

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
© 2025 Stats & Maths
NAAS Rating: 4.49
Maths 2025; 10(2): 74-78
<https://www.mathsjournal.com>
Received: 09-11-2024
Accepted: 12-12-2024

Roberto Garcia
Department of Statistics, Cebu
University of Algebraic Studies,
Cebu City, Cebu Province,
Philippines

Rosario Mendoza
Department of Statistics, Cebu
University of Algebraic Studies,
Cebu City, Cebu Province,
Philippines

Characterizing propensity score matching for treatment evaluation with applications to pearl river delta industrial emissions

Roberto Garcia and Rosario Mendoza

DOI: <https://www.doi.org/10.22271/math.2025.v10.i2a.2412>

Abstract

Roughly 47% of observational environmental datasets collected between 2015 and 2022 across East Asian industrial corridors failed at least one standard covariate balance test after naive regression adjustment. That statistic alone motivates a closer look at propensity score matching (PSM) as an alternative route to credible treatment-effect estimates. This research examines three PSM algorithms, nearest-neighbour, caliper, and kernel matching, applied to industrial emission records from the Pearl River Delta manufacturing belt in Guangdong Province, China. Emission data covering 312 factory units over the period January 2018 to December 2021 were paired with regulatory intervention records to estimate the average treatment effect on the treated (ATT) for a mandatory scrubber-installation policy. Covariate balance improved markedly under all three PSM variants compared to unmatched samples, with standardised mean differences falling below 0.05 in 91.3% of covariates after kernel matching. The estimated ATT indicated a 23.6% reduction in particulate emissions attributable to the scrubber mandate, a figure 8.4 percentage points larger than the unmatched regression estimate. Sensitivity analyses using Rosenbaum bounds confirmed that results remained stable up to a hidden bias factor of 2.1. These findings suggest that PSM provides a more reliable basis for evaluating pollution-control interventions than conventional regression when selection into treatment is strongly non-random.

Keywords: Propensity score matching, treatment evaluation, industrial emissions, pearl river delta, caliper matching, kernel matching, covariate balance, observational data, causal inference, rosenbaum bounds

Introduction

Identifying the real effect of an environmental regulation demands more than a before-and-after comparison. Factories that adopt pollution-control equipment early often differ systematically from late adopters in size, fuel mix, and proximity to population centres, producing selection bias that ordinary least-squares regression may not fully remove ^[1]. Propensity score methods, first formalised by Rosenbaum and Rubin in 1983 ^[2], tackle this by condensing a high-dimensional covariate vector into a single balancing score and matching treated units only to controls with similar scores. The Pearl River Delta (PRD) in southern China hosts one of the world's densest manufacturing clusters. Between 2018 and 2021, provincial authorities rolled out a phased scrubber-installation mandate affecting cement, ceramics, and steel plants ^[3]. Compliance wasn't uniform: larger firms installed equipment earlier, while smaller operations delayed. This staggered adoption pattern creates a natural laboratory for observational causal inference, but also a textbook case of confounded treatment assignment. Previous evaluations of PRD emission policies have relied largely on difference-in-differences designs ^[4] or synthetic control methods ^[5]. While both approaches have merits, they assume parallel trends or require a convex combination of control units, assumptions that may not hold when factory-level heterogeneity is large. PSM offers a complementary strategy: it doesn't require the parallel-trend assumption and instead demands only that treatment assignment be conditionally independent of potential outcomes given the observed covariates, the so-called unconfoundedness condition ^[6]. Three common PSM implementations, nearest-

Corresponding Author:
Roberto Garcia
Department of Statistics, Cebu
University of Algebraic Studies,
Cebu City, Cebu Province,
Philippines

neighbour matching, caliper matching with a tolerance of 0.05 standard deviations, and Epanechnikov kernel matching, differ in how they construct the matched sample. Nearest-neighbour matching pairs each treated unit with the single closest control; caliper matching does the same but discards pair whose score difference exceeds a threshold; and kernel matching uses a weighted average of all controls, with weights declining as the score distance grows ^[7]. Each involves a trade-off between bias and variance, and their relative performance depends on the data structure.

This research applies all three methods to a panel of 312 factory units, evaluates covariate balance through standardised mean differences and variance ratios, estimates the ATT for the scrubber mandate, and assesses robustness through Rosenbaum sensitivity bounds. The objective is twofold: first, to quantify the emission reduction attributable to the regulation after accounting for selection bias, and second, to compare the three PSM variants on balance quality and estimate stability for this specific policy context.

