Matching on generalized propensity scores with continuous exposures

X. Wu, F. Mealli, M.-A. Kioumourtzoglou, F. Dominici, D. Braun

https://arxiv.org/abs/1812.06575

Harvard University, University of Florence, Columbia University

July 31, 2019

Motivation

- In many observational studies, the treatment (or exposure) is continuous in nature.

Motivation

- In many observational studies, the treatment (or exposure) is continuous in nature.
- Estimating the causal effects in those studies is challenging
  1. there is a large set of covariates that are associated with both the exposure and outcome of interest (potential confounders).
  2. one has to allow for flexible estimation of the exposure-response functions on a continuous scale.

Previous Studies

Existing causal inference approaches including,

- HI-GPS estimator : adjusting the estimated propensity socre (PS) as a covariate in the outcome model (Hirano & Imbens 2004).
- IPTW estimator : using the estimated PS for inverse probability of treatment weighting (Robins et al. 2000).
- Doubly robust (DR) estimator : an augmentation of the IPTW estimator that is more robust to model misspecification (Bang & Robins 2005, Kennedy et al. 2017).

Previous Studies

Existing causal inference approaches including,

- HI-GPS estimator : adjusting the estimated propensity socre (PS) as a covariate in the outcome model (Hirano & Imbens 2004).
- IPTW estimator : using the estimated PS for inverse probability of treatment weighting (Robins et al. 2000).
- Doubly robust (DR) estimator : an augmentation of the IPTW estimator that is more robust to model misspecification (Bang & Robins 2005, Kennedy et al. 2017).

None have been applied to air pollution epidemiology context.

1. They require that either the PS model or the outcome model, or both, are correctly specified.
2. Both IPTW and DR approaches rely on weighting and are therefore sensitive to extreme values of PS.
3. Assessing covariate balance when using these approaches is not straightforward.

Objectives

To develop new framework for generalized propensity score caliper **matching** with replacement to estimate the causal effect using observational data in the settings with **continuous exposure**.

Objectives

To develop new framework for generalized propensity score caliper **matching** with replacement to estimate the causal effect using observational data in the settings with **continuous exposure**.
Why matching ?

1. it is more robust to misspecifications of the generalized propensity score model (Waernbaum 2012).
2. it is free of dependence on the outcome model specification.
3. it allows for the transparent assessment of covariate balance.

Why matching with generalized propensity score ?

1. it is consistent and asymptotically normal when matching on a scalar function, such as the (generalized) propensity score (Abadie & Imbens 2006).

Objectives

To develop new framework for generalized propensity score caliper **matching** with replacement to estimate the causal effect using observational data in the settings with **continuous exposure**.

Why matching ?

1. it is more robust to misspecifications of the generalized propensity score model (Waernbaum 2012).
2. it is free of dependence on the outcome model specification.
3. it allows for the transparent assessment of covariate balance.

Why matching with generalized propensity score ?

1. it is consistent and asymptotically normal when matching on a scalar function, such as the (generalized) propensity score (Abadie & Imbens 2006).

Why matching with replacement ?

1. It eases the computational burden, by avoiding computationally cumbersome optimal/sequential greedy matching algorithms that are needed for matching without replacement (Imbens & Rubin 2015).

Notations

- N denote the study sample size.
- $\mathbf{C}_j$ denote the pretreatment covariates for unit j, which is characterized by a M-vector $(C_{1j}, \ldots, C_{Mj})$, For each unit $j \in \{1, 2, \ldots, N\}$.
- W_j denote the continuous exposure for unit j, $W_j \in \mathbb{W}$ with a range $[w^0, w^1]$.
- $Y_j(w)$ denote the counterfactual outcome for unit j at the exposure level w.
- The target estimand is $\forall w \in [w^0, w^1]$,

$$\mu(w) = E[Y_j(w)].$$

Assumptions

- **Assumption 1** (Consistency) : $W = w$ implies $Y = Y(w)$.
- **Assumption 2** (Overlap) : For all values of $\mathbf{c}$, the density function of receiving any possible exposure $w \in \mathbb{W} = [w^0, w^1]$ is positive :

