Caliper Synthetic Matching:

Radius matching with adaptive calipers and local synthetic controls

Jonathan Che & Luke Miratrix, Harvard University

Main idea

We propose a matching method that:
- **Finds local matches** for each treated unit
- **Corrects local bias** using synthetic controls

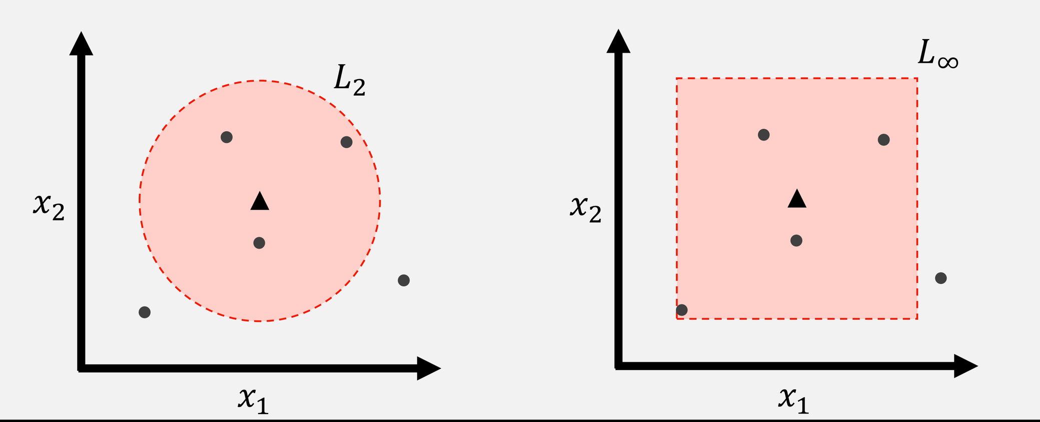
Background

Matching is a **simple and transparent** method for designing observational studies. Exact matches **balance the joint dist. of observed covariates** and allow straightforward causal inferences under conditional ignorability, but are rarely possible to find in practice. To approximate exact matching, we use **local synthetic controls** in the spirit of [1] to directly **improve match quality and control bias.**

Proposed method

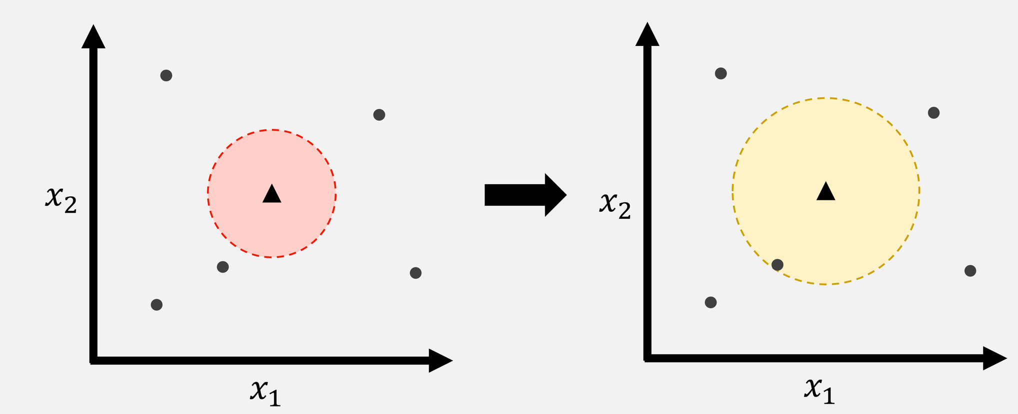
Radius matching bounds bias

- Define donor pool for each treated unit using radius matching [2]
- Radius bounds bias for a Lipschitz class of potential outcome functions [3]