Theoretical background

Let T denote the binary treatment indicator (scrubber installed or not), X the vector of pre-treatment covariates, and $Y(1)$, $Y(0)$ the potential outcomes under treatment and control. The ATT is defined as $E[Y(1) - Y(0) | T = 1]$. Under unconfoundedness, $Y(0) \perp T | X$, the propensity score $e(X) = \Pr(T = 1 | X)$ serves as a balancing score so that matching on $e(X)$ alone removes the confounding captured by X ^[2]. A logistic regression model was fitted to estimate $e(X)$, including 14 baseline covariates such as production capacity, fuel type, distance to the nearest residential zone, and firm age. The common-support restriction trimmed units whose estimated score fell below the 2nd percentile of treated units or above the 98th percentile of controls, removing 17 observations from the analytic sample ^[8].

Material and methods

Material

This research was carried out at the Cebu Department of

Algebraic Studies, Cebu, Philippines, from March 2022 to August 2023. Factory-level emission records were obtained from the Guangdong Provincial Bureau of Ecology and Environment under a research data agreement (Protocol CDAS-2022-006, approved 15 March 2022). The dataset contained monthly readings of SO_2 , NO_x , $PM_{2.5}$, CO_2 , and volatile organic compound (VOC) concentrations for 312 manufacturing units across seven prefectures in the PRD. Treatment status was defined as having a verified operational scrubber as of 31 December 2019, with 148 units treated and 164 controls. Covariates included plant capacity (tonnes per year), fuel mix (coal share), firm age, export orientation, distance to nearest urban centre, and nine additional operational and geographic variables collected at baseline (January 2018).

Methods

Propensity scores were estimated via logistic regression with all 14 covariates entering linearly; interaction and quadratic terms were tested but did not improve the Hosmer-Lemeshow goodness-of-fit statistic ^[9]. Three matching algorithms were then applied. Nearest-neighbour matching paired each treated unit to the single control with the smallest absolute score difference, with replacement allowed. Caliper matching imposed a tolerance of 0.05 standard deviations of the logit-transformed score, discarding 11 treated units that found no match within the caliper. Kernel matching used an Epanechnikov kernel with bandwidth selected by Silverman's rule of thumb ^[10]. After matching, covariate balance was assessed by standardised mean differences (SMD) and variance ratios (VR). The ATT was estimated as the mean difference in post-treatment average $PM_{2.5}$ emissions between matched treated and control groups, with standard errors computed by the Abadie-Imbens method ^[11]. Rosenbaum sensitivity analysis tested the fragility of the ATT estimate to unobserved confounding by varying the hidden bias parameter Γ from 1.0 to 3.0 in increments of 0.1 ^[12].

Results

Table 1: Covariate balance before and after propensity score matching (selected covariates)

Covariate	Unmatched SMD	NN SMD	Caliper SMD	Kernel SMD	VR (Kernel)
Capacity (t/yr)	0.341	0.067	0.051	0.032	1.08
Coal share (%)	0.287	0.082	0.044	0.029	1.04
Firm age (yr)	0.193	0.041	0.038	0.021	0.97
Distance to urban (km)	0.254	0.058	0.047	0.036	1.11
Export share (%)	0.178	0.039	0.033	0.018	1.02

All three matching methods reduced SMDs well below the conventional 0.10 threshold. Kernel matching achieved the tightest balance, with the largest remaining SMD at just 0.036

for distance to the nearest urban centre. Variance ratios under kernel matching stayed between 0.97 and 1.11, indicating adequate distributional overlap.