$$f(w|\mathbf{c}) > 0 \text{ for all } w, \ \mathbf{c}.$$

Assumptions

- **Assumption 3** (Approximate Weak Unconfoundedness) : Let $I_j(.)$ be the indicator variable indicating if exposure level $W_j = \tilde{w}$ or not for $\tilde{w} \in [w - \delta, w + \delta]$, where δ is the caliper defined as the radius of the neighborhood around w, and it follows a positive sequence tending to zero as $N \to \infty$. For all $w \in \mathbb{W}$, in which w is continuously distributed with respect to $\mathbb{W} = [w^0, w^1]$, then

$$\{I_j(\tilde{w})\}_{\tilde{w} \in [w-\delta, w+\delta]} \perp\!\!\!\perp Y_j(w) \mid \mathbf{C}_j.$$

Assumptions

- **Assumption 4** (Smoothness) : Suppose the average exposure-response function $E\{Y_j(w)\}$ is continuous with respect to w, and $h \geq \delta$, where h is a sequence tending to zero as $N \to \infty$ and δ as previously defined, then

$$\lim_{h \to 0} E\{Y_j(w-h)\} = \lim_{h \to 0} E\{Y_j(w+h)\} = E\{Y_j(w)\}.$$

Generalized Propensity Scores

Definition. The generalized propensity scores (GPS) is the conditional density function of the exposure given pretreatment covariates : $\mathbf{e}(\mathbf{c}_j) = \{f_{W|\mathbf{C}_j}(w|\mathbf{c}_j), \forall w \in [w^0, w^1]\}$. The individual $e(w, \mathbf{c}_j) = f_{W|\mathbf{C}_j}(w|\mathbf{c}_j)$ are called realizations of $\mathbf{e}(\mathbf{c}_j)$.

Matching Function

We propose a one-to-one nearest neighbor caliper matching procedure with replacement, which jointly matches both on estimated GPS and exposure values.

Matching Function

We propose a one-to-one nearest neighbor caliper matching procedure with replacement, which jointly matches both on estimated GPS and exposure values.

1. We standardize both quantities via a standardized Euclidean transformation, i.e,

$$w_j^* = \frac{w_j - \min_j w_j}{\max_j w_j - \min_j w_j}, \; e^*(w_j, \mathbf{c_j}) = \frac{e(w_j, \mathbf{c_j}) - \min_j e(w_j, \mathbf{c}_j)}{\max_j e(w_j, \mathbf{c}_j) - \min_j e(w_j, \mathbf{c}_j)},$$

Matching Function

We propose a one-to-one nearest neighbor caliper matching procedure with replacement, which jointly matches both on estimated GPS and exposure values.

1. We standardize both quantities via a standardized Euclidean transformation, i.e,

$$w_j^* = \frac{w_j - \min_j w_j}{\max_j w_j - \min_j w_j}, \; e^*(w_j, \mathbf{c_j}) = \frac{e(w_j, \mathbf{c_j}) - \min_j e(w_j, \mathbf{c}_j)}{\max_j e(w_j, \mathbf{c}_j) - \min_j e(w_j, \mathbf{c}_j)},$$

2. Based on the standardized quantities, we propose the specifications of the matching function, i.e. caliper metric matching as follows :

$$m_{gps}(w, e) = \arg \min_{j: w_j \in [w-\delta, w+\delta]} ||(e^*(w_j, \mathbf{c}_j), w_j^*) - (e^*, w^*)||_{(\lambda, 1-\lambda)},$$

where $||.||_{(\lambda,1-\lambda)}$ is a pre-specified two-dimensional metric with weights $(\lambda, 1-\lambda)$ for the corresponding dimensions.

Select λ

$$m_{gps}(e, w) = \arg \min_{j: w_j \in [w-\delta, w+\delta]} ||(e^*(w_j, \mathbf{c}_j), w_j^*) - (e^*, w^*)||_{(\lambda, 1-\lambda)},$$

- The tuning parameter λ is to control the relative weight that is attributed to the distance measures of the exposure versus the GPS estimates.
- In practice, λ could be specified in the range $[0, 1]$ depending on the prioritization of adjusting for potential confounders and the potential effects heterogeneity due to different levels of exposures.
- The optimal λ could be specified by minimizing a utility function that measures the degree of covariate balance.