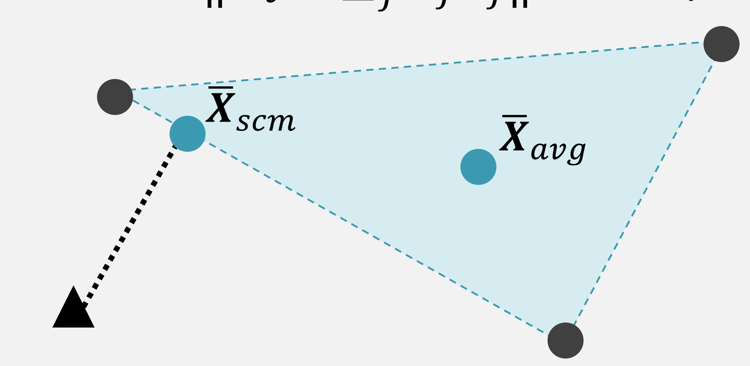
Calipers adapt to the data

- Adapt radius size (i.e., caliper) to data to ensure sufficiently rich donor pool
  - Design choice: e.g., ≥ 1 donor unit, $\geq p+1$ donor units



Synthetic controls minimize local extrapolation bias [4]

- Construct synthetic control for each treated unit using its donor pool
  - SCM: $\min \lVert \boldsymbol{X}_t - \sum_j w_j \boldsymbol{X}_j \rVert$ for simplex w_j



Simulations

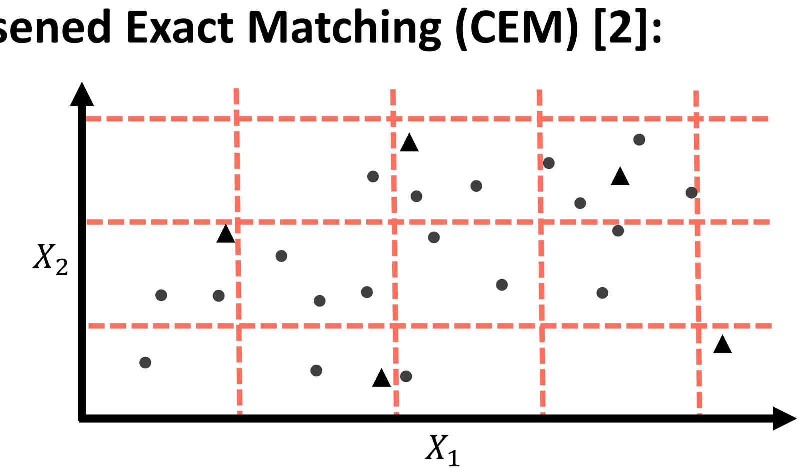
Dataset

Simulation based on 2016 ACIC data challenge [5]
- 200 1,000 x 10 subsets
- Vary response models and levels of treatment-effect heterogeneity/overlap

Methods

		Donor pool method		
		Caliper (L_∞)	Coarsen	PS 1-NN
Estimation method	Average	Radial	CEM	N/A
	SCM	CSM	CEM + SCM	

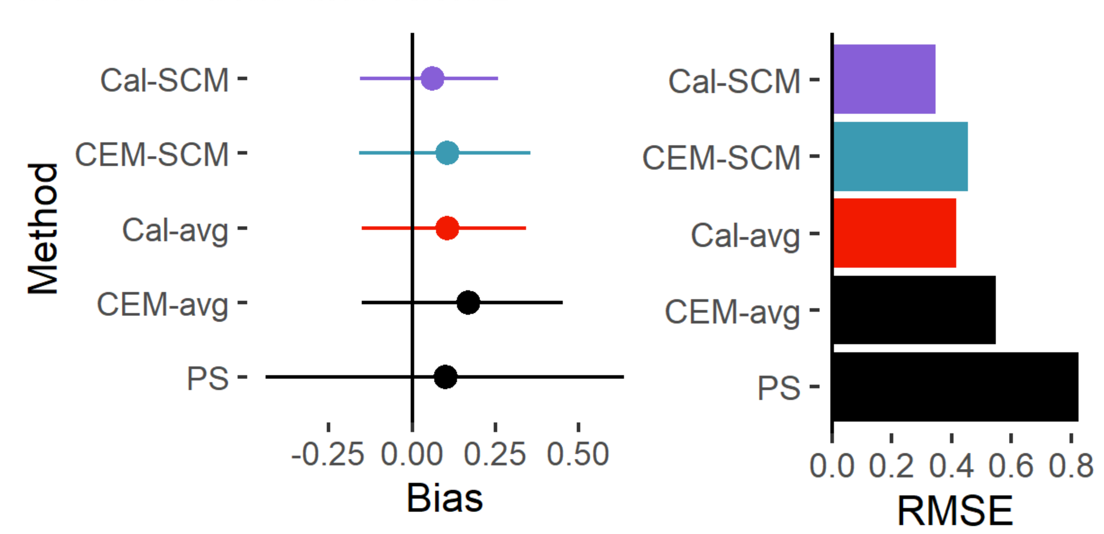
Coarsened Exact Matching (CEM) [2]:



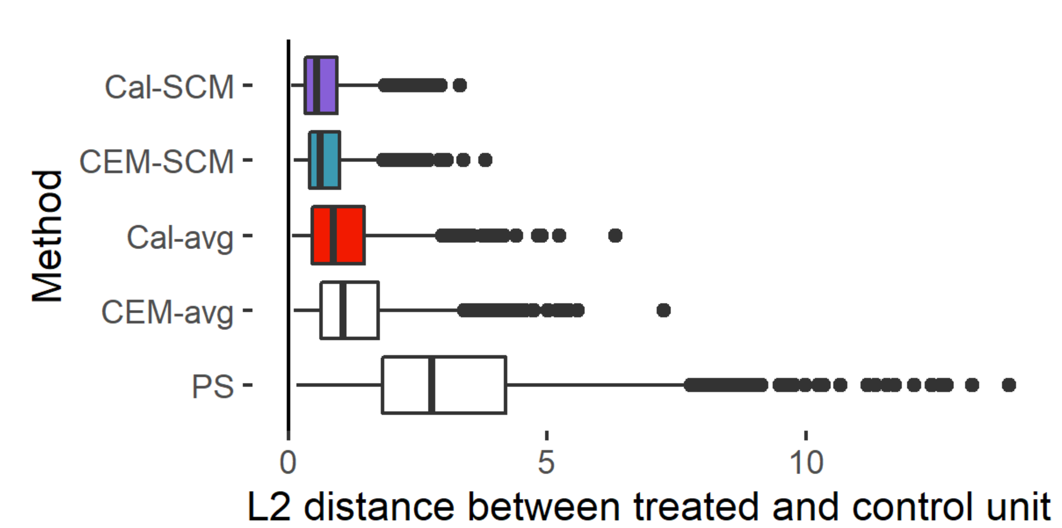
Results

Radius matching and synthetic controls…

…reduce bias and RMSE

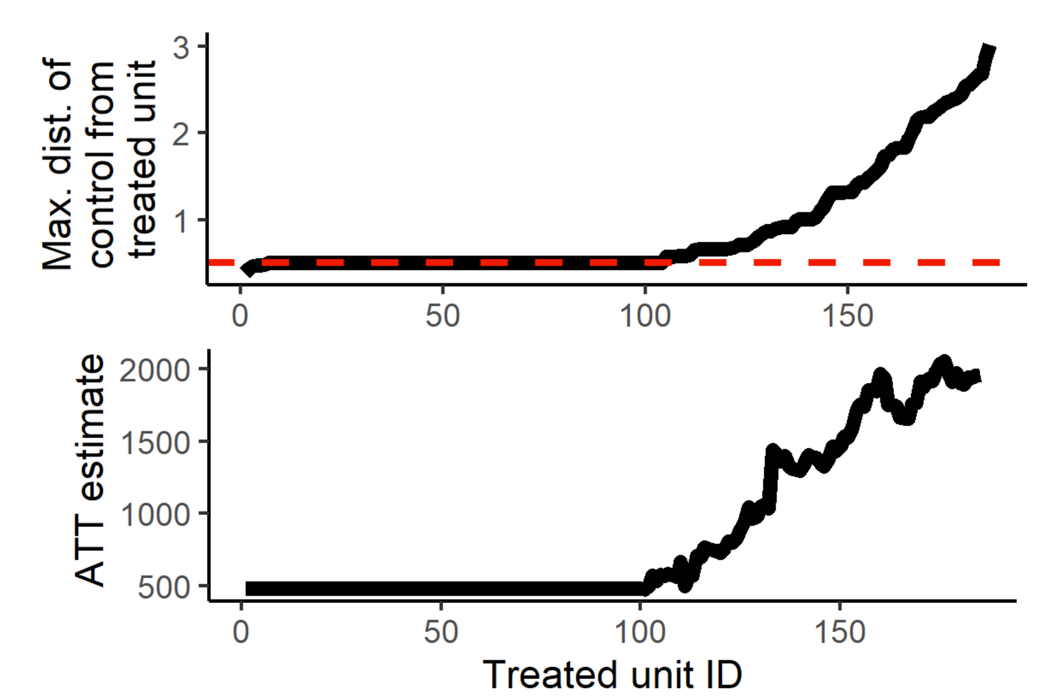


…reduce distances between treated and counterfactual units (potential extrapolation bias)



Example

We demonstrate the diagnostic tools available to CSM using the Lalonde (1999) dataset.



As caliper increases…
- …distance ↗ ⇒ potential bias ↗
- …estimand changes

Z	ID	age	educ	…	re74	re75	Y
Tx unit	1	33	12		0	0	16k
SC unit	1	33	12		0	0	23k
Tx unit	2	26	12		8.4k	5.8k	1.4k
SC unit	2	26	12.1		8.4k	5.8k	11.6k
	…						
Tx unit	185	27	9		0	0.9k	1.8k
SC unit	185	36	12		0	8.9k	11.8k

Conclusion:
Caliper diagnostics indicate that the standard \$1700-\$2000 effect estimate depends on a number of relatively poor matches

Mathematical intuition

Expected bias of weighting estimator using control units j within caliper δ of treated unit t:

$$\mathbb{E}\left[\sum_j w_j Y_j - Y_t(0)\right]$$
$$= \sum_j w_j f_0(\boldsymbol{X}_j) - f_0(\boldsymbol{X}_t)$$
$$= \sum_j w_j d(\boldsymbol{X}_j, \boldsymbol{X}_t) \nabla_{\boldsymbol{v}_j} f(\boldsymbol{X}_t) + o(\delta)$$

for directional deriv. $\nabla_{\boldsymbol{v}_j}$, unit vector $\boldsymbol{v}_j = \frac{\boldsymbol{X}_j - \boldsymbol{X}_t}{d(\boldsymbol{X}_j, \boldsymbol{X}_t)}$

→ Calipers bound $d(\boldsymbol{X}_j, \boldsymbol{X}_t) \leq \delta$

→ Lipschitz assumption bounds $\nabla_{\boldsymbol{v}_j} f(\cdot)$

