SUPPLEMENTARY MATERIAL: CONTINUAL LIFELONG CAUSAL EFFECT INFERENCE WITH REAL WORLD EVIDENCE

Anonymous authors
Paper under double-blind review

A SIMULATION PROCEDURE

Our synthetic data include confounders, instrumental, adjustment, and irrelevant variables. The interrelations among these variables, treatments, and outcomes are illustrated in Figure 1. The number of observed variables in the vector $X = (C^\intercal, Z^\intercal, I^\intercal, A^\intercal)^\intercal$ is set to 100, including 35 confounders in C, 35 adjustment variables in A, 10 instrumental variables in Z, and 20 irrelevant variables in I. The model used to generate the continuous outcome variable Y in this simulation is the partially linear regression model, extending the ideas described in Robinson (1988); Jacob et al. (2019); Chu et al. (2020):

$$Y = \tau((C^\intercal, A^\intercal)^\intercal)T + g((C^\intercal, A^\intercal)^\intercal) + \epsilon, \tag{1}$$

where ϵ are unobserved covariates, which follow a standard normal distribution $N(0, 1)$ and $E[\epsilon|C, A, T] = 0$. $T \overset{ind.}{\sim}$ Bernoulli$(e_0((C^\intercal, Z^\intercal)^\intercal))$ and $e_0((C^\intercal, Z^\intercal)^\intercal)$ is the propensity score, which represents the treatment selection bias based on their own confounders C and instrumental variables Z. Because we aim to simulate multiple data sources $\{\mathcal{D}_d; d = 1, ..., D\}$, the vector of all observed covariates $X = (C^\intercal, Z^\intercal, I^\intercal, A^\intercal)^\intercal$ is sampled from different multivariate normal distribution with mean vector $\mu_C^d, \mu_Z^d, \mu_I^d,$ and μ_A^d and different random positive definite covariance matrices Σ^d.

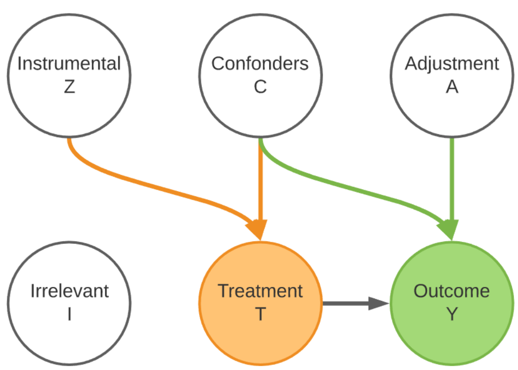

Figure 1: The types of variables.

For each data source, except for the different magnitude of mean vector and structure of covariance matrix, the simulation procedure is the same. Let D be the diagonal matrix with the square roots of the diagonal entries of Σ on its diagonal, i.e., $D = \sqrt{diag(\sigma)}$, then the correlation matrix is given as:

$$R = D^{-1}\Sigma D^{-1}. \tag{2}$$

We use algorithm 3 in Hardin et al. (2013) to simulate positive definite correlation matrices consisting of different types of variables. Our correlation matrices are based on the hub correlation structure which has a known correlation between a hub variable and each of the remaining variables (Zhang

& Horvath, 2005; Langfelder et al., 2008). Each variable in one type of variables is correlated to the hub-variable with decreasing strength from specified maximum correlation to minimum correlation, and different types of variables are generated independently or with weaker correlation among variable types. Defining the first variable as the hub, for the ith variable $(i = 2, 3, ..., n)$, the correlation between it and the hub-variable in one type of variables is given as:

$$R_{i,1} = \rho_{\max} - \left(\frac{i-2}{d-2}\right)^{\gamma} (\rho_{\max} - \rho_{\min}), \tag{3}$$

where $\rho_{\max}$ and $\rho_{\min}$ are specified maximum and minimum correlations, and the rate γ controls rate at which correlations decay.

After specifying the relationship between the hub variable and the remaining variables in the same type of variables, we use Toeplitz structure to fill out the remainder of the hub correlation matrix and get the hub-Toeplitz correlation matrix R_{type} for other type of variables. Here, R is the $n \times n$ matrix having the blocks R_Z, R_C, R_A, and R_I along the diagonal and zeros at off-diagonal elements. This yields a correlation matrix with nonzero correlations within the same type and zero correlation among other types. The amount of correlations among types which can be added to the positive-definite correlation matrix R is determined by its smallest eigenvalue.

The function $\tau((C^\intercal, A^\intercal)^\intercal)$ describes the true treatment effect as a function of the values of adjustment variables A and confounders C; namely $\tau((C^\intercal, A^\intercal)^\intercal) = (\sin((C^\intercal, A^\intercal)^\intercal \times b_\tau))^2$ where b_τ represents weights for every covariate in the function, which is generated by uniform$(0, 1)$. The variable treatment effect implies that its strength differs among the units and is therefore conditioned on C and A. The function $g((C^\intercal, A^\intercal)^\intercal)$ can have an influence on outcome regardless of treatment assignment. It is calculated via a trigonometric function to make the covariates nonlinear, which is defined as $g((C^\intercal, A^\intercal)^\intercal) = (\cos((C^\intercal, A^\intercal)^\intercal \times b_g))^2$. Here, b_g represents a weight for each covariate in this function, which is generated by uniform$(0, 1)$. The bias is attributed to unobserved covariates which follow a random normal distribution $N(0, 1)$. The treatment assignment T follows the Bernoulli distribution, i.e., $T \overset{ind.}{\sim}$ Bernoulli$(e_0((C^\intercal, Z^\intercal)^\intercal))$ with probability $e_0((C^\intercal, Z^\intercal)^\intercal) = \Phi(\frac{a-\mu(a)}{\sigma(a)})$, where $e_0((C^\intercal, Z^\intercal)^\intercal)$ represents the propensity score, which is the cumulative distribution function for a standard normal random variable based on confounders C and instrumental variables Z, i.e., $a = \sin((C^\intercal, Z^\intercal)^\intercal \times b_a)$, where b_a is generated by uniform$(0, 1)$.

We totally simulate five different data sources with five different multivariate normal distributions to represent the incrementally available observational data. In each data source, we randomly draw 10000 samples including treatment units and control units. Therefore, for five datasets, they have different selection bias, magnitude of covariates, covariance matrices for variables, and number of treatment and control units. To ensure a robust estimation of model performance, for each data source, we repeat the simulation procedure 10 times and obtain 10 synthetic datasets.

REFERENCES

Zhixuan Chu, Stephen L Rathbun, and Sheng Li. Matching in selective and balanced representation space for treatment effects estimation. *arXiv preprint arXiv:2009.06828*, 2020.

Johanna Hardin, Stephan Ramon Garcia, and David Golan. A method for generating realistic correlation matrices. *The Annals of Applied Statistics*, pp. 1733–1762, 2013.

Daniel Jacob, Wolfgang Karl Härdle, and Stefan Lessmann. Group average treatment effects for observational studies. *arXiv preprint arXiv:1911.02688*, 2019.

Peter Langfelder, Bin Zhang, and Steve Horvath. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics*, 24(5):719–720, 2008.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pp. 931–954, 1988.

Bin Zhang and Steve Horvath. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1), 2005.