Granger mediation analysis of multiple time series with an application to functional magnetic resonance imaging.

Zhao, Y., Luo, X. (2019). Biometrics, 75(3), 788–798.

Yiqiao Jin

Supervisor: Bo Fu

Fudan University School of Data Science

2023.10.20

Rough Overview

1. Experiment
   - Functional magnetic resonance imaging (fMRI) experiment.
   - To study how the effect of a randomized stimulus on the Blood-oxygen-level dependent (BOLD) activity of one outcome region is mediated by the activity of the other mediator region.
   - To quantify the causal effects through a brain pathway, while capturing the dynamic dependence between two brain regions.
2. Granger mediation analysis
   - A new framework for causal mediation analysis in multiple time-series settings.
   - Vector autoregressive (VAR) models across the temporal correlations.
   - Individual variability and correlated errors between the mediator and the outcome variables.

Functional magnetic resonance imaging experiment

1. Experiment
   - Each participant performs a motor conflict task, responding to randomized STOP/GO experimental stimuli in a sequence of trials. Participants are instructed to press buttons when seeing the GO stimulus, and to withhold from pressing under the STOP stimulus.
   - Brain activities are measured by fMRI using the blood-oxygen level dependent (BOLD) contrast. The GO stimulus, compared with STOP, will increase brain activation in the primary motor cortex (M1), a brain region responsible for finger movements.
   - The presupplementary motor area (preSMA), was hypothesized to be one of the primary areas for processing the stimuli and mediating the M1 response.
2. Question
   - Causal mediation analysis with time series data.
   - Unclear to what extent preSMA mediates the stimulus effect on M1.

