Revisiting and Improvement of Endogenous Stratification In Random Experiments

Shuo Wang, Nancy Huang, Yaze Gao

4/24/2022

Overview

- Bring up the current concerns in the Endogenous Stratification Method in random experiments
- Demonstrate the issue in endogenous stratification in random experiments through mathematical calculations and code simulations.
- Present our method which can help to resolve the concerns in current endogenous stratification in random experiments, and illustrate our method by code simulations.

Endogenous Stratification Method

- Advocated by David Kent.
- Increasingly utilized in medical research recently.
- General idea:
  1. Regress the outcome variable on baseline characteristics by using the full sample of experimental controls;
  2. Generate predicted potential outcomes $Y(0)$ for all sample units in both treatment and control groups.
  3. Stratify all observations based on predicted $Y(0)$ and calculate treatment effect within each strata.

Concerns of Endogenous Stratification Method

- Problematic for studies with small sample sizes and large number of predictor variables which the predictor of the outcome without treatment may be severely overfitted in the control sample.
- Contrary to the idea of exogenous stratification which will stratify units based on their pre-treatment covariates.
- Proof:

$$
\begin{aligned}
& {x_j}^T \hat{\beta} - {x_j}^T \beta \\
=& \hat{y}_j - {x_j}^T \beta \\
=& (Hy)_j - {x_j}^T \beta \\
=& H_{jj} y_j + \sum_{i \neq j} H_{ji} y_i - {x_j}^T \beta
\end{aligned}
$$

- Main idea: y with large negative regression error is not in the same strata as y with large positive regression error, and thus can not be averaged.

Alternative

- In 2014, Abadie, Chingos, and West purposed an alternative way of endogenous stratification in order to mitigate the existing problem as described previously.
- Abadie *et al.* conclude that the alternative endogenous stratification estimators can show extensively improved small sample behavior in simulation based on leave-one-out cross-validation and repeated split sample techniques.
- We will use **jackknife bias correction** and **elastic-net regularization** in this project for illustration purpose.

Alternative 1: Delete-m Jackknife Bias Correction

The procedures of Jackknife bias correction are the following:

- Construct $\hat{\beta}$:
  $\hat{\beta} = (\sum_{i=1}^{n} x_i(1-z_i)x_i')^{-1} \sum_{i=1}^{n} x_i(1-z_i)y_i \triangleq f_n(x_1, \ldots, x_n)$
- Generate Jackknife replicates
- Bias estimation: $\hat{\text{bias}}(\hat{\beta})_{jack} = (m-1)(\bar{\beta}_{(m)} - \hat{\beta})$, where
  $\bar{\beta}_{(m)} = m^{-1} \sum_{j=1}^{m} \hat{\beta}_{(j)}$
- Estimator after bias correction:
  $\tilde{\beta}_{jack} = \hat{\beta} - \hat{\text{bias}}(\hat{\beta})_{jack} = m\hat{\beta} - (m-1)\bar{\beta}_{(m)}$
- Estimated average treatment effect:
  $\hat{\tau}_k^{jack} = \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < x_i^T \tilde{\beta}_{jack} \leq c_k]$

Alternative 2: Elastic-net Regularization

The procedures of Elastic-net Regularization are the following:

- Estimator for Elastic-net Regularization: $\hat{\beta}_{elastic} = \arg\min_{\beta_{elastic}}(\|y - X\beta_{elastic}\|^2 + \lambda_2\|\beta_{elastic}\|_2^2 + \lambda_1\|\beta_{elastic}\|_1)$.
- Estimated average treatment effect: $\hat{\tau}_k^{elastic} = \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < x_i^T\hat{\beta}_{elastic} \leq c_k]$

JTPA Data

- From National JTPA study, which is a large experimental evaluation of a job training program commissioned by the United States Department of Labor in 1980.
- 19 variables with 11204 observations, details are discussed in section **Solution and Simulation**, subsection **Data Description** in the paper
- In our project, only use the observations of male and discard the rest of the observations. The total number of observations is around 4000 to 5000. Our model includes all the variables expect `sex`, `id`, and `train`.

