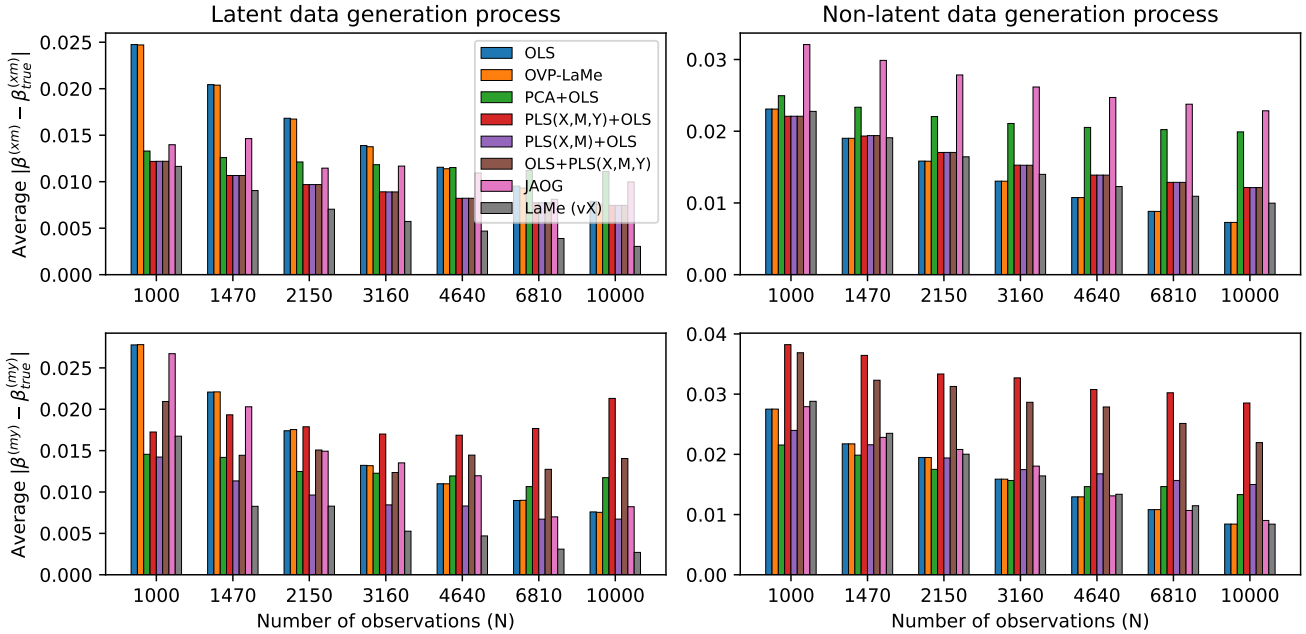


Figure 2. LaMe accurately retrieves underlying parameters — LaMe and baselines when $I = 100$, $J = 100$, $L = 1$, and a signal-to-noise ratio from S to M and Y equal to 0.1 for varying N. **(Top)** Mean absolute error of the estimated coefficients $\beta^{(xm)}$ relative to the true coefficients $\beta^{(xm)}_{true}$ | **(Bottom)** Mean absolute error of the estimated coefficients $\beta^{(my)}$ relative to the true coefficients $\beta^{(my)}_{true}$ | **(left)** Data generated using a latent process ($K = 10$, see Eq. (2)) with the true K provided to all latent models | **(right)** Data generated using a non-latent process (see Eq. (1)) with $K = \frac{\min(J,I+L)}{2}$ for all latent models except for OVP-LaMe where $K = \frac{I+J+L}{2} = 100$. Alt text: on latent data, LaMe's estimates decrease with N toward smaller mean absolute errors on the true parameters whereas other models do not for $N \in [1000; 10000]$; on non-latent data, estimates from OLS, OVP-LaMe, LaMe and JAOG decrease with N toward a smaller mean absolute error on $\beta^{(my)}_{true}$ whereas other models do not for $N \in [1000; 10000]$.

$SNR(Y|X, S) = 0.1$ and $SNR(S|X) = 1$. Data generated from the latent mediation system in Eq. (2) uses $K_{true} = 10$ latent dimensions. For all latent models (OVP-LaMe aside), K is set to K_{true} when data are generated under a latent model. When data are generated under a non-latent model, we set $K = \frac{\min(J,I+L)}{2}$ to test latent models' ability to compress a full-rank covariance matrix between X and M, Y through a matrix of half this rank. Regularized models (LASSO, ElasticNet, Pathway LASSO, PCA+LASSO) are not considered because the data in this experiment were generated without sparse coefficients.

Fig. 2 reports the mean absolute error between the inferred $(\beta^{(xm)}, \beta^{(my)})$ and the true $(\beta^{(xm)}_{true}, \beta^{(my)}_{true})$, for varying numbers of observations N, and for data generated under latent and non-latent models. First, LaMe consistently outperforms all baselines in the latent case (Fig. 2-left). In particular, because $K_{true} \ll \min(J, I + L)$, LaMe reduces estimation error through its low-rank restriction where OLS cannot. Second, and as expected, OVP-LaMe performs similarly to OLS, confirming that the violation of Eq. (5) allows OVP-LaMe to converge to a solution equivalent to the OLS solution. In particular, OVP-LaMe is prone to overfitting, just as OLS, when data are generated under a latent model. In the non-latent case, LaMe outperforms latent baselines and performs nearly as well as the non-latent models, despite the constraint $K = \frac{\min(J,I+L)}{2}$.

Additional results for $\{I = 50, J = 200\}$, as well as for varying signal-to-noise ratios are presented and discussed in Web Appendix 21.

4.2.3 *Causal effect recovery.* We recall from Eq. (3) that $NIE = \beta^{(xs)}\beta^{(sy)}$, $NDE = \beta^{(xy)}$ and $TE = \beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}$ (Albert et al., 2016; Derkach et al., 2019). We observe that LaMe correctly retrieves the ground-truth natural effects coefficients overall in Fig. 3, when optimized with $N = 1\ 000$ observations on data generated under the exact recoverability assumption (see Proposition 1) with $I = 10$, $J = 50$, $K = 5$ and $L = 1$, unless specified otherwise. The normalization using the true total effect provides a shared comparison ground while avoiding artificial errors when either the NDE or NIE coefficient is small. Fig. 3-top in particular shows that natural effects estimation error is stable when K is larger than the true number of latent dimensions K_{true} used for data generation. When $K < K_{true}$ however, the NDE and NIE estimates are not accurate, as the under-parametrized model must express effects through a smaller number of latent variables. Taken together, these findings hint at a possible model selection strategy based on parameter stability rather than information criteria. Estimation

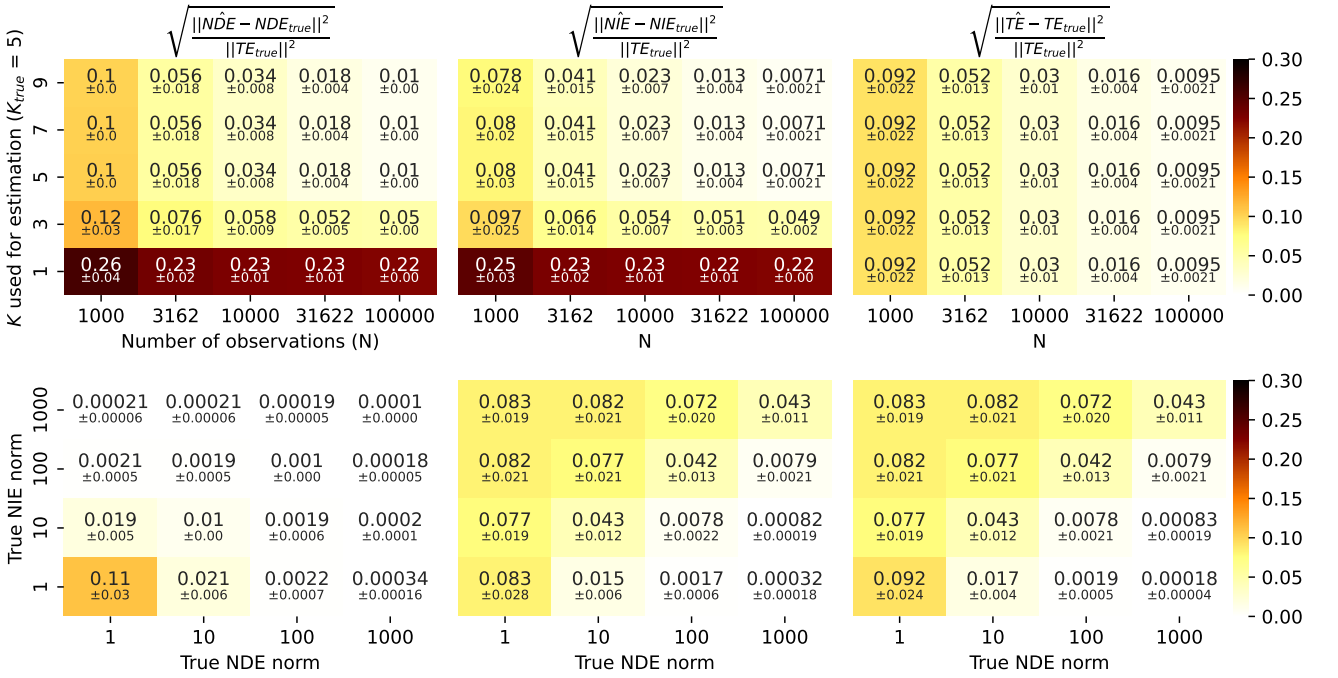


Figure 3. LaMe accurately recovers the natural effects — Root mean squared error (RMSE) of estimated natural effect coefficients, normalized by the norm of the true total-effect coefficient, averaged over 100 runs, along with the associated standard deviation, for LaMe optimized with $N = 1\ 000$ observations on data generated with $I = 10$, $J = 50$, $K = 5$ and $L = 1$, unless specified otherwise. **(Top)** Varying the number of observations and the number of latent variables used for estimation when the true number of latent variables used for data generation is fixed at $K_{true} = 5$ | **(Bottom)** Varying the values of NIE and NDE coefficients norm used for data generation | Root mean squared error relative to the true total effect between the true and estimated **(Left)** natural direct effect, **(Middle)** natural indirect effect and **(Right)** total effect. Alt text: Heatmap representing the RMSE of LaMe with respect to the true NIE, NDE and TE path coefficients relative to the true TE. The RMSE of NIE, NDE and TE decreases with N, with $K - K_{true}$ (plateauing once $K \geqslant K_{true}$) and with $\frac{||NDE_{true}||^2}{||NIE_{true}||^2}$, with a relative RMSE remaining below 0.1 in most cases.

error also decreases over the sample sizes considered. Finally, note that the direct path ensures that TE is equal to the OLS regression coefficient X and Y, namely $TE = (X^TX)^{-1}X^TY$, explaining why it does not depend on K. Fig. 3-bottom shows that LaMe estimates the NDE, NIE and TE coefficients more accurately when the norm of the NDE coefficient matrix exceeds that of the NIE coefficient matrix: in this case, most of the observed covariance between X and Y is expressed through the NDE, whose estimation does not require a fine approximation of the unobserved variable S. Conversely, this explains why the relative error on the NDE, NIE and TE coefficients increases when the norm of the NIE coefficient is larger than that of the NDE. Overall, LaMe accurately retrieves the true natural effects across a range of direct and indirect effect magnitudes, supporting the applicability of the LaMe model to various settings.

4.2.4 *Retrieving the manifest variables involved in mediation.* We show here that LaMe efficiently identifies the entries of M that are both influenced by X through S and informative about Y through S given X. These are not observed mediators, since there is no direct path between M and X or Y; M is a descendant of the latent mediator rather than a causal intermediate on the indirect path. However, they are the closest analogue of observed mediating variables and are used as such in (Derkach et al., 2019); for simplicity, we refer to them as mediating manifest features. Building on the definition proposed in (Derkach et al., 2019), we define a global mediating manifest feature M_j as the j^{th} entry of M under the exact recoverability assumption such that both:

$$\sum_i^I |C_{ij}^{XM}| := \sum_i^I |\sum_k^K \beta_{ik}^{(xs)}\beta_{kj}^{(sm)}| \neq 0 \qquad \text{and} \qquad \sum_l^L |C_{jl}^{MY}| := \sum_l^L |\sum_k^K \beta_{jk}^{(ms)}\beta_{kl}^{(sy)}| \neq 0$$

A mediating manifest feature is therefore an entry of M, noted M_j, such that $C_j := \sum_i^I \sum_l^L |C_{ij}^{XM}||C_{jl}^{MY}| \neq 0$. This new definition is intended to generalize that proposed by (Derkach et al., 2019), in particular to make it invariant to linear transformations of the latent space and to be robust to cases where manifest variables are correlated (see Web Appendix 20).

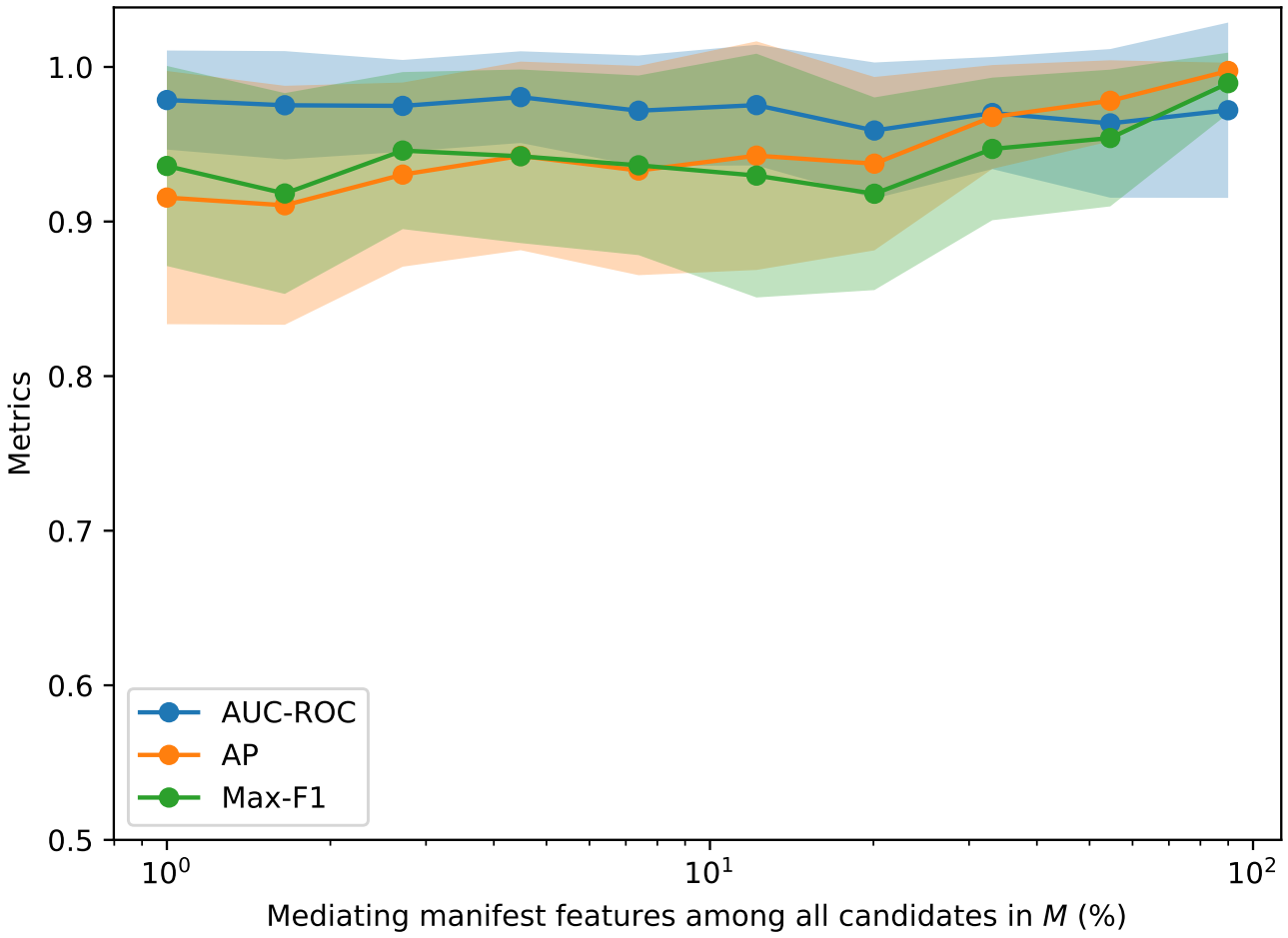


Figure 4. LaMe accurately recovers the observed "mediating" set of manifest variables in M among a large set of "non-mediating" candidates – Data are generated from Eq. (2) with $I = 10, J = 10 + \Delta J, K = 2, L = 1, N = 5\,000$ without confounding, where $\Delta J \in [1, 990]$ is the number of features in M that do not contribute to mediation mechanisms (that are either not influenced by X, or not informative about Y given X). The shaded area represents the standard deviation over 50 datasets generated under the same parameter setting. The AUC-ROC, AP and Max-F1 all range from 0 (worst) to 1 (best), and quantify how well the model recovers the true set of mediating manifest features among all possible candidates. Alt text: The three metrics remain close to 1 over the whole range of candidate mediators, with standard deviations around 0.05 across all experiments.

We generate data from Eq. (2) such that there are exactly $J = 10$ observed mediating manifest variables M_j with $C_j \neq 0$ and $\Delta J \in [1, 990]$ non-mediating manifest variables with $C_j = 0$. In practice, we generate full-rank parameters (which ensures that $C_j \neq 0$) for the system with J mediating manifest variables, expand $\beta^{(sm)}$ and $\Sigma_{m|s}$ with ΔJ columns (and rows for $\Sigma_{m|s}$) such that $C_j = 0$, and generate 5 000 observations according to the resulting set of parameters to train LaMe. We then evaluate whether the model accurately distinguishes the ΔJ manifest variables with a true $C_j = 0$ from the J manifest variables with $C_j \neq 0$ using the area under the ROC curve (AUC-ROC), the Average Precision (AP) and the maximum F1 across classification thresholds (Max-F1). These metrics estimate how well the inferred C_j (continuous value) distinguishes the nonzero entries in $C_j^{(true)}$ (binary value based on its null/non-null entries) without the need to specify a threshold between scores associated with nonzero and near-zero values of C_j. For AUC-ROC, AP and Max-F1, a value of 1 means the model achieves perfect separation between mediating and non-mediating variables in M.

The results presented in Fig. 4 show that LaMe consistently distinguishes the ten true mediating manifest features even in situations where these represent only 1% of the total number of manifest variables.

Table 1

Out-of-sample and in-sample performance metrics of LaMe (version X and M, with and without $\ell 1$ regularization, assuming full noise covariances) and all baseline models. The last lines report the average improvement of models' metrics over OLS across all datasets –hence the null line for OLS. NLPD is the Negative Log-Predictive Density, R^2 the fraction of explained variance and AIC the Akaike Information Criterion. When applicable, the number of latent variables is reported as K. Errors correspond to the standard error across five folds. The colour scale indicates relative performance within each dataset and metric ; (orange=best).