Rewriting the linear bias term:

$$\sum_j w_j d(\boldsymbol{X}_j, \boldsymbol{X}_t) \nabla_{\boldsymbol{v}_j} f(\boldsymbol{X}_t)$$
$$= \sum_j w_j (\boldsymbol{X}_j - \boldsymbol{X}_t)^T \nabla f(\boldsymbol{X}_t)$$
$$= \sum_k m_k \left(\sum_j w_j X_{jk} - X_{tk}\right)$$

- SCM minimizes $\sum_k v_k \left(\sum_j w_j X_{jk} - X_{tk}\right)^2$
- An exact SC unit satisfies $\sum_j w_j X_{jk} = X_{tk}$

References

[1] Abadie, Alberto, and Jérémy L'Hour. "A penalized synthetic control estimator for disaggregated data." *Journal of the American Statistical Association* (2021).

[2] Deheja, Rajeev and Sadek Wabha. "Propensity Score-Matching Methods For Nonexperimental Causal Studies." *The Review of Economics and Statistics* (2002).

[3] Iacus, Stefano, Gary King, and Giuseppe Porro. "Multivariate Matching Methods that are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* (2011).

[4] Kellogg, Maxwell, Magne Mogstad, Guillaume A. Pouliot , and Alexander Torgovitsky. "Combining Matching and Synthetic Control to Tradeoff Biases From Extrapolation and Interpolation." *Journal of the American Statistical Association* (2021).

[5] Dorie, Vincent, et al. "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." *Statistical Science* (2019).

Take-home message:

We propose a matching method that transparently improves match quality and controls bias by directly using local matches to create synthetic controls.