Revisiting and Improvement of Endogenous Stratification In Random Experiments

Shuo Wang, Nancy Huang, Yaze Gao

April 25, 2022

1 Introduction

The Endogenous Stratification Method advocated by David Kent is increasingly utilized in medical research recently. The general idea of this method is to first regress the outcome variable on baseline characteristics by using the full sample of experimental controls and then use the estimated coefficients from first stage to generate predicted potential outcomes without treatment for all sample units. However, this method is considered 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. This idea is contrary to the idea of exogenous stratification which will stratify units based on their pre-treatment covariates. In 2014, Abadie, Chingos, and West [1] 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. In our project, we are going to bring up the current concerns in the endogenous stratification in random experiments and demonstrate the issue of endogenous stratification in random experiments by digging into the theory which means we will go through the calculation of the idea. Then we will present our method which can help to resolve the concerns in current endogenous stratification in random experiments, and illustrate our method by simulating data.

2 Method

2.1 Exogenous Stratification

Before discussing endogenous stratification in random experiment, we are going to take a look of exogenous stratification. We want to get the stratification based on some exogenous variables. The assumptions for exogenous stratification are regular causal assumptions, including SUTVA about no interference and no hidden variations of treatment, unconfoundedness assumption, and overlapping assumption. Assume the assumptions are satisfied, in exogenous stratification, we estimate average treatment effect based on exogenous variables by applying the following equation:

$$\begin{aligned}\tau_k &= \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < f(X) \leq c_k] \\ &= \mathbb{E}[Y^{obs} \mid Z = 1, X, c_{k-1} < f(X) \leq c_k] - \mathbb{E}[Y^{obs} \mid Z = 0, X, c_{k-1} < f(X) \leq c_k]\end{aligned}$$

If the exact exogenous variable X is known before experiment, the strata will not be affected by the assignment mechanism. Within each strata, there are a lot of methods such as stratified random experiment or weighting methods that can be utilized to get unbiased estimator for τ_k under some particular designs. For example, in stratified or block random experiment, by using Neyman's Repeated Sampling Approach, we not only can get the within stratum unbiased estimator for average treatment effect τ_k but also obtain the variance and interval for the unbiased estimator. Moreover, for weighting methods such as inverse weighting probability and overlap weights, if the covariates between different groups exhibit sufficient overlap and the procedure is designed correctly, the average treatment effect τ_k will be an unbiased estimator of the estimand of interest. The correct procedure, where the correct propensity score takes an important role, can be explored following the procedure of Homework 2 Question 2. Based on the discussion above, we can see that exogenous stratification could be an appropriate method to estimate the average treatment effect τ_k.

2.2 Endogenous Stratification

In some situation, we want to estimate treatment effect of the subgroup where the potential outcome in the absence of treatment is much lower or higher than average. In this case, we need to estimate the potential outcome without treatment for all units first, and get the stratification based on predicted $Y(0)$. If we do not have an alternative way to get $Y(0)$ for all units externally, we can only in-sample observed outcome to proceed prediction. Hence, we can only get which strata does one treated unit belong to after the experiment is done, which is definitely influenced by assignment mechanism, and many other randomness. In this endogenous stratification, the equation we are trying to estimate is shown below.

$$\mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < Y(0) \leq c_k]$$

In order to estimate the above equation, the potential outcome Y(0) needs to be estimated because for each unit, only one of the potential outcomes is observable. In order to derive the estimation, conducting a regression model of Y(0) on X would be a nice way. However, this method exists some concerns. There is no guarantee that the assumptions for this method can be satisfied. In addition, it is impossible to evaluate whether the model is correct or not. Moreover, the exact strata cannot be estimated by this method. Hence, instead of estimating the above equation, to find out the true value of strata, we are now focusing on the equation below:

$$\tau_k = \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < X^T\beta \leq c_k], \tag{1}$$

