Nonconstant Error Variance in Generalized Propensity Score Model

Doyoung Kim, Chanmin Kim
Department of Statistics, Sungkyunkwan University

Introduction

Although several generalized propensity score (GPS) models have been proposed to balance covariates among treatment groups in observational study, **the existence of heteroskedasticity in treatment variable distributions** has not been considered much in parametric forms.

We propose a novel GPS method to handle non-constant variance depending on certain covariates in the treatment model by extending Xiao et al. (2020) with weighted least squares method.

We first consider circumstances when heteroskedasticity in GPS model occurs with parametric propensity score model procedures, suggest existing variance function estimation, and propose modified caliper metric matching methods with simulation results.

Preliminaries

Annotation

- $\boldsymbol{X}$: Covariate
- T : Treatment

Parametric Generalized Propensity Score Model

- $T_i|\boldsymbol{X}_i \sim N(\boldsymbol{X}_i^T\beta, \sigma^2)$
- Propensity score $R_i = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(T_i-\boldsymbol{X}_i^T\beta)^T(T_i-\boldsymbol{X}_i^T\beta)}{2\sigma^2})$

Caliper Metric Matching[1]

$$m_{GPS}(e,w) = \underset{j:w_j \in [w-\delta, w+\delta]}{argmin}\left\|\left(\lambda e^*\left(w_j, \boldsymbol{X}_j\right), (1-\lambda)w_j^*\right) - \left(\lambda e^*, (1-\lambda)w^*\right)\right\|$$

- w_j^* : jth standardized treatment
- $e^*(w_j^*, \boldsymbol{X}_j)$: Standardized propensity score with jth treatment & jth covariate
- δ : Caliper parameter
- λ : Scale parameter between 0 & 1

Covariate Balance (Global Measure)[1]

$$\left|\sum_{i=1}^{I}\sum_{k=1}^{m_i} n_{ik}\boldsymbol{X}_{ik}^* \mathrm{W}_{ik}^*\right| < \boldsymbol{\varepsilon}_1$$

- $\boldsymbol{X}_{ik}^* = \mathbf{S}_X^{\frac{1}{2}}(\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{ik})$, $\mathrm{W}_{ik}^* = \mathrm{S}_W^{\frac{1}{2}}(\mathrm{W}_{ik} - \overline{\mathrm{W}}_{ik})$, $\boldsymbol{\varepsilon_1}$ is a pre-specified threshold (0.1)[2]
- $I = \left\lfloor\frac{w^1-w^0}{2\delta} + \frac{1}{2}\right\rfloor$ where $[w^0, w^1]$ is range of treatment of interest
- m_i : the number of units within the block $[w_i - \delta, w_i + \delta]$ where $\{w_1 = w^0 + \delta, \ldots, w_I = w^0 + (2I-1)\delta\}$

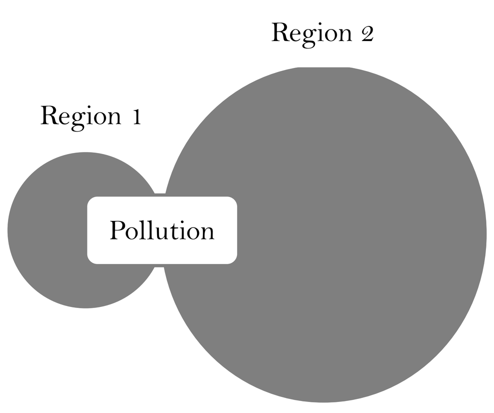
1) Wu, X., Mealli, F., Kioumourtzoglou, M.-A., Dominici, F. & Braun, D. (2020), 'Matching on Generalized Propensity Scores with Continuous Exposures'
2) Zhu, Y. Coffman, D. L. & Ghosh, D. (2015), 'A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments', *Journal of causal inference*

Motivation

Existence of heteroskedasticity

- Variance of parametric GPS model depends on certain $\boldsymbol{X}$
- Range of treatment the observation takes varies

e.g.
$\boldsymbol{X}$: Age, Race, Sex, Region …
T : Air pollution exposure
Y : Mortality among 65+ years old



Variance Function Model[3]

$$T_i = \mathrm{f}(\boldsymbol{X}_i, \beta) + g(\mu_i, z_i, \theta)\varepsilon_i\ ,\ \varepsilon_i \sim \mathrm{N}(0, \sigma^2) \quad \leftrightarrow \quad T_i|\boldsymbol{X}_i \sim N(\boldsymbol{X}_i^T\beta, \sigma^2 g^2(\mu_i, z_i, \theta))$$