Select δ

$$m_{gps}(e, w) = \arg \min_{j: w_j \in [w-\delta, w+\delta]} ||(e^*(w_j, \mathbf{c}_j), w_j^*) - (e^*, w^*)||_{(\lambda, 1-\lambda)},$$

- The caliper δ is defined as the radius of the neighborhood around w, which means for any target exposure level w, we only allow for matches with an observed unit j satisfying $|| \ W_j - w \ || \leq \delta$.
- To achieve asymptotic properties for the matching estimator, $\delta = O(1/\sqrt{N})$ is chosen which guarantees consistency.
- In practice, theoretical justifications guide us to choose $\delta \approx \epsilon/\sqrt{N}$, and we set ϵ to be the range of the exposure levels, i.e. $\epsilon = \max_j w_j - \min_j w_j$.
- The optimal δ could be also specified by minimizing a utility function that measures the degree of covariate balance.

Proposed Method : GPS Matching

1. Design Stage : Fit a GPS model relating w to $\mathbf{C}$, $\hat{e}(w_j, \mathbf{c}_j) = \hat{g}_{\Phi}^{-1}(\mathbf{c}_j)$, where g can be either parametric or non-parametric.

Proposed Method : GPS Matching

1. Design Stage : Fit a GPS model relating w to $\mathbf{C}$, $\hat{e}(w_j, \mathbf{c}_j) = \hat{g}_{\Phi}^{-1}(\mathbf{c}_j)$, where g can be either parametric or non-parametric.
2. Analysis Stage : Define the suitable caliper matching function by specifying the desired metric, scale parameter λ, and caliper δ. Match individuals based on the matching function. Impute $Y_j(w)$ as : $\hat{Y}_j(w) = Y^{obs}_{m_{gps}(e(w,\mathbf{c}_j),w)}$ for $j = 1, 2, \ldots, N$ successively. The matching estimator $\hat{\mu}(w)$ is equal to the overall average $\hat{E}[Y_j(w)]$ for each exposure level w.

Proposed Method : GPS Matching

1. Design Stage : Fit a GPS model relating w to $\mathbf{C}$, $\hat{e}(w_j, \mathbf{c}_j) = \hat{g}_{\Phi}^{-1}(\mathbf{c}_j)$, where g can be either parametric or non-parametric.
2. Analysis Stage : Define the suitable caliper matching function by specifying the desired metric, scale parameter λ, and caliper δ. Match individuals based on the matching function. Impute $Y_j(w)$ as : $\hat{Y}_j(w) = Y^{obs}_{m_{gps}(e(w,\mathbf{c}_j),w)}$ for $j = 1, 2, \ldots, N$ successively. The matching estimator $\hat{\mu}(w)$ is equal to the overall average $\hat{E}[Y_j(w)]$ for each exposure level w.
3. Estimated average causal exposure-response function : Fit a normal-kernel smoother of $\mu(w)$ on w to estimate the exposure-response function.

Covariate Balance

In theory, a matched dataset aimed at mimicking a randomized experiment is well balanced if $\forall w$,

$$I(W_j = w) \perp\!\!\!\perp \mathbf{C}_j.$$

In practice, we introduce two new measures of covariate balance ;

1. Global Measure : The absolute correlation between exposures and each pretreatment covariate.
2. Local Measure : The absolute standardized bias between $W_j \in [w - \delta, w + \delta]$ v.s. $W_j \notin [w - \delta, w + \delta]$ for every single exposure level w.

Simulation Settings : Data Generating Mechanism

1. Six confounders $(C_1, C_2, ..., C_6)$, which include a combination of continuous and categorical variables

$$C_1, \ldots, C_4 \sim N(0, \boldsymbol{I}_4), C_5 \sim U\{-2, 2\}, C_6 \sim U(-3, 3)$$

2. Three specifications of generalized propensity score models

$$\begin{aligned}
&1)\ \ W = 9\{-0.8 + (0.1, 0.1, -0.1, 0.2, 0.1, 0.1)\boldsymbol{C}\} - 3 + N(0, 5)\\
&2)\ \ W = 15\{-0.8 + (0.1, 0.1, -0.1, 0.2, 0.1, 0.1)\boldsymbol{C}\} + 2 + 2T(4)\\
&3)\ \ W = 15\{-0.8 + (0.1, 0.1, -0.1, 0.2, 0.1, 0.1)\boldsymbol{C}\} + 3^{1/2}C_3^2 + T(4)
\end{aligned}$$