		$NLPD(Y\|X,M)\downarrow$	$R^2(Y\|X,M)\uparrow$	AIC $\downarrow$	$K\downarrow$
NHANES	LaMe (vM)	1.356 ± 0.020	0.151 ± 0.009	97.404 ± 1.110	1.000 ± 0.001
	LaMe (vM) (reg)	1.355 ± 0.020	0.154 ± 0.007	98.270 ± 1.257	1.000 ± 0.001
	LaMe (vX)	1.356 ± 0.020	0.151 ± 0.009	97.404 ± 1.110	1.000 ± 0.001
	LaMe (vX) (reg)	1.355 ± 0.020	0.154 ± 0.007	98.449 ± 1.325	1.000 ± 0.001
	PCA+OLS	1.386 ± 0.022	0.114 ± 0.023	99.382 ± 1.086	24.000 ± 0.001
	PCA+LASSO	1.357 ± 0.020	0.150 ± 0.012	99.872 ± 1.077	24.000 ± 0.001
	PLS(X,M,Y)+OLS	1.502 ± 0.035	-0.148 ± 0.104	97.406 ± 1.107	3.600 ± 0.456
	PLS(X,M)+OLS	1.358 ± 0.020	0.150 ± 0.008	97.147 ± 1.058	3.400 ± 0.358
	OVP-LaMe	1.389 ± 0.023	0.110 ± 0.024	98.597 ± 1.087	19.000 ± 0.001
	JAOG	1.367 ± 0.019	0.136 ± 0.011	97.761 ± 1.120	1.800 ± 0.179
EPIC CAN	LaMe (vM)	0.724 ± 0.003	0.752 ± 0.001	899.770 ± 1.959	1.000 ± 0.001
	LaMe (vM) (reg)	0.719 ± 0.003	0.754 ± 0.001	900.111 ± 2.035	1.000 ± 0.001
	LaMe (vX)	0.720 ± 0.002	0.754 ± 0.001	899.784 ± 1.962	1.000 ± 0.001
	LaMe (vX) (reg)	0.719 ± 0.003	0.755 ± 0.001	902.733 ± 1.817	1.000 ± 0.001
	PCA+OLS	0.779 ± 0.009	0.738 ± 0.001	993.797 ± 1.271	130.600 ± 1.972
	PCA+LASSO	0.714 ± 0.003	0.757 ± 0.001	998.631 ± 1.333	130.600 ± 1.972
	PLS(X,M,Y)+OLS	0.791 ± 0.017	0.720 ± 0.009	898.288 ± 1.961	11.600 ± 0.219
	PLS(X,M)+OLS	0.753 ± 0.008	0.738 ± 0.005	898.268 ± 1.964	11.600 ± 0.219
	OVP-LaMe	1.392 ± 0.036	0.609 ± 0.004	1046.354 ± 1.913	206.000 ± 0.001
	JAOG	1.173 ± 0.086	0.622 ± 0.018	899.466 ± 1.745	4.400 ± 1.486
PISA - I	LaMe (vM)	1.068 ± 0.001	0.504 ± 0.001	102.335 ± 0.098	7.000 ± 0.001
	LaMe (vM) (reg)	1.069 ± 0.001	0.504 ± 0.001	102.337 ± 0.106	7.000 ± 0.001
	LaMe (vX)	1.068 ± 0.001	0.504 ± 0.001	102.336 ± 0.098	7.000 ± 0.001
	LaMe (vX) (reg)	1.068 ± 0.001	0.504 ± 0.001	102.328 ± 0.112	7.000 ± 0.001
	PCA+OLS	1.077 ± 0.001	0.496 ± 0.001	102.353 ± 0.099	19.000 ± 0.001
	PCA+LASSO	1.077 ± 0.001	0.496 ± 0.001	102.367 ± 0.099	19.000 ± 0.001
	PLS(X,M,Y)+OLS	1.438 ± 0.108	-0.181 ± 0.298	103.065 ± 0.202	12.000 ± 0.001
	PLS(X,M)+OLS	1.419 ± 0.109	-0.149 ± 0.308	103.023 ± 0.204	12.000 ± 0.001
	OVP-LaMe	1.068 ± 0.001	0.505 ± 0.001	102.365 ± 0.099	23.000 ± 0.001
	JAOG	1.162 ± 0.001	0.402 ± 0.001	103.858 ± 0.108	1.000 ± 0.001
CRIME	LaMe (vM)	1.055 ± 0.011	0.575 ± 0.004	79.560 ± 1.031	7.000 ± 0.001
	LaMe (vM) (reg)	1.037 ± 0.008	0.578 ± 0.003	92.369 ± 1.104	1.000 ± 0.001
	LaMe (vX)	1.055 ± 0.011	0.575 ± 0.004	79.560 ± 1.031	7.000 ± 0.001
	LaMe (vX) (reg)	1.037 ± 0.008	0.578 ± 0.003	93.572 ± 1.722	1.000 ± 0.001
	PCA+OLS	1.135 ± 0.021	0.556 ± 0.008	81.309 ± 1.005	31.600 ± 0.219
	PCA+LASSO	1.044 ± 0.014	0.582 ± 0.004	86.625 ± 0.927	31.600 ± 0.219
	PLS(X,M,Y)+OLS	1.035 ± 0.007	0.564 ± 0.006	86.660 ± 1.065	24.800 ± 2.704
	PLS(X,M)+OLS	1.037 ± 0.007	0.566 ± 0.005	86.642 ± 1.070	25.400 ± 2.586
	OVP-LaMe	1.149 ± 0.023	0.550 ± 0.010	90.303 ± 0.991	50.000 ± 0.001
	JAOG	1.184 ± 0.036	0.516 ± 0.022	83.379 ± 0.901	5.000 ± 0.001

4.3 *Real-world data*

4.3.1 *Data.* The **NHANES** dataset (National Center for Health Statistics, 2024) examines whether participants' ($N = 2,710$) glycohemoglobin levels (Y) are a direct consequence of lifestyle habits ($I = 14$; hours of sleep, alcohol intake, sunscreen use, etc.) or are mediated by biomarkers ($J = 25$; blood levels of lead, insulin, cholesterol, etc.).

The **EPIC-CAN** dataset (International Agency for Research on Cancer, 2024) examines whether the age of a cancer event ($N = 4,993$) is directly influenced by socio-demographic and lifestyle features ($I = 33$; weight, height, gender, age, smoking habits, physical activity index, diet, etc.) or whether this influence is mediated by blood levels of a high-dimensional collection of candidate proteomic features measured at inclusion in the study ($J = 379$).

The **CRIME** dataset (Redmond, 2002) examines whether cities' ($N = 1,993$) crime rates (output Y) are a direct consequence of population characteristics ($I = 67$ representing the number of inhabitants, unemployment rate, percentage of divorce, ethnicity, etc.) or are mediated by city-level contextual factors ($J = 33$ representing median rent, public transportation, number of vacant houses, etc.).

The **PISA-I** dataset comes from the PISA corpus (Organisation for Economic Co-operation and Development, 2023). It examines whether students' ($N = 52,028$) mathematics proficiency score is a direct consequence of their home environment ($I = 27$; indexes quantifying home possessions, availability of electronic devices, mental health, body image, etc.) or is mediated by the school environment ($J = 20$; indexes quantifying the sense of belonging, safety perception, school quality, etc.).

The complete list of variables is provided in Web Appendix 24. The applications evaluate whether LaMe can summarize exposure–biomarker–outcome associations using compact latent representations while retaining predictive performance; these applications are not intended to provide substantive causal mediation analyses. In particular, for EPIC-CAN, our analysis focused on individuals who developed cancer during follow-up, precluding a straightforward causal interpretation of the fitted relationships. We therefore do not report natural direct and indirect effects for these applications, and instead assess predictive performance, goodness of fit, and the ability to summarize observed relationships through parsimonious latent representations.

4.3.2 *Method.* We empirically observe that most models reach a plateau in predictive performance (the addition of new data no longer improves predictive accuracy) when trained on 80% of the data, and converge to similar fitted solutions. We instead report the results for a small-training-sample stress test by using an outer 5-fold cross validation procedure: we train the models on 20% of the available data, and evaluate them on the 80% remaining observations; this is iterated five times with different training data. The results from the usual 5-fold cross validation are provided in Web Appendix 22. Note that this procedure serves for unbiased evaluation only, and is not involved in any hyper-parameter tuning: test data are systematically isolated from the training procedure.

We optimize LaMe for $K \in [1, 2, 3, 4, 5, 7, 10, 15, 20, 25]$ and select the version with the lowest Bayesian Information Criterion within each fold; for PCA, K has been chosen using scikit-learn's maximum likelihood estimator for the number of components; for PLS and JAOG, K has been chosen to maximize their respective scoring function on a validation set (10% randomly sampled from the training set).

4.3.3 *Results.* We report our results for LaMe and latent baselines in Table 1. LaMe achieves predictive performance comparable to that of the strongest latent baselines, sometimes improving on them, while using parsimonious representations. Visualizations of the inferred systems are provided in Web Appendix 23. In particular, these experiments show that LaMe achieves competitive predictive performance with compact latent representations. These results support a favorable trade-off between predictive performance and latent dimensionality under the stated model-selection procedures. On the other hand, over-parametrized models such as OVP-LaMe converge to solutions that are equivalent to the OLS solution. A similar case arises for PCA-based methods here, for which the maximum likelihood estimator selected models with large K –thus forfeiting any low-dimensional covariance structure. For regularized LaMe on CRIME, and for NHANES and EPIC-CAN, LaMe uses a single latent dimension where other latent approaches used many more to achieve comparable performance in terms of prediction and goodness-of-fit.

5. Conclusion

We have introduced LaMe as a general framework for joint estimation of multivariate latent mediation systems that makes the conditions for dimension reduction and identification of the effect decomposition explicit. The simulations demonstrate gains from low-rank estimation when the data contain shared latent structure and recovery of natural effects under the stipulated measurement model. They also show that accurate estimation of the total effect can coexist with persistent errors in its direct and indirect components when too few latent dimensions are fitted. The real-world applications complement these findings by demonstrating competitive prediction with compact latent representations. The present evidence supports LaMe as a scalable, parsimonious and accurate option for high-dimensional mediation analysis across a wide variety of situations.

Possible directions for future work include incorporating mixed data types such as categorical exposures, jointly inferring the number of latent components K (e.g. through point-process priors), relaxing linearity assumptions and extending the framework to survival models with censored data.

Acknowledgments. This work was supported by a grant from the French National Research Agency as part of the "Investissements d'Avenir ExcellencES" program from France 2030 (SHAPE-Med@Lyon; [ANR-22-EXES-0012]). Funding for grant [IIG_FULL_2022_013] was obtained from Wereld Kanker Onderzoek Fonds (WKOF) as part of the World Cancer Research Fund International grant programme. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Conflicts of interest. The authors declare no conflict of interest.

Data availability statement. The NHANES August 2021–August 2023 data are publicly available at `https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?Cycle=2021-2023`. The CRIME dataset (DOI: 10.24432/C53W3X) is publicly available at `https://doi.org/10.24432/C53W3X`. The PISA-I data were derived from the PISA 2022 database, publicly available at `https://www.oecd.org/en/data/datasets/pisa-2022-database.html`. The data underlying EPIC-CAN is based on data from the European Prospective Investigation into Cancer and Nutrition (EPIC) and cannot be shared publicly due to privacy considerations. EPIC data are available by application to the EPIC Steering Committee (`https://epic.iarc.fr/access/`).

References

Albert, J. M., Geng, C., and Nelson, S. (2016). Causal mediation analysis with a latent mediator. *Biometrical Journal* **58,** 535–548.

Anderson, T. W., Rubin, H., et al. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, pages 111–150.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Cai, X., Wang, Q., and Zhu, Y. (2023). Mediation analysis with latent factors using simultaneous group-wise and parameter-wise penalization. *Stat* **12,** 630.

Chechik, G., Globerson, A., Tishby, N., and Weiss, Y. (2005). Information bottleneck for gaussian variables. *Journal of Machine Learning Research* **6,** 165–188.

Chén, O. Y., Crainiceanu, C., Ogburn, E. L., Caffo, B. S., Wager, T. D., and Lindquist, M. A. (2017). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics* **19,** 121–136.

Derkach, A., Pfeiffer, R. M., Chen, T.-H., and Sampson, J. N. (2019). High dimensional mediation analysis with latent variables. *Biometrics* **75,** 745–756.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Hu, J. Y., DeSimone, M., and Wang, Q. (2024). Utilizing latent connectivity among mediators in high-dimensional mediation analysis. *Stat* **13,** 675.

International Agency for Research on Cancer (2024). European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. (accessed June 26, 2025).

Kawano, S., Fujisawa, H., Takada, T., and Shiroishi, T. (2018). Sparse principal component regression for generalized linear models. *Computational Statistics & Data Analysis* **124,** 180–196.

Nath, T., Caffo, B., Wager, T., and Lindquist, M. A. (2023). A machine learning based approach towards high-dimensional mediation analysis. *NeuroImage* **268,** 119843.

National Center for Health Statistics (2024). National Health and Nutrition Examination Survey, August 2021–August 2023. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available at `https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?Cycle=2021-2023` (accessed June 26, 2025).

Neal, R. M. and Hinton, G. E. (1998). *A View of the Em Algorithm that Justifies Incremental, Sparse, and other Variants*, pages 355–368. Springer Netherlands.

Organisation for Economic Co-operation and Development (2023). PISA 2022 Database. OECD. Available at `https://www.oecd.org/en/data/datasets/pisa-2022-database.html` (accessed June 26, 2025).

Redmond, M. (2002). Communities and Crime. UCI Machine Learning Repository. Available at `https://doi.org/10.24432/C53W3X` (accessed June 26, 2025).

Rijnhart, J. J. M., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W. R., and Heymans, M. W. (2021). Mediation analysis methods used in observational research: a scoping review and recommendations. *BMC Med Res Methodol* **21,**.

Wang, H., Lengerich, B. J., Aragam, B., and Xing, E. P. (2018). Precision lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* **35,** 1181–1187.

Ye, Z., Zhu, Y., and Coffman, D. L. (2021). Variable selection for causal mediation analysis using lasso-based methods. *Statistical methods in medical research* **30,** 1413–1427.

Zeng, P., Shao, Z., and Zhou, X. (2021). Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Computational and Structural Biotechnology Journal* **19,** 3209–3224.

Zhao, Y., Lindquist, M. A., and Caffo, B. S. (2020). Sparse principal component based high-dimensional mediation analysis. *Computational Statistics and Data Analysis* **142,** 106835.

Zhao, Y. and Luo, X. (2022). Pathway lasso: Pathway estimation and selection with high-dimensional mediators. *Statistics and its interface* **15,** 39–50.

Zhu, Y., He, Y., Ma, J., Hu, M., Li, S., and Li, J. (2024). Causal inference with latent variables: Recent advances and future prospectives. In *Proceedings of the 30th ACM SIGKDD*, KDD '24, page 6677–6687, New York, NY, USA. Association for Computing Machinery.

Supplementary Materials. Web Appendices are available with this paper at the Biometrics website on Oxford Academic. The Python code for performing the simulations and data analysis in this paper is available at `https://github.com/GaelPouxMedard/LaMe-repr/releases/tag/v1.1.0-paper`. Code is permanently archived in Zenodo at `https://doi.org/10.5281/zenodo.23264883`.

6. Observed confounding between exposures, mediators and outcomes

Based on (Derkach et al., 2019), observed confounding C enters latent mediation systems as follows:

$$\begin{cases} C = \varepsilon_c \\ X = C\beta^{(cx)} + \varepsilon_x \\ S = X\beta^{(xs)} + C\beta^{(cs)} + \varepsilon_s \\ M = S\beta^{(sm)} + \varepsilon_m \\ Y = X\beta^{(xy)} + S\beta^{(sy)} + C\beta^{(cy)} + \varepsilon_y \end{cases} \tag{14}$$

Observed confounding C enters the non-latent mediation system as follows:

$$\begin{cases} C = \varepsilon_c \\ X = C\beta^{(cx)} + \varepsilon_x \\ M = X\beta^{(xm)} + C\beta^{(cm)} + \varepsilon_m \\ Y = X\beta^{(xy)} + M\beta^{(my)} + C\beta^{(cy)} + \varepsilon_y \end{cases} \tag{15}$$

Here, we demonstrate that observed exposures X and observed confounders C can be jointly represented as a single composite observed variable $\xi := [X, C]$ from a likelihood perspective. As such, the entire technical demonstration in the main text can be conducted by considering $\xi = [X]$ without loss of generality. Modelling confounding simply requires replacing exposures by their augmented counterpart ξ of dimension $I + \dim(C)$, and performing the separation step below after parameters estimation.

For latent systems (Eq. (14)), we start from any joint distribution $p(X, S, M, Y, C)$ along with the conditional independence relationships corresponding to confounded LaMe (see Eq. (14)), namely $M \perp (X, C) \mid S$ and $Y \perp M \mid (X, C, S)$. Developing $p(X, S, M, Y, C)$ using the chain rule, we show that X and C can be treated jointly as a single observed variable from a statistical standpoint:

$$\begin{aligned} & p(X, S, M, Y, C) \\ =& p(X, C)p(S|X, C)p(M|S)p(Y|X, C, S) \\ :=& p(\xi)p(S|\xi)p(M|S)p(Y|\xi, S), \end{aligned} \tag{16}$$

where we defined $\xi := [X, C]$ as a composite observed variable. The last factorization is identical in form to that of unconfounded LaMe-vX in the main text, with ξ replacing X; maximum likelihood estimates for all these parameters are therefore obtained from the same inference procedure, treating ξ as an augmented observed exposure variable of dimension $I + \dim(C)$.

For non-latent systems (Eq. (15)), the same demonstration immediately follows from the chain rule noting that $p(\xi, M, Y) = p(\xi)p(M|\xi)p(Y|\xi, M)$.

The path coefficients from X and C to the mediator and the outcome are then obtained by appropriate slicing of the coefficients associated with ξ e.g. in the latent case $\beta^{(\xi s)} = [\beta^{(xs)}, \beta^{(cs)}]^T$ and $\beta^{(\xi y)} = [\beta^{(xy)}, \beta^{(cy)}]^T$. In both latent and non-latent cases, the path coefficient from C to X is obtained from the off-diagonal blocks in the inferred covariance of ξ, namely Σ_ξ. After appropriate slicing of Σ_ξ in terms of exposures X and confounders C:

$$\Sigma_\xi = \begin{pmatrix} \Sigma_x & \Sigma_{xc} \\ \Sigma_{cx} & \Sigma_c \end{pmatrix}, \tag{17}$$

then the path coefficient from C to X is $\beta^{(cx)} = \Sigma_c^{-1}\Sigma_{cx}$.

7. Exact recoverability error

We recall that:

$$\beta^{(ms)} = \underbrace{((\beta^{(sm)})^T\Sigma_s\beta^{(sm)} + \Sigma_{m|s})^{-1}}_{\Sigma_m^{-1}}(\beta^{(sm)})^T \underbrace{(\Sigma_{s|x} + (\beta^{(xs)})^T\Sigma_x\beta^{(xs)})}_{\Sigma_s}$$

$\beta^{(ms)} = \Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s$.

The expected difference between S and S^* is zero following the tower property of nested expectations:

$$\begin{aligned}\mathbb{E}[S^* - S] &= \mathbb{E}[\mathbb{E}[S^* - S|M]] \\ &= 0\end{aligned}$$

The variance of the difference is:

$$\begin{aligned}Var(S^* - S) &= Var(S) + (1-2)Var(\mathbb{E}[S|M]) \\ &= \Sigma_s - (\beta^{(ms)})^T\Sigma_m\beta^{(ms)} \\ &= \Sigma_s(I - \beta^{(sm)}\beta^{(ms)})\end{aligned}$$

From the definition of $\beta^{(ms)}$:

$$\begin{aligned}\beta^{(ms)} &= \underbrace{((\beta^{(sm)})^T\Sigma_s\beta^{(sm)} + \Sigma_{m|s})^{-1}}_{\Sigma_m^{-1}}(\beta^{(sm)})^T\Sigma_s \\ &= \Sigma_{m|s}^{-1}((\beta^{(sm)})^T\Sigma_s - (\beta^{(sm)})^T\Sigma_s\beta^{(sm)}\beta^{(ms)}) \\ &= \Sigma_{m|s}^{-1}(\beta^{(sm)})^T\Sigma_s(I - \beta^{(sm)}\beta^{(ms)}) \\ &= \Sigma_{m|s}^{-1}(\beta^{(sm)})^T Var(S^* - S)\end{aligned}$$

Therefore:

$$\begin{aligned}Var(S^* - S) &= \Sigma_s(I - \beta^{(sm)}\beta^{(ms)}) \\ &= \Sigma_s(I - \beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T Var(S^* - S)) \\ &= (\Sigma_s^{-1} + \beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T)^{-1}\end{aligned}$$

The result from the main text follows:

$$\begin{aligned}\mathbb{E}[||S^* - S||^2] &= Tr(Var(S^* - S)) + ||\mathbb{E}[S^* - S]||^2 \\ &= Tr(\left(\Sigma_s^{-1} + \beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T\right)^{-1}) + 0 \\ &= \mathcal{O}(J^{-1}),\end{aligned} \tag{18}$$

8. Bottleneck condition

In the most general setting, we assume that $I \geqslant 1$, $J \geqslant 1$, $K \geqslant 1$, $L \geqslant 1$, that all noise covariance matrices are full, and that variables are centred on their average value so that intercepts can be omitted. We show that under these conditions, the optimal solution of a model that violates the bottleneck condition $K < \min(I + L, J)$ corresponds to a reparametrization of the Ordinary Least Squares (OLS) optimum.