An example of fMRI experiment

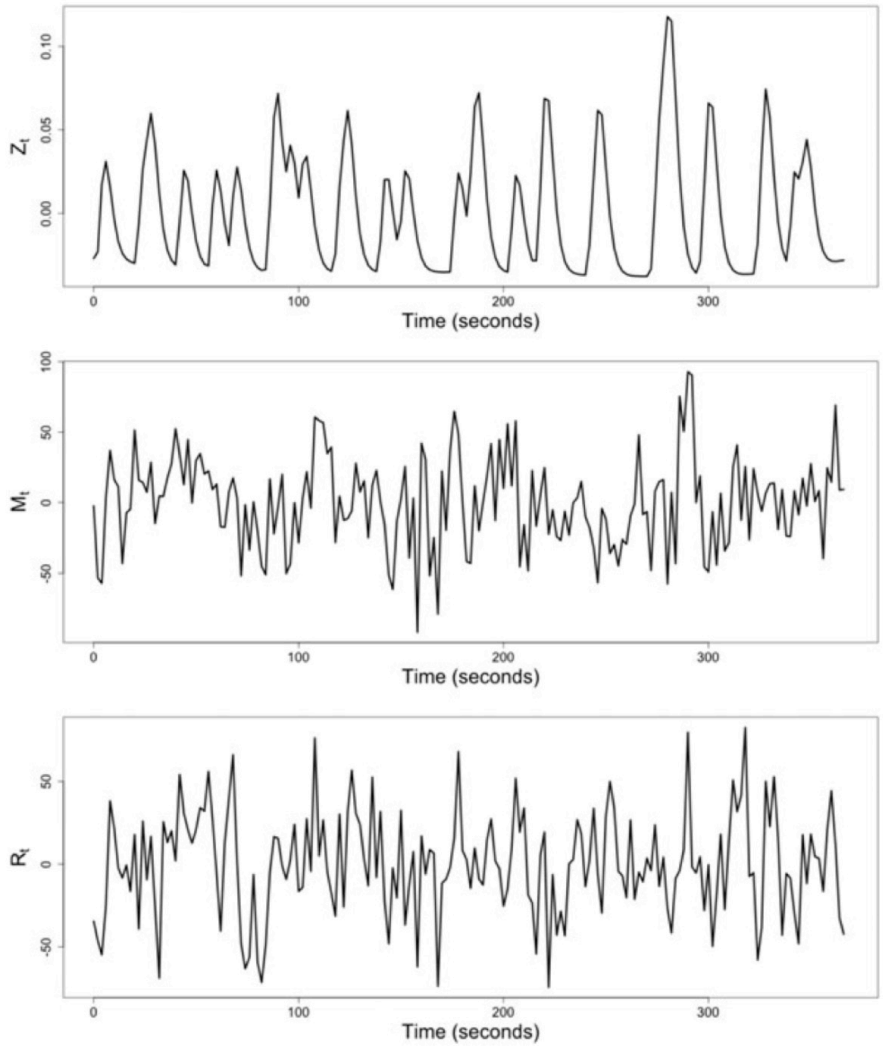

Figure: An example of time-series data from one participant

Notation

- $R =$ outcome of interest for each individual
- $Z =$ exposure or treatment of interest for each individual
- $M =$ post-treatment intermediate(s) for each individual (potentially on the pathway between A and Y)
- $C =$ set of baseline covariates for each individual that are not affected by the exposure
- $R_{zm} =$ counterfactual outcome R for each indiviual when intervening to set Z to z and M to m
- $M_z =$ counterfactual post-treatment intermediate(s) M for each indiviual when intervening to set Z to z
- Total effect decomposition

$$\begin{aligned}\textbf{TE} &= \mathbb{E}[R_1] - \mathbb{E}[R_0] \\ &= \{\mathbb{E}[R_1] - \mathbb{E}[R_{1M_0}]\} + \{\mathbb{E}[R_{1M_0}] - \mathbb{E}[R_0]\} \\ &= \textbf{NIE} + \textbf{NDE}\end{aligned}$$

Natural (in)direct effects

Let $s_t = (s_{1t}, s_{2t})$ be the stimulus assignment at time t, and $\bar{s}_t = (s_1, \cdots, s_t)$ be the historical stimulus assignment up to time t, where $s_{qt} = 1$ if stimulus q is applied at time t and 0 otherwise for $t = 1, \cdots, T$ equally spaced time periods, and only two randomized stimuli: $q = 1$ (GO) and $q = 2$ (STOP). Consider the following model

$$M_t(\bar{s}_t) = Z_t(\bar{s}_t)\alpha + \epsilon_t^{(M)}(\bar{s}_t), \tag{1}$$

$$R_t(\bar{s}_t, M_t(\bar{s}_t^*)) = Z_t(\bar{s}_t)\gamma + M_t(\bar{s}_t^*)\beta + \epsilon_t^{(R)}(\bar{s}_t, \bar{s}_t^*), \tag{2}$$

Following the standard causal mediation definitions we have

$$\mathbf{TE}(\bar{s}_t, \bar{s}_t^*) = \{Z_t(\bar{s}_t) - Z_t(\bar{s}_t^*)\}(\gamma + \alpha\beta), \tag{3}$$

$$\mathbf{NDE}(\bar{s}_t, \bar{s}_t^*) = \{Z_t(\bar{s}_t) - Z_t(\bar{s}_t^*)\}\gamma, \tag{4}$$

$$\mathbf{NIE}(\bar{s}_t, \bar{s}_t^*) = \{Z_t(\bar{s}_t) - Z_t(\bar{s}_t^*)\}\alpha\beta. \tag{5}$$

Difficulties

i. Widely available neuroimaging analysis tools are not apropos, because they usually analyze either the stimulus activationd or the connectivity (correlations) between region.

ii. Specialized causal mediation analyses for fMRI to infer stimulus effects on BOLD responses remain limitations.

- Lindquist (2012) proposed a functional mediation model with fMRI mediators and a scalar outcome.
- Zhao and Luo (2014) proposed a multilevel causal mediation framework that addresses the issues related to unmeasured confounding and individual variation, but did not directly model the temporal dependence in fMRI time series.
- Chén et al. (2017) recently proposed multiple mediator models with a scalar behavioral outcome where none of the mediators is modeled as time series.
- ...