3. The outcome model is a cubical function of the exposure and additive terms for the confounders and interactions between exposure and confounders

$$\begin{aligned}
Y \mid W, \mathbf{C} &\sim N(\mu(W, \mathbf{C}), 10)\\
\mu(W, \mathbf{C}) &= -10 - (2, 2, 3, -1, 2, 2)\mathbf{C} + 0.13^2 W^3\\
&\quad - W(0.1 - 0.1C_1 + 0.1C_4 + 0.1C_5 + 0.1C_3^2)
\end{aligned}$$

Simulation Settings : Alternative Approaches

There are three existing approaches based on GPS aim at estimating causal effect of a continuous exposure in observational studies, named

1. Hirano and Imbens (HI-GPS) estimator (Hirano & Imbens 2004).
2. IPTW estimator (Robins et al. 2000).
3. Doubly robust (DR) estimator (Bang & Robins 2005, Kennedy et al. 2017).

Simulation Settings : Performance

Performance of each estimator is evaluated using absolute bias and mean squared error (MSE).

$$\widehat{\text{Absolute Bias}} = \int_{\hat{\mathcal{W}}^*} |\frac{1}{S}\sum_{s=1}^{S} \hat{Y}_s(w) - Y(w)| f_W(w) dw$$

$$\widehat{\text{MSE}} = \int_{\hat{\mathcal{W}}^*} [\frac{1}{S}\sum_{s=1}^{S} \{\hat{Y}_s(w) - Y(w)\}^2]^{1/2} f_W(w) dw.$$

where $\hat{\mathcal{W}}^*$ denotes a trimmed version of the support of $\hat{\mathcal{W}}$ excluding 10% of mass at the boundaries.

Simulation Results : Absolute Bias and MSE

TABLE – Absolute Bias and Mean Squared Error (MSE) based on sample size 5000

GPS model fit	**GPS Matching**	HI-GPS	IPTW	Kennedy's DR	True Reg
1) Correctly Specified, No Extreme GPS	0.51 (1.18)	0.80 (0.94)	0.08 (0.59)	0.34 (1.41)	0.01 (0.24)
2) Correctly Specified, Extreme GPS	0.89 (2.17)	1.67 (1.78)	$*(*)$	1.13 (3.39)	0.02 (0.33)
3) Misspecified	0.69 (2.26)	2.07 (2.21)	33.86 $(*)$	0.86 (2.71)	0.03 (0.56)

Notes : Matching = the proposed generalized propensity score caliper matching ; HI-GPS = outcome regression that adds generalized propensity score as covariates with correctly specified model ; IPTW = inverse probability of treatment weighted ; DR = doubly robust ; True Reg = regression using true model. * represented values larger than 100.