It has already been established that in the fully unconstrained case, OLS is the maximum likelihood estimator for a linear model of the observed variables, and that it exactly matches the first and second population moments (Hayashi, 2000). Although the latent system defined in the main text assumes that linear relations between observed variables are structured by a latent variable S, the relations between observed variables remain linear. Therefore, whenever the latent model can attain (a given reparametrization of) the OLS solution, this solution also corresponds to a maximum likelihood estimator of the latent model.

We therefore derive the value of K for which the latent model will necessarily produce a solution equivalent to the OLS solution, while using its own parametric representation. Equivalence here means that the two models reproduce exactly the same population moments through their respective parametrizations. Here, the latent model is identified from the population

moments $\Sigma_{XX}, \Sigma_{MM}, \Sigma_{YY}, \Sigma_{XM}, \Sigma_{XY}, \Sigma_{MY}$ by the following system of equations:

$$\begin{aligned}
\Sigma_{XX} &= \Sigma_x \\
\Sigma_{MM} &= \Sigma_{m|s} + (\beta^{(sm)})^T \Sigma_{s|x} \beta^{(sm)} + (\beta^{(xs)}\beta^{(sm)})^T \Sigma_x \beta^{(xs)}\beta^{(sm)} \\
\Sigma_{YY} &= \Sigma_{y|x,s} + (\beta^{(sy)})^T \Sigma_{s|x} \beta^{(sy)} + (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)})^T \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}) \\
\Sigma_{XY} &= \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}) \\
\Sigma_{XM} &= \Sigma_x \beta^{(xs)}\beta^{(sm)} \\
\Sigma_{MY} &= (\beta^{(sm)})^T \Sigma_{s|x} \beta^{(sy)} + (\beta^{(xs)}\beta^{(sm)})^T \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)})
\end{aligned} \tag{19}$$

The OLS solution is reachable when all these equations can be exactly satisfied simultaneously. The equations involving $\Sigma_{XX}, \Sigma_{MM}, \Sigma_{YY}$, and Σ_{XY} are straightforward, as these population moments are always reachable from the latent model's parameters in the fully unconstrained case: Σ_{XX}, Σ_{MM} and Σ_{YY} can always be exactly modelled by the full matrices Σ_x, $\Sigma_{m|s}$ and $\Sigma_{y|x,s}$; Σ_{XY} can always be modelled exactly by $\beta^{(xy)}$ because its contribution to Σ_{YY} can always be compensated by the full matrix $\Sigma_{y|x,s}$, and because it does not appear in Σ_{MY} once Σ_{XY} is known. The rank constraint on the joint system Eq. (19) induced by the value of K can therefore be expressed in terms of Σ_{XM} and Σ_{MY} only.

Replacing the expression for Σ_{XM} in Σ_{MY}, then $\Sigma_{MY} = (\beta^{(sm)})^T \Sigma_{s|x} \beta^{(sy)} + (\Sigma_{XM})^T \Sigma_{XX}^{-1} \Sigma_{XY}$; the above joint system then reduces to:

$$C := \begin{pmatrix} \Sigma_{XX}^{-1}\Sigma_{XM} \\ \Sigma_{MY}^T - \Sigma_{XY}^T \Sigma_{XX}^{-1} \Sigma_{XM} \end{pmatrix} = \begin{pmatrix} \beta^{(xs)} \\ (\beta^{(sy)})^T \Sigma_{s|x} \end{pmatrix} \beta^{(sm)} \tag{20}$$

The matrix of population moments has $I + L$ rows and J columns ; its rank is therefore upper-bounded by $\min(J, I + L)$. Its parametric representation instead has rank upper-bounded by $\min(J, I + L, K)$. Therefore, a sufficient condition for it to be represented exactly by the latent model is $K \geqslant \min(I + L, J)$. Conversely, a necessary but not sufficient condition for an approximate solution is $K < \min(J, I + L)$. It follows that the globally optimal solution of an unconstrained latent linear model with $K \geqslant \min(J, I + L)$ must necessarily be a reparametrization of the optimal OLS solution.

Note that the exact condition on K for the latent model to be able to reparametrize the OLS solution is $K \geqslant rank(C)$, with C defined above ; however, the sufficient condition based on rank upper bounds is clearer and sufficient for our demonstration, and is the only one we use in the main text.

9. Dependence of S on the scale of the outcome in sum-of-variance objectives

Let θ denote the set of parameters to estimate. Omitting the interprets for simplicity, the joint objective of the model in (Nath et al., 2023) is:

$$\arg\min_{\theta} ||Y - X\beta^{(xy)} - S\beta^{(sy)}||^2 + ||S - X\beta^{(xs)}||^2$$
$$\text{where } S = \Phi(M)$$

After profiling S, this can equivalently be written as an inference problem on S directly:

$$\arg\min_{S=\Phi(M)} Var(Y|X,S) + Var(S|X)$$

The sensitivity of the optimal representation S to the scale of Y is immediate, since the transformation $Y \to cY$ makes $Var(Y|X,S) \to c^2 Var(Y|X,S)$:

$$\arg\min_{S=\Phi(M)} Var(S|X) + Var(Y|X,S) \neq \arg\min_{S=\Phi(M)} Var(S|X) + c^2 Var(Y|X,S)$$

More importantly, this implies that $\hat{NIE}(cY) \neq c\hat{NIE}(Y)$, where $\hat{NIE}(Y)$ denotes the estimated natural indirect effect associated with a one-unit increase in the exposures; the NIE also depends on the scale of Y.

10. Constrained LVMA is a factor analysis model

We show that under the assumptions of the latent linear model from (Derkach et al., 2019), namely $\Sigma_{m|s} = diag(\sigma_{m_1|s}, ..., \sigma_{m_J|s})$, $I = 1$ and $L = 1$, a correctly specified LVMA model (full rank coefficient matrices and positive semi-definite covariances) recovers the same latent subspace as a Factor Analysis (FA) model for M whenever the latent space dimension is larger than $K = 1$, that is large enough not to constrain the population cross-covariances of the system (see Sec. 8).

Although the authors in (Derkach et al., 2019) work under the assumption that $\Sigma_{s|x} = I$, we chose a representation where the unconditional variance $\Sigma_s = I$, without loss of generality. These formulations are equivalent due to the usual invariance of linear latent-variable models to rescaling and rotations of the latent space ; the system can be equivalently defined in terms of

$\Sigma_s = I$ or $\Sigma_{s|x} = I$ after applying the appropriate transformations of the latent space. Here, we choose the convention where $\Sigma_s = I$ to simplify the derivations below. Then, the model implies the following population moment equations:

$$\Sigma_{XX} = \Sigma_x \tag{21}$$

$$\Sigma_{MM} = \Sigma_{m|s} + (\beta^{(sm)})^T \underbrace{(\Sigma_{s|x} + (\beta^{(xs)})^T \Sigma_x \beta^{(xs)})}_{\Sigma_s = I} \beta^{(sm)}$$

$$= \Sigma_{m|s} + (\beta^{(sm)})^T \beta^{(sm)} \tag{22}$$

$$\Sigma_{YY} = \Sigma_{y|x,s} + (\beta^{(sy)})^T \Sigma_{s|x} \beta^{(sy)} + (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)})^T \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}) \tag{23}$$

$$\Sigma_{XY} = \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}) \tag{24}$$

$$\Sigma_{XM} = \Sigma_x \beta^{(xs)} \beta^{(sm)} \tag{25}$$

$$\Sigma_{MY} = (\beta^{(sm)})^T \Sigma_{s|x} \beta^{(sy)} + (\beta^{(xs)}\beta^{(sm)})^T \Sigma_x (\beta^{(xs)}\beta^{(sy)} + \beta^{(xy)}) \tag{26}$$

Equation (22) is exactly the equation for an FA model applied to M, which uniquely identifies the latent subspace. The cross-moment equations can then determine the other population parameters according to this subspace without altering it. In particular, the FA equation can be solved independently from the rest of the system when the model is correctly specified.

As shown in Sec. 8, $K < rank(C)$ induce low-rank constraints, where $rank(C) \leqslant \min(I+L, J)$, meaning the model can only approximate the population moments $Var(M)$, $Cov(X, M)$ and $Cov(M, Y)$ through joint optimization. Therefore, Eq. (22) cannot be solved analytically in isolation: its estimation results from a minimization problem defined by Eq. (22), Eq. (25) and Eq. (26), and the resulting latent space no longer matches with that of an FA solution.

In (Derkach et al., 2019) however, $K > I = 1$ is large enough that the model is not rank-constrained. Then, $\beta^{(sm)}$ and $\Sigma_{m|s}$ can be obtained analytically from Eq. (22) as the solution of a FA model, under suitable identification conditions (Anderson, Rubin, et al., 1956). Once these are known, deriving a closed-form solution for $\beta^{(xs)}, \beta^{(sy)}, \beta^{(xy)}, \Sigma_x, \Sigma_{y|x,s}$ is straightforward from the remaining population moment equations, provided the model remains correctly specified.

We support this observation numerically in Fig. 5-left, where the blue curve shows that the model in (Derkach et al., 2019) recovers exactly the same solution as a simple FA model on M when $K \geqslant K_{true} = rank(C)$, and starts to yield different solutions when $K < K_{true}$. The orange curve shows that the same model with unconstrained noise covariances does not yield the same solution as a FA model, regardless of the rank of C. This further demonstrates the importance of choosing K based on the system's dimensions in a way that guarantees the model does not reduce to either FA or ordinary least squares problems. For exhaustiveness, we also report results when both models are incorrectly specified, that is when the true $\Sigma_{m|s}$ used for data generation is full, in Fig. 5-right; results are similar.

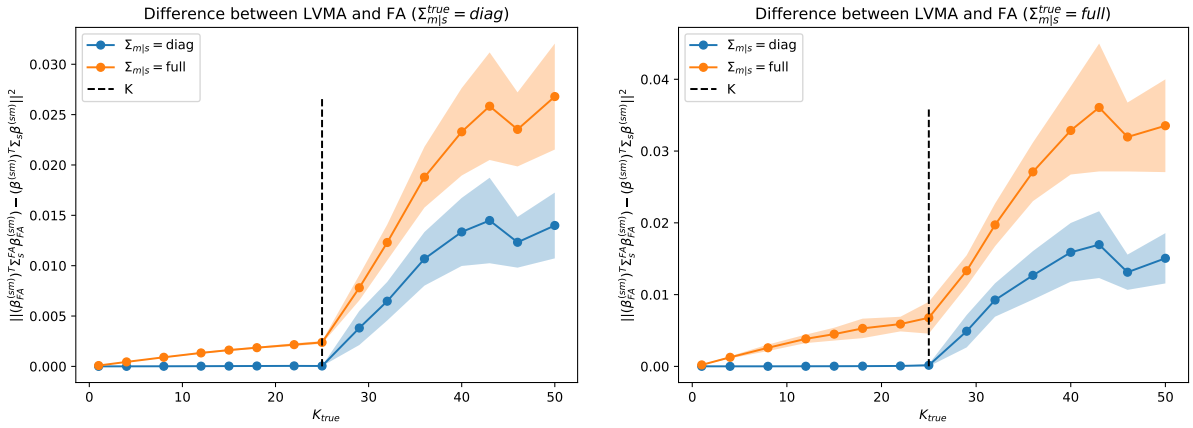


Figure 5. LVMA is similar to a factor analysis — **(Left)** Data generated using a diagonal $\Sigma_{m|s}$; **(Right)** Data generated using a full $\Sigma_{m|s}$. We report the MSE between the covariance of M explained by S from a factor analysis model and the same quantity estimated from one of two latent models, one with diagonal conditional covariance $\Sigma_{m|s}$, and one with full conditional covariance $\Sigma_{m|s}$. We generate $N = 1000$ data points with $I = J = 50$, $L = 1$ and $K_{true} = rank(C)$ (see Eq. (20)) denoting the varying rank of the system used for data generation. We fit the two latent models with $K = 25$. Note that $I > 1$ was not considered in LVMA, although it provides more informative visualizations for our demonstration. The dashed black line indicates when $K_{true} = K$. The shaded region indicates the standard deviation over 50 runs. This shows that forcing $\Sigma_{m|s}$ to be diagonal makes the latent model equivalent to a factor analysis model when $K \geqslant K_{true}$.

11. E-step developments

11.1 *Jensen's inequality*

Here, we derive an expression for a lower bound for $\log p(X, M, Y|\theta)$. Let $q(S)$ denote any distribution of S:

$$\begin{aligned}
\log p(X, M, Y|\theta) &= \log p(X, M, Y|\theta) \underbrace{\int_S q(S) dS}_{=1} \\
&= \int_S q(S) \log p(X, M, Y|\theta) dS \\
&= \int_S q(S) \log \left(\frac{q(S) p(S, X, M, Y|\theta)}{q(S) p(S, X, M, Y|\theta)} p(X, M, Y|\theta) \right) dS \\
&= \int_S q(S) \log \frac{p(S, X, M, Y|\theta)}{q(S)} dS + \int_S q(S) \log \frac{q(S) p(X, M, Y|\theta)}{p(S, X, M, Y|\theta)} dS \\
&= \int_S q(S) \log \frac{p(S, X, M, Y|\theta)}{q(S)} dS + \underbrace{\int_S q(S) \log \frac{q(S)}{p(S|X, M, Y, \theta)} dS}_{KL(q(S)||p_\theta(S|X,M,Y)) \geqslant 0} \\
&\geqslant \int_S q(S) \log \frac{p(S, X, M, Y|\theta)}{q(S)} dS
\end{aligned} \tag{27}$$

where the Kullback-Leibler divergence is $KL(q_{\theta,X,M,Y}(\cdot)||p_\theta(\cdot|X, M, Y)) := \int_S q(S) \log \frac{q_{\theta,o}(s)}{p_\theta(s|o)} dS \geqslant 0$ and $q(S)$ is any distribution of S.

Alternatively, using Jensen's inequality ($\mathbb{E}[\log x] \leqslant \log \mathbb{E}[x]$) directly:

$$\log p(X, M, Y|\theta) = \underbrace{\log \int_S q(S) \frac{p(S, X, M, Y|\theta)}{q(S)} dS}_{\log \mathbb{E}[\frac{p(S,X,M,Y|\theta)}{q(S)}]} \geqslant \underbrace{\int_S q(S) \log \frac{p(S, X, M, Y|\theta)}{q(S)} dS}_{\mathbb{E}[\log \frac{p(S,X,M,Y|\theta)}{q(S)}]}$$

11.2 *Lower-bound maximization*

Here, we demonstrate that the lower bound in Eq. (27) is equal to $\log p(X, M, Y|\theta)$ when $q(S) = q_{\theta,X,M,Y}(S) = \frac{p(S,X,M,Y|\theta)}{p(X,M,Y|\theta)}$ –it is maximized.

$$\begin{aligned}
\int_S q_{\theta,X,M,Y}(S) \log \frac{p(S, X, M, Y|\theta)}{q_{\theta,X,M,Y}(S)} dS &= \int_S \frac{p(S, X, M, Y|\theta)}{p(X, M, Y|\theta)} \log \frac{p(S, X, M, Y|\theta)}{\frac{p(S,X,M,Y|\theta)}{p(X,M,Y|\theta)}} dS \\
&= \int_S \frac{p(S, X, M, Y|\theta)}{p(X, M, Y|\theta)} \log p(X, M, Y|\theta) dS \\
&= \frac{\log p(X, M, Y|\theta)}{p(X, M, Y|\theta)} \underbrace{\int_S p(S, X, M, Y|\theta) dS}_{:= p(X,M,Y|\theta)} \\
&= \log p(X, M, Y|\theta)
\end{aligned} \tag{28}$$

The equality naturally follows (because a lower bound cannot be larger than the bounded quantity):

$$\begin{aligned}
\log p(X, M, Y|\theta) &= \int_S q_{\theta,X,M,Y}(S) \log \frac{p(S, X, M, Y|\theta)}{q_{\theta,X,M,Y}(S)} dS \\
\text{when} \qquad q_{\theta,X,M,Y}(S) &= \frac{p(S, X, M, Y|\theta)}{p(X, M, Y|\theta)}
\end{aligned} \tag{29}$$

12. Full M-step derivation

Intercepts are omitted for notational clarity; they can be included in the usual way by augmenting the relevant design matrix with a column of ones.

12.1 *LaMe-vX*

The solutions for the LaMe-vX model are summarized below:

$$\begin{aligned}
\beta^{(xs)} &= \left(X^T X \right)^{-1} X^T \overline{S} & \beta^{(sm)} &= \left(\overline{S^T S} \right)^{-1} \overline{S}^T M \\
\beta^{(xy)} &= \left(X^T (I - P_S) X \right)^{-1} X^T (I - P_S) Y & \beta^{(sy)} &= \left(\overline{S^T S} \right)^{-1} \overline{S}^T (Y - X\beta^{(xy)})
\end{aligned} \tag{30a}$$

$$\Sigma_x = \frac{1}{N} X^T X \tag{30b}$$
$$\Sigma_{s|x} = \frac{1}{N}\left(\overline{S^T S} + \left(\overline{S} - X\beta^{(xs)}\right)^T \left(\overline{S} - X\beta^{(xs)}\right) - \overline{S}^T\overline{S}\right)$$
$$\Sigma_{m|s} = \frac{1}{N}\left((\beta^{(sm)})^T(\overline{S^T S})\beta^{(sm)} + \left(M - \overline{S}\beta^{(sm)}\right)^T \left(M - \overline{S}\beta^{(sm)}\right) - (\overline{S}\beta^{(sm)})^T(\overline{S}\beta^{(sm)})\right)$$
$$\Sigma_{y|x,s} = \frac{1}{N}\left((\beta^{(sy)})^T(\overline{S^T S})\beta^{(sy)} + (Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)})^T(Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)}) - (\overline{S}\beta^{sy})^T(\overline{S}\beta^{sy})\right)$$

where $P_S := \overline{S}\left(\overline{S^T S}\right)^{-1}\overline{S}^T$ is the estimated latent projection-like matrix.