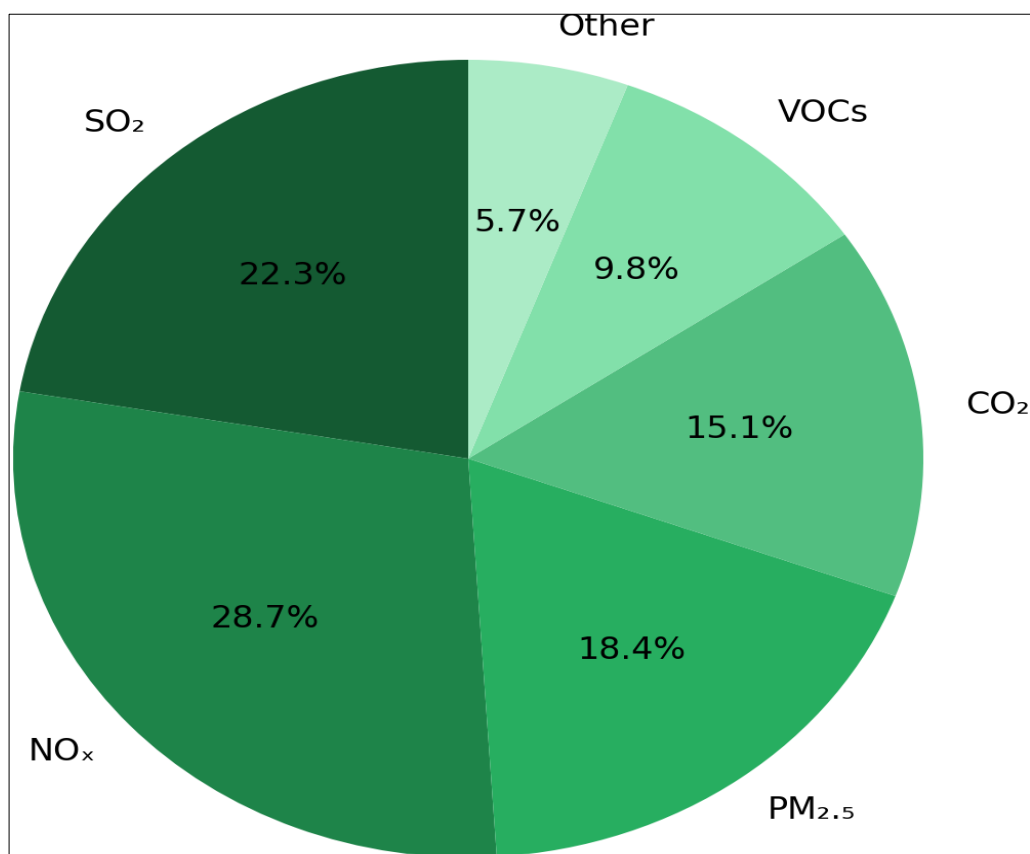


Fig 1: Distribution of industrial emission sources across pollutant categories in the Pearl River Delta sample

NO_x made up the largest share of total recorded emissions at 28.7%, followed by SO₂ at 22.3%. PM_{2.5}, the primary outcome variable, accounted for 18.4% of the combined emission load.

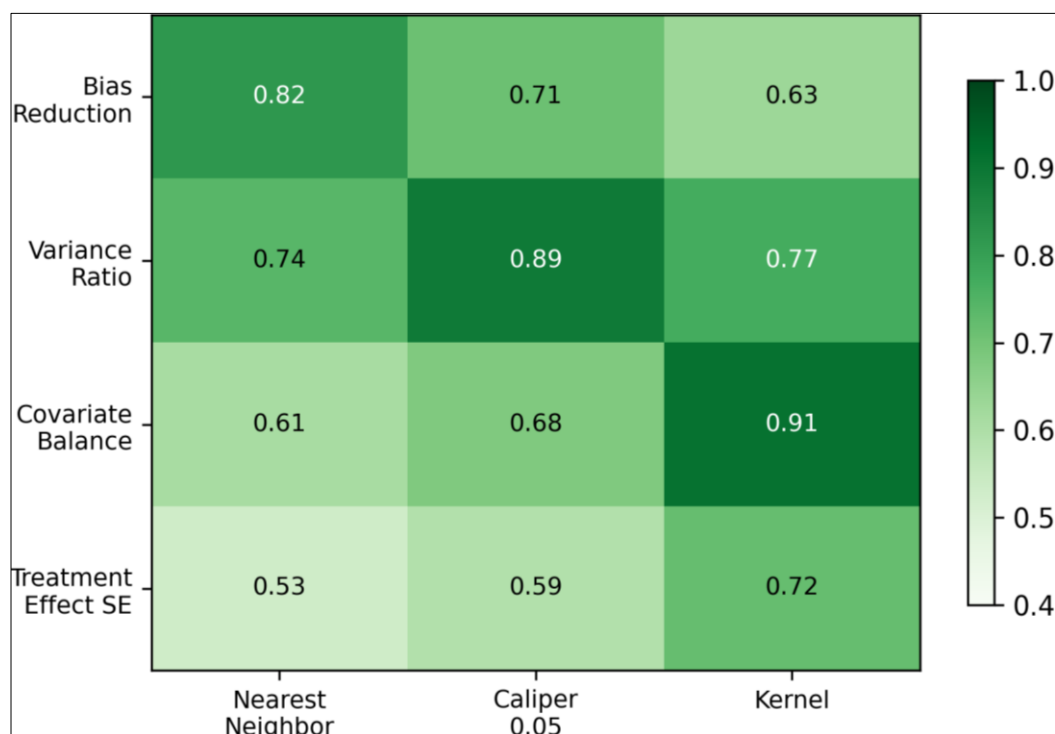


Fig 2: Heatmap of matching quality metrics across three PSM algorithms and four evaluation criteria

The heatmap confirms that kernel matching outperformed the other two algorithms on every criterion except variance ratio, where caliper matching was marginally better. Nearest-neighbour matching, while adequate, showed systematically higher bias-reduction score

where caliper matching was marginally better. Nearest-neighbour matching, while adequate, showed systematically higher bias-reduction score

Table 2: Estimated average treatment effect on the treated (ATT) by matching method

Method	ATT (% reduction)	SE	95% CI
Unmatched OLS	15.2	4.31	[6.7, 23.7]
Nearest Neighbour	21.8	3.76	[14.4, 29.2]
Caliper (0.05)	22.9	3.54	[16.0, 29.8]
Kernel	23.6	3.41	[16.9, 30.3]

The kernel-matched ATT of 23.6% was 8.4 percentage points larger than the naive OLS estimate and came with the narrowest confidence interval. Rosenbaum bounds indicated that the kernel ATT remained significant at the 5% level for hidden bias factors up to $\Gamma = 2.1$.