Simulation Results : Covariate Balance

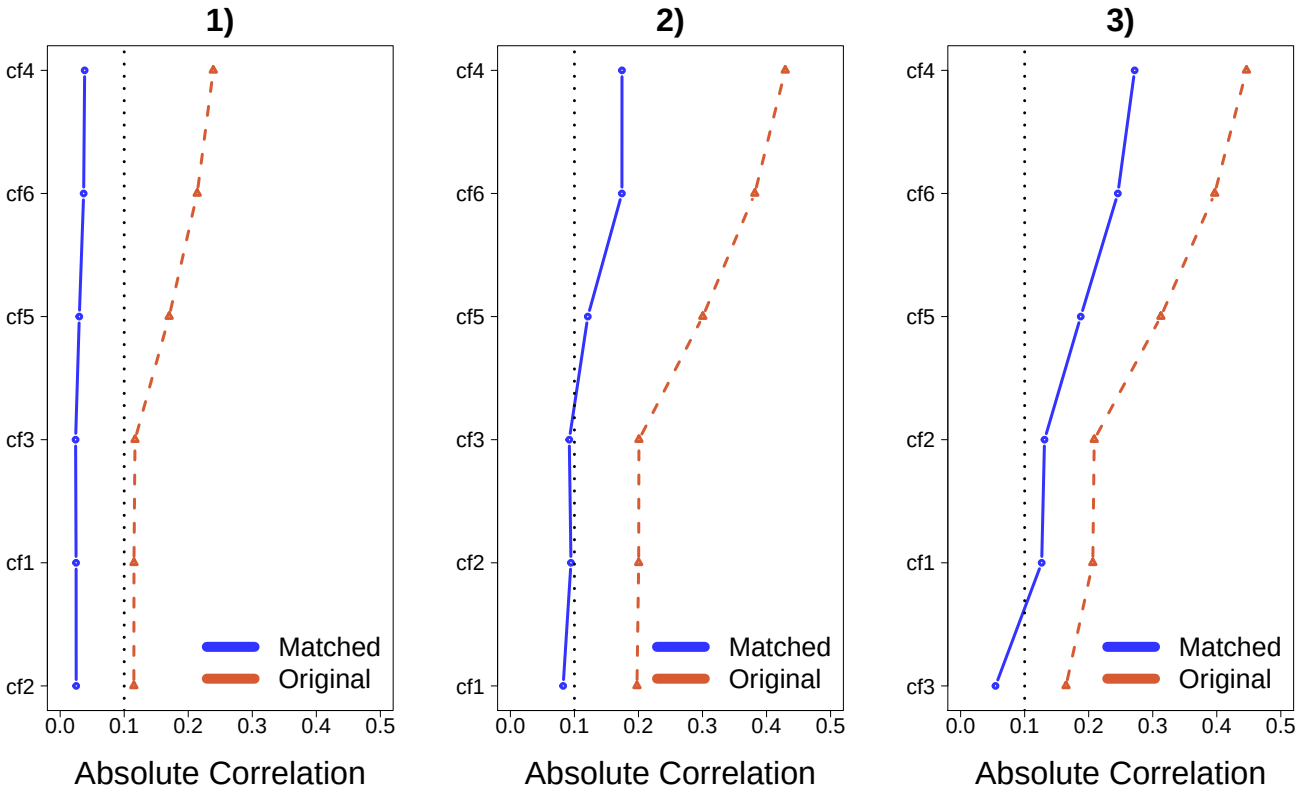

FIGURE – Absolute correlations (ACs) between exposures and each confounders.

Data Application : Background

We apply the proposed matching method to estimate the effect of long-term $PM_{2.5}$ exposure on mortality.

- To this end, we use information on Medicare enrollees across New England (VT, NH, CT, MA, RI and ME) from 2000 to 2012.
- This study population includes a total of 3.3 million individuals with 24.5 million person-years of follow up, who reside in 2,202 zip codes
- We constructed counts corresponding to the mortality for Medicare enrollees for each zip code per year across New England.
- $PM_{2.5}$ exposures were estimated at a 1km $\times$ 1km grid cell resolution using a spatio-temporal prediction model with excellent predictive accuracy (cross validation R^2 = 0.84)(Di et al. 2016).
- To obtain annual average $PM_{2.5}$ at each zip code, we aggregate the gridded concentrations using area-weighted averages. The range of annual average $PM_{2.5}$ across New England from 2000 to 2012 was $2.05 - 15.43\ \mu g/m^3$.

Data Application : Results

We found each 1 $\mu g/m^3$ increase of exposure level of annual average $PM_{2.5}$ causes an approximately 7.4×10^{-4} increase in all-cause mortality.