12.1.1 *A, B, C.*

$$p(S, X, M, Y|\theta) := e^{-\frac{1}{2}\langle S^T S, A\rangle + \langle S, B\rangle + C}$$
$$A := \Sigma_{s|x}^{-1} + \beta^{(sm)}\Sigma_{m|s}^{-1}(\beta^{(sm)})^T + \beta^{(sy)}\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T$$
$$B := Y\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T + X\beta^{(xs)}\Sigma_{s|x}^{-1} - X\beta^{(xy)}\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T + M\Sigma_{m|s}^{-1}(\beta^{(sm)})^T$$
$$C := -\frac{1}{2}\langle X^T X, \left(\Sigma_x^{-1} + \beta^{(xs)}\Sigma_{s|x}^{-1}(\beta^{(xs)})^T + \beta^{(xy)}\Sigma_{y|x,s}^{-1}(\beta^{(xy)})^T\right)\rangle$$
$$- \frac{1}{2}\langle M^T M, \Sigma_{m|s}^{-1}\rangle - \frac{1}{2}\langle Y^T Y, \Sigma_{y|x,s}^{-1}\rangle$$
$$+ \langle X^T Y, \beta^{(xy)}\Sigma_{y|x,s}^{-1}\rangle - \frac{N}{2}\log\left((2\pi)^{I+J+K+L}|\Sigma_x||\Sigma_{s|x}||\Sigma_{m|s}||\Sigma_{y|x,s}|\right)$$

Let $\bar{\ell}_\theta$ denote the expected incomplete log-likelihood (see the main text):

$$\bar{\ell}_\theta = \int_S q_{\theta,X,M,Y}(S)\log\frac{p(S, X, M, Y|\theta)}{q_{\theta,X,M,Y}(S)}dS$$

12.1.2 $x \to s$. For $\beta^{(xs)}$, $\Sigma_{s|x}$:

$$\partial_{\beta^{(xs)}}\bar{\ell}_\theta = \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\partial_{\beta^{(xs)}}\left(-\frac{1}{2}(s_i - x_i\beta^{(xs)})\Sigma_{s|x}^{-1}(s_i - x_i\beta^{(xs)})^T\right)ds_i = 0$$
$$= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\partial_{\beta^{(xs)}}\left((s_i - x_i\beta^{(xs)})\Sigma_{s|x}^{-1}(s_i - x_i\beta^{(xs)})^T\right)ds_i = 0$$
$$= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\left((-x_i^T)(s_i - x_i\beta^{(xs)})\Sigma_{s|x}^{-1}\right)ds_i = 0$$
$$\Leftrightarrow \beta^{(xs)} = \left(\sum_i^N x_i^T x_i\right)^{-1}\sum_i^N x_i^T\overline{s_i}$$
$$= (X^T X)^{-1}X^T\overline{S}$$
$$\partial_{\Sigma_{s|x}}\bar{\ell}_\theta = \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\partial_{\Sigma_{s|x}}\left(-\frac{1}{2}(s_i - x_i\beta^{(xs)})\Sigma_{s|x}^{-1}(s_i - x_i\beta^{(xs)})^T - \frac{1}{2}\ln(|\Sigma_{s|x}|)\right)ds_i = 0$$
$$= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\left(\Sigma_{s|x}^{-1}(s_i - x_i\beta^{(xs)})^T(s_i - x_i\beta^{(xs)})\Sigma_{s|x}^{-1} - \Sigma_{s|x}^{-1}\right)ds_i = 0$$
$$\Leftrightarrow \Sigma_{s|x} = \frac{1}{N}\sum_i^N \int_S q_{\theta,x_i,m_i,y_i}(s_i)\left((s_i - x_i\beta^{(xs)})^T(s_i - x_i\beta^{(xs)})\right)ds_i$$
$$= \frac{1}{N}\sum_i^N \overline{s_i^T s_i} - \overline{s_i}^T(x_i\beta^{(xs)}) - (x_i\beta^{(xs)})^T\overline{s_i} + (x_i\beta^{(xs)})^T(x_i\beta^{(xs)})$$
$$= \frac{1}{N}\left(\overline{S^T S} - \overline{S}^T X\beta^{(xs)} - (X\beta^{(xs)})^T\overline{S} + (X\beta^{(xs)})^T(X\beta^{(xs)})\right)$$
$$= \frac{1}{N}\left(\overline{S^T S} + \left(\overline{S} - X\beta^{(xs)}\right)^T\left(\overline{S} - X\beta^{(xs)}\right) - \overline{S}^T\overline{S}\right)$$

12.1.3 $s \to m$. For $\beta^{(sm)}$, $\Sigma_{m|s}$:

$$\begin{aligned}
\partial_{\beta^{(sm)}}\bar{\ell}_\theta &= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)(-2s_i^T)(m_i - s_i\beta^{(sm)})\Sigma_{m|s}^{-1} ds_i = 0 \\
\Leftrightarrow \beta^{(sm)} &= \left(\sum_i^N \overline{s_i^T s_i}\right)^{-1} \sum_i^N \overline{s_i}^T m_i \\
&= \left(\overline{S^T S}\right)^{-1} \overline{S}^T M \\
\partial_{\Sigma_{m|s}}\bar{\ell}_\theta &= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i) \left(-\Sigma_{m|s}^{-1}(m_i - s_i\beta^{(sm)})^T(m_i - s_i\beta^{(sm)})\Sigma_{m|s}^{-1} + \Sigma_{m|s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \Sigma_{m|s} &= \frac{1}{N}\sum_i^N \int_S q_{\theta,x_i,m_i,y_i}(s_i)(m_i - s_i\beta^{(sm)})^T(m_i - s_i\beta^{(sm)}) ds_i \\
&= \frac{1}{N}\sum_i^N m_i^T m_i - m_i^T(\overline{s_i}\beta^{(sm)}) - (\overline{s_i}\beta^{(sm)})^T m_i + \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)(s_i\beta^{(sm)}))^T(s_i\beta^{(sm)}) ds \\
&= \frac{1}{N}\sum_i^N m_i^T m_i - m_i^T(\overline{s_i}\beta^{(sm)}) - (\overline{s_i}\beta^{(sm)})^T m_i + (\beta^{(sm)})^T\overline{S_i^2}\beta^{(sm)} \\
&= \frac{1}{N}\left(M^T M - M^T\overline{S}\beta^{(sm)} - (\overline{S}\beta^{(sm)})^T M + (\beta^{(sm)})^T(\overline{S^T S})\beta^{(sm)}\right) \\
&= \frac{1}{N}\left((\beta^{(sm)})^T(\overline{S^T S})\beta^{(sm)} + \left(M - \overline{S}\beta^{(sm)}\right)^T\left(M - \overline{S}\beta^{(sm)}\right) - (\overline{S}\beta^{(sm)})^T(\overline{S}\beta^{(sm)})\right)
\end{aligned}$$

12.1.4 $x, s \to y$. For $\beta^{(xy)}$, $\beta^{(sy)}$:

$$\begin{aligned}
\partial_{\beta^{(xy)}}\bar{\ell}_\theta &= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i) \left(-2x_i^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \beta^{(xy)} &= \left(\sum_i^N x_i^T x_i\right)^{-1}\left(\sum_i^N x_i^T Y_i - x_i^T\overline{s_i}\beta^{(sy)}\right) \\
&= \left(X^T X\right)^{-1} X^T Y - \left(X^T X\right)^{-1}\left(X^T\overline{S}\right)\beta^{(sy)} \\
\partial_{\beta^{(sy)}}\bar{\ell}_\theta &= \sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i) \left(-2s_i^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \beta^{(sy)} &= \left(\sum_i^N \overline{s_i^T s_i}\right)^{-1}\left(\sum_i^N \overline{s_i}^T Y_i - \overline{s_i}^T x_i\beta^{(xy)}\right) \\
&= \left(\overline{S^T S}\right)^{-1}\overline{S}^T Y - \left(\overline{S^T S}\right)^{-1}\left(\overline{S}^T X\right)\beta^{(xy)}
\end{aligned}$$

Thus:

$$\underbrace{\begin{pmatrix} X^T X & X^T\overline{S} \\ \overline{S}^T X & \overline{S^T S}\end{pmatrix}}_{A}\underbrace{\begin{pmatrix}\beta^{(xy)} \\ \beta^{(sy)}\end{pmatrix}}_{\Theta} = \underbrace{\begin{pmatrix} X^T Y \\ \overline{S}^T Y\end{pmatrix}}_{r}$$

The solution is $\Theta = A^{-1}r$.

Finally for $\Sigma_{y|x,s}$:

$$\begin{aligned}
&\partial_{\Sigma_{y|x,s}}\bar{\ell}_\theta = \\
&\sum_i^N \int_{\mathcal{S}} q_{\theta,x_i,m_i,y_i}(s_i)\left(\Sigma_{y|x,s}^{-1}(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1} - \Sigma_{y|x,s}^{-1}\right)ds_i \\
&= 0 \\
\Leftrightarrow \Sigma_{y|x,s} &= \frac{1}{N}\sum_i^N \int_S q_{\theta,x_i,m_i,y_i}(s_i)(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})ds_i \\
&= \frac{1}{N}((Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)})^T(Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)}) - (\overline{S}\beta^{(sy)})^T(\overline{S}\beta^{(sy)})) \\
&+ \frac{1}{N}(\beta^{(sy)})^T\overline{S^TS}\beta^{(sy)}
\end{aligned}$$

12.2 *LaMe-vM*

12.2.1 *A, B, C.*

$$\begin{aligned}
p(S,X,M,Y|\theta) &:= e^{-\frac{1}{2}\langle S^TS,A\rangle + \langle S,B\rangle + C} \\
A &:= \Sigma_{s|m}^{-1} + \beta^{(sx)}\Sigma_{x|s}^{-1}(\beta^{(sx)})^T + \beta^{(sy)}\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T \\
B &:= M\beta^{(ms)}\Sigma_{s|m}^{-1} + X\Sigma_{x|s}^{-1}(\beta^{(sx)})^T - X\beta^{(xy)}\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T + Y\Sigma_{y|x,s}^{-1}(\beta^{(sy)})^T \\
C &:= -\frac{1}{2}\langle X^TX, \left(\Sigma_{x|s}^{-1} + \beta^{(xy)}\Sigma_{y|x,s}^{-1}(\beta^{(xy)})^T\right)\rangle \\
&\quad - \frac{1}{2}\langle M^TM, \left(\Sigma_m^{-1} + \beta^{(ms)}\Sigma_{s|m}^{-1}(\beta^{(ms)})^T\right)\rangle - \frac{1}{2}\langle Y^TY, \Sigma_{y|x,s}^{-1}\rangle \\
&\quad + \langle X^TY, \beta^{(xy)}\Sigma_{y|x,s}^{-1}\rangle - \frac{N}{2}\log\left((2\pi)^{I+J+K+L}|\Sigma_m||\Sigma_{s|m}||\Sigma_{x|s}||\Sigma_{y|x,s}|\right)
\end{aligned}$$

12.2.2 $m \to s$. For $\beta^{(ms)}$, $\Sigma_{s|m}$:

$$\begin{aligned}
\partial_{\beta^{(ms)}}\bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i)\partial_{\beta^{(ms)}}\left(-\frac{1}{2}(s_i - m_i\beta^{(ms)})\Sigma_{s|m}^{-1}(s_i - m_i\beta^{(ms)})^T\right)ds_i = 0 \\
&= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i)\left((m_i^T)(s_i - m_i\beta^{(ms)})\Sigma_{s|m}^{-1}\right)ds_i = 0 \\
&= (\sum_i^N m_i^Tm_i)\beta^{(ms)} - \sum_i^N m_i^T(\overline{s_i}) = 0 \\
\Leftrightarrow \beta^{(ms)} &= \left(\sum_i^N m_i^Tm_i\right)^{-1}\sum_i^N m_i^T(\overline{s_i}) \\
&= (M^TM)^{-1}M^T\overline{S} \\
\partial_{\Sigma_{s|m}}\bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i)\partial_{\Sigma_{s|m}}\left(-(s_i - m_i\beta^{(ms)})\Sigma_{s|m}^{-1}(s_i - m_i\beta^{(ms)})^T - \ln(|\Sigma_{s|m}|)\right)ds_i = 0 \\
&= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i)\left(-\Sigma_{s|m}^{-1}(s_i - m_i\beta^{(ms)})^T(s_i - m_i\beta^{(ms)})\Sigma_{s|m}^{-1} + \Sigma_{s|m}^{-1}\right)ds_i = 0 \\
\Leftrightarrow \Sigma_{s|m} &= \frac{1}{N}\sum_i^N \int_S q_{\theta,o_i}(s_i)\left((s_i - m_i\beta^{(ms)})^T(s_i - m_i\beta^{(ms)})\right)ds_i \\
&= \frac{1}{N}\sum_i^N \overline{s_i^Ts_i} + (\overline{s_i} - m_i\beta^{(ms)})^T(\overline{s_i} - m_i\beta^{(ms)}) - \overline{s_i}^T\overline{s_i} \\
&= \frac{1}{N}\left(\overline{S^TS} + (\overline{S} - M\beta^{(ms)})^T(\overline{S} - M\beta^{(ms)}) - \overline{S}^T\overline{S}\right)
\end{aligned}$$

12.2.3 $s \to x$. For $\beta^{(sx)}$, $\Sigma_{x|s}$:

$$\begin{aligned}
\partial_{\beta^{(sx)}} \bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i)(-2s_i^T)(x_i - s_i\beta^{(sx)})\Sigma_{x|s}^{-1} ds_i = 0 \\
\Leftrightarrow \beta^{(sx)} &= \left(\sum_i^N \overline{s_i^T s_i}\right)^{-1} \sum_i^N \overline{s_i}^T(x_i) \\
&= \left(\overline{S^T S}\right)^{-1} \overline{S}^T X
\end{aligned}$$

$$\begin{aligned}
\partial_{\Sigma_{x|s}} \bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i) \left(-\Sigma_{x|s}^{-1}(x_i - s_i\beta^{(sx)})^T(x_i - s_i\beta^{(sx)})\Sigma_{x|s}^{-1} + \Sigma_{x|s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \Sigma_{x|s} &= \frac{1}{N}\sum_i^N \int_S q_{\theta,o_i}(s_i)(x_i - s_i\beta^{(sx)})^T(x_i - s_i\beta^{(sx)}) ds_i \\
&= \frac{1}{N}\sum_i^N (\beta^{(sx)})^T \overline{s_i^T s_i}\beta^{(sx)} + (x_i - \overline{s_i}\beta^{(sx)})^T(x_i - \overline{s_i}\beta^{(sx)}) - (\overline{s_i}\beta^{(sx)})^T(\overline{s_i}\beta^{(sx)}) \\
&= \frac{1}{N}\left((\beta^{(sx)})^T(\overline{S^T S})\beta^{(sx)} + (X - \overline{S}\beta^{(sx)})^T(X - \overline{S}\beta^{(sx)}) - (\overline{S}\beta^{(sx)})^T(\overline{S}\beta^{(sx)})\right)
\end{aligned}$$

12.2.4 $x, s \to y$. For $\beta^{(xy)}$, $\beta^{(sy)}$:

$$\begin{aligned}
\partial_{\beta^{(xy)}} \bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i) \left(-2x_i^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \beta^{(xy)} &= \left(\sum_i^N x_i^T x_i\right)^{-1} \left(\sum_i^N x_i^T Y_i - x_i^T \overline{s_i}\beta^{(sy)}\right) \\
&= \left(X^T X\right)^{-1} X^T Y - \left(X^T X\right)^{-1} \left(X^T \overline{S}\right) \beta^{(sy)} \\
\partial_{\beta^{(sy)}} \bar{\ell}_\theta(O) &= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i) \left(-2s_i^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1}\right) ds_i = 0 \\
\Leftrightarrow \beta^{(sy)} &= \left(\sum_i^N \overline{s_i^T s_i}\right)^{-1} \left(\sum_i^N \overline{s_i}^T Y_i - \overline{s_i}^T x_i\beta^{(xy)}\right) \\
&= \left(\overline{S^T S}\right)^{-1} \overline{S}^T Y - \left(\overline{S^T S}\right)^{-1} \left(\overline{S}^T X\right) \beta^{(xy)}
\end{aligned}$$

Thus:

$$\underbrace{\begin{pmatrix} X^T X & X^T\overline{S} \\ \overline{S}^T X & \overline{S^T S} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \beta^{(xy)} \\ \beta^{(sy)} \end{pmatrix}}_{\Theta} = \underbrace{\begin{pmatrix} X^T Y \\ \overline{S}^T Y \end{pmatrix}}_{r}$$

The solution is $\Theta = A^{-1}r$.
Finally for $\Sigma_{y|x,s}$:

$$\begin{aligned}
&\partial_{\Sigma_{y|x,s}} \bar{\ell}_\theta(O) \\
&= \sum_i^N \int_{\mathcal{S}} q_{\theta,o_i}(s_i) \left(\Sigma_{y|x,s}^{-1}(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})\Sigma_{y|x,s}^{-1} - \Sigma_{y|x,s}^{-1}\right) ds_i \\
&= 0 \\
\Leftrightarrow \Sigma_{y|x,s} &= \frac{1}{N}\sum_i^N \int_S q_{\theta,o_i}(s_i)(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)})^T(Y_i - x_i\beta^{(xy)} - s_i\beta^{(sy)}) ds_i \\
&= \frac{1}{N}((Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)})^T(Y - X\beta^{(xy)} - \overline{S}\beta^{(sy)})) \\
&- \frac{1}{N}((\overline{S}\beta^{(sy)})^T(\overline{S}\beta^{(sy)}) + (\beta^{(sy)})^T\overline{S^T S}\beta^{(sy)})
\end{aligned}$$

13. Model's features and remarks

13.0.1 *Rotational invariance.* Latent mediation models exhibit rotational invariance: for any invertible linear transformation $Q \in \mathbb{R}^{K\times K}$, the sets $(S,\ \beta^{(xs)},\ \beta^{(sm)},\ \beta^{(sy)},\ \Sigma_{s|x})$ and $(SQ,\ \beta^{(xs)}Q,\ Q^{-1}\beta^{(sm)},\ Q^{-1}\beta^{(sy)},\ Q^T\Sigma_{s|x}Q)$ induce exactly the same likelihood. Although ordering conventions can fix a canonical rotation post-hoc (e.g. in PCA), the above rotational invariance remains. Note that, despite the name, rotational invariance is not restricted to orthogonal rotations; any invertible linear transform Q used in the transformations above leaves the likelihood unchanged.

This implies that latent components (and their corresponding parameters) are not uniquely defined. Importantly, once the noise-signal invariance has been fixed by choosing a canonical representation, this non-identifiability is confined to the latent representation itself. The induced mappings between observed variables, such as the path coefficients $\beta^{(xm)} = \beta^{(xs)}\beta^{(sm)}$, remain identifiable. Consequently, the marginalized mediation model

$$p(X, M, Y) = \int_S p(S, X, M, Y)dS$$

remains identifiable.

13.0.2 *Regularized versions.* Regularization can be introduced by changing the prior $p(\theta)$ on the parameters in the log-likelihood, which amounts to adding penalty terms. Here we present ℓ_1 and ℓ_2 regularizations, which both modify the definition of C in the E-step as well as the optimization problem solved in the M-step. We describe below how such penalties can be applied to any coefficient matrix $\beta^{(uv)} \in \{\beta^{(xs)}, \beta^{(sm)}, \beta^{(sy)}, \beta^{(xy)}\}$:

ℓ_2 **(Ridge) regularization** corresponds to placing a zero-mean Gaussian prior with covariance $\Sigma_{v|u}$ on $\beta^{(uv)}$, which amounts to setting $p(\beta^{(uv)}) \propto e^{-\frac{\lambda_2}{2}\langle(\beta^{(uv)})^T\beta^{(uv)},\Sigma^{-1}_{v|u}\rangle}$. As a result, the term C in the log-likelihood becomes $C_{reg} = C - \frac{\lambda_2}{2}\langle(\beta^{(uv)})^T\beta^{(uv)}, \Sigma^{-1}_{v|u}\rangle$, and the update equations become $\beta^{(uv)}_{reg} = (U^TU + \lambda_2 I)^{-1}U^TV$ with corresponding updates of the noise covariance matrices. Importantly, the M-step remains available in closed form under ℓ_2 regularization. Note that this is equivalent to simply adding a penalty term $-\frac{\lambda_2}{2}\langle(\beta^{(uv)})^T\beta^{(uv)}, \Sigma^{-1}_{v|u}\rangle$ to the log-likelihood.

ℓ_1 **(LASSO) regularization** corresponds to placing a zero-mean Laplace prior on $\beta^{(uv)}$, which amounts to setting $p(\beta^{(uv)}) \propto e^{-\lambda_1\sum_{ij}|\beta^{(uv)}_{ij}|}$. As a result, the term C in the log-likelihood becomes $C_{reg} = C - \lambda_1\sum_{ij}|\beta^{(uv)}_{ij}|$. In contrast to the ℓ_2 case, the update equations have no analytical solution and must be solved numerically. Notably, ℓ_1 regularization can partially break rotational invariance by favouring sparse orientations of the latent space. Note that this is equivalent to simply adding a penalty term $-\lambda_1\sum_{ij}|\beta^{(uv)}_{ij}|$ to the log-likelihood.

13.0.3 *Log-normally distributed variables.* Real-world data are not always well approximated by Gaussian distributions. When outcomes are strictly positive and exhibit right skewness, as in survival analysis for example, the log-normal distribution can be a more appropriate modelling choice. The LaMe framework naturally accommodates log-normally distributed variables through a simple reparametrization. Let $Z \in \{X, M, Y\}$ denote a strictly positive variable assumed to follow a log-normal distribution. The associated transformations are then $Z \to \log Z$ throughout the model, and $\log p(X, M, Y|\theta) \to \log p(X, M, Y|\theta) - \sum_{i,j}\log Z_{ij}$, where the summation runs over all the elements of Z.

14. Likelihood-equivalent reparametrizations

Starting from the observed moments parametrization from Section 8, we allow for arbitrary reparametrizations of the parameters related to S, namely $(\beta^{(xs)}, \beta^{(sm)}, \beta^{(sy)}, \Sigma_{s|x})$, noted $(\beta^{(xs)}_*, \beta^{(sm)}_*, \beta^{(sy)}_*, \Sigma^*_{s|x})$. These transformed parameters describe every possible reparametrization, including translations and changes of rank. Remaining parameters $\beta^{(xy)}$, Σ_x, $\Sigma_{m|s}$ and $\Sigma_{y|x,s}$ must be transformed accordingly to preserve the observed moments.