Simulated Data

- Procedure:

(1) For each $i = 1, ..., N$ where $N = 200, 500, 1000,$ generate

$$\epsilon_{i,j} \overset{iid}{\sim} \begin{cases} \text{Normal}(0,1) & j = 2m-1 \text{ for } m = 1,2,...,10 \text{ or } j = 21,22,...,40 \\ \text{Poisson}(7) & j = 2m \text{ for } m = 1,2,...,10 \end{cases}$$

(2) Let $X_{i,j} = \epsilon_{i,j}$ for $j = 1, ..., K$ where $K = 10, 20, 30$

(3) Generate $y_i = 3 + \sqrt{2}X_{i,1} + ... + \sqrt{2}X_{i,\frac{k}{2}} + 0.01X_{i,\frac{k}{2}+1} + ... + 0.01X_{i,k} + \eta$ where $\eta \sim N(0, \sigma^2)$ such that $Var(\mathbf{y}) = 100$

(4) Generate $Z \overset{iid}{\sim}$ Bernoulli(0.5)

(5) If $Z_i = 1,$ then set $y_i^{new} = y_i + 3$ so that the true ATE is 3

- We repeat the above procedure 100 times and obtain the average over the 100 ATEs within each strata using the jackknife bias correction method and elastic-net regularization technique.

Result: JTPA Data

Table 1: JTPA Estimation Results

Panel A: Average treatment effect

	Unadjusted	Adjusted
$\hat{\tau}$	4160.576	3936.766

Panel B: Average treatment effect by predicted outcome group

	Unadjusted			Adjusted		
	low	medium	high	low	medium	high
$\hat{\tau}_k$	5290.796	3712.424	3697.412	4748.074	3925.481	3738.127
$\hat{\tau}_k^{jack}$	5271.727	3742.237	3732.222	4775.589	3906.496	3711.809
$\hat{\tau}_k^{elastic}$	4673.707	2934.488	3997.914	4526.049	3042.961	3911.252

Result: Simulated Data

Table 2: Simulated Data Results

	K = 10						K = 20						K = 30					
	Unadjusted			Adjusted			Unadjusted			Adjusted			Unadjusted			Adjusted		
	low	medium	high	low	medium	high	low	medium	high	low	medium	high	low	medium	high	low	medium	high
N = 200																		
$\hat{\tau}_k$	4.7064	2.4211	0.8237	4.8020	2.4453	0.9036	5.3482	3.2470	1.1857	5.4805	3.1134	-0.8478	4.5077	2.8399	1.2149	7.5284	2.9084	7.6316
$\hat{\tau}_k^{jack}$	4.6438	2.4520	1.2153	4.9662	2.4046	2.4340	5.3385	3.1961	2.2860	5.5742	3.1570	2.7558	4.4986	2.8779	1.2291	7.5051	2.8536	7.0418
$\hat{\tau}_k^{elastic}$	3.9625	2.8153	1.0915	4.3960	2.6198	2.0923	4.5318	3.1585	1.3928	5.0192	3.0829	2.3105	3.8641	2.8681	2.1007	6.9259	2.8497	6.8980
N = 500																		
$\hat{\tau}_k$	3.9767	2.7491	2.5944	4.0144	2.7336	2.1890	3.7902	3.0922	2.0648	3.8523	3.0276	3.1751	3.5563	2.9116	2.4241	3.5158	2.9305	-16.3013
$\hat{\tau}_k^{jack}$	3.9622	2.7576	2.6186	3.9732	2.7363	2.1603	3.7402	3.1045	2.0808	3.8511	3.0360	3.8060	3.5181	2.8825	2.3062	3.4660	2.9444	8.4470
$\hat{\tau}_k^{elastic}$	3.5657	2.7670	3.2986	3.5073	2.7177	0.8024	3.5744	3.1066	2.1916	3.6309	3.0477	4.3711	3.3043	2.9691	2.5661	4.5414	2.9841	6.4364
N = 1000																		
$\hat{\tau}_k$	3.3631	3.0428	2.5680	3.3143	3.0398	2.8845	3.5074	2.8559	2.0718	3.4653	2.8258	2.3604	3.3045	2.9476	2.8662	3.3148	2.9406	2.8087
$\hat{\tau}_k^{jack}$	3.3342	3.0563	2.5459	3.2858	3.0523	2.8874	3.5218	2.8478	2.0739	3.4818	2.8241	3.6730	3.2706	2.9424	2.8759	3.3156	2.9417	2.1025
$\hat{\tau}_k^{elastic}$	3.2321	3.0135	2.8247	3.1950	3.0073	2.8654	3.3192	2.9003	2.3171	3.2423	2.8429	1.2068	3.1363	2.9746	3.0375	3.1761	2.9610	3.5205

Conclusion:

- In this project, we demonstrate the over-fitting issue with the endogenous stratification in random experiments through mathematical calculations and data simulations.
- Then we introduce and show how the Jackknife and elastic-net estimators could help to resolve the concerns with simulations using first JTPA data and then simulated data.
- With the results, we conclude that the Jackknife and elastic-net techniques can substantially improve the small sample behavior by the full-sample endogenous stratification estimator.