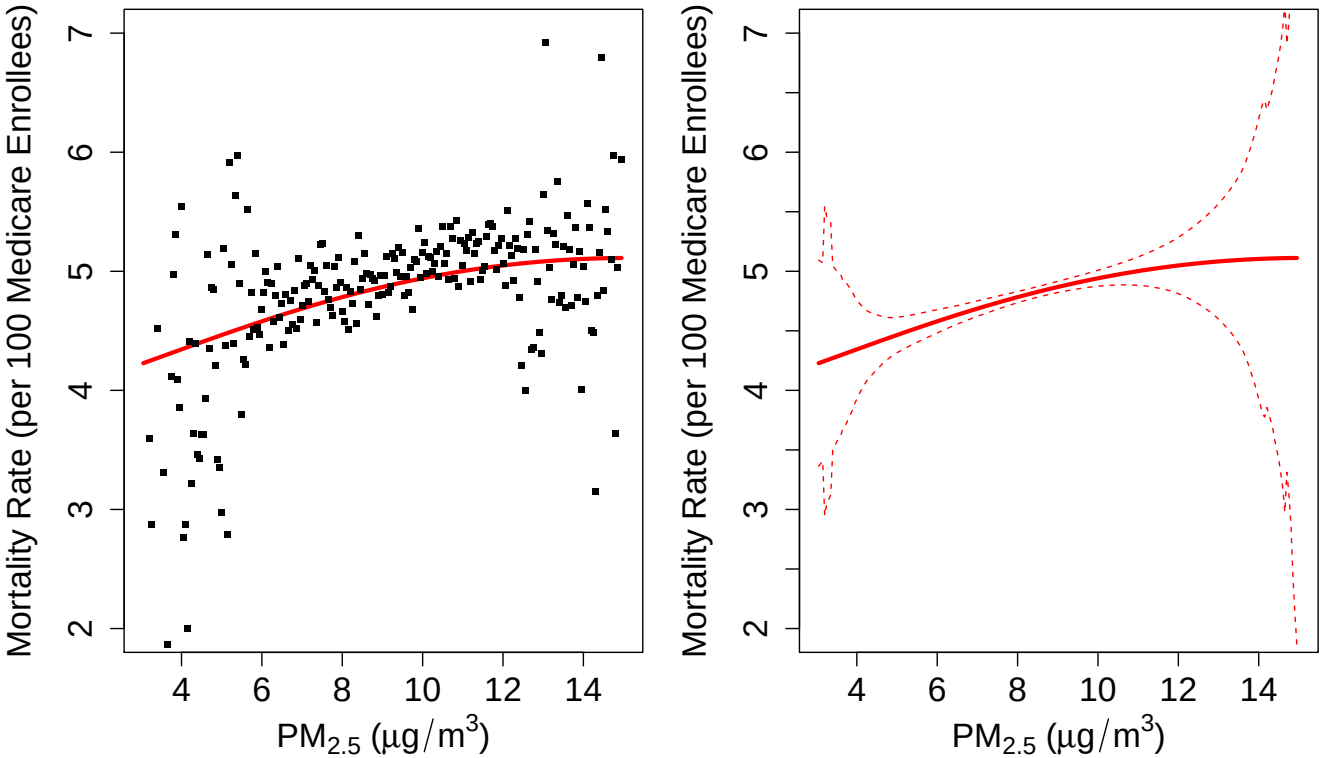

FIGURE – The causal exposure-response function relating all-cause mortality to long-term $PM_{2.5}$ exposure.