Now, we assume that the bottleneck assumption holds (see Section 8), meaning:

$$\operatorname{rank}(C) = \operatorname{rank}\begin{pmatrix}\Sigma^{-1}_{XX}\Sigma_{XM}\\ \Sigma^T_{MY} - \Sigma^T_{XY}\Sigma^{-1}_{XX}\Sigma_{XM}\end{pmatrix} \geqslant \operatorname{rank}\begin{pmatrix}\beta^{(xs)}\beta^{(sm)}\\ (\beta^{(sy)})^T\Sigma_{s|x}\beta^{(sm)}\end{pmatrix} = K \tag{31}$$

This constraint means that all latent dimensions K are used to approximate the observed moments in C. Therefore, both factors $(\beta^{(xs)}, (\beta^{(sy)})^T\Sigma_{s|x})^T$ and $\beta^{(sm)}$ above must have rank K. Any other factorization of the same C through K dimensions must consequently differ only by an invertible transformation.

In particular, $\beta^{(sm)}_*$ must also have rank K from the condition above. It follows that $(\beta^{(sm)}_*(\beta^{(sm)}_*)^T)$ is invertible, meaning $R := (\beta^{(sm)}_*)^T(\beta^{(sm)}_*(\beta^{(sm)}_*)^T)^{-1}$ exists and satisfies $\beta^{(sm)}_*R = I_K$. Therefore, right-multiplying the parametric approximation of C by R and reminding that likelihood equivalence implies that this representation is preserved by equivalent transformations, then:

$$\begin{pmatrix}\beta^{(xs)}\\ (\beta^{(sy)})^T\Sigma_{s|x}\end{pmatrix}\beta^{(sm)}R = \begin{pmatrix}\beta^{(xs)}_*\\ (\beta^{(sy)}_*)^T\Sigma^*_{s|x}\end{pmatrix}\beta^{(sm)}_*R = \begin{pmatrix}\beta^{(xs)}_*\\ (\beta^{(sy)}_*)^T\Sigma^*_{s|x}\end{pmatrix}$$

Defining $H_M = \beta^{(sm)}R$, which must therefore be an invertible $K \times K$ matrix to preserve the rank K of the transformed expression of C, it follows parametrizations preserving observed moments when the bottleneck condition is satisfied are:

$$\begin{aligned}\beta_*^{(xs)} &= \beta^{(xs)}H_M \\ \beta_*^{(sm)} &= H_M^{-1}\beta^{(sm)} \\ (\beta_*^{(sy)})^T\Sigma_{s|x}^* &= (\beta^{(sy)})^T\Sigma_{s|x}H_M\end{aligned}$$

The last equation preserved the covariance between M and Y given X, and defines a condition up to an invertible linear transformation $K \times K$ matrix noted H_Y:

$$\Sigma_{s|x}^*\beta_*^{(sy)} = H_M^T\Sigma_{s|x}\beta^{(sy)}$$

Since $\Sigma_{s|x}^*$ must remain positive-definite, we can invert it, which yields the transformation of $\beta^{(sy)}$ implied by the conservation of observed moments parametrization:

$$\beta_*^{(sy)} = \underbrace{(\Sigma_{s|x}^*)^{-1}H_M^T\Sigma_{s|x}}_{:=H_Y^{-1}}\beta^{(sy)}$$

Noting that covariances are positive-definite and H_M is invertible, H_Y is also guaranteed to be invertible. Finally, from the definition of H_Y:

$$\Sigma_{s|x}^* = H_M^T\Sigma_{s|x}H_Y$$

Remaining expressions for the transformed parameters follows as in the main text.

15. Supporting derivations for identifiability

15.1 *Transformed covariances remain positive semi-definite*

We recall:

$$\begin{aligned}H_YH_M^{-1} &= \Sigma_{s|x}^{-1}\left((\beta^{(sm)}\Sigma_m^{-1}(\beta^{(sm)})^T)^{-1} - (\beta^{(xs)})^T\Sigma_x\beta^{(xs)}\right) \\ \Sigma_{s|x}^* &= H_M^T\Sigma_{s|x}H_Y \\ \Sigma_{m|s}^* &= \Sigma_{m|s} + (\beta^{(sm)})^T\Sigma_{s|x}\beta^{(sm)} - (\beta_*^{(sm)})^T\Sigma_{s|x}^*\beta_*^{(sm)} \\ \Sigma_{y|x,s}^* &= \Sigma_{y|x,s} + (\beta^{(sy)})^T\Sigma_{s|x}\beta^{(sy)} - (\beta_*^{(sy)})^T\Sigma_{s|x}^*\beta_*^{(sy)}\end{aligned}$$

The equations above guarantee positive semi-definite conditional covariances after transformation. Note that in general, $A^TGA \succeq 0$ when $G \succeq 0$ for any transformation of A.

For convenience of notation, note that the noise covariance of M given X is $\Sigma_{m|x} = \Sigma_m - (\beta^{(xs)}\beta^{(sm)})^T\Sigma_x\beta^{(xs)}\beta^{(sm)}$. Therefore:

$$\begin{aligned}H_YH_M^{-1} &= \Sigma_{s|x}^{-1}\left((\beta^{(sm)}\Sigma_m^{-1}(\beta^{(sm)})^T)^{-1} - (\beta^{(xs)})^T\Sigma_x\beta^{(xs)}\right) \\ &= \Sigma_{s|x}^{-1}(\beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)^{-1} \\ H_MH_Y^{-1} &= \beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T\Sigma_{s|x}\end{aligned}$$

The result is immediate since $\Sigma_{s|x}^*$ since $\Sigma_{m|x} \succeq 0$:

$$\begin{aligned}\Sigma_{s|x}^* &= H_M^T(\beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)^{-1}H_M \\ &\succeq 0\end{aligned}$$

Developing $\Sigma_{m|s}^*$, we can show it also is a form $A^TGA \succeq 0$:

$$\begin{aligned}\Sigma_{m|s}^* &= \Sigma_{m|x} - (\beta^{(sm)})^T((\beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)^{-1})\beta^{(sm)} \succeq 0 \\ &= (I - (\beta^{(sm)})^T((\beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)^{-1})\beta^{(sm)}\Sigma_{m|x}^{-1}) \\ &\times \Sigma_{m|x} \\ &\times (I - (\beta^{(sm)})^T((\beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)^{-1})\beta^{(sm)}\Sigma_{m|x}^{-1})^T \\ &\succeq 0\end{aligned}$$

Finally, let $\xi := (\Sigma_{s|x}^{-1} - \beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)$ and note that:

$$\begin{aligned} \xi &= \xi\Sigma_{s|x}\xi^T + \beta^{(sm)}\Sigma_{m|x}^{-1}\Sigma_{m|s}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T \\ \Leftrightarrow \xi &\succeq 0 \end{aligned}$$

Then positive semi-definiteness of $\Sigma_{y|x,s}^*$ immediately follows:

$$\begin{aligned} \Sigma_{y|x,s}^* &= \Sigma_{y|x,s} + (\beta^{(sy)})^T\Sigma_{s|x}(\Sigma_{s|x}^{-1} - \beta^{(sm)}\Sigma_{m|x}^{-1}(\beta^{(sm)})^T)\Sigma_{s|x}\beta^{(sy)} \\ &= \Sigma_{y|x,s} + (\beta^{(sy)})^T\Sigma_{s|x}\xi\Sigma_{s|x}\beta^{(sy)} \\ &\succeq 0 \end{aligned}$$

15.2 *Invariance of parametric covariance*

$$\begin{aligned} \Sigma_m^* &= \Sigma_{m|s}^* + (\beta_*^{(sm)})^T\Sigma_s^*\beta_*^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T\Sigma_{s|x}\beta^{(sm)} - (\beta_*^{(sm)})^T\Sigma_{s|x}^*\beta_*^{(sm)} + (\beta_*^{(sm)})^T(\Sigma_{s|x}^* + (\beta_*^{(xs)})^T\Sigma_x^*\beta_*^{(xs)})\beta_*^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T\Sigma_{s|x}\beta^{(sm)} + (\beta_*^{(xs)}\beta_*^{(sm)})^T\Sigma_x^*\beta_*^{(xs)}\beta_*^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T\Sigma_{s|x}\beta^{(sm)} + (\beta^{(xs)}H_MH_M^{-1}\beta^{(sm)})^T\Sigma_x^*\beta^{(xs)}H_MH_M^{-1}\beta^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T(\Sigma_{s|x} + (\beta^{(xs)})^T\Sigma_x^*\beta^{(xs)})\beta^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T(\Sigma_{s|x} + (\beta^{(xs)})^T\Sigma_x\beta^{(xs)})\beta^{(sm)} \\ &= \Sigma_{m|s} + (\beta^{(sm)})^T\Sigma_s\beta^{(sm)} \\ &= \Sigma_m \end{aligned}$$

15.3 *Relation under exact recoverability*

$$\begin{aligned} \Sigma_{m|s}\Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s &= 0 \\ \Leftrightarrow (\Sigma_m - (\beta^{(sm)})^T\Sigma_s\beta^{(sm)})\Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s &= 0 \\ \Leftrightarrow (\beta^{(sm)})^T\Sigma_s - (\beta^{(sm)})^T\Sigma_s\beta^{(sm)}\Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s &= 0 \\ \Leftrightarrow (\beta^{(sm)})^T\Sigma_s(I - \beta^{(sm)}\Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s) &= 0 \\ \Leftrightarrow \beta^{(sm)}\Sigma_m^{-1}(\beta^{(sm)})^T\Sigma_s &= I \end{aligned}$$

16. Inverting the model's parameters

Parameter estimates for any factorization of the joint linear Gaussian distribution $p(X, S, M, Y)$ can be recovered from the corresponding covariance representation. For instance, if the optimal parameter set θ_{vM} is obtained from the LaMe-vM factorization, we can easily derive the equivalent parameter set $\theta_{vX} = f(\theta_{vM})$ used in the LaMe-vX factorization. For illustration, consider the expectation of X under the LaMe-vM factorization $\mathbb{E}[X|M] = M\beta^{(ms)}\beta^{(sx)} := M\beta^{(mx)}$, which has no causal interpretation. Instead, under the LaMe-vX factorization, $\mathbb{E}[M|X] = X\beta^{(xm)}$ depends on the causal path coefficient $\beta^{(xm)}$, and can be recovered as follows (Bishop, 2006):

$$\begin{aligned} \beta^{(xm)} &= \Sigma_{xx}^{-1}\Sigma_{xm} \\ &= \Sigma_{xx}^{-1}\Sigma_{mx}^T \\ &= \Big(\underbrace{(\beta^{(sx)})^T(\beta^{(ms)})^T\Sigma_m\beta^{(ms)}\beta^{(sx)}}_{\text{Noise propagating from } M \text{ to } X} + \underbrace{(\beta^{(sx)})^T\Sigma_{s|m}\beta^{(sx)}}_{\text{Noise from } S \text{ to } X} + \underbrace{\Sigma_{x|s}}_{\text{Intrinsic noise}}\Big)^{-1} \\ &\times \Big(\underbrace{\Sigma_m\beta^{(ms)}\beta^{(sx)}}_{\text{Signal from } M \text{ to } X}\Big)^T \end{aligned} \tag{32}$$

For any generic unconditioned linear coefficient $\beta^{(ab)}$, the following relation holds:

$$\beta^{(ab|a)} = \Sigma_{aa}^{-1}\Sigma_{ab} \tag{33}$$

When conditioning on additional variables, the coefficient reads:

$$\beta^{(ab|a,c)} = (\Sigma_{aa} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{ca})^{-1}(\Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}) \tag{34}$$

Finally, we recall below all covariances and cross-covariances needed to compute the regression parameters in various models and factorizations. Note that in the main text, we defined the shorthand notation $\beta^{(my)} := \beta^{(my|m,x)} = (\Sigma_{mm} - \Sigma_{mx}\Sigma_{xx}^{-1}\Sigma_{xm})^{-1}(\Sigma_{my} - \Sigma_{mx}\Sigma_{xx}^{-1}\Sigma_{xy})$ and $\beta^{(xm)} := \beta^{(xm|x)} = \Sigma_{xx}^{-1}\Sigma_{xm}$.

16.1 *Non-latent model*

Standard mediation models can be expressed as:

$$p(x, m, y) = p(y|x, m)p(m|x)p(x)$$

We have:

$$\begin{aligned}
\Sigma_{xx} &= \Sigma_x \\
\Sigma_{mm} &= (\beta^{(xm)})^T \Sigma_{xx} \beta^{(xm)} + \Sigma_m \\
\Sigma_{mx} &= (\beta^{(xm)})^T \Sigma_{xx} \\
\Sigma_{xm} &= \Sigma_{xx} \beta^{(xm)} \\
\Sigma_{my} &= (\beta^{(xm)})^T \Sigma_x \beta^{(xy)} + \Sigma_{mm} \beta^{(my)} \\
\Sigma_{xy} &= \Sigma_{xx} (\beta^{(xy)} + \beta^{(xm)} \beta^{(my)})
\end{aligned}$$

16.2 *Latent model*

Latent mediation models can be expressed as:

$$p(x, m, y, s) = p(y|x, s)p(s|x)p(m|s)p(x) = p(y|x, s)p(s|m)p(x|s)p(m)$$

The equations for inverting the parameters remain the same, but covariance matrices are now expressed as:
Model vX.

$$\begin{aligned}
\Sigma_{xx} &= \Sigma_x \\
\Sigma_{ss} &= \Sigma_{s|x} + (\beta^{(xs)})^T \Sigma_{xx} \beta^{(xs)} \\
\Sigma_{mm} &= \Sigma_{m|s} + (\beta^{(sm)})^T \Sigma_{ss} \beta^{(sm)} \\
\Sigma_{xm} &= \Sigma_{xx} \beta^{(xs)} \beta^{(sm)} \\
\Sigma_{mx} &= (\beta^{(sm)})^T (\beta^{(xs)})^T \Sigma_{xx} \\
\Sigma_{my} &= (\beta^{(sm)})^T (\beta^{(xs)})^T \Sigma_{xx} \beta^{(xy)} + (\beta^{(sm)})^T \Sigma_{ss} \beta^{(sy)} \\
\Sigma_{xy} &= \Sigma_{xx} (\beta^{(xy)} + \beta^{(xs)} \beta^{(sy)})
\end{aligned}$$

Model vM.

$$\begin{aligned}
\Sigma_{mm} &= \Sigma_m \\
\Sigma_{ss} &= \Sigma_{s|m} + (\beta^{(ms)})^T \Sigma_{mm} \beta^{(ms)} \\
\Sigma_{xx} &= \Sigma_{x|s} + (\beta^{(sx)})^T \Sigma_{ss} \beta^{(sx)} \\
\Sigma_{mx} &= \Sigma_{mm} \beta^{(ms)} \beta^{(sx)} \\
\Sigma_{xm} &= (\beta^{(sx)})^T (\beta^{(ms)})^T \Sigma_{mm} \\
\Sigma_{my} &= \Sigma_{mm} \beta^{(ms)} (\beta^{(sx)} \beta^{(xy)} + \beta^{(sy)}) \\
\Sigma_{xy} &= \Sigma_{xx} \beta^{(xy)} + (\beta^{(sx)})^T \Sigma_{ss} \beta^{(sy)}
\end{aligned}$$

17. Signal-to-noise ratios

17.1 *LaMe*

17.1.1 *LaMe-vX.* By definition, signal-to-noise ratios are:

$$SNR(S|X) = \frac{Tr((\beta^{(xs)})^T\Sigma_x\beta^{(xs)})}{Tr(\Sigma_{s|x})} = \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xs)}||_F^2}{||\Sigma_{s|x}^{\frac{1}{2}}||_F^2}$$

$$SNR(M|S) = \frac{Tr((\beta^{(sm)})^T((\beta^{(xs)})^T\Sigma_x\beta^{(xs)} + \Sigma_{s|x})\beta^{(sm)})}{Tr(\Sigma_{m|s})} = \frac{||((\beta^{(xs)})^T\Sigma_x\beta^{(xs)} + \Sigma_{s|x})^{\frac{1}{2}}\beta^{(sm)}||_F^2}{||\Sigma_{m|s}^{\frac{1}{2}}||_F^2}$$

$$\begin{aligned} SNR(Y|X,S) &= \frac{Tr((\beta^{(xy)} + \beta^{(xs)}\beta^{(sy)})^T\Sigma_x(\beta^{(xy)} + \beta^{(xs)}\beta^{(sy)}) + (\beta^{(sy)})^T\Sigma_{s|x}\beta^{(sy)})}{Tr(\Sigma_{y|x,s})} \\ &= \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xs)}\beta^{(sy)})||_F^2 + ||\Sigma_{s|x}^{\frac{1}{2}}\beta^{(sy)}||_F^2}{||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2} \end{aligned}$$

The scaling factor can be obtained by simple algebra so that $\Sigma'_A = s_A\Sigma_A$ so that the signal-to-noise ratio of A given its parent nodes is R_A, for $A \in [X, S, M, Y]$. Thus introducing s_m, s_y and s_s:

$$\begin{aligned} SNR(S|X) := R_s &= \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xs)}||_F^2}{s_s||\Sigma_{s|x}^{\frac{1}{2}}||_F^2} \\ \Leftrightarrow s_s &= \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xs)}||_F^2}{R_s||\Sigma_{s|x}^{\frac{1}{2}}||_F^2} \\ SNR(M|S) := R_m &= \frac{||((\beta^{(xs)})^T\Sigma_x\beta^{(xs)} + s_s\Sigma_{s|x})^{\frac{1}{2}}\beta^{(sm)}||_F^2}{s_m||\Sigma_{m|s}^{\frac{1}{2}}||_F^2} \\ \Leftrightarrow s_m &= \frac{||((\beta^{(xs)})^T\Sigma_x\beta^{(xs)} + s_s\Sigma_{s|x})^{\frac{1}{2}}\beta^{(sm)}||_F^2}{R_m||\Sigma_{m|s}^{\frac{1}{2}}||_F^2} \\ SNR(Y|X,S) := R_y &= \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xs)}\beta^{(sy)})||_F^2 + s_s||\Sigma_{s|x}^{\frac{1}{2}}\beta^{(sy)}||_F^2}{s_y||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2} \\ \Leftrightarrow s_y &= \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xs)}\beta^{(sy)})||_F^2 + s_s||\Sigma_{s|x}^{\frac{1}{2}}\beta^{(sy)}||_F^2}{R_y||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2} \end{aligned}$$

17.1.2 *LaMe-vM.* By definition, signal-to-noise ratios are:

$$SNR(S|M) = \frac{Tr((\beta^{(ms)})^T\Sigma_m\beta^{(ms)})}{Tr(\Sigma_{s|m})} = \frac{||\Sigma_m^{\frac{1}{2}}\beta^{(ms)}||_F^2}{||\Sigma_{s|m}^{\frac{1}{2}}||_F^2}$$

$$SNR(X|S) = \frac{Tr((\beta^{(sx)})^T((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + \Sigma_{s|m})\beta^{(sx)})}{Tr(\Sigma_{x|s})} = \frac{||((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + \Sigma_{s|m})^{\frac{1}{2}}\beta^{(sx)}||_F^2}{||\Sigma_{x|s}^{\frac{1}{2}}||_F^2}$$

$$\begin{aligned} SNR(Y|X,S) &= \frac{Tr((\beta^{(sy)})^T\Sigma_s\beta^{(sy)} + (\beta^{(xy)})^T\Sigma_x\beta^{(xy)} + 2(\beta^{(sx)}\beta^{(xy)})^T\Sigma_s\beta^{(sy)})}{Tr(\Sigma_{y|x,s})} \\ &= \frac{||((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + \Sigma_{s|m})^{\frac{1}{2}}(\beta^{(sx)}\beta^{(xy)} + \beta^{(sy)})||_F^2 + ||\Sigma_{x|s}^{\frac{1}{2}}\beta^{(xy)}||_F^2}{||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2} \end{aligned}$$

where the unconditional variances are $\Sigma_s = ((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + \Sigma_{s|m})$ and $\Sigma_x = \Sigma_{x|s} + (\beta^{(sx)})^T\Sigma_s\beta^{(sx)}$ in the LaMe-vM parametrization.