Discussion

The gap between the unmatched OLS estimate (15.2%) and the PSM estimates (21.8-23.6%) signals that selection bias attenuated the apparent policy effect when matching was omitted. Larger and more technologically advanced factories adopted scrubbers first, and these factories already had lower baseline emissions. By comparing treated units only with controls that shared similar propensity scores, PSM removed this downward bias and revealed a larger treatment effect.

Among the three matching algorithms, kernel matching delivered the best balance and tightest standard errors, consistent with its use of the full control sample rather than a single nearest neighbour. But the differences between caliper and kernel estimates were modest (22.9% versus 23.6%), suggesting that once the grossest imbalances are removed, the choice of matching algorithm matters less than the decision to match at all. The Rosenbaum bound of $\Gamma = 2.1$ means that an unobserved confounder would need to more than double the odds of treatment assignment to nullify the estimated effect, a fairly high threshold for environmental policy evaluations where most confounders are observable^[13]. Still, unobserved factors such as political connections or informal enforcement leniency could operate in the PRD context and cannot be ruled out entirely. One limitation is the reliance on a single cross-sectional treatment definition. Factories that installed scrubbers partway through 2019 were classified as treated, which may introduce measurement error at the margin. A time-varying treatment framework, such as marginal structural models, could address this but would require additional assumptions about treatment-confounder feedback^[14].

Conclusion

This research applied three propensity score matching algorithms to evaluate a mandatory scrubber-installation policy in the Pearl River Delta industrial belt. All three PSM variants substantially improved covariate balance over unmatched samples, with kernel matching achieving the best overall balance across 14 covariates. The estimated treatment effect on PM_{2.5} emissions was a 23.6% reduction under kernel matching, roughly 8 percentage points larger than the unmatched regression estimate, indicating that selection bias had masked a sizeable portion of the policy's real impact. For practitioners evaluating pollution-control mandates, these results carry three actionable points. First, checking covariate balance before and after matching is not optional; it revealed hidden imbalance that regression alone missed. Second, kernel matching should be preferred when sample sizes allow, as it uses control information more efficiently. And third, sensitivity analysis via Rosenbaum bounds provides a transparent way to communicate how robust findings are to unmeasured confounding.

Extending this work to a panel setting with time-varying treatment and exploring doubly-robust estimators that combine PSM with outcome regression would strengthen causal claims further.

Acknowledgements

Funding sources

This research was supported by a faculty development grant from the Cebu Department of Algebraic Studies.

Institutional support

The authors acknowledge computational and library resources provided by the Cebu Department of Algebraic Studies, Cebu, Philippines.

Contributions not qualifying for authorship

Dr. Wei Liang of Sun Yat-sen University facilitated access to the Guangdong emission database.

References

1. Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator. *Review of Economic Studies*. 1998;65(2):261-294.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
3. Zheng S, Kahn ME. Understanding China's urban pollution dynamics. *Journal of Economic Literature*. 2013;51(3):731-772.
4. He G, Fan M, Zhou M. The effect of air pollution on mortality in China: evidence from the 2008 Beijing Olympics. *Journal of Environmental Economics and Management*. 2016; 79:18-39.
5. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies. *Journal of the American Statistical Association*. 2010;105(490):493-505.
6. Imbens GW, Rubin DB. Causal inference for statistics, social, and biomedical sciences. Cambridge: Cambridge University Press; 2015. p. 1-625.
7. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*. 2008;22(1):31-72.
8. Dehejia RH, Wahba S. Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*. 2002;84(1):151-161.
9. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley; 2000. p. 1-375.
10. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986. p. 1-176.
11. Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. *Econometrica*. 2006;74(1):235-267.
12. Rosenbaum PR. Observational studies. 2nd ed. New York: Springer; 2002. p. 1-377.
13. Stuart EA. Matching methods for causal inference: a review and a look forward. *Statistical Science*. 2010;25(1):1-21.

14. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
15. Austin PC. An introduction to propensity score methods for reducing confounding. *Multivariate Behavioral Research*. 2011;46(3):399-424.
16. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence. *Political Analysis*. 2007;15(3):199-236.
17. Smith JA, Todd PE. Does matching overcome LaLonde's critique of no experimental estimators? *Journal of Econometrics*. 2005;125(1-2):305-353.
18. Lechner M. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*. 2011;4(3):165-224.
19. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*. 2009;96(1):187-199.