where we hope the linear assumption $\mathbb{E}(Y(0)|X) = X^T\beta$ holds.
In general, for the endogenous stratification, according to Abadie *et al.*, it is popular to practice the experimental data by groups constructed on the basis of the predicted values from a regression of the outcome on baseline covariates for the full sample of experimental controls. We can estimated β by utilizing $\hat{\beta} = (\sum_{i=1}^{N} x_i(1-z_i)x_i')^{-1}\sum_{i=1}^{N} x_i(1-z_i)y_i$, where x_i is the vectors of baseline characteristics for both treatment and control group; z_i is the treatment for each units; and y_i is the potential outcomes for both treatment and control group. However, since we may never know what the true β is and we use in-sample estimation, the overfitting issue in regression error, $\epsilon_i = y_i - x_i'\beta$, for untreated units will result in positive bias in ATE for units with low predicted outcomes, and negative bias in ATE for units with high predicted outcomes eventually. Using the data on out-of-sample untreated units to estimate a prediction model for the potential outcome without treatment, the estimation of β might result in smaller bias than the in-sample estimation. However, this out-of-sample implementation makes sense when there exists externally validated models to predict the potential outcome without treatment for all observations. If there is no externally validated models, the alternative is to use in-sample data on the relationship between the interested outcome and covariates for the experimental controls to estimate potential outcomes without treatment for all experimental units. Nevertheless, the alternative exists some concerns on over-fitting issue we mentioned before. The following section will go through the further analysis based on equation 1.

3 Analysis

According to regression analysis, we can know that $\hat{\tau_k} = \frac{\sum_{i=1}^{n} YB_i(k)Z_i}{\sum_{i=1}^{n} B_i(k)Z_i} - \frac{\sum_{i=1}^{n} YB_i(k)(1-Z_i)}{\sum_{i=1}^{n} B_i(k)(1-Z_i)}$, where $B_i(k) = \mathbb{1}_{\{c_{k-1} < x_i^T\hat{\beta} \leq c_k\}}$, is converge to $\mathbb{E}[Y^{obs} \mid Z = 1, c_{k-1} < X^T\beta \leq c_k] - \mathbb{E}[Y^{obs} \mid Z = 0, c_{k-1} < X^T\beta \leq c_k]$.
Under random experiment design, we can see that

$$\begin{aligned}
&\mathbb{E}[Y^{obs} \mid Z = 1, c_{k-1} < X^T\beta \leq c_k] - \mathbb{E}[Y^{obs} \mid Z = 0, c_{k-1} < X^T\beta \leq c_k] \\
=&\mathbb{E}[Y(1) \mid Z = 1, c_{k-1} < X^T\beta \leq c_k] - \mathbb{E}[Y(0) \mid Z = 0, c_{k-1} < X^T\beta \leq c_k] \\
=&\mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < X^T\beta \leq c_k] \\
=&\tau_k
\end{aligned}$$

So $\hat{\tau_k} \to \tau_k$. However, according to Abadie *et al.*, the estimation of average treatment effect in finite sample and in-sample data will lead to the bias with the patterns that units with low predicted outcomes will have upward bias while units with high predicted outcomes will have downward bias. The reason for displaying this pattern is shown below.

In the model, the population counterpart of $\hat{\beta}$ can be considered as

$$\beta = (\mathbb{E}[xx^T \mid Z=0])^{-1}\mathbb{E}[xy \mid Z=0] \text{ and } \hat{\beta} = (\sum_{i=1}^{N} x_i(1-z_i)x_i^T)^{-1}(\sum_{i=1}^{N} x_i(1-z_i)y_i)$$

Now, let's take the minimum strata

$$c_0 < X^T\beta \le c_1, \text{ where } x^T\beta = \mathbb{E}[Y(0) \mid Z=0, X=x]$$