Difficulties

iii. Standard causal mediation methods usually impose strong assumptions.

(1) Sequential ignorability assumption

The most used and flexible version is given by Imai et al.(2010)

$$\{R_i(z,m), M_i(z^*)\} \perp\!\!\!\perp Z_i \mid C_i = c, \tag{6}$$

$$R_i(z,m) \perp\!\!\!\perp M_i(z^*) \mid Z_i = z^*, C_i = c, \tag{7}$$

where $z = 1, z^* = 0$ and m, c in the support of M, C and

$$Pr(Z_i = z \mid M_i = m, C_i = c) > 0. \tag{8}$$

Equivalently,

- the effect of Z on R is unconfounded conditional on C,
- the effect of Z on M is unconfounded conditional on C,
- the effect of M on R is unconfounded conditional on (C, Z).

However, as Imai et al. (2010) wrote, "the proposed assumption may be too strong for the typical situations", especially when an mediator-outcome confounder U exists.

Difficulties

(2) Mediator-outcome confounder U
U will introdue a residual correlation between M and R. Particularly, if there do exist an exposure-induced mediator-outcome confounder L, then natural (in)direct effects will not be identified from data (Avin et al., 2005) without an additional assumption as equation (4)

$$R_i(z, m) \perp\!\!\!\perp M_i(z^*) \mid C_i = c. \tag{9}$$

(3) Independent observation assumption(unrealistic in time-series data)
For example, the estimation of the ACME is straightforward since the error terms are independent of each other under the LSEM framework given by the sequential ignorability and equations below (Baron and Kenny, 1986)