- μ_i : $\mathrm{f}(\boldsymbol{X}_i, \beta)$
- z_i : Subset of $\boldsymbol{X}_i$
- θ : Coefficient(s) of variance function

- Graphical methods for defining the form of variance function

e.g.

$$g(\mu_i\ z_i, \theta) = 1 + \theta_1 z_i + \theta_2 z_i^2$$

if absolute residuals $|r_i|$ & z_i shows quadratic form

3) Carroll, R. J., Ruppert, D., (1988), 'Transformation and Weighting in Regression', Chapman and Hall

Model Specification & Proposal[2)3)]

1) Estimating $\boldsymbol{\theta}$ & $\boldsymbol{\beta}$

(1) Obtain LS estimator $\hat{\beta}_{LS}$, $\hat{\beta}^{(0)} = \hat{\beta}_{LS}$, $k = 0$
(2) Calculate

$$\sum_{i=1}^{N}\left\{\frac{r_i^2 \prod_{j=1}^{N} g\left(\mu_j\left(\hat{\beta}^{(k)}\right), z_j, \theta\right)^{\frac{1}{n}}}{g\left(\mu_i\left(\hat{\beta}^{(k)}\right), z_i, \theta\right)}\right\}^2$$

to obtain $\hat{\theta}$
(3) Use the estimated weights $\widehat{w}_i = \frac{1}{g^2(\mu_i(\hat{\beta}^{(k)}), z_i, \hat{\theta})}$ to obtain $\hat{\beta}_{GLS}$
(4) Set $k = k+1$, let $\hat{\beta}^{(k)} = \hat{\beta}_{GLS}$ and return to (2)

2) Implement GPS matching approach

(1) Estimate GPS assuming $T_i|\boldsymbol{X}_i \sim N(\boldsymbol{X}_i^T\beta, \sigma^2 g^2(\mu_i, z_i, \theta))$
Propensity score $R_i = \frac{1}{\sqrt{2\pi\sigma^2 g^2(\mu_i, z_i, \theta)}}\exp(-\frac{(T_i-\boldsymbol{X}_i^T\beta)^T(T_i-\boldsymbol{X}_i^T\beta)}{2\sigma^2 g^2(\mu_i, z_i, \theta)})$
(2) Select parameters (λ, δ) minimizing covariate balance through grid search
(3) Match individuals based on one of proposed caliper metric matching methods

<Proposed Method 1>

$$m_{GPS}(e,w) = \underset{j:w_j \in [w-\delta, w+\delta]}{argmin}\left\|\left(\alpha v_j^*, \beta w_j^*, (1-\alpha-\beta)\left(w_j, \boldsymbol{X}_j\right)\right) - \left(\alpha v^*, \beta w^*, (1-\alpha-\beta)e^*\right)\right\|$$

<Proposed Method 2>

$$d_j = |\sigma^2 g(\mu, z, \theta) - \sigma^2 g(\mu_j, z_j, \theta)|$$

$$m_{GPS}(e,w) = \underset{j:w_j \in [w-\delta, w+\delta]}{argmin}\left\|\left(\lambda e^*\left(w_j, \boldsymbol{X}_j\right), (1-\lambda)w_j^*\right) - \left(\lambda e^*, (1-\lambda)w^*\right)\right\| \text{ for } d_j^* < \gamma$$

(4) Impute $\hat{Y}_j(w) = Y_{matched}^{obs}$ for $j = 1, \ldots, \mathrm{N}$ for all predetermined treatment levels w.
(5) With obtained $\hat{\mu}(w) = \hat{E}\{\hat{Y}_j(w)\}$, get a smoothed average causal treatment-response function

Simulation

Settings

- $n = 200$, simulation replicates $= 100$
- Covariate : $\boldsymbol{X}_1 - \boldsymbol{X}_4 \sim \mathrm{N}(0, I_4)$, $\boldsymbol{X}_5 \sim U(-2,2)$, $\boldsymbol{X}_6 \sim U(-3,3)$
- Treatment : $10 + 0.9\boldsymbol{X}_1 + 1.8\boldsymbol{X}_2 - 0.9\boldsymbol{X}_3 + 0.9\boldsymbol{X}_4 + 0.9\boldsymbol{X}_5 + 0.9\boldsymbol{X}_6$
- Variance function : $\sigma^2 g^2(\mu_i, z_i, \theta) = 1^2(0.3\boldsymbol{X}_4^2 + 1.2\boldsymbol{X}_4 + 1.2\boldsymbol{X}_6)^2$
- Outcome : $Y \mid W, \boldsymbol{X} \sim N\{\mu(W, \boldsymbol{X}), 1.5^2\}$
$\mu(W, \boldsymbol{X}) = 15 + 1.2\boldsymbol{X}_1 + 1.2\boldsymbol{X}_2 + 1.2\boldsymbol{X}_3 + 1.2\boldsymbol{X}_4 + 1.2\boldsymbol{X}_5 + \boldsymbol{X}_6 + 0.1W + 0.1W^2 + 0.25\boldsymbol{X}_1 W + 0.1\boldsymbol{X}_6 W$