Then the average treatment effect τ_1 is

$$\tau_1 = \mathbb{E}[Y_i(1) \mid Z_i = 1, c_0 < x_i^T\beta \le c_1] - \mathbb{E}(Y_i(0) \mid Z_i = 0, c_0 < x^t\beta \le c_1) \text{ and}$$

$$\hat{\tau_1} = \frac{\sum_{i=1}^n YB_i(1)Z_i}{\sum_{i=1}^n B_i(1)Z_i} - \frac{\sum_{i=1}^n YB_i(1)(1-Z_i)}{\sum_{i=1}^n B_i(1)(1-Z_i)}, \text{ where } B_i(1) = \mathbb{1}_{\{c_0 < x_i^T\hat{\beta} \le c_1\}}$$

Assume that there exist j, where $Z_j = 0$ and $\epsilon_j = y_j - x_j^T\beta = y_j - \mathbb{E}(y|X = x_j) \ll 0$.

Let $H = X(X^TX)^{-1}X^T$, which is the hat matrix in linear regression. One of the properties of H is $H^2 = H$, which can be written as

$$\begin{aligned}
&\sum_{k=1}^{n} H_{ik}H_{ik} = H_{ii} \\
\Rightarrow &\sum_{k=1}^{n} H_{ik}^2 = H_{ii} \qquad\qquad (H^T = H) \\
\Rightarrow &H_{ii} = H_{ii}^2 + \sum_{k\neq i} H_{ik}^2 \ge H_{ii}^2 \\
\Rightarrow &H_{ii} \in [0,1]
\end{aligned} \tag{2}$$

When we take intercept into account,

$$H_{ii} = \frac{1}{n} + (x_i - \bar{x})^T[(\tilde{X} - \mathbb{1}_n\bar{x}^T)^T(\tilde{X} - \mathbb{1}_n\bar{x}^T)]^{-1}(x_i - \bar{x}),$$

where $\tilde{X}$ is the design matrix for all control units, but does not include intercept for now.$\mathbb{1}_n$ is a column vector with dimension n. Now let us prove that the matrix in the middle is a positive definite matrix.

$\forall v \in \mathbb{R}^p$,

$$\begin{aligned}
&v^T(\tilde{X} - \mathbb{1}_n\bar{x}^T)^T(\tilde{X} - \mathbb{1}_n\bar{x}^T)v \\
=&[(\tilde{X} - \mathbb{1}\bar{x})v]^T[(\tilde{X} - \mathbb{1}\bar{x}^T)v] \\
=&\sum_{m=1}^{p} {v'_m}^2 \ge 0,
\end{aligned}$$

where $v'_m = [(X - \mathbb{1}\bar{x}^T)v]_m$.

Since we require the design matrix to be invertible, we can say that $(\tilde{X} - \mathbb{1}_n\bar{x}^T)^T(\tilde{X} - \mathbb{1}_n\bar{x}^T)$ is positive definite, and the inverse of it is also positive definite. This implies that $(x_i - \bar{x})^T[(\tilde{X} - \mathbb{1}_n\bar{x}^T)^T(\tilde{X} - \mathbb{1}_n\bar{x}^T)]^{-1}(x_i - \bar{x})$ should be a positive scalar. Therefore, we can say that $H_{ii} \in [\frac{1}{n}, 1]$.

Continuing on our analysis,

$$\begin{aligned}
&x_j^T\hat{\beta} - x_j^T\beta \\
=&\hat{y_j} - x_j^T\beta \\
=&(Hy)_j - x_j^T\beta \\
=&H_{jj}y_j + \sum_{i\neq j} H_{ji}y_i - x_j^T\beta
\end{aligned} \tag{3}$$

When y_j is positive,

$$\text{equation (3)} \ \leq y_j - x_j^T\beta + \sum_{i\neq j} H_{ji}y_j, \qquad (H_{ii} \leq 1)$$

As long as y_j is substantially less than $x_j^T\beta$, this term will be less than zero.
When y_j is negative,

$$\text{equation (3)} \ \leq \frac{1}{n}y_j - x_j^T\beta + \sum_{i\neq j} H_{ji}y_i, \qquad (H_{ii} \geq \frac{1}{n})$$

Since $y_j \ll x_j^T\beta$, we can expect that this term will be less than zero. Therefore, we can expect that when $y_j - x_j^T\beta$ is substantially less than zero, $x_j^T\hat{\beta} - x_j^T\beta$ is less than zero. Thus we show that if there exist such units, they will more likely to fall into interval with small predicted value, which means they will tend to cluster at the left strata with small value of $x^T\hat{\beta}$, and they have very small y^{obs}.

