

Bayesian Estimation of Causal Effects Using Proxies of a Latent Interference Network

Bar Weinstein Daniel Nevo

Department of Statistics and Operations Research, Tel Aviv University



Introduction

- Interference occurs when the outcome of a unit depends on treatments assigned to other units.
- Interference structure represented by a network.
- Accurately measuring networks of social interactions is a formidable task.
- Researchers frequently only have access to proxy measurements
 - Measurement error.
 - Networks from multiple sources.
 - Multilayer networks.

Setup and Structural Causal Model

- Finite population of N units. \mathbf{Z} treatments; \mathbf{Y} outcomes; \mathbf{X} covariates.
 - True (latent) interference network \mathbf{A}^* .
 - Observed proxy networks $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_B\}$.
 - Observed data $\mathcal{D} = (\mathbf{Y}, \mathbf{Z}, \mathcal{A}, \mathbf{X})$.
1. True network model $\mathbf{A}^* = f_{A^*}(\mathbf{X}, \mathbf{U}, \varepsilon_{A^*})$.
 2. Treatments model $Z_i = f_Z(\mathbf{A}^*, \mathcal{A}, \mathbf{X}_i, \mathbf{X}_{-i}, \varepsilon_{Z_i})$.
 3. Outcome model $Y_i = f_Y(Z_i, \mathbf{X}_i, \mathbf{Z}_{-i}, \mathbf{X}_{-i}, \mathbf{A}^*, \mathbf{U}_i, \varepsilon_{Y_i})$.

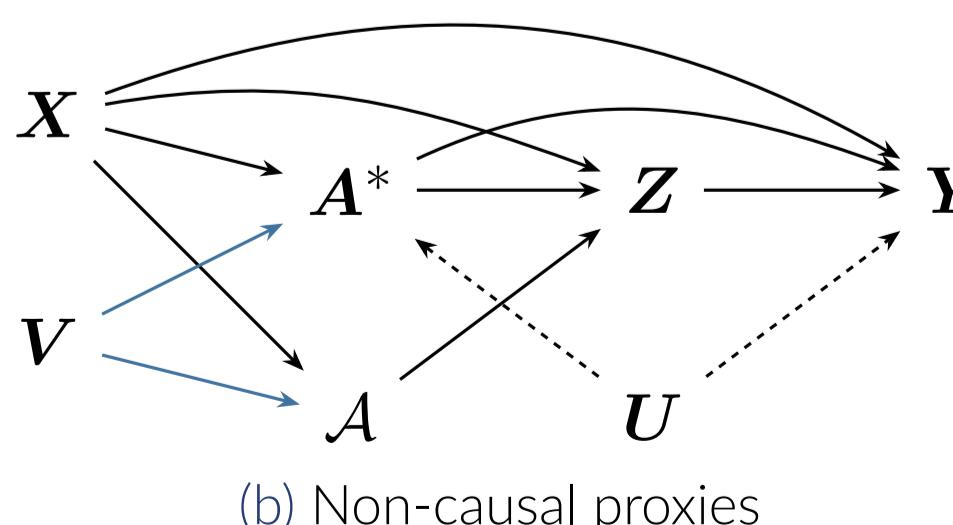
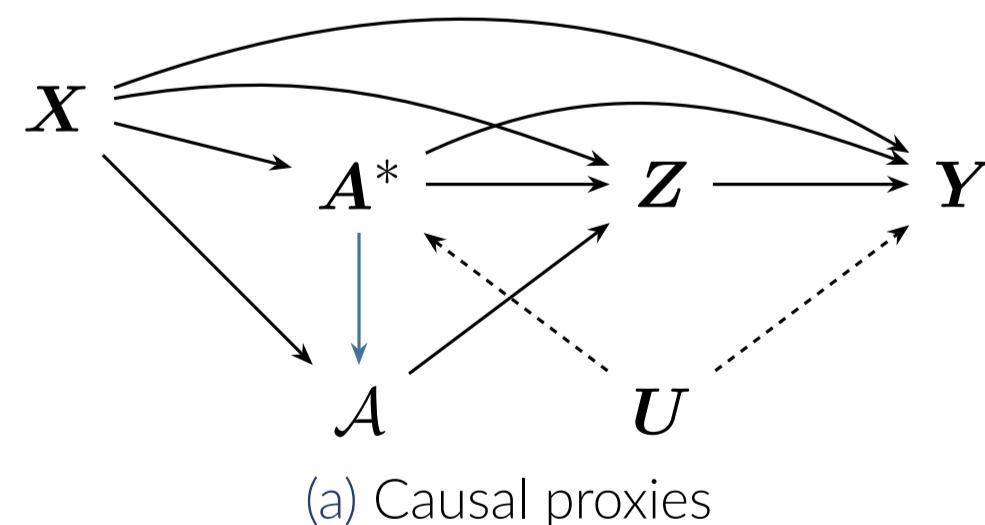