$$M_i = \alpha_1 + \beta_1 Z_i + \epsilon_{i1}, \tag{10}$$

$$R_i = \alpha_2 + \beta_2 Z_i + \gamma M_i + \epsilon_{i2} \tag{11}$$

A conceptual diagram for the time-series data

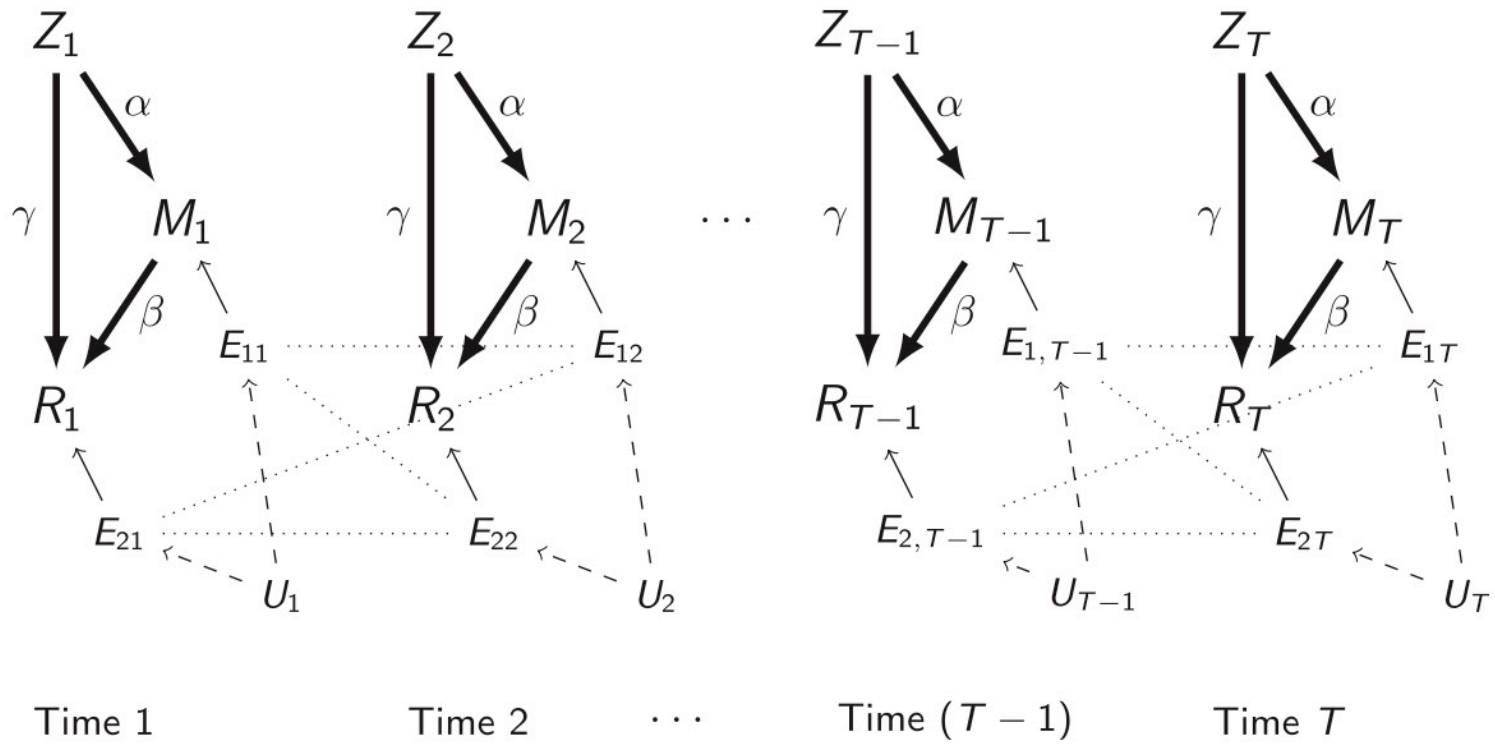

Figure: At each time point, the bold arrows between the convolved stimulus time series $\mathrm{Z_t}$, the mediator $\mathrm{M_t}$, and the outcome $\mathrm{R_t}$ depict the causal mediation mechanism we aim to study. $\mathrm{E_{1t}}$ and $\mathrm{E_{2t}}$ are random autoregressive errors. An unmeasured confounding variable $\mathrm{U_t}$ influences both errors. Dotted lines represent the autoregressive dependence in the model.

Granger mediation model

For the observed data $(Z_t, M_t, R_t), t = 1, \cdots, T$, the following LSEM model is proposed after all variables are centered

$$M_t = Z_t\alpha + E_{1t}, \tag{12}$$

$$R_t = Z_t\gamma + M_t\beta + E_{2t}, \tag{13}$$

where E_{1t} and E_{2t} are two zero-mean error processes.

To account for the spatiotemporal dependence between the two error processes, E_{1t} and E_{2t} are assumed to follow a VAR model of order p

$$E_{1t} = \sum_{j=1}^{p}(\omega_{11_j}E_{1,t-j} + \omega_{21_j}E_{2,t-j}) + \epsilon_{1t}, \tag{14}$$

$$E_{2t} = \sum_{j=1}^{p}(\omega_{12_j}E_{1,t-j} + \omega_{22_j}E_{2,t-j}) + \epsilon_{2t}, \tag{15}$$

Granger mediation model

where the error vector $(\epsilon_{1t}, \epsilon_{2t})^T$ is presumed to be a bivariate Gaussian white nosie process as

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{16}$$

- $(\epsilon_{1t}, \epsilon_{2t})^T$ is independent of $(\epsilon_{1u}, \epsilon_{2u})^T$ for $t \neq u$,
- $\mathbf{\Sigma}$ and $\mathrm{VAR}(P)$ are introduced for the spatiotemporal correlations, where p is usually small (1 or 2) for fMRI data,
- The correlation parameter δ is introduced to model the instantaneous mediator-outcome dependence, meanwhile, it can be interpreted as the magnitude of the unmeasured confounding effect as an sensitivity parameter of assumption (9).

Identification assumptions

(A1) The exposure randomization regime is the same across time and participants.

(A2) Models are correctly specified with no exposure-mediator interaction.