In finite sample, if $x_j^T\hat{\beta} \leq c_1$ and $x_j^T\beta > c_1$, then it will leads to bias in strata $(c_0, c_1]$, since this extreme value for y^{obs} is used to calculate the treatment effect because of in-sample setting. On the other hand, suppose we know true β and we define the stratification based on $X^T\beta$, with a large number of samples, when we calculate treatment effect using the observed outcome, the outcome of unit with large negative regression error will compensate with the one with large positive regression error, and finally the mean of all observed outcome will approximately equal to $\mathbb{E}[Y_j(0) \mid Z_j = 0, c_0 < x_j^T\beta \leq c_1]$. But if we use the predicted outcome $X^T\hat{\beta}$ to define the stratification, we do not have this compensating benefit. Since the units with $y_j \ll x_j^T\beta = \mathbb{E}(Y_j(0)|X_j = x_j)$ are clustered at the left strata, and the units with $y_j \gg x_j^T\beta = \mathbb{E}(Y_j(0)|X_j = x_j)$ are clustered at the right strata, these two y_j can not be averaged and compensated to approximate the population expectation under stratification setting. Hence, the treatment effect, basically the average of observed outcome, will be biased within each strata.

More concretely about the direction of bias within each strata, we still focus on the left strata with small $x^T\hat{\beta}$. For the unit with large negative regression error and thus very small observed outcome, it will be used to calculate the treatment effect in strata $(c_0, c_1]$. Nevertheless, the small observed value of y_j will generate negative bias in $\frac{\sum_{i=1}^N y_i(1-z_i)\mathbb{1}_{\{c_0<x_i^T\hat{\beta}\leq c_1\}}}{\sum_{i=1}^N (1-z_i)\mathbb{1}_{\{c_0<x_i^T\hat{\beta}\leq c_1\}}}$, and thus result in positive bias in $\hat{\tau}_1$, which means $\hat{\tau}_1 > \tau_1$. Similarly, we can get the conclusion that $\hat{\tau}_1 < \tau$ for the right strata with large $x^T\hat{\beta}$ values.

Further more, from equation (2) and equation (3), we can see that when H_{jj} is close to 1, or equivalently, the unit j's covariates are far from the mean of covariates across all units, we will have $H_{ji}, \forall i \neq j$ close to 0, then the influence of y_j on $\hat{\beta}$ is very significant. It indicates that if there exists one observation with large regression error ϵ_j and large leverage score H_{jj}, the $\hat{\beta}$ estimated from finite samples including this observation will be much biased from the true β, which will lead to large bias of treatment effect within each strata stratified according to in-sample potential outcome prediction.

In this section, the reason for causing the issue in endogenous stratification in random experiment is explained. The solution for the over-fitting issue will be discussed in the following section.

4 Solution and Simulation

4.1 Method Introductions

In order to avoid the over-fitting issue, instead of utilizing out-of-sample information, which might be hard to get, there are several ways to prevent over-fitting issue. For example, before we fit the model, we can use feature selection first. However, this might violate the assumptions for causal inference or lead to lost of information. During the model-fitting stage, we can utilized methods such as cross-validation, regularization, Bayesian shrinkage method, data augmentation, and sample splitting to restrain over-fitting issue. After we fit the model, techniques of bias correction such

as jackknife and bootstrap resampling can be used for preventing over-fitting. In this project, we are going to examine performances of one of the regularization methods and one of the bias correction method. The methods we chose are jackknife bias correction and elastic-net regularization.

