INTEGRATING SPATIAL CLUSTERING AND WEIGHTED DID FOR RETAIL INTERVENTION ASSESSMENT: A FRAMEWORK FOR SPILLOVER CONTROL AND STRATEGIC RISK EVALUATION

Banani Mohapatra[1], Bhavnish Walia[2]

Abstract

Quantifying the causal impact of spatial treatments such as retail pricing changes, logistics incentives, or geo-promotions is challenging due to spatial interference, spillovers, and non-random treatment assignment. This paper develops a geo-based causal inference framework consisting of unsupervised clustering, spatial buffering, and econometric estimation for such challenges. Using hybrid PCA-based K-Means and DBSCAN clustering, covariate-balanced treatment-control spatial contiguous pairs are created using Haversine distance buffers for spillover control. The framework is illustrated on a large-scale retail intervention for which a national retailer removed pickup fees for store pickup orders. Analysis integrates store metadata and transaction data across several regions, encompassing over 10,000 retail outlets. Treatment impacts are estimated based on an Inverse Propensity Weighting (IPW) weighted Difference-in-Differences (DiD) model controlling time-invariant confounders. We find a 3.79% lift in sales, corroborating significant post-intervention gains and robust parallel trends. The paper contributes a spatial causal inference framework with a scalable framework for marketing, strategy, strategic risk, and policy evaluation, with accompanying spatial spillover management and market risk management insights for localised retail decision-making.

Keywords: Difference-in-Differences, hybrid clustering, inverse propensity weighting, spatial causal inference, treatment spillover mitigation

---

[1] Walmart
Email: bananimohapatra29@gmail.com*

[2] Amazon
Email: bhavnish.walia@gmail.com

DOI:

Introduction

Background and motivation

For organisations operating geographically distributed services such as omni-channel retail chains, logistics networks, advertising platforms, or public infrastructure systems, localised programs are a strategic lever for driving revenue growth, customer loyalty, and competitive differentiation. For example, a large U.S.-based omni-retailer recently waived store pickup fees on online orders picked up in-store, with the intent to increase order frequency and bias customer preference toward low-cost fulfilment channels. Operationally simple, it is challenging and directly financially consequential for pricing, promotions, and supply chain planning to causally estimate the true effect of such initiatives. In practice, business leaders must confront difficulties stemming from spatial interference, market heterogeneity, as well as the absence of randomised treatment assignment (Reich et al., 2021; Rubin, 1974; Keele & Titiunik, 2015). For example, waiving pickup fees at select stores can shift demand not only at treated locations but also at nearby untreated stores-attracting spillover effects that violate the Stable Unit Treatment Value Assumption (SUTVA) (Card & Krueger, 1994). This contamination renders classical estimators such as Difference-in-Differences (DiD) or Inverse Propensity Weighting (IPW) (Pollmann, 2024; Rosenbaum & Rubin, 1983). While the literature has addressed spatially adjusted matching, hierarchical spatial models, and clustering-based methods, existing methodologies often lack built-in spillover prevention measures, i.e., adaptive geographic buffers, to ensure business insights remain uncontaminated and actionable (Anselin, 1988; Gibbons, 2022; Sinnott, 1984). This work closes a prominent gap in spatial causal inference with the development of an applied, large-scale framework for estimating localised treatment effects within real- world retailing environments, such as those in which spillovers, spatial dependencies invalidate traditional estimators.

Scope, objectives, and research questions

This study proposes a generalised geo-based causal inference framework that integrates hybrid clustering, adaptive haversine buffering, and IPW-weighted DiD estimation, supported by diagnostics such as standardised mean differences, placebo tests, and event studies. The research addresses four key questions:

RQ1: How can hybrid clustering improve covariate balance and spatial coherence?
RQ2: Can dynamic geographic buffers effectively reduce spillover contamination?
RQ3: Does IPW-adjusted DiD enhance robustness against confounding?
RQ4: How well do diagnostics validate assumptions and strengthen stakeholder confidence?