Fig.1 Directed acyclic graphs representing the assumed SCMs.

Causal Estimands

- Effects of hypothetical interventions on population-level treatments.
- $Y_i(\mathbf{z})$ outcomes under the intervention $do(\mathbf{Z} = \mathbf{z})$.
- Treatment assignment policies:
 1. **Static policy.** Set \mathbf{Z} to some fixed value \mathbf{z} .
 2. **Dynamic policy.** Set \mathbf{Z} as a function of covariates $\mathbf{z} = h(\mathbf{X})$.
 3. **Stochastic policy.** $do(\mathbf{Z} = \mathbf{z})$ occurs with probability $\pi_\alpha(\mathbf{z})$.
- Estimands are contrasts of $N^{-1} \sum_{i=1}^N \mathbb{E}[Y_i(\mathbf{z}) | \mathbf{A}^*]$ for different policies (in stochastic we marginalize over π_α).

Numerical Illustration

Proxies \mathcal{A} from auto-regressive measurement error model of \mathbf{A}^* . Outcomes \mathbf{Y} from Gaussian Markov Random Field model.

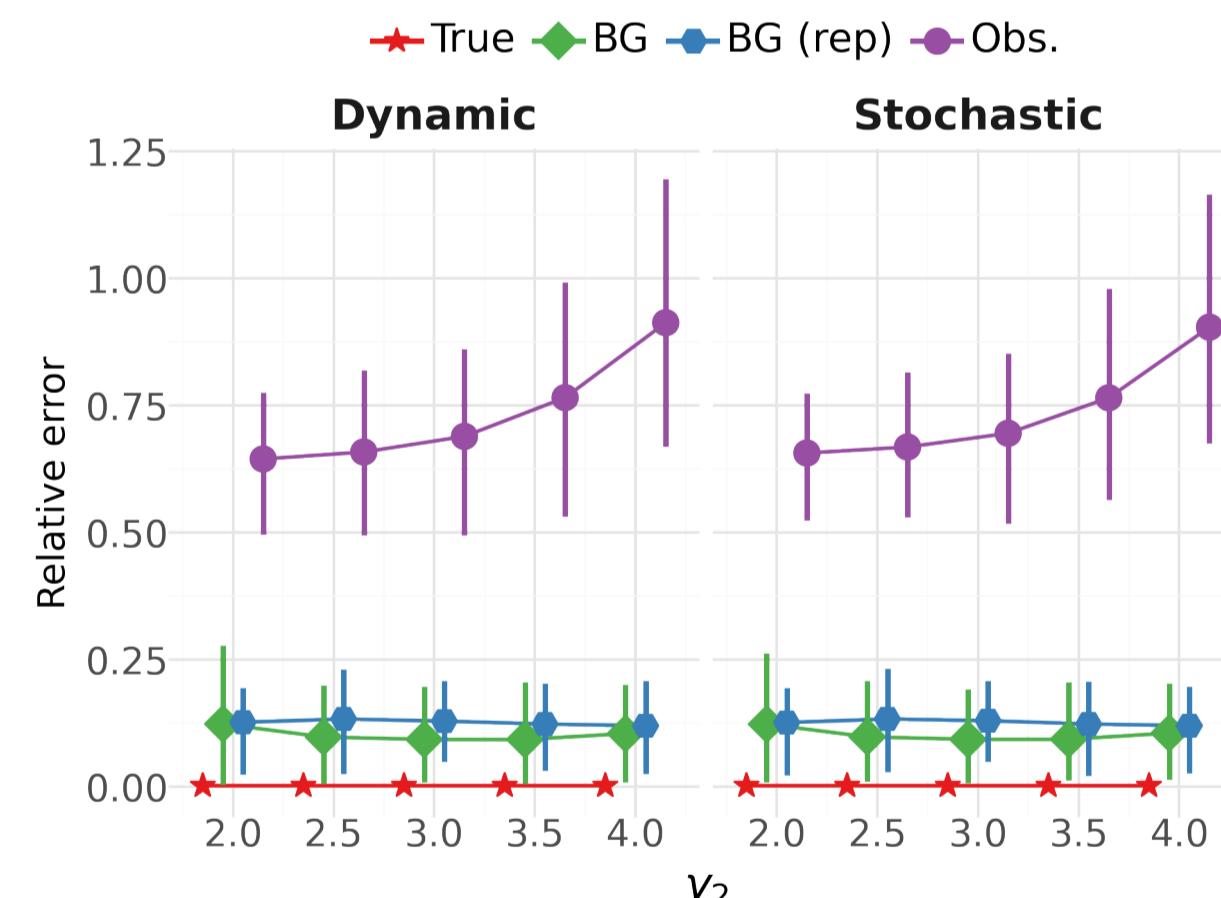


Fig.2 Relative error in estimating dynamic and stochastic estimands. Proxies \mathcal{A} are weaker as γ_2 (x-axis) increases.

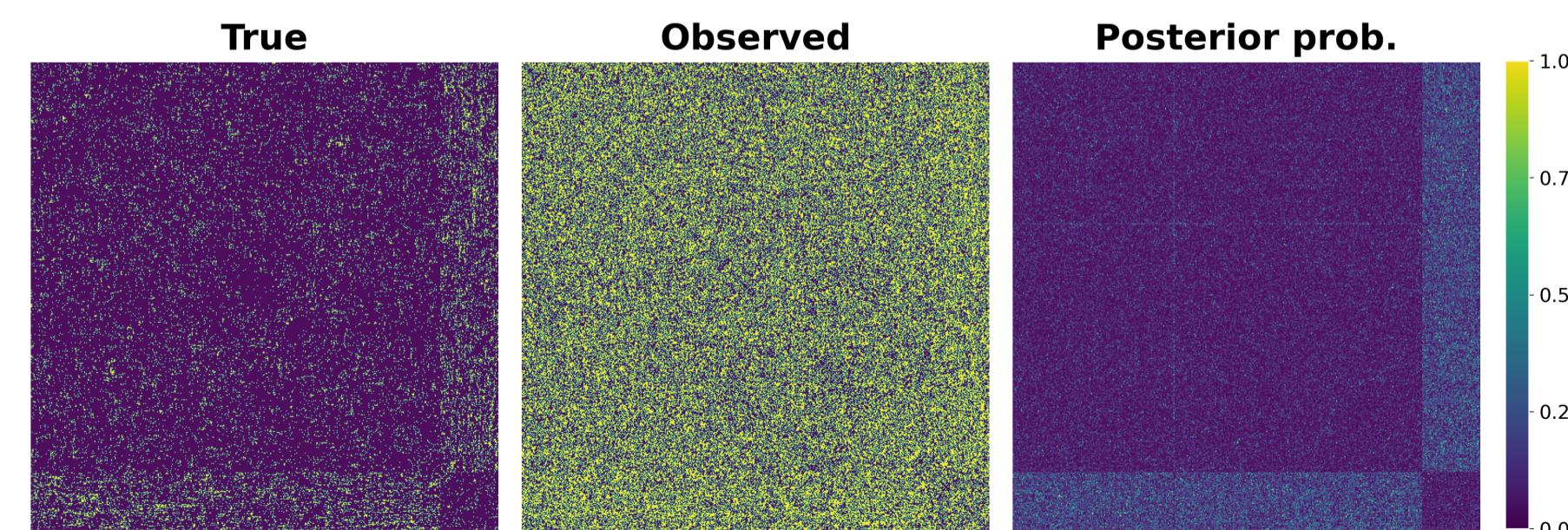


Fig.3 True network, observed proxy, and posterior probabilities in one iteration.