The jackknife bias correction can be viewed as a linear combination approximation, aiming at making the biases of estimators with different sample sizes cancel each other [2]. In our project, we will utilize the delete-m jackknife. The elastic-net regularization is a combination of l_1 and l_2 penalties.

The procedures of delete-m jackknife bias correction is described below [3]. In our setting, the estimator $\hat{\beta}$ is constructed as

$$\hat{\beta} = (\sum_{i=1}^{n} x_i(1-z_i)x_i')^{-1} \sum_{i=1}^{n} x_i(1-z_i)y_i \triangleq f_n(x_1, \ldots, x_n).$$

Let m be the number of mutually exclusive groups with size of g. For example, if $g = 2$, $m = \frac{n}{g} = \frac{n}{2}$. In order to estimate the bias for $\hat{\beta}$, the jackknife resampling technique is involved. When $g = 2$, the jackknife replicates are constructed [4]:

$$\begin{aligned}
\hat{\beta}_{(1)} &= f_{n-2}(x_3, x_4, x_5, \ldots, x_n) \\
\hat{\beta}_{(2)} &= f_{n-2}(x_1, x_2, x_5, x_6, \ldots, x_n) \\
&\vdots \\
\hat{\beta}_{(m)} &= f_{n-2}(x_1, x_2, \ldots, x_{n-2})
\end{aligned}$$

where each replicate is a "leave-g-out" estimate based on the jackknife subsampling consisting of all but g of the data points. Then we can estimate the bias with

$$\hat{\text{bias}}(\hat{\beta})_{jack} = (m-1)(\bar{\beta}_{(m)} - \hat{\beta}),$$

where $\bar{\beta}_{(m)} = m^{-1} \sum_{j=1}^{m} \hat{\beta}_{(j)}$, and $\hat{\beta}$ is calculated using all samples. The estimator after bias correction is

$$\tilde{\beta}_{jack} = \hat{\beta} - \hat{\text{bias}}(\hat{\beta})_{jack} = m\hat{\beta} - (m-1)\bar{\beta}_{(m)}.$$

Then we can estimate $\hat{\tau_k}$ by $\hat{\tau}_k^{jack} = \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < x_i^T \tilde{\beta}_{jack} \leq c_k]$.

The estimator for elastic-net regularization [5] is defined as

$$\hat{\beta}_{elastic} = \underset{\beta_{elastic}}{\arg\min}(\|y - X\beta_{elastic}\|^2 + \lambda_2 \|\beta_{elastic}\|_2^2 + \lambda_1 \|\beta_{elastic}\|_1).$$

There are two stages for the procedure. First for each fixed λ_2 it finds the ridge regression coefficients, and then does a LASSO type shrinkage. In our project, we define $\alpha = 0.5$ which implies that λ_1 equals to λ_2. Then we utilize cross-validation to choose the best λ; the result could be optimized.
Then we can estimate $\hat{\tau_k}$ by $\hat{\tau}_k^{elastic} = \mathbb{E}[Y(1) - Y(0) \mid c_{k-1} < x_i^T \hat{\beta}_{elastic} \leq c_k]$.

4.2 Data Description

The data set utilized in our project is the JTPA data set. JTPA data set is from the National JTPA study which is a large experimental evaluation of a job training program commissioned by the United States Department of Labor in 1980. The data set contains 19 variables with 11204 observations. The variables are listed below.

