arXiv

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

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

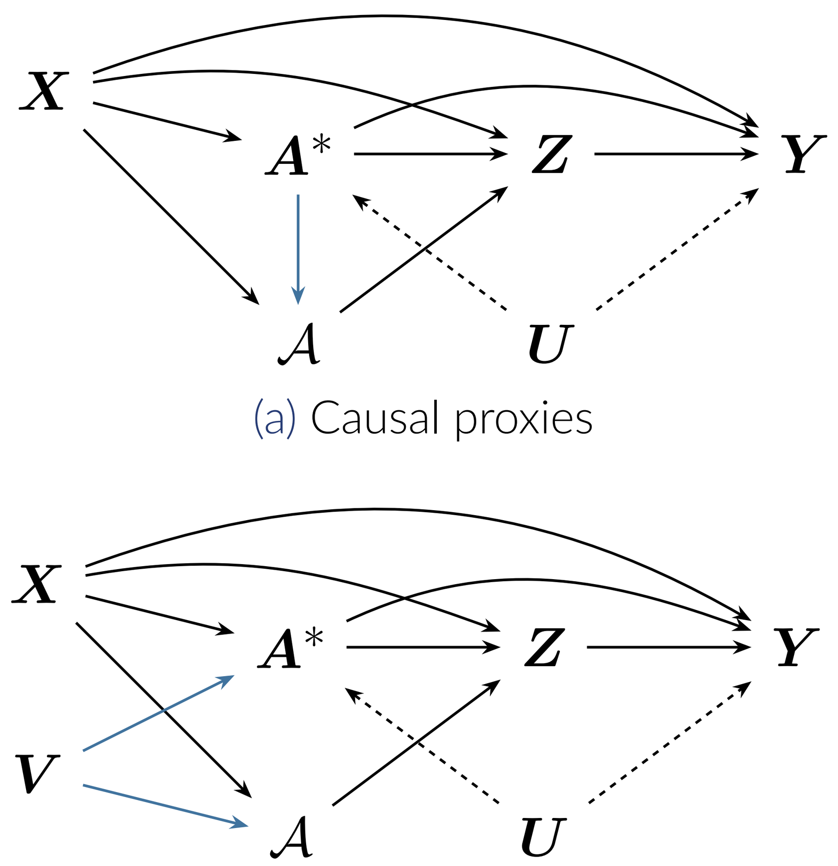
Introduction

- 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. $\boldsymbol{Z}$ treatments; $\boldsymbol{Y}$ outcomes; $\boldsymbol{X}$ covariates.
- True (latent) interference network $\boldsymbol{A}^*$.
- Observed proxy networks $\mathcal{A} = \{\boldsymbol{A}_1, \dots, \boldsymbol{A}_B\}$.
- Observed data $\boldsymbol{\mathcal{D}} = (\boldsymbol{Y}, \boldsymbol{Z}, \mathcal{A}, \boldsymbol{X})$.

1. True network model $\boldsymbol{A}^* = f_{A^*}(\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{\varepsilon}_{A^*})$.
2. Treatments model $Z_i = f_Z(\boldsymbol{A}^*, \mathcal{A}, \boldsymbol{X}_i, \boldsymbol{X}_{-i}, \varepsilon_{Z_i})$.
3. Outcome model $Y_i = f_Y(Z_i, \boldsymbol{X}_i, \boldsymbol{Z}_{-i}, \boldsymbol{X}_{-i}, \boldsymbol{A}^*, \boldsymbol{U}_i, \varepsilon_{Y_i})$.



(a) Causal proxies

(b) Non-causal proxies

Fig.1 Directed acyclic graphs representing the assumed SCMs.