(A3) Consistency assumption.

(A4) The exposure assignment is fully random across time.

(A5) The causal parameters are time-invariant in the model (12) and (13).

(A6) The time-invariant covariance matrix of the Gaussian errors in the model (1) and (2) is not affected by the exposure assignments

$$\mathrm{Cov}\{\epsilon_t^{(M)}(\bar{s}_t^*), \epsilon_t^{(R)}(\bar{s}_t)\} = \mathrm{Cov}\{\epsilon_t^{(M)}(\bar{s}_t), \epsilon_t^{(R)}(\bar{s}_t)\} = \mathbf{\Sigma}. \qquad (17)$$

(A7) The eigenvalues of the companion matrix have modulus less than one, called Stationarity condition for model (14) and (15).

(A8) The constant effect of unmeasured mediator-outcome confounder U_t is fully characterized by the error correlation δ.

The equivalent models

The equivalent formulation for models (12) to (15) of the observed data

$$M_t = Z_t\alpha + \sum_{j=1}^{p}(\phi_{1j}Z_{t-j} + \psi_{11_j}M_{t-j} + \psi_{21_j}R_{t-j}) + \epsilon_{1t}, \tag{18}$$

$$R_t = Z_t\gamma + M_t\beta + \sum_{j=1}^{p}(\phi_{2j}Z_{t-j} + \psi_{12_j}M_{t-j} + \psi_{22_j}R_{t-j}) + \epsilon_{2t}. \tag{19}$$

Thus we propose to estimate the parameters $\{\alpha, \beta, \gamma, \omega_{11_j}, \omega_{21_j}, \omega_{12_j}, \omega_{22_j}\}$ by transforming the parameters $\{\phi_{1j}, \phi_{1j}, \psi_{11_j}, \psi_{21_j}, \psi_{12_j}, \psi_{22_j}\}$ from models (12-15) to models (18) and (19).

However, the equivalent formulation is linear structural equation models with correlated errors, therefore, one cannot fit them separately.

The maximum (conditional) likelihood estimation

- Let $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}, \boldsymbol{\phi}_1^T, \boldsymbol{\psi}_{11}^T, \boldsymbol{\psi}_{21}^T)^T$, $\boldsymbol{\theta}_2 = (\boldsymbol{\gamma}, \boldsymbol{\phi}_2^T, \boldsymbol{\psi}_{12}^T, \boldsymbol{\psi}_{22}^T)^T$, where $\boldsymbol{\psi}_{jk} = (\psi_{jk_1}, \cdots, \psi_{jk_p})$, $\boldsymbol{\psi}_{jk} = (\psi_{jk_1}, \cdots, \psi_{jk_p})$, for $j, k = 1, 2$.
- Let $\boldsymbol{X}_t = (Z_t, \boldsymbol{Z}_{t-1}^{(p)T}, \boldsymbol{M}_{t-1}^{(p)T}, \boldsymbol{R}_{t-1}^{(p)T})^T$, where $\boldsymbol{Z}_{t-1}^{(p)T} = (Z_{t-1}, \cdots, Z_{t-p})^T$, defined analogously for $\boldsymbol{M}_{t-1}^{(p)T}$ and $\boldsymbol{R}_{t-1}^{(p)T}$.
- Let $\boldsymbol{\Theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \beta, \sigma_1, \sigma_2)$ be all the model parameters except δ.

Given the initial p time periods, the conditional log-likelihood is

$$\begin{aligned}\ell(\boldsymbol{\Theta}, \delta \mid \boldsymbol{Z}, \mathscr{I}_p) &= \sum_{t=p+1}^{T} \log \mathscr{P}((M_t, R_t) \mid \boldsymbol{X}_t) \\ &= -\frac{T-p}{2} \log \sigma_1^2 \sigma_2^2 (1-\delta^2) - \frac{1}{2\sigma_1^2} \|\boldsymbol{M} - \boldsymbol{X}\boldsymbol{\theta}_1\|_2^2 \\ &\quad - \frac{1}{2\sigma_2^2(1-\delta^2)} \|(\boldsymbol{R} - \boldsymbol{M}\beta - \boldsymbol{X}\boldsymbol{\theta}_2) - \kappa(\boldsymbol{M} - \boldsymbol{X}\boldsymbol{\theta_1})\|_{\boldsymbol{2}}^{\boldsymbol{2}},\end{aligned} \tag{20}$$