Data Application : Covariate Balance

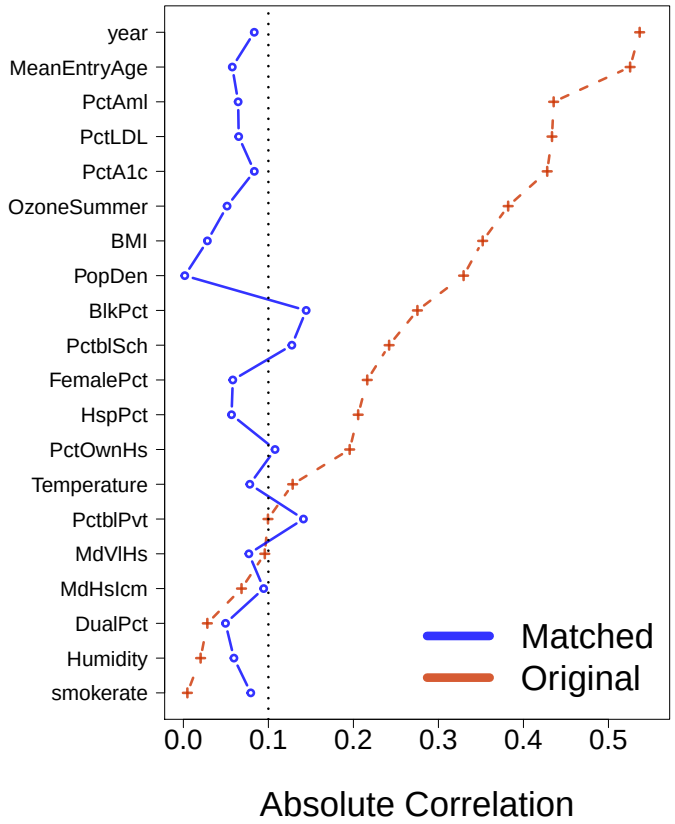

FIGURE – Absolute correlations (ACs) between exposures and each potential confounders. Importantly, time trend (year) has a strong imbalance before matching, yet is balanced after matching.

Conclusions

1. We developed an innovative approach for GPS caliper matching for estimating causal effects using observational data in settings with continuous exposures.

Conclusions

1. We developed an innovative approach for GPS caliper matching for estimating causal effects using observational data in settings with continuous exposures.
2. This approach fills an important gap by extending the matching framework to continuous exposures, which are very common in observational studies.

Conclusions

1. We developed an innovative approach for GPS caliper matching for estimating causal effects using observational data in settings with continuous exposures.
2. This approach fills an important gap by extending the matching framework to continuous exposures, which are very common in observational studies.
3. We prove under the approximate weak unconfoundedness assumption that our proposed matching estimators attain $(N\delta)^{1/2}$-asymptotic normality if the caliper δ is well chosen $(= O(N^{-1/2}))$.

Conclusions

1. We developed an innovative approach for GPS caliper matching for estimating causal effects using observational data in settings with continuous exposures.
2. This approach fills an important gap by extending the matching framework to continuous exposures, which are very common in observational studies.
3. We prove under the approximate weak unconfoundedness assumption that our proposed matching estimators attain $(N\delta)^{1/2}$-asymptotic normality if the caliper δ is well chosen $(= O(N^{-1/2}))$.
4. We show by simulation studies that the proposed matching framework shares advantages that, 1) robustness to GPS model misspecification, especially in the presence of extreme values of GPS (Waernbaum 2012), 2) elimination of outcome model dependency, 3) straightforward assessment of covariate balance.

Conclusions

1. We developed an innovative approach for GPS caliper matching for estimating causal effects using observational data in settings with continuous exposures.
2. This approach fills an important gap by extending the matching framework to continuous exposures, which are very common in observational studies.
3. We prove under the approximate weak unconfoundedness assumption that our proposed matching estimators attain $(N\delta)^{1/2}$-asymptotic normality if the caliper δ is well chosen $(= O(N^{-1/2}))$.
4. We show by simulation studies that the proposed matching framework shares advantages that, 1) robustness to GPS model misspecification, especially in the presence of extreme values of GPS (Waernbaum 2012), 2) elimination of outcome model dependency, 3) straightforward assessment of covariate balance.
5. In data application, we can detect causal effects of long-term exposure to $PM_{2.5}$ on all-cause mortality.

References I

Abadie, A. & Imbens, G. W. (2006), 'Large sample properties of matching estimators for average treatment effects', *econometrica* **74**(1), 235–267.

Bang, H. & Robins, J. M. (2005), 'Doubly robust estimation in missing data and causal inference models', *Biometrics* **61**(4), 962–973.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y. & Schwartz, J. (2016), 'Assessing pm2. 5 exposures with high spatiotemporal resolution across the continental united states', *Environmental science & technology* **50**(9), 4712–4721.

Hirano, K. & Imbens, G. W. (2004), 'The propensity score with continuous treatments', *Applied Bayesian modeling and causal inference from incomplete-data perspectives* **226164**, 73–84.

Imbens, G. W. & Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

Kennedy, E. H., Ma, Z., McHugh, M. D. & Small, D. S. (2017), 'Non-parametric methods for doubly robust estimation of continuous treatment effects', *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* **79**(4), 1229–1245.

Robins, J. M., Hernan, M. A. & Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology'.

Waernbaum, I. (2012), 'Model misspecification and robustness in causal inference : comparing matching with doubly robust estimation', *Statistics in medicine* **31**(15), 1572–1581.