The scaling factor can be obtained by simple algebra so that $\Sigma'_A = s_A\Sigma_A$ so that the signal-to-noise ratio of A given its

parent nodes is R_A, for $A \in [X, S, M, Y]$. Thus introducing s_x, s_y and s_s:

$$SNR(S|M) := R_s = \frac{||\Sigma_m^{\frac{1}{2}}\beta^{(ms)}||_F^2}{s_s||\Sigma_{s|m}^{\frac{1}{2}}||_F^2} \Leftrightarrow s_s = \frac{||\Sigma_m^{\frac{1}{2}}\beta^{(ms)}||_F^2}{R_s||\Sigma_{s|m}^{\frac{1}{2}}||_F^2}$$

$$SNR(X|S) := R_x = \frac{||((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + s_s\Sigma_{s|m})^{\frac{1}{2}}\beta^{(sx)}||_F^2}{s_x||\Sigma_{x|s}^{\frac{1}{2}}||_F^2}$$

$$\Leftrightarrow s_x = \frac{||((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + s_s\Sigma_{s|m})^{\frac{1}{2}}\beta^{(sx)}||_F^2}{R_x||\Sigma_{x|s}^{\frac{1}{2}}||_F^2}$$

$$SNR(Y|X, S) := R_y = \frac{||\Sigma_{ss}^{\frac{1}{2}}(s_s)(\beta^{(sx)}\beta^{(xy)} + \beta^{(sy)})||_F^2 + s_x||\Sigma_{x|s}^{\frac{1}{2}}\beta^{(xy)}||_F^2}{s_y||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2}$$

$$\Leftrightarrow s_y = \frac{||\Sigma_{ss}^{\frac{1}{2}}(s_s)(\beta^{(sx)}\beta^{(xy)} + \beta^{(sy)})||_F^2 + s_x||\Sigma_{x|s}^{\frac{1}{2}}\beta^{(xy)}||_F^2}{R_y||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2}$$

$$= \frac{||((\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + s_s\Sigma_{s|m})^{\frac{1}{2}}(\beta^{(sx)}\beta^{(xy)} + \beta^{(sy)})||_F^2 + s_x||\Sigma_{x|s}^{\frac{1}{2}}\beta^{(xy)}||_F^2}{R_y||\Sigma_{y|x,s}^{\frac{1}{2}}||_F^2}$$

where $\Sigma_{ss}(s_s) = (\beta^{(ms)})^T\Sigma_m\beta^{(ms)} + s_s\Sigma_{s|m}$ and $\Sigma_{xx}(s_x, s_s) = (\beta^{(sx)})^T(\beta^{(ms)})^T\Sigma_m\beta^{(ms)}\beta^{(sx)} + (\beta^{(sx)})^T s_s\Sigma_{s|m}\beta^{(sx)} + s_x\Sigma_{x|s}$.

17.2 *Non-latent models*

By definition, signal-to-noise ratios are:

$$SNR(M|X) = \frac{Tr((\beta^{(xm)})^T\Sigma_x\beta^{(xm)})}{Tr(\Sigma_{m|x})} = \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xm)}||_F^2}{||\Sigma_{m|x}^{\frac{1}{2}}||_F^2}$$

$$SNR(Y|X, M) = \frac{Tr((\beta^{(xy)} + \beta^{(xm)}\beta^{(my)})^T\Sigma_x(\beta^{(xy)} + \beta^{(xm)}\beta^{(my)}) + (\beta^{(my)})^T\Sigma_{m|x}\beta^{(my)})}{Tr(\Sigma_{y|x,m})}$$

$$= \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xm)}\beta^{(my)})||_F^2 + ||\Sigma_{m|x}^{\frac{1}{2}}\beta^{(my)}||_F^2}{||\Sigma_{y|x,m}^{\frac{1}{2}}||_F^2}$$

The scaling factor can be obtained by simple algebra so that $\Sigma'_A = s_A\Sigma_A$ so that the signal-to-noise ratio of A given its parent nodes is R_A, for $A \in [X, M, Y]$. Thus introducing s_m and s_y:

$$SNR(M|X) := R_m = \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xm)}||_F^2}{s_m||\Sigma_{m|x}^{\frac{1}{2}}||_F^2}$$

$$\Leftrightarrow s_m = \frac{||\Sigma_x^{\frac{1}{2}}\beta^{(xm)}||_F^2}{R_m||\Sigma_{m|x}^{\frac{1}{2}}||_F^2}$$

$$SNR(Y|X, M) := R_y = \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xm)}\beta^{(my)})||_F^2 + s_m||\Sigma_{m|x}^{\frac{1}{2}}\beta^{(my)}||_F^2}{s_y||\Sigma_{y|x,m}^{\frac{1}{2}}||_F^2}$$

$$\Leftrightarrow s_y = \frac{||\Sigma_x^{\frac{1}{2}}(\beta^{(xy)} + \beta^{(xm)}\beta^{(my)})||_F^2 + s_m||\Sigma_{m|x}^{\frac{1}{2}}\beta^{(my)}||_F^2}{R_y||\Sigma_{y|x,m}^{\frac{1}{2}}||_F^2}$$

18. Baselines implementation details

JAOG. We extended (Nath et al., 2023)'s model to handle multivariate latent variables ($K \geqslant 1$). In the article, the authors posit a deterministic dimension reduction model $S = \varphi(M)$, whose output, S, is used in a standard LASSO mediation model. To do so, the authors define an optimization target both for inferring the parameters of $\varphi(M)$ and for inferring the structural parameters of the mediation model. Adapting their notation, we define $E = Y - X\beta^{(xy)}$ and $H = X\beta^{(xs)}$. The target of the multivariate extension of (Nath et al., 2023) then becomes $||\varphi(M) - (E(\beta^{(sy)})^T + H)(\beta^{(sy)}(\beta^{(sy)})^T + I_K)^{-1}||^2$. Finally, we choose $\varphi(M)$ to be a multivariate linear Gaussian regression model so that (Nath et al., 2023)'s approach is the closest to the data generation process used in our experiments.

LVMA. We considered a direct comparison with (Derkach et al., 2019), whose model assumes a diagonal manifest residual covariance $\Sigma_{m|s}$, whereas LaMe and other baselines impose the perfect-measurement restriction $\Sigma_{m|s}\beta^{(ms)} = 0$ for effect identification. In preliminary experiments, an adaptation accommodating multivariate exposures and outcomes performed poorly when the data-generating process violated the diagonal residual-covariance assumption. Such a comparison would therefore combine differences in latent dimensionality with differences in covariance restrictions, making it difficult to isolate the contribution of dimension reduction. Moreover, a direct benchmark of (Derkach et al., 2019) would require extending its scalar exposure-outcome formulation while preserving its covariance restrictions, penalization, and model-selection procedure. The resulting comparison would depend on additional methodological choices beyond the scope of the present evaluation. We therefore avoid attributing the performance of a substantially modified implementation to the original method. Instead, we use an overparametrized LaMe comparator with $K = \frac{I+J+L}{2}$, satisfying $K \geqslant \min(I+L, J)$. This comparator retains the same covariance flexibility while removing the rank restriction responsible for compression, which closely relates to the point we discussed about (Derkach et al., 2019) in the background section.

All remaining baselines were implemented using scikit-learn.

19. Scalability

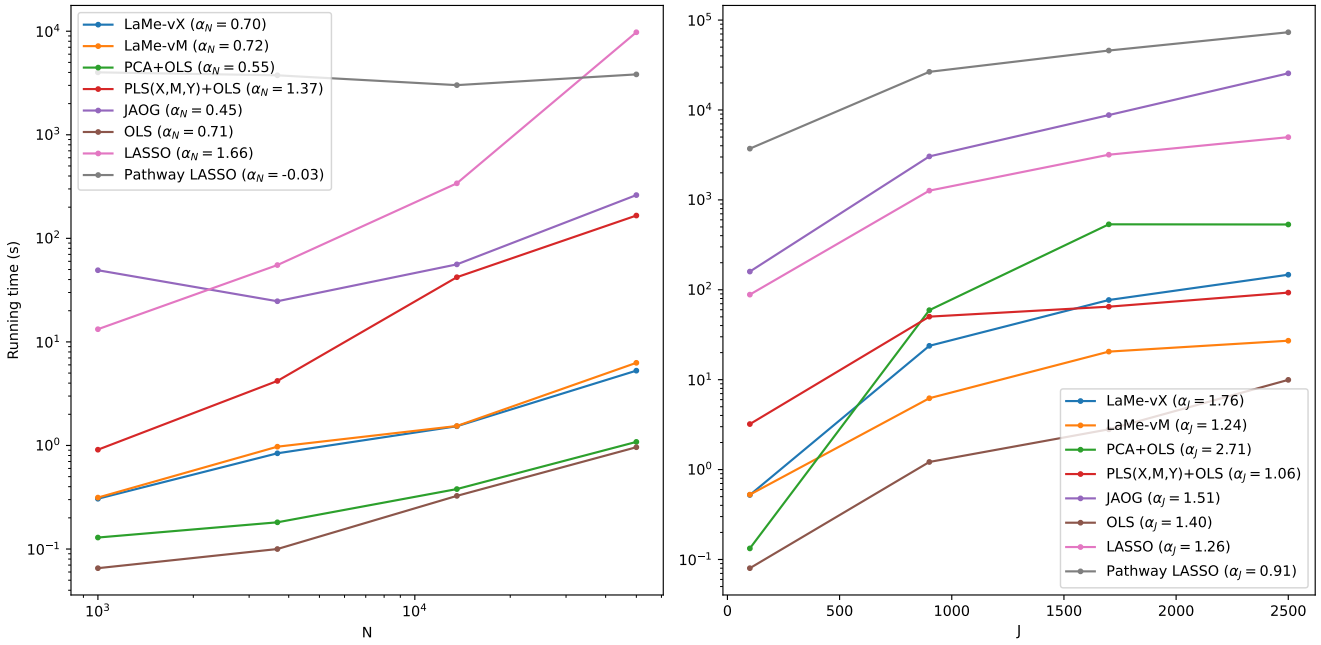


Figure 6. Scalability of the models — For a system with $I = 100$, $J = 100$, $L = 1$, $N = 3000$, with one full inference cycle per model. Each curve is normalized by its smallest entry to show how running time evolves as a function of N (left) and J (right). α denotes the power-law exponent.

In Fig. 6, we plot the normalized running time of LaMe and the baseline methods for a system with $I = 100$, $J = 100$, $L = 1$, $N = 3000$, unless specified otherwise. The models are run for a full inference procedure, meaning LaMe's EM algorithm is run 100 times, LASSO, PLS and PCA select their hyperparameters by cross-validation, etc. Refer to the main text for more details.

We see that LaMe scales similarly to other methods with respect to N, and better than most with respect to J. Pathway LASSO seems to scale well, but has the longest absolute convergence time. This figure also highlights the benefits of choosing the LaMe-vM factorization over the LaMe-vX factorization when J is large. Absolute running times must be interpreted with caution due to their dependence on experimental parameters (number of runs, number of cross-validation folds, convergence criterion, etc.). Overall, this further supports the versatility of the LaMe framework, which can handle a range of situations efficiently and accurately.

20. Correlation and identifiability issues in retrieving the observed measurement set

In (Derkach et al., 2019), the authors define the mediating set between exposure X_i and outcome Y_l as: $\mathcal{M}^{il}_{LVMA} := \{M_j | C^{LVMA}_{ijl} := \sum_k |\beta^{(xs)}_{ik}\beta^{(sm)}_{kj}\beta^{(sy)}_{kl}| \neq 0\}$, where C^{LVMA}_{ijl} is a shorthand notation. This definition, however, does not account for candidate mediators' correlation and non-identifiability, which raises interpretation issues.

Correlation matters for interpretation, because $\beta^{(sm)}_{kj} \neq 0$ in $\mathcal{M}^{il}_{LVMA}$ does not necessarily imply that M_j informs latent variables S acting as mediators between X_i and Y_l. For instance, $\beta^{(sm)}_{kj} \neq 0$ could reflect variation in M_j along a latent component associated with X_i, without implying that this component contributes to Y_l. For M_j to be interpreted as a mediator, a proper definition should therefore account for both a non-zero $\beta^{(xs)}\beta^{(sm)}$ (M varies with X through S) and a non-zero $\beta^{(ms)}\beta^{(sy)}$ (M informs Y through S); the expression of $\beta^{(ms)}$ in particular involves $\Sigma_{m|s}$, thus accounting for correlation between manifest variables.

The non-identifiability issue arises in two cases when rotational and scaling invariance exists. Scaling invariance can be removed by forcing $\Sigma_{s|x} = I$, as in (Derkach et al., 2019); rotational invariance however, remains. Rotational invariance states that $(\beta^{(xs)}, \beta^{(sm)}, \beta^{(sy)}, \theta(I))$ and $(\beta^{(xs)}R, R^T\beta^{(sm)}, R^T\beta^{(sy)}, \theta(R))$ define an equivalent reparametrization of the latent mediation model, where $\theta(R)$ denotes the set of other parameters related to S after an orthogonal rotation R. The issue arises because $\mathcal{M}^{il}_{LVMA}$ is not invariant to these likelihood-equivalent parametrizations: factorizing $C^{LVMA}_{ijl} = |\beta^{(xs)}_{i\cdot} diag(\beta^{(sm)}_{\cdot j})\beta^{(sy)}_{\cdot l}| \neq |\beta^{(xs)}_{i\cdot} R \cdot diag(R^T\beta^{(sm)}_{\cdot j}) \cdot R^T\beta^{(sy)}_{\cdot l}|$ shows that R does not cancel out. As such, mediating biomarkers can be arbitrarily added or removed from the set $\mathcal{M}^{il}_{LVMA}$ through likelihood-equivalent rotations of the latent space, hence the need for a rotationally invariant definition of the "mediating" manifest set of variables.

21. Additional synthetic experiments

We explore in Fig. 7 how models perform when there is an imbalance between the dimensionality of X and M. Our results show that models perform similarly as in the balanced case, although with a notable increase in accuracy of two-step latent approaches (PCA, PLS) for latent data. Increasing J means the mediators carry more information about the whole system, thus alleviating PCA's sequential estimation. On the other hand, lowering I reduces the amount of information flowing through the direct path $\beta^{(xy)}$, which also alleviates PLS's sequential estimation bias.

In Fig. 9, we compare models against extreme signal-to-noise ratios from S to M and Y on latent data. We use cosine similarity as a normalized measure for absolute comparison. Overall, we observe a sharp decrease in accuracy for $\beta^{(my)}$ once $SNR(Y|S) < 3\%$ whereas $\beta^{(xm)}$ estimates remain accurate for even smaller $SNR(M|S)$ –which can be explained by a greater amount of information available due to $J \gg L$. Overall, LaMe still outperforms baselines and retains its accuracy for lower values of $SNR(M, Y|S)$.

Fig. 8 shows that LaMe's performance is unaffected by the dimension of M once $SNR(M|S)$ is fixed. We use cosine similarity as a normalized measure for absolute comparison.

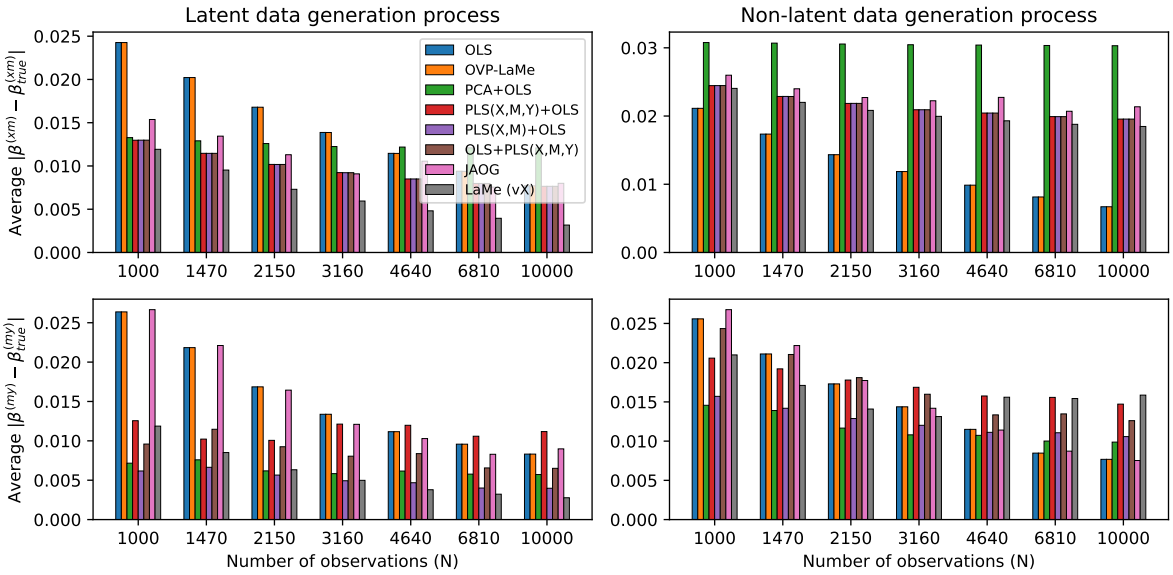


Figure 7. LaMe accurately retrieves underlying parameters — LaMe and baselines when $I = 50$, $J = 200$, $L = 1$, and a signal-to-noise ratio from S to Y and M equal to 0.1. **(Top)** Mean absolute error over all parameters in $\beta^{(xm)}_{inf}$ and $\beta^{(xm)}_{true}$; **(Bottom)** Mean absolute error over all parameters $\beta^{(my)}_{inf}$ and $\beta^{(my)}_{true}$; **(left)** data are generated using a latent process ($K = 10$) with the true K provided to all latent models ; **(right)** data are generated using a non-latent process with $K = \frac{\min(I+L,J)}{2}$ for all latent models except for OVP-LaMe where $K = \frac{I+J}{2} = 100$.

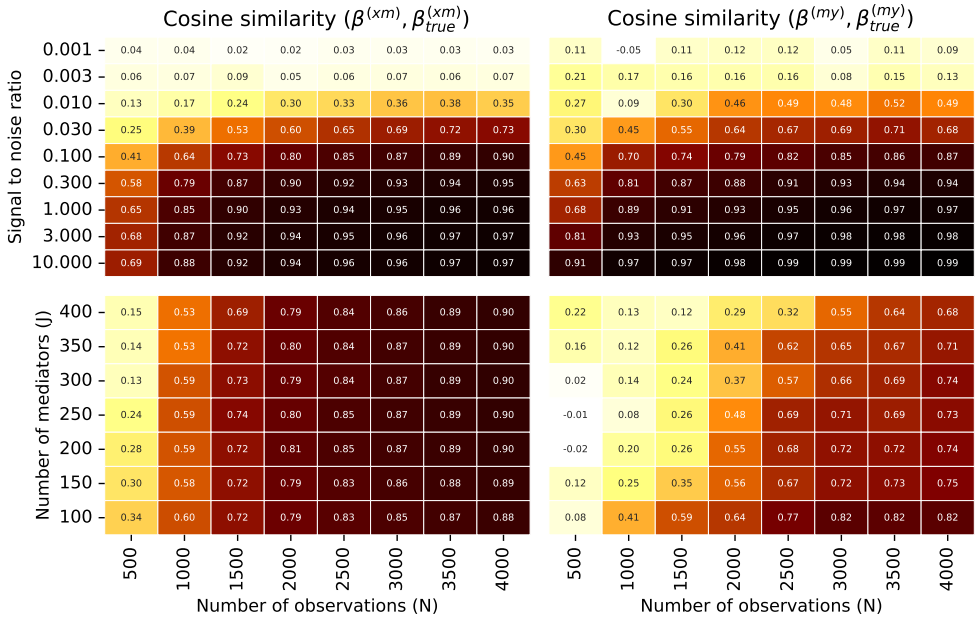


Figure 8. LaMe's parameter retrieval in challenging situations — LaMe's cosine similarity when $I = 100$, $J = 100$, $K = 10$, $L = 1$ and $SNR(Y|S) = SNR(M|S) = 0.1$, unless specified otherwise, with a varying number of observations N | (top) with respect to $SNR(M|S) = SNR(Y|S)$ | (bottom) with respect to the number of mediators J ; a dataset is generated once for $J = 400$, and models are trained on subsets of M, which can cause $SNR(Y|S), SNR(M|S)$ to fluctuate around their target value across different J | **(left)** between estimated $\beta^{(xm)}$ and $\beta^{(xm)}_{true}$ | **(right)** between estimated $\beta^{(my)}$ and $\beta^{(my)}_{true}$.