where $\mathscr{I}_p = \{(Z_1, M_1, R_1), \cdots, (Z_p, M_p, R_p)\}$ is the initial p observations; $\boldsymbol{R} = (R_{p+1}, \cdots, R_T)^T$, similarly for $\boldsymbol{M}$ and $\boldsymbol{X}$; and $\kappa = \delta\sigma_2/\sigma_1$.

The non-identifiability of the error correlation δ

In the model (18) and (19), a nonzero correlation parameter δ can be interpreted as the existence of omitted variables that are related to both the observed value of the mediator M and the potential outcomes Y even after conditioning on the exposure Z and the observed covariates C.

Although δ cannot be estimated from the conditional likelihood of single-level data, estimators for β and γ are expressed as functions of δ, and consistent after correcting for δ from

$$\begin{pmatrix} M_t(\bar{s}_t{}^*) \\ R_t(\bar{s}_t) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} Z_t(\bar{s}_t{}^*)\alpha \\ Z_t(\bar{s}_t)(\gamma + \alpha\beta) \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \beta\sigma_1^2 + \delta\sigma_1\sigma_2 \\ \beta\sigma_1^2 + \delta\sigma_1\sigma_2 & \beta^2 + \sigma_1^2 + 2\beta\delta\sigma_1\sigma_2 + \sigma_2^2 \end{pmatrix} \right)$$

Let $\gamma' = \gamma + \alpha\beta$, $\rho = \beta\sigma_1^2 + \delta\sigma_1\sigma_2$ and $\tau = \beta^2 + \sigma_1^2 + 2\beta\delta\sigma_1\sigma_2 + \sigma_2^2$. We cannot solve for $(\beta, \gamma, \delta, \sigma_2)$ from the five identifiable parameters $(\alpha, \gamma', \sigma_1^2, \rho\tau^2)$ because there are six unknown parameters in $(\alpha, \beta, \gamma, \sigma_1, \sigma_2, \delta)$.

The population level coefficients

For the time series of participant i ($i = 1, \cdots, N$), we model the first-level scan-time data by single-level Granger mediation model (12) to (15), and all the modeling parameters could be denoted with subscript i $(\alpha_i, \beta_i, \gamma_i)$.

In order to estimate the population averages of the causal effects and account for the between-participant variations, consider the following multivariate linear model

$$\boldsymbol{\vartheta}_i = \boldsymbol{\vartheta} + \boldsymbol{\eta}_i \tag{21}$$

where $\boldsymbol{\vartheta}_i = (\alpha_i, \beta_i, \gamma_i)^T$; $\boldsymbol{\vartheta} = (\alpha, \beta, \gamma)^T$ denotes the population level coefficients; and $\eta_{\mathrm{i}} = (\epsilon_{\mathrm{i}}^{\alpha}, \epsilon_{\mathrm{i}}^{\beta}, \epsilon_{\mathrm{i}}^{\gamma})^{\mathrm{T}}$ is the random error of participant i. which is assumed to be iid from a trivariate normal distribution with mean zero and covariance matrix Λ.

At the population level, the population direct effect is γ, and the population indirect effect is $\alpha\beta$ by the product method or the difference method.

The population level error correlation δ

Let δ_i be the error correlation between ϵ_{i_1t} and ϵ_{i_2t} for participant i.

- As shown before, δ_i is not identifiable from the individual level conditional likelihood function for each participant i.
- Since the joint likelihood of N independent participants is simply a product of individual likelihood functions, one cannot estimate different δ_i from the joint likelihood function either.