- `id`: Observation id
- `earnings`: Total earnings over the 30 months period following the assignment into the treatment or control group
- `offer`: Randomly assigned offer of job training services
- `train`: Receipt of job training
- sex: 0 represents female, and 1 represents male.
- `hsorged`: High school graduate or GED
- black: 0 represents non-black people, and 1 represents black people
- `hispanic`: 0 represents non-Hispanic people, and 1 represents Hispanic people

- married: 0 represents non-married people, and 1 represents married people
- wkless13: 0 represents the applicant worked 13 or more weeks in the 12 months prior to the assignment, and 1 represents the applicant worked less than 13 weeks in the 12 months prior to the assignment
- `afdc`: 0 represents non-afdc, and 1 represents afdc
- age2225: 0 represents otherwise, and 1 represents age ranging from 22 to 25
- `age2629`: 0 represents otherwise, and 1 represents age ranging from 26 to 29
- age3035: 0 represents otherwise, and 1 represents age ranging from 30 to 35
- age3644: 0 represents otherwise, and 1 represents age ranging from 36 to 44
- age4554: 0 represents otherwise, and 1 represents age ranging from 45 to 54
- class_tr: recommended service strategy – class
- `ojt_jsa`: recommended service strategy – on-the-job-training
- f2sms: 0 represents otherwise, and 1 represents earning data are from a second follow-up survey

In our project, we only use the observations of male and discard the rest of the observations. Hence, the total number of observations utilized is around 4000 to 5000. Our model includes all the variables except `sex`, id, and train.

Besides of the actual data set, we also utilize the simulated data set. The next section will include the detailed procedures of our simulation.

4.3 Simulation

Similar to what Abadie *et al.* did in their paper, we also compare the jackknife bias corrected estimator and elastic-net regularized estimator with the simulated data which is generated as follows:

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

(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. The result is in Table 2.

5 Result

The estimation results of JTPA data are reported in Table 1. Panel A shows the unadjusted and adjusted estimated ATE without stratification and Panel B demonstrates the ATE by predicted outcome group with stratification. Both the unadjusted and adjusted estimator in Panel A suggest a positive treatment effect. In Panel B, all of the unadjusted and the adjusted endogenous stratification estimators suggest positive effect as Panel A does. The magnitude of ATE, however, varies by predicted outcome group and by the correction / regularization techniques. For the low predicted outcome group, both the unadjusted and adjusted full-sample estimators $\hat{\tau}_k$ have much larger magnitude than in Panel A while the estimators are slightly smaller than those in Panel A for the medium and high outcome group. The ATE of high outcome group is the closest to that in the first row of the table. With Jackknife correction, the results are very similar to that of the full-sample estimator. The elastic regularization method produces relatively smaller treatment effect than the other two estimators in general, with the low predicted outcome group still having a larger ATE while the medium and high group have smaller ATE than the ones in Panel A.

The results of simulated data as described in Section 4.3 are shown in Table 2. The true treatment effect should be around 3 but the full-sample stratification estimator $\hat{\tau}_k$ will sometimes produce very small or even negative treatment effect, and is particularly severe when the sample size is small ($N = 200$) or the number of regressors is large ($K = 30$) which indicates the over-fitting

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

issue as discussed in previous sections. The Jackknife and elastic-net estimators produce mostly similar results and tend to correct the estimators so that the estimated value is closer to the true value than the full-sample stratification estimator. As K increases and N decreases, the bias of the full-sample stratification estimator increases, while the Jackknife and elastic-net estimators are less sensitive to K and N and therefore can produce more stable estimates. In summary, the Jackknife and elastic-net estimators outperform the full-sample stratification estimator.

6 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.

Appendix

Code: https://github.com/star7878/STA-640-final/blob/main/STA640_final.Rmd

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

References

[1] A. Abadie, M. Chingos, and M. West, "Endogenous stratification in randomized experiments," *The Review of Economics and Statistics*, vol. 100, no. 4, pp. 567–580, 2018.

[2] J. Jiao and Y. Han, "Bias correction with jackknife, bootstrap, and taylor series," 2017. [Online]. Available: https://arxiv.org/abs/1709.06183

[3] F. Busing, E. Meijer, and R. Leeden, "Delete-m jackknife for unequal m," *Statistics and Computing*, no. 3, pp. 3–8, 1999.

[4] A. Cameron and P. Trivedi, *Microeconometrics: Methods and Applications*, 05 2005.

[5] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 2005.