Metrics[2]

- Absolute Bias

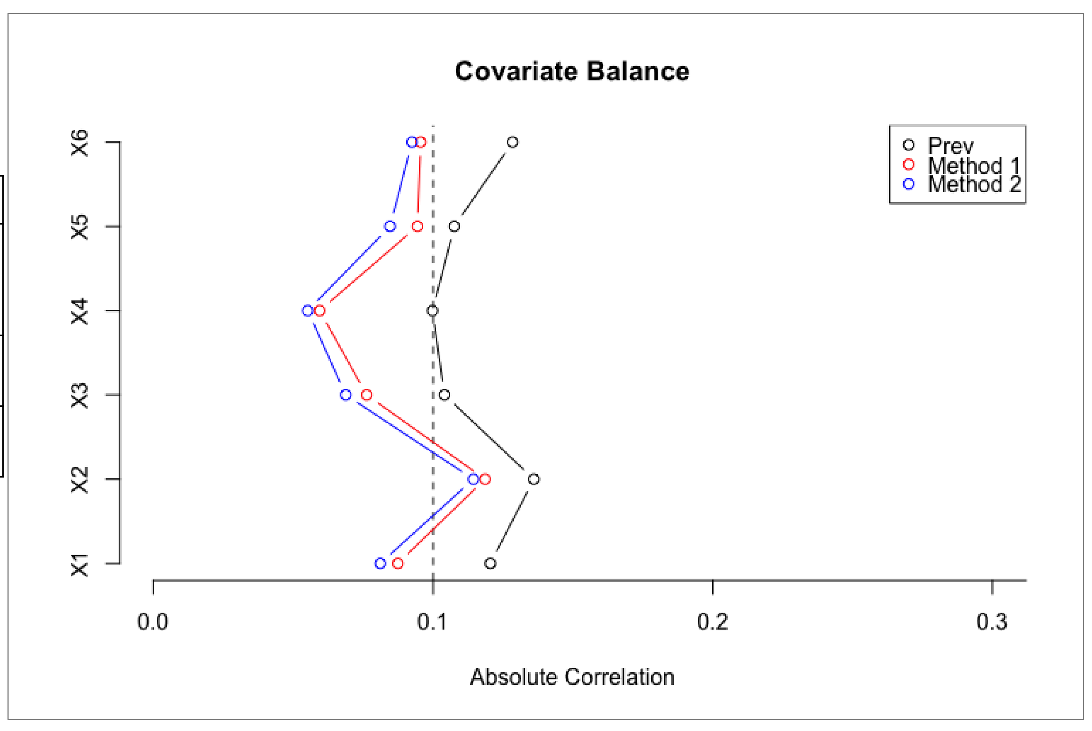
$$\int_{\widehat{W}^*}\left|\frac{1}{S}\sum_{s=1}^{S}\hat{Y}_s(w) - Y(w)\right| f_w(w)dw \quad \text{replaced by} \quad \frac{1}{M}\sum_{i=1}^{M}\left|\frac{1}{S}\sum_{s=1}^{S}\hat{Y}_s(w_i) - Y(w_i)\right|,\ i = 1, \ldots, M$$

- MSE

$$\int_{\widehat{W}^*}\left[\frac{1}{S}\sum_{s=1}^{S}\{\hat{Y}_s(w) - Y(w)\}^2\right]^{1/2} f_w(w)dw \quad \text{replaced by} \quad \frac{1}{M}\sum\left[\frac{1}{S}\sum_{s=1}^{S}\{\hat{Y}_s(w_i) - Y(w_i)\}^2\right]^{1/2}$$

Results

	Absolute Bias	MSE
Previous Method	9.728	13.014
Method 1	9.743	12.041
Method 2	9.665	12.025



Future Study

- Simulation metrics computation with marginal probabilities of treatment & full range of continuous treatment through kernel smoothing
- Application with heteroskedastic data