(A8) δ_i is constant across participants, that is $\delta_i = \delta$ for all i.

- Assumption (A8) reduces the number of parameters in the model and pool information across subjects to estimate a single *delta*.
- A two-stage algorithm and a block coordinate-descent algorithm are developed to estimate δ by maximizing the conditional log-likelihood function.

The joint log-likelihood for maximizing

Let $\Upsilon = (\delta, \boldsymbol{\vartheta}, \boldsymbol{\Lambda}, (\boldsymbol{\theta}_{i_1}, \boldsymbol{\theta}_{i_2}, \beta_i), (\sigma_{1_i}, \sigma_{2_i})$, the conditional log-likelihood function (conditioning on the initial p time points of each subject's data) is written as

$$\begin{aligned} h(\boldsymbol{\Upsilon}) &= \sum_{i=1}^{N} \sum_{t=p+1}^{T_i} \log \mathcal{P}(R_{i_t}, M_{i_t} \mid \boldsymbol{X_{i_t}}, \boldsymbol{\theta}_{i_1}, \boldsymbol{\theta}_{i_2}, \beta_i, \sigma_{1_i}, \sigma_{2_i}) \\ &\quad + \sum_{i=1}^{N} \log(\boldsymbol{\vartheta_i} \mid \boldsymbol{\vartheta}, \boldsymbol{\lambda}) \\ &= h_1 + h_2, \end{aligned} \tag{22}$$

where $\boldsymbol{\vartheta} = (\alpha_i, \beta_i, \gamma_i)$, α_i and β_i are the first element of $\boldsymbol{\theta}_{i_t}$ and $\boldsymbol{\theta}_{i_2}$; T_i is the number of time points of subject i; h_1 is the sum of N log=likelihood functions (20); h_2 is the log-likelihood function of model (21).

Computationally light and asymptotically consistent

This method splits the computation cost by two steps.

- In the first stage, the coefficients in the single-level model with a given δ for each participant i by

$$\beta = \frac{\rho}{\sigma_1^2} - \frac{\delta}{\sigma_1^2\sqrt{1-\delta^2}}\sqrt{\sigma_1^2\tau^2 - \rho^2}, \tag{23}$$

$$\sigma_2^2 = \frac{1}{\sigma_1^2(1-\delta^2)}[\sigma_1^2\tau^2 - \rho^2], \tag{24}$$

$$\gamma = \gamma^{'} - \alpha\beta. \tag{25}$$

- In the second stage, the estimated coefficients from the first stage was plugged into model (21) $\boldsymbol{\vartheta}_i = \boldsymbol{\vartheta} + \boldsymbol{\eta}_i$.

To estimate δ, different δ are given to repeat the two-stage computation, and then use a one-dimensional optimization algorithm to find the δ that yields the maximum joint likelihood h.

Improve the finite sample performance

A block coordinate-descent algorithm is proposed to maximize h_1 and h_2 jointly that could improve the finite sample performance.

Consider the following optimization problem

$$\min_{\boldsymbol{\Upsilon}:((\sigma_{i_1},\sigma_{i_2}),\boldsymbol{\Lambda})\in\mathscr{S}} h(\boldsymbol{\Upsilon}), \tag{26}$$

where is a constraint set for the variance components (a positive constraint on $(\sigma_{i_1}, \sigma_{i_2})$, and a positive-definite constraint on $\boldsymbol{\Lambda}$).

Optimize blocks of variables $(\sigma_{i_1}^{-1}, \sigma_{i_2}^{-1})$, $(\boldsymbol{\theta}_{i_1}, \boldsymbol{\theta}_{i_1}, \beta_i)$, $\boldsymbol{\vartheta}$, $\boldsymbol{\Lambda}^{-1}$ iteratively, since the optimizers for each block of variables are given in explicit forms, conditional on all other variables. After obtaining the profile likelihood value for each δ, δ is estimated by a one-dimensional optimization algorithm.