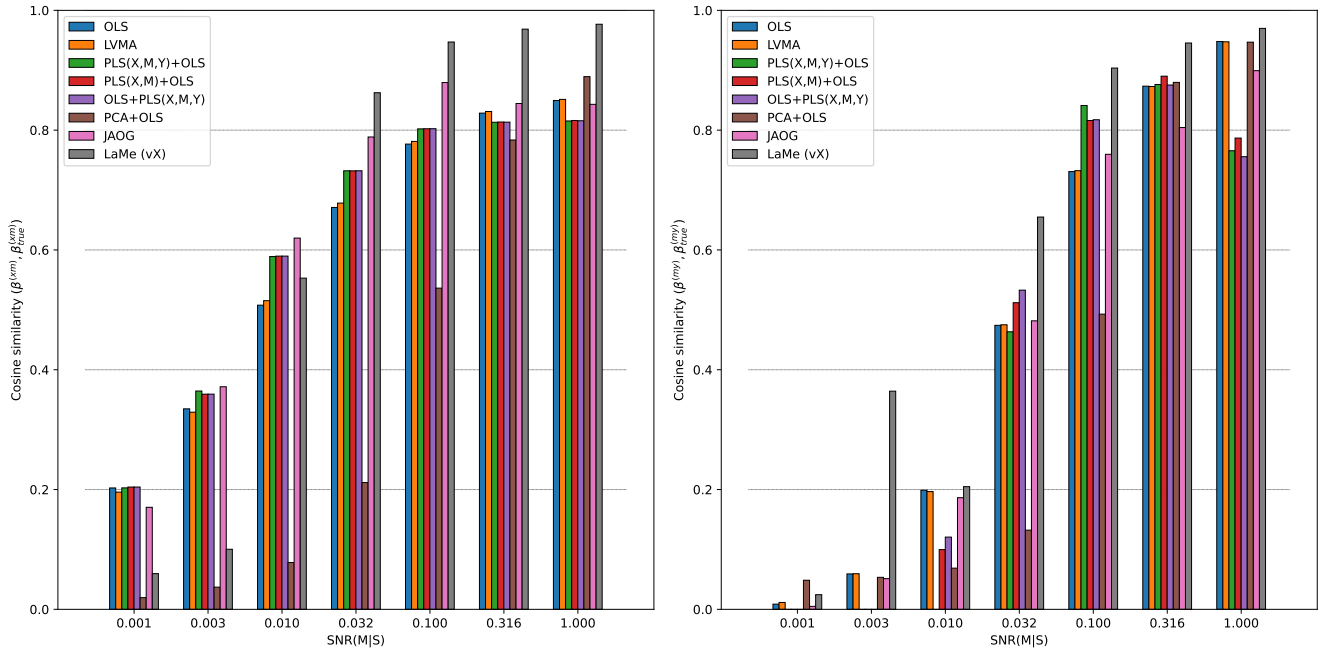


Figure 9. LaMe accurately retrieves underlying parameters over a range of signal-to-noise ratios — LaMe and baselines when $I = 100$, $J = 100$, $K = 10$, $L = 1$, and a varying signal-to-noise ratio from S to Y and M. **(left)** Cosine similarity between $\beta^{(xm)}_{inf}$ and $\beta^{(xm)}_{true}$; **(right)** Cosine similarity between $\beta^{(my)}_{inf}$ and $\beta^{(my)}_{true}$; data are generated using a latent process ($K = 10$).

22. Real-world data - Standard 5-fold CV

The results for a standard 5-fold cross-validation procedure are provided in Table 2. The data are split into 80% training data and 20% test data with no overlap between testing samples. We stress that PCA+LASSO's overall good predictive performances are consistently achieved at the cost of high-dimensional latent mediators, whereas models such as LaMe or JAOG seem to efficiently compress the relevant information into a smaller number of latent variables.

Table 2

Out-of-sample and in-sample performance metrics of LaMe (version X and M, with and without $\ell 1$ regularization, assuming full noise covariances) and all baseline models. The last lines report the average improvement of models' metrics over OLS across all datasets –hence the null line for OLS. NLPD is the Negative Log-Predictive Density, R^2 the fraction of explained variance and AIC the Akaike Information Criterion. When applicable, the number of latent variables is reported as K. Errors correspond to the standard error across five folds. The colour scale indicates relative performance within each dataset and metric ; (orange=best).

		$NLPD(Y\|X,M)\downarrow$	$R^2(Y\|X,M)\uparrow$	AIC $\downarrow$	$K\downarrow$
NHANES	LaMe (vM)	1.326 ± 0.051	0.186 ± 0.010	98.318 ± 0.320	3.000 ± 0.001
	LaMe (vM) (reg)	1.326 ± 0.052	0.186 ± 0.010	98.374 ± 0.310	3.000 ± 0.001
	LaMe (vX)	1.326 ± 0.051	0.186 ± 0.010	98.318 ± 0.320	3.000 ± 0.001
	LaMe (vX) (reg)	1.326 ± 0.052	0.187 ± 0.010	98.322 ± 0.317	3.000 ± 0.001
	PCA+OLS	1.326 ± 0.053	0.187 ± 0.012	98.809 ± 0.321	24.000 ± 0.001
	PCA+LASSO	1.320 ± 0.051	0.194 ± 0.009	98.893 ± 0.322	24.000 ± 0.001
	PLS(X,M,Y)+OLS	1.484 ± 0.023	-0.139 ± 0.070	98.610 ± 0.368	6.000 ± 0.001
	PLS(X,M)+OLS	1.325 ± 0.051	0.186 ± 0.009	98.319 ± 0.320	6.000 ± 0.001
	OVP-LaMe	1.326 ± 0.052	0.187 ± 0.012	98.612 ± 0.321	19.000 ± 0.001
	JAOG	1.325 ± 0.052	0.189 ± 0.011	99.219 ± 0.320	1.800 ± 0.179
EPIC CAN	LaMe (vM)	0.701 ± 0.007	0.762 ± 0.001	861.197 ± 0.607	4.000 ± 0.001
	LaMe (vM) (reg)	0.701 ± 0.007	0.762 ± 0.001	861.029 ± 0.495	5.000 ± 0.001
	LaMe (vX)	0.701 ± 0.007	0.762 ± 0.001	861.197 ± 0.607	4.000 ± 0.001
	LaMe (vX) (reg)	0.701 ± 0.007	0.762 ± 0.001	866.853 ± 0.463	1.000 ± 0.001
	PCA+OLS	0.747 ± 0.007	0.746 ± 0.002	933.025 ± 0.613	377.600 ± 0.219
	PCA+LASSO	0.692 ± 0.008	0.766 ± 0.001	935.368 ± 0.635	377.600 ± 0.219
	PLS(X,M,Y)+OLS	0.776 ± 0.018	0.723 ± 0.009	861.128 ± 0.611	15.000 ± 0.001
	PLS(X,M)+OLS	0.720 ± 0.009	0.752 ± 0.004	861.032 ± 0.610	15.000 ± 0.001
	OVP-LaMe	0.747 ± 0.006	0.746 ± 0.002	897.439 ± 0.602	206.000 ± 0.001
	JAOG	0.736 ± 0.013	0.750 ± 0.003	863.432 ± 0.638	3.400 ± 0.358
PISA - I	LaMe (vM)	1.066 ± 0.004	0.506 ± 0.004	102.216 ± 0.025	10.000 ± 0.001
	LaMe (vM) (reg)	1.066 ± 0.004	0.506 ± 0.004	102.217 ± 0.027	10.000 ± 0.001
	LaMe (vX)	1.066 ± 0.004	0.506 ± 0.004	102.216 ± 0.025	10.000 ± 0.001
	LaMe (vX) (reg)	1.066 ± 0.004	0.506 ± 0.004	102.232 ± 0.023	10.000 ± 0.001
	PCA+OLS	1.075 ± 0.004	0.497 ± 0.004	102.227 ± 0.025	19.000 ± 0.001
	PCA+LASSO	1.075 ± 0.004	0.497 ± 0.004	102.231 ± 0.025	19.000 ± 0.001
	PLS(X,M,Y)+OLS	1.197 ± 0.005	0.358 ± 0.007	102.496 ± 0.025	12.000 ± 0.001
	PLS(X,M)+OLS	1.189 ± 0.003	0.368 ± 0.005	102.477 ± 0.024	12.000 ± 0.001
	OVP-LaMe	1.066 ± 0.004	0.506 ± 0.004	102.213 ± 0.025	23.000 ± 0.001
	JAOG	1.162 ± 0.003	0.402 ± 0.003	103.830 ± 0.028	1.000 ± 0.001
CRIME	LaMe (vM)	0.894 ± 0.014	0.651 ± 0.014	72.937 ± 0.252	15.000 ± 0.001
	LaMe (vM) (reg)	0.896 ± 0.016	0.649 ± 0.015	75.055 ± 0.330	10.000 ± 0.001
	LaMe (vX)	0.895 ± 0.015	0.650 ± 0.015	72.937 ± 0.251	15.000 ± 0.001
	LaMe (vX) (reg)	0.898 ± 0.017	0.648 ± 0.016	74.700 ± 0.246	10.000 ± 0.001
	PCA+OLS	0.889 ± 0.015	0.655 ± 0.014	73.022 ± 0.246	32.000 ± 0.001
	PCA+LASSO	0.885 ± 0.016	0.657 ± 0.016	73.542 ± 0.243	32.000 ± 0.001
	PLS(X,M,Y)+OLS	0.927 ± 0.014	0.626 ± 0.013	76.419 ± 0.222	32.000 ± 0.001
	PLS(X,M)+OLS	0.925 ± 0.014	0.627 ± 0.013	76.401 ± 0.224	32.000 ± 0.001
	OVP-LaMe	0.889 ± 0.015	0.655 ± 0.014	75.213 ± 0.247	50.000 ± 0.001
	JAOG	0.893 ± 0.014	0.652 ± 0.015	80.429 ± 0.255	5.000 ± 0.001

23. Full mediation graphs

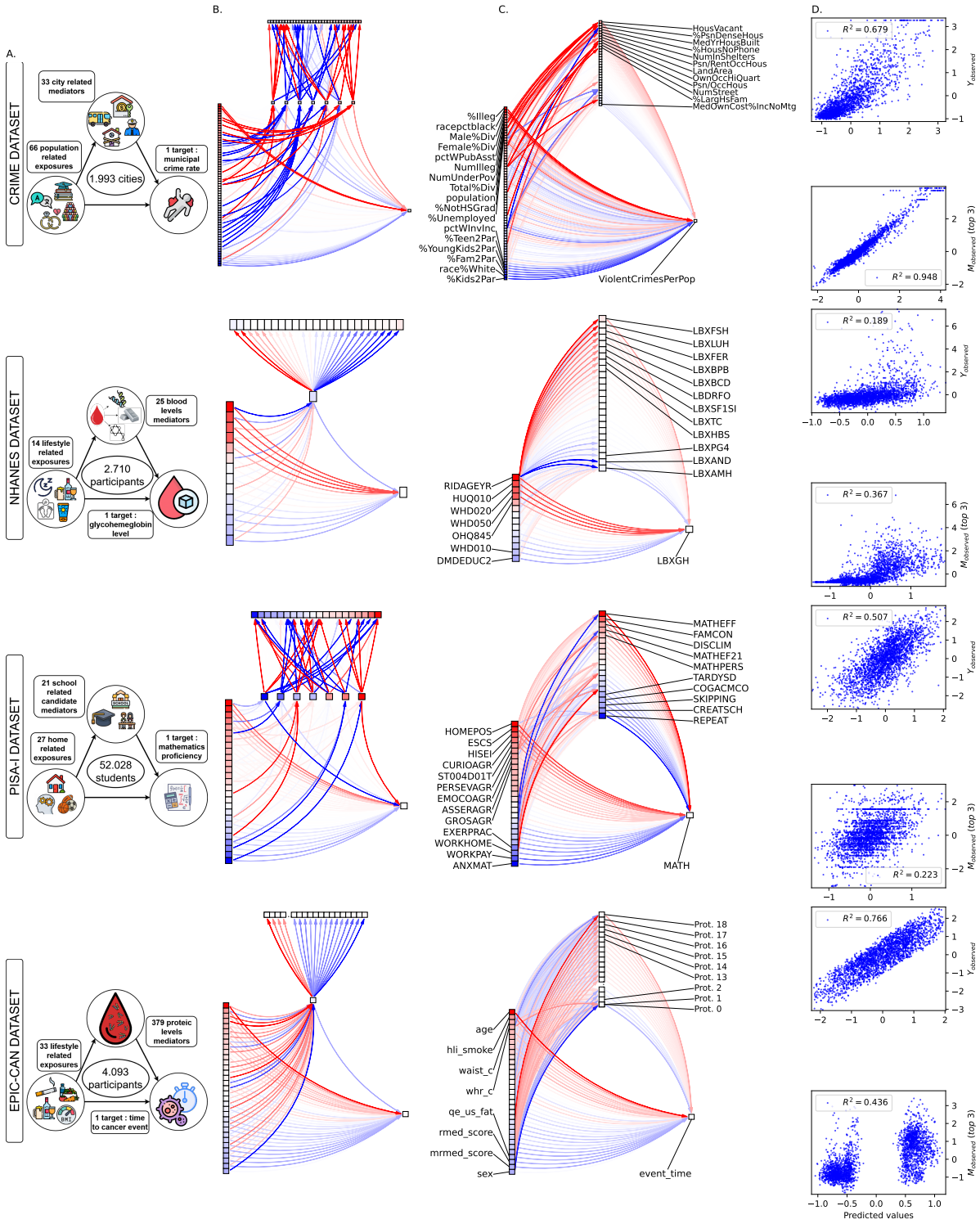


Figure 10. LaMe's results on real-world data — **(A)** Schematic illustration of the dataset and of the mediated system at hand. **(B,C)** Results from the LaMe-vX model, assuming either the causal structural model with latent mediators with K chosen for the lowest BIC (B) or without latent mediators (C). Arrows represent the standardized structural coefficients $\tilde{\beta}^{(uv)} = \Sigma_u^{\frac{1}{2}} \beta^{(uv)} \Sigma_v^{-\frac{1}{2}}$; for variables with a causal path to the outcome, squares' colour represents the total causal effect of a unit intervention on the corresponding variable on the outcome Y under the model specified with latent mediators–in this case, the sum of all causal paths leading to Y; in the latent graph, mediator colours represent marginal associations with Y, not causal effects. For both squares and arrows, blue represents a negative relation, red represents a positive relation, and colour intensity represents its strength with extremal values at ± 0.25 s.d. Annotations of (C) match those in (B). **(D)** Plot of the predicted values of Y and entries of M (the three with highest R^2) against their true values, along with the associated coefficient of determination R^2.

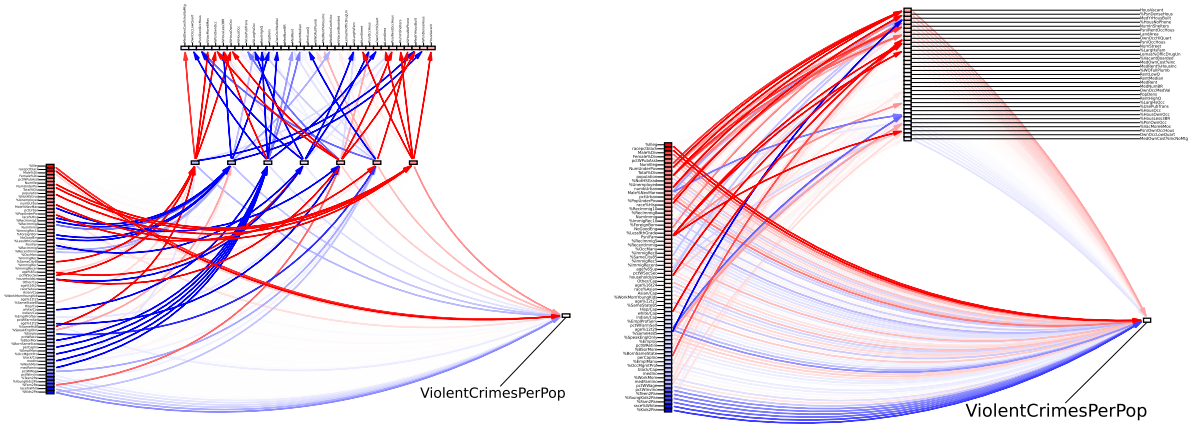


Figure 11. Full graphs for the CRIME dataset

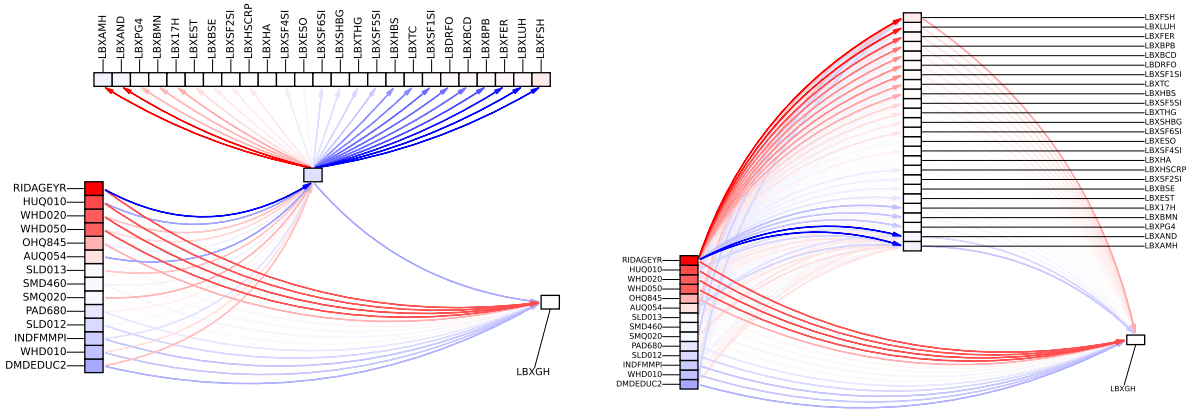


Figure 12. Full graphs for the NHANES dataset

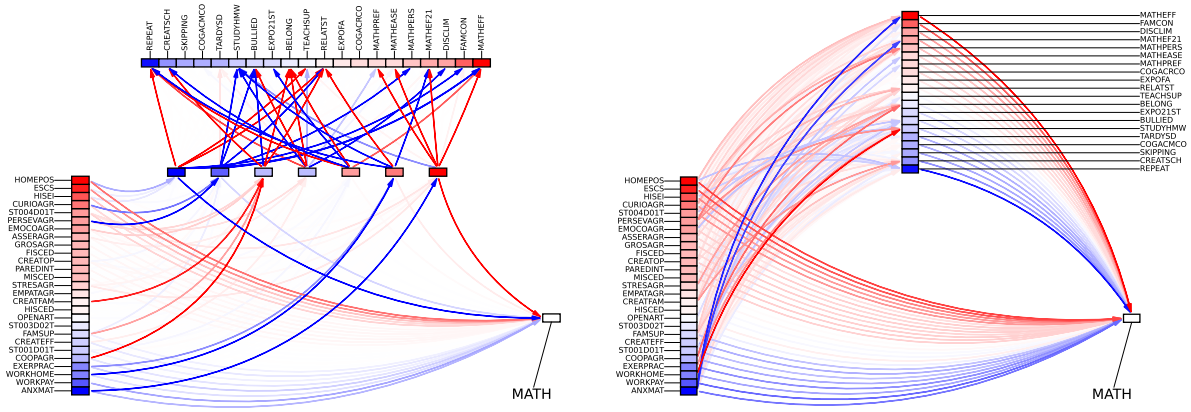


Figure 13. Full graphs for the PISA-I dataset

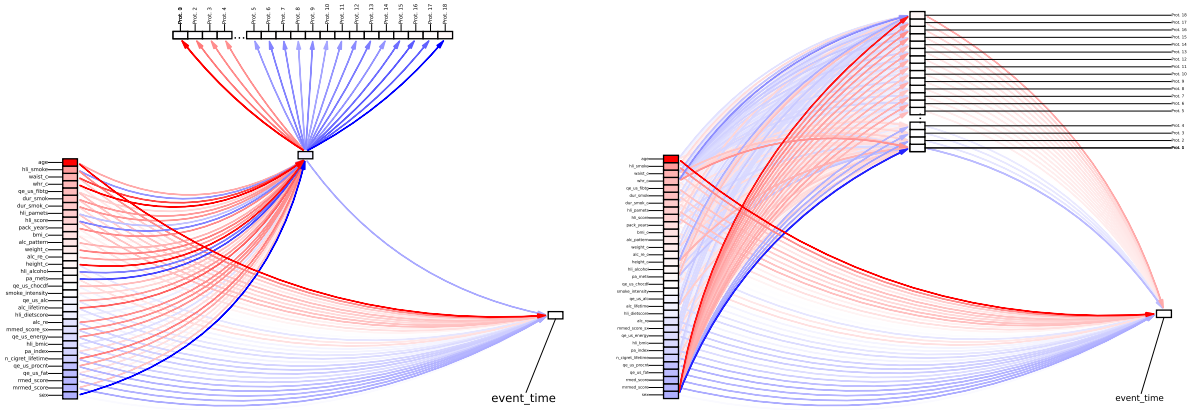


Figure 14. Full graphs for the EPIC-CAN dataset

24. Real-world data labels

List of all source variables considered in the proposed real-world datasets. Not all variables are being used in the main experiments, some being removed to avoid collinearity, reduce noise and/or due to irrelevance regarding the task at hand.