Bayesian Estimation and Inference

The joint posterior distribution is

$$p(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{A}^* | \mathcal{D}) \propto p(\mathbf{Y} | \mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\ \times p(\mathcal{A} | \mathbf{A}^*, \mathbf{X}, \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) \\ \times p(\mathbf{A}^* | \mathbf{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

- Mixed space of continuous parameters $(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ and discrete latent variables (\mathbf{A}^*) . Discrete space contains $\mathcal{O}(2^{N^2})$ terms.
- Marginalization over \mathbf{A}^* is not possible. Random-walk MH or other MCMC methods do not scale well.
- We propose a Block Gibbs MCMC algorithm that iterates between continuous $\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}$ and discrete \mathbf{A}^* updates.

Locally Informed Proposals

- Update \mathbf{A}^* by flipping edges $A_{ij}^* = 0/1 \rightarrow A_{ij}^* = 1/0$, favoring flips with larger increases in the log-conditional posterior $\lambda(\mathbf{A}^* | \cdot) = \log(p(\mathbf{A}^* | \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathcal{D}))$ via locally balanced weights.
- At iteration t , let $\mathbf{A}_{t+1}^*(ij) = \text{FlipEntries}(\mathbf{A}_t^*, ij)$ and define $\Delta(ij, t | \cdot) = \lambda(\mathbf{A}_{t+1}^*(ij) | \cdot) - \lambda(\mathbf{A}_t^* | \cdot)$. Approximate with one gradient (computed once per iteration):

$$\tilde{\Delta}(ij, t | \cdot) = -(2A_{t,ij}^* - 1) \frac{\partial \lambda(\mathbf{A}_t^* | \cdot)}{\partial A_{t,ij}^*} \approx \Delta(ij, t | \cdot).$$

Algorithm 1 A Single \mathbf{A}^* Update with Locally Informed Proposals

Input: Data \mathcal{D} , parameters $(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma})$, current state \mathbf{A}_t^* , log-posterior $\lambda(\cdot)$, # entries to update $K \geq 1$.

- 1: Compute differences $\tilde{\Delta}(\mathbf{A}_t^* | \cdot) \equiv \tilde{\Delta}(\mathbf{A}_t^* | \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathcal{D})$.
- 2: Compute $q(ij | \mathbf{A}_t^*, \cdot) = \text{Softmax}\left(\frac{1}{2}\tilde{\Delta}(\mathbf{A}_t^* | \cdot)\right)$.
- 3: Sample without replacement K edges \mathcal{I} using Gumbel-Top-K trick with probabilities $q(ij | \mathbf{A}_t^*, \cdot)$.
- 4: Forward probability $q(\mathcal{I} | \mathbf{A}_t^*, \cdot) = \prod_{ij \in \mathcal{I}} q(ij | \mathbf{A}_t^*, \cdot)$.
- 5: $\mathbf{A}_{t+1}^* \leftarrow \text{FlipEntries}(\mathbf{A}_t^*, \mathcal{I})$.
- 6: Compute backward probability $q(\mathcal{I} | \mathbf{A}_{t+1}^*, \cdot)$.
- 7: Accept with probability

$$\min \left(\exp(\Delta(\mathbf{A}_{t+1}^*, \mathbf{A}_t^* | \cdot)) \frac{q(\mathcal{I} | \mathbf{A}_{t+1}^*, \cdot)}{q(\mathcal{I} | \mathbf{A}_t^*, \cdot)}, 1 \right).$$