Experiment and Results

- OpenfMRI database with accession number ds000030,
- $N = 121$ right handed participants,
- Randomly intermixed with 96 GO and 32 STOP stimuli,
- the lag parameter $p = 2$,
- 200 bootstrap samples for inference.

Table: Average estimates and 95% confidence intervals from GMA-*h*, GMA-ts, MACC-*h*, KKB, and BK for the fMRI data set using 200 bootstrap samples

Method	δ	γ	α	β	$\alpha\beta_p$
GMA-h	−0.370 (−0.530, −0.156)	−1.729 (−2.445, −0.964)	−0.739 (−1.487, 0.035)	0.838 (0.644, 0.999)	−0.623 (−1.239, 0.033)
GMA-ts	−0.343 (−0.501, −0.163)	−1.722 (−2.461, −0.904)	−0.740 (−1.442, 0.080)	0.810 (0.656, 0.965)	−0.604 (−1.234, 0.055)
MACC-h	−0.762 (−0.799, −0.721)	−2.310 (−2.958, −1.641)	−0.259 (−0.810, 0.303)	1.511 (1.410, 1.619)	−0.391 (−1.271, 0.465)
KKB	...	−2.513 (−2.922, −2.073)	−0.225 (−0.772, 0.326)	0.617 (0.589, 0.647)	−0.140 (−0.467, 0.196)
BK	...	−2.583 (−3.023, −2.142)	−0.235 (−0.774, 0.352)	0.616 (0.588, 0.647)	−0.146 (−0.467, 0.211)

GMA, Granger mediation analysis method proposed in this paper, using either the hierarchical likelihood algorithm (GMA-h) or the two-stage algorithm (GMA-ts); MACC-h, mediation analysis with correlated errors by Zhao and Luo (2014); KKB, multilevel mediation by Kenny et al. (2003); BK, single-level mediation by Baron and Kenny (1986).

Discussion

Recapitulation

- Granger mediation analysis framework is proposed for time series data.
- CMA across the mediation variables, and VAR models across the temporal correlations are integrated.
- Individual variability and correlated errors between the mediator and the outcome variables are modeled.
- The error correlation δ is considered as the sensitivity parameter of the unmeasured mediator-outcome confounding effect.
- Asymptotically unique and consistent estimates can be obtained or finite sample performance is improved by proposed algorithms.

Deficiencies

1. Focus on randomized treatment, observational studies are not apropos.
2. Interaction between the mediator and the outcome are not allowed.
3. Time-invariant and time-varing covariates are not allowed.
4. Model misspecification (if nonlinear effects exist).
5. The error correlation δ is assumed to be constant across participants

Reference

[1] Baron, R.M. and Kenny, D.A. (1986) The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51, 1173.

[2] 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.

[3] Imai, K., Keele, L. and Yamamoto, T. (2010) Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science, 51–71.

[4] Kenny, D.A., Korchmaros, J.D. and Bolger, N. (2003) Lower level mediation in multilevel models. Psychological Methods, 8, 115.

[5] Lindquist, M.A. (2012) Functional causal mediation analysis with an application to brain connectivity. Journal of the American Statistical Association, 107, 1297–1309.

Reference

[6] Sobel, M.E. and Lindquist, M.A. (2014) Causal inference for fmri time series data with systematic errors of measurement in a balanced on/off study of social evaluative threat. Journal of the American Statistical Association, 109, 967–976.

[7] VanderWeele, T.J. (2015) Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press.

[8] VanderWeele, T.J. and TchetgenTchetgen, E.J. (2017) Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society, 79, 917–938.

[9] Zhao, Y., and Luo, X. (2014) Estimating mediation effects under correlated errors with an application to fmri. [Preprint] Available at: arXiv:1410.7217 (Accessed May 4, 2017).

[10] Zhao, Y., & Luo, X. (2019). Granger mediation analysis of multiple time series with an application to functional magnetic resonance imaging. Biometrics, 75(3), 788–798. https://doi.org/10.1111/biom.13056.