24.1 *PISA-I*

24.1.1 *X.* **ST003D02T** = Birth month 1-12 , **ST004D01T** = Gender 1-2 , **ST001D01T** = class level , **STUBMI** = Body Mass Index (BMI) , **MISCED** = Mother's level of education (ISCED) , **FISCED** = Father's level of education (ISCED) , **HISCED** = Highest level of education of parents (ISCED) , **PAREDINT** = Index highest parental education (international years of schooling scale) , **BMMJ1** = Mother's occupational status (ISEI) based on 4-digit human coded ISCO , **BFMJ2** = Father's occupational status (ISEI) based on 4-digit human coded ISCO , **HISEI** = Highest parental occupational status (ISEI) based on 4-digit human coded ISCO , **ESCS** = Index of economic, social and cultural status , **FAMSUP** = Family support (WLE) , **FAMSUPSL** = Family support for self-directed learning (WLE) , **CURSUPP** = Current parental/guardian Support (WLE) , **SOCCON** = Social Connections: Ease of Communication About Worries and Concerns (WLE) , **SOCONPA** = Social connection to Parents (WLE) , **ICTRES** = ICT Resources (WLE) , **HOMEPOS** = Home possessions (WLE) , **ICTHOME** = ICT availability outside of school (WLE) , **ICTAVHOM** = Availability and Usage of ICT at Home , **ICTQUAL** = Quality of access to ICT (WLE) , **ICTEFFIC** = Self-efficacy in digital competencies (WLE) , **ICTWKDY** = Frequency of ICT activity (weekday) (WLE) , **ICTWKEND** = Frequency of ICT activity (weekend) (WLE) , **PARINVOL** = Parental Involvement (WLE) , **PQMIMP** = Parent Attitudes Toward Mathematics (WLE) , **PQMCAR** = Mathematics Career (WLE) , **PAREXPT** = Parent's expectations in child's career , **ATTIMMP** = Parents' Attitudes towards Immigrants (WLE) , **EXERPRAC** = Exercise or practice a sport before or after school , **WORKPAY** = Working for pay before or after school , **WORKHOME** = Working in household/take care of family members before or after school , **PERSEVAGR** = Perseverance (agreement) (WLE) , **CURIOAGR** = Curiosity (agreement) (WLE) , **COOPAGR** = Cooperation (agreement) (WLE) , **EMPATAGR** = Empathy (agreement) (WLE) , **ASSERAGR** = Assertiveness (agreement) (WLE) , **STRESAGR** = Stress resistance (agreement) (WLE) , **EMOCOAGR** = Emotional control (agreement) (WLE) , **GROSAGR** = Growth Mindset (WLE) , **INFOSEEK** = Information seeking about future careers (WLE) , **PROBSELF** = Problems with self-directed learning (WLE) , **FEELLAH** = Feelings about learning at home (WLE) , **SDLEFF** = Self-directed learning self-efficacy (WLE) , **ACCESSFP** = Access to money and financial products, sources of money (WLE) , **FLFAMILY** = Parental involvement in matters of FL (WLE) , **FLCONFIN** = Confidence about financial matters (WLE) , **FLCONICT** = Confidence about financial matters using digital devices (WLE) , **ACCESSFA** = Access to money and financial products, financial activities (WLE) , **ATTCONFM** = Attitudes towards and confidence about financial matters (WLE) , **FRINFLFM** = Friends' influence on financial matters (WLE) , **FCFMLRTY** = Familiarity with concepts of finance , **OPENART** = Openness to Art and Reflection TBD (WLE) , **IMAGINE** = Imagination and Adventurousness TBD (WLE) , **CREATHME** = Creative Home Environment (WLE) , **CREATACT** = Participation in creative activities outside of school (WLE) , **CREATOPN** = Creativity and Openness to Intellect (WLE) , **CREATOR** = Openness to creativity: Other's report (WLE) , **CREATEFF** = Creative self-efficacy (WLE) , **CREATFAM** = Creative peers and family environment (WLE) , **CREATOOS** = Creative Activities outside of school (WLE) , **CREATOP** = Creativity and Openness to Intellect TBD (WLE) , **ICTDISTR** = Distress from online content and cyberbullying , **ANXMAT** = Mathematics Anxiety (WLE) , **BODYIMA** = Body Image (WLE) , **LIFESAT** = Students' Life Satisfaction across Domains (WLE) , **PSYCHSYM** = Psychosomatic Symptoms (WLE) , **EXPWB** = Experienced Well-being (Previous Day) (WLE)

24.1.2 *M.* **REPEAT** = Grade repetition , **MISSSC** = Missing school for more than 3 months , **SKIPPING** = Skipping classes or days of school , **TARDYSD** = Arriving late for school stricter definition , **DURECEC** = Duration in early childhood education and care , **STUDYHMW** = Studying for school or homework before or after school , **EXPECEDU** = Highest expected educational level , **RELATST** = Quality of student-teacher relationships (WLE) , **BELONG** = Sense of belonging (WLE) , **BULLIED** = Being bullied (WLE) , **FEELSAFE** = Feeling safe (WLE) , **SCHRISK** = School safety risks (WLE) , **SCHSUST** = School actions/activities to sustain learning (WLE) , **LEARRES** = Types of learning resources used while school was closed (WLE) , **ICTSCH** = ICT availability at school (WLE) , **PQSCHOOL** = School quality (WLE) , **ICTAVSCH** = Availability and Usage of ICT at School , **PASCHPOL** = School Policies for Parental Involvement (WLE) , **MATHPREF** = Prefers maths , **MATHEASE** = Ease at maths , **MATHMOT** = Motivation in maths , **CREATSCH** = Creative school and class environment (WLE) , **CREATAS** = Creative Activities at school (WLE) , **ICTSUBJ** = Subject-related ICT Use During Lessons (WLE) , **ICTENQ** = Use of ICT in enquiry-based learning activities (WLE) , **ICTFEED** = Support or feedback via ICT (WLE) , **ICTOUT** = Use of ICT for school activities outside of the classroom (WLE) , **ICTREG** = Views of regulated ICT use in school (WLE) , **ICTINFO** = Students' practices regarding online information (WLE) , **FLSCHOOL** = Financial education in school lessons (WLE) , **FLMULTSB** = Financial education in school lessons, multiple subjects (WLE) , **COGACRCO** = Cognitive activation in mathematics: Foster reasoning Version B (WLE) , **COGACMCO** = Cognitive activation in mathematics: Encourage mathematical thinking Version B (WLE) , **EXPOFA** = Exposure to Formal and Applied Mathematics Tasks (WLE) , **EXPO21ST** = Exposure to Mathematical Reasoning and 21st century mathematics tasks (WLE) , **MATHEFF** = Mathematics self-efficacy: formal and applied mathematics - response options reversed in 2022 (WLE) , **MATHEF21** = Mathematics self-efficacy: mathematical reasoning and 21st century skills (WLE) , **FAMCON** = Subjective familiarity with mathematics concepts (WLE) , **MATHPERS** = Effort and Persistence in Mathematics (WLE) , **TEACHSUP** = Mathematics Teacher Support (WLE) , **DISCLIM** = Disciplinary climate in mathematics (WLE)

24.1.3 *Y.* **MATH** = Mathematics proficiency score

24.2 *NHANES*

24.2.1 *X.* **RIAGENDR** = Gender , **RIDAGEYR** = Age at screening , **DMDEDUC2** = Education level [1-5] , **AUQ054** = General condition of hearing (audiometry) [1-6] from [excellent-deaf] , **ALQ130** = Avg number alcoholic drinks/day/past 12 mos [1-15] , **ALQ280** = Number times 8+ drinks in 1 day/past 12 mos , **DEQ034A** = Stay in the shade when sunny outside [1-5] from [always-never] in the shadows , **DEQ034C** = Wear a long sleeved shirt [1-5] from [always-never] , **DEQ034D** = Use sunscreen [1-5] from [always-never] , **ECD070A** = Weight at birth, pounds , **HUQ010** = General health condition [1-5] from [good-poor] , **HOD051** = Number of rooms in home , **INDFMMPI** = Family monthly poverty level index , **KIQ481** = How many times wake up and urinate? , **OCQ180** = Number of hours worked in the last week , **OHQ845** = Rate the health of your teeth and gums [1-5] from [good-poor] , **PAD680** = Minutes sedentary activity/day (sitting at school, work, bus, etc) , **RXQ050** = Number of prescription medicines taken last 30 days [1-5] , **SLD012** = Sleep hours - weekdays or workdays , **SLD013** = Sleep hours - weekends , **SMQ020** = Smoked at least 100 cigarettes in life , **SMD460** = Number of people who live here smoke tobacco? [0-2] , **WHD010** = Current self-reported height (inches) [40-83] , **WHD020** = Current self-reported weight (pounds) [60-530] , **WHD050** = Self-reported weight - 1 yr ago (pounds) [60-530] , **FNQ450** = >17yo - Difficulty communicating [1-4] from [no difficulty-cannot at all] , **FNQ460** = >17yo - Difficulty remembering things/concentrating [1-4] from [no difficulty-cannot at all] , **FNQ530** = >17yo - How often very sad/depressed [1-5] from [daily-never]

24.2.2 *M.* **LBXAGP** = alpha-1-acid glycoprotein (g/L) , **LBXTC** = Total Cholesterol (mg/dL) , **LBXFER** = Ferritin(ng/mL) , **LBDRFO** = RBC folate (ng/mL) , **LBXHSCRP** = HS C-Reactive Protein (mg/L) - Linked to risk of CVD , **LBXIN** = Insulin (uU/mL) , **LBXBPB** = Blood lead (ug/dL) , **LBXBCD** = Blood cadmium (ug/L) , **LBXTHG** = Blood mercury, total (ug/L) , **LBXBSE** = Blood selenium (ug/L) , **LBXBMN** = Blood manganese (ug/L) , **LBXSF1SI** = 5-Methyl-tetrahydrofolate (nmol/L) , **LBXSF2SI** = Folic acid (nmol/L) , **LBXSF3SI** = 5-Formyl-tetrahydrofolate (nmol/L) , **LBXSF4SI** = Tetrahydrofolate (nmol/L) , **LBXSF5SI** = 5,10-Methenyl-tetrahydrofolate (nmol/L) , **LBXSF6SI** = Mefox oxidation product (nmol/L) , **LBX17H** = 17alpha-hydroxyprogesterone (ng/dL) , **LBXAND** = Androstenedione (ng/dL) , **LBXAMH** = Anti-Mullerian hormone (ng/mL) , **LBXDHE** = DHEAS (µg/dL) , **LBXEST** = Estradiol (pg/mL) , **LBXESO** = Estrone (ng/dL) , **LBXFSH** = Follicle Stimulating Hormone (mIU/mL) , **LBXLUH** = Luteinizing Hormone (mIU/mL) , **LBXPG4** = Progesterone (ng/dL) , **LBXTST** = Testosterone, total (ng/dL) , **LBXSHBG** = SHBG (nmol/L) , **LBXTFR** = Transferrin receptor (mg/L) , **LBXVIDMS** = 25OHD2+25OHD3 (nmol/L) , **LBXHA** = Hepatitis A antibody [1,2] for [positive,negative] , **LBXHBS** = Hepatitis B surface antibody [1,2] for [positive,negative] , **LBXGLU** = Fasting Blood/Plasma Glucose (mg/dL)

24.2.3 *Y.* **LBXGH** = Glycohemoglobin (%) - Glucose in blood (%)

24.3 *EPIC-CAN*

Acronyms: whr: waist to height ratio, rmed: Mediterranean diet, alc: alcohol, _c: category, pa: physical activity, aggr: aggregated, qe_us: USDA, hli: Healthy Lifestyle Index

24.3.1 *X.* **age** , **sex** , **whr_c** , **height_c** , **weight_c** , **waist_c** , **bmi_c** , **l_school** , **mrmed_score** , **mmed_score_sx** , **rmed_score** , **mrmed_score** , **pa_mets** , **pa_index** , **alc_re** , **alc_re_c** , **alc_lifetime** , **alc_pattern** , **dur_smok** , **dur_smok_c** , **n_cigret_aggr** , **n_cigret_lifetime** , **smoke_intensity** , **pack_years** , **qe_us_alc** , **qe_us_chocdf** , **qe_us_energy** , **qe_us_fat** , **qe_us_fibtg** , **qe_us_procnt** , **hli_dietscore** , **hli_pamets** , **hli_smoke** , **hli_alcohol** , **hli_bmic** , **hli_score** , **systol** , **diastol**

24.3.2 *M.* Protein identifiers

24.3.3 *Y.* Age at cancer event

24.4 *CRIME*

24.4.1 *X.* **population** population for community , **householdsize** mean people per household , **racepctblack** percentage of population that is african american , **racePctWhite** percentage of population that is caucasian , **racePctAsian** percentage of population that is of asian heritage , **racePctHisp** percentage of population that is of hispanic heritage , **agePct12t21** percentage of population that is 12-21 in age , **agePct12t29** percentage of population that is 12-29 in age , **agePct16t24** percentage of population that is 16-24 in age , **agePct65up** percentage of population that is 65 and over in age , **numbUrban** number of people living in areas classified as urban , **pctUrban** percentage of people living in areas classified as urban , **PctForeignBorn** percent of people foreign born , **PctBornSameState** percent of people born in the same state as currently living , **PctSameHouse85** percent of people living in the same house as in 1985 (5 years before) , **PctSameCity85** percent of people living in the same city as in 1985 (5 years before) , **PctSameState85** percent of people living in the same state as in 1985 (5 years before) , **MalePctDivorce** percentage of males who are divorced , **MalePctNevMarr** percentage of males who have never married , **FemalePctDiv** percentage of females who are divorced , **TotalPctDiv** percentage of population who are divorced , **PersPerFam** mean number of people per family , **PctFam2Par** percentage of families (with kids) that are headed by two parents , **PctKids2Par** percentage of kids in family housing with two parents , **PctYoungKids2Par** percent of kids 4 and under in two parent households , **PctTeen2Par** percent of kids age 12-17 in two parent households , **PctWorkMomYoungKids** percentage of moms of kids 6 and under in labor force , **PctWorkMom** percentage of moms of kids under 18 in labor force , **NumIlleg** number of kids born to never married , **PctIlleg** percentage of kids born to never married , **NumImmig** total number of people known to be foreign born , **PctImmigRecent** percentage of immigrants who immigrated within last 3 years , **PctImmigRec5** percentage of immigrants who immigated within last 5 years , **PctImmigRec8** percentage of immigrants who immigated within last 8 years , **PctImmigRec10** percentage of immigrants who immigated within last 10 years , **PctRecentImmig** percent of population who have immigrated within the last 3 years , **PctRecImmig5** percent of population who have immigrated within the last 5 years , **PctRecImmig8** percent of population who have immigrated within the last 8 years , **PctRecImmig10** percent of population who have immigrated within the last 10 years , **PctSpeakEnglOnly** percent of people who speak only English , **PctNotSpeakEnglWell** percent of people who do not speak English well , **PctLess9thGrade** percentage of people 25 and over with less than a 9th grade education , **PctNotHSGrad** percentage of people 25 and over that are not high school graduates , **PctBSorMore** percentage of people 25 and over with a bachelor's degree or higher education , **medIncome** median household income , **pctWWage** percentage of households with wage or salary income in 1989 , **pctWFarmSelf** percentage of households with farm or self employment income in 1989 , **pctWInvInc** percentage of households with investment / rent income in 1989 , **pctWSocSec** percentage of households with social security income in 1989 , **pctWPubAsst** percentage of households with public assistance income in 1989 , **pctWRetire** percentage of households with retirement income in 1989 , **medFamInc** median family income (differs from household income for non-family households) , **perCapInc** per capita income , **whitePerCap** per capita income for caucasians , **blackPerCap** per capita income for african americans , **indianPerCap** per capita income for native americans , **AsianPerCap** per capita income for people with asian heritage , **OtherPerCap** per capita income for people with 'other' heritage , **HispPerCap** per capita income for people with hispanic heritage , **NumUnderPov** number of people under the poverty level , **PctPopUnderPov** percentage of people under the poverty level , **PctUnemployed** percentage of people 16 and over, in the labor force, and unemployed , **PctEmploy** percentage of people 16 and over who are employed , **PctEmplManu** percentage of people 16 and over who are employed in manufacturing , **PctEmplProfServ** percentage of people 16 and over who are employed in professional services , **PctOccupManu** percentage of people 16 and over who are employed in manufacturing , **PctOccupMgmtProf** percentage of people 16 and over who are employed in management or professional occupations

24.4.2 *M.* **LemasSwornFT** number of sworn full time police officers , **LemasSwFTPerPop** sworn full time police officers per 100K population , **LemasSwFTFieldOps** number of sworn full time police officers in field operations (on the street as opposed to administrative etc) , **LemasSwFTFieldPerPop** sworn full time police officers in field operations (on the street

as opposed to administrative etc) per 100K population , **LemasTotalReq** total requests for police , **LemasTotReqPerPop** total requests for police per 100K population , **PolicReqPerOffic** total requests for police per police officer , **PolicPerPop** police officers per 100K population , **RacialMatchCommPol** a measure of the racial match between the community and the police force. High values indicate proportions in community and police force are similar , **PctPolicWhite** percent of police that are caucasian , **PctPolicBlack** percent of police that are african american , **PctPolicHisp** percent of police that are hispanic , **PctPolicAsian** percent of police that are asian , **PctPolicMinor** percent of police that are minority of any kind , **OfficAssgnDrugUnits** number of officers assigned to special drug units , **NumKindsDrugsSeiz** number of different kinds of drugs seized , **PolicAveOTWorked** police average overtime worked , **PolicCars** number of police cars , **PolicOperBudg** police operating budget , **LemasPctPolicOnPatr** percent of sworn full time police officers on patrol , **LemasGangUnitDeploy** gang unit deployed (numeric - decimal - but really ordinal - 0 means NO, 1 means YES, 0.5 means Part Time) , **LemasPctOfficDrugUn** percent of officers assigned to drug units , **PolicBudgPerPop** police operating budget per population , **PctLargHouseFam** percent of family households that are large (6 or more) , **PctLargHouseOccup** percent of all occupied households that are large (6 or more people) , **PersPerOccupHous** mean persons per household , **PersPerOwnOccHous** mean persons per owner occupied household , **PersPerRentOccHous** mean persons per rental household , **PctPersOwnOccup** percent of people in owner occupied households , **PctPersDenseHous** percent of persons in dense housing (more than 1 person per room) , **PctHousLess3BR** percent of housing units with less than 3 bedrooms , **MedNumBR** median number of bedrooms , **HousVacant** number of vacant households , **PctHousOccup** percent of housing occupied , **PctHousOwnOcc** percent of households owner occupied , **PctVacantBoarded** percent of vacant housing that is boarded up , **PctVacMore6Mos** percent of vacant housing that has been vacant more than 6 months , **MedYrHousBuilt** median year housing units built , **PctHousNoPhone** percent of occupied housing units without phone (in 1990, this was rare!) , **PctWOFullPlumb** percent of housing without complete plumbing facilities , **OwnOccLowQuart** owner occupied housing - lower quartile value , **OwnOccMedVal** owner occupied housing - median value , **OwnOccHiQuart** owner occupied housing - upper quartile value , **RentLowQ** rental housing - lower quartile rent , **RentMedian** rental housing - median rent (Census variable H32B from file STF1A) , **RentHighQ** rental housing - upper quartile rent , **MedRent** median gross rent (Census variable H43A from file STF3A - includes utilities) , **MedRentPctHousInc** median gross rent as a percentage of household income , **MedOwnCostPctInc** median owners cost as a percentage of household income - for owners with a mortgage , **MedOwnCostPctIncNoMtg** median owners cost as a percentage of household income - for owners without a mortgage , **NumInShelters** number of people in homeless shelters , **NumStreet** number of homeless people counted in the street , **LandArea** land area in square miles , **PopDens** population density in persons per square mile , **PctUsePubTrans** percent of people using public transit for commuting

24.4.3 *Y.* **ViolentCrimesPerPop** total number of violent crimes per 100K population GOAL attribute (to be predicted)