Causal Estimands

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

Numerical Illustration

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

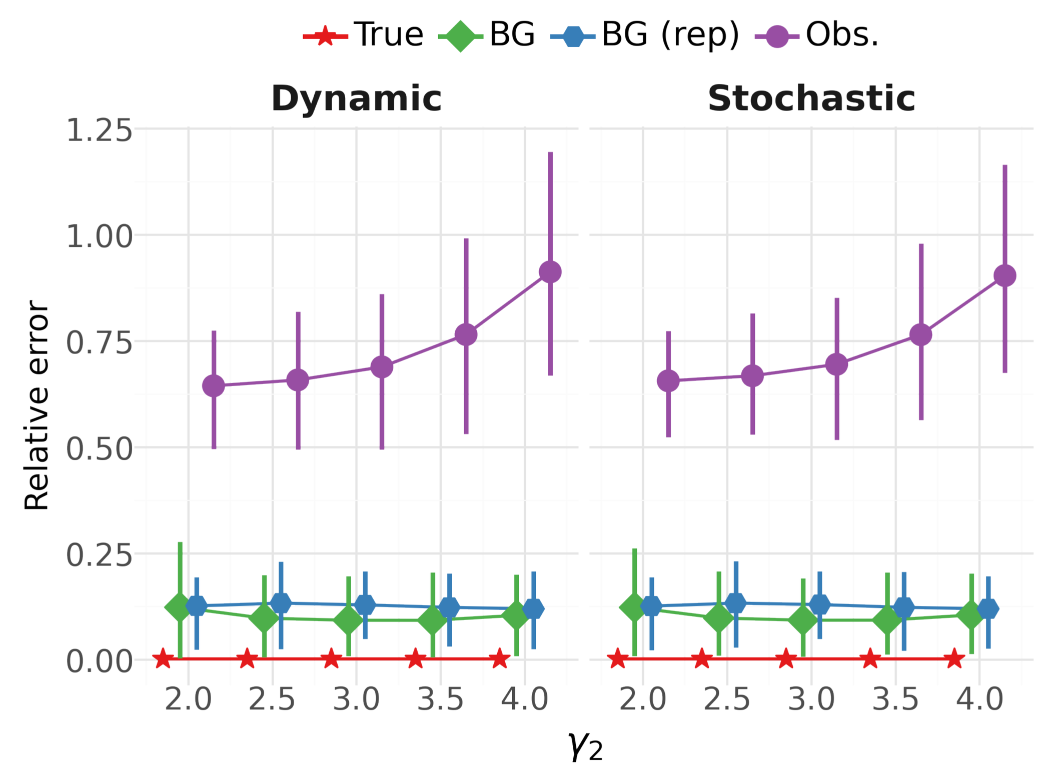


Fig.2 Relative error in estimating dynamic and stochastic estimands. Proxies $\mathcal{A}$ are weaker as γ_2 (x-axis) increases.

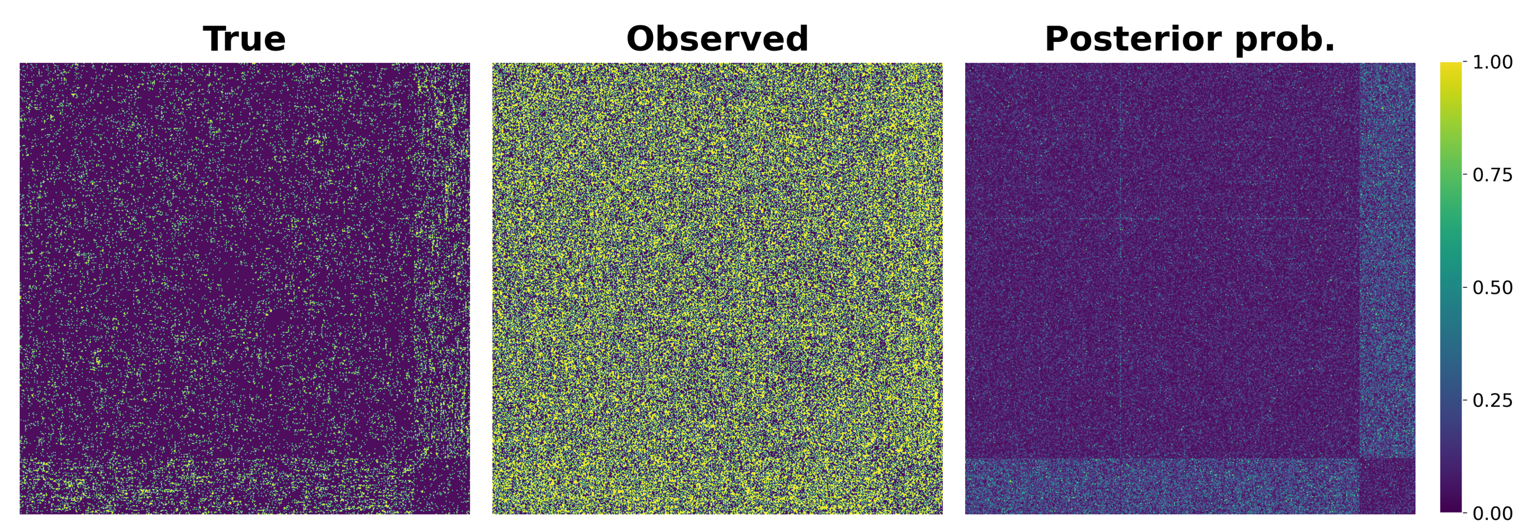


Fig.3 True network, observed proxy, and posterior probabilities in one iteration.

Bayesian Estimation and Inference

The joint posterior distribution is

$$
\begin{aligned}
p(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}) \propto\ & p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\
& \times p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) \\
& \times p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).
\end{aligned}
$$

- Mixed space of continuous parameters $(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma})$ and discrete latent variables $(\boldsymbol{A}^*)$. Discrete space contains $\mathcal{O}\left(2^{N^2}\right)$ terms.
- Marginalization over $\boldsymbol{A}^*$ 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 $\boldsymbol{A}^*$ updates.

Locally Informed Proposals

- Update $\boldsymbol{A}^*$ by flipping edges $A^*_{ij} = 0/1 \rightarrow A^*_{ij} = 1/0$, favoring flips with larger increases in the log-conditional posterior $\lambda(\boldsymbol{A}^* \mid \cdot) = \log\left(p(\boldsymbol{A}^* \mid \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\mathcal{D}})\right)$ via locally balanced weights.
- At iteration t, let $\boldsymbol{A}^*_{t+1}(ij) = \texttt{FlipEntries}(\boldsymbol{A}^*_t, ij)$ and define $\Delta(ij, t \mid \cdot) = \lambda(\boldsymbol{A}^*_{t+1}(ij) \mid \cdot) - \lambda(\boldsymbol{A}^*_t \mid \cdot)$. Approximate with one gradient (computed once per iteration): $\widetilde{\Delta}(ij, t \mid \cdot) = -(2A^*_{t,ij} - 1)\frac{\partial \lambda(\boldsymbol{A}^*_t \mid \cdot)}{\partial A^*_{t,ij}} \approx \Delta(ij, t \mid \cdot)$.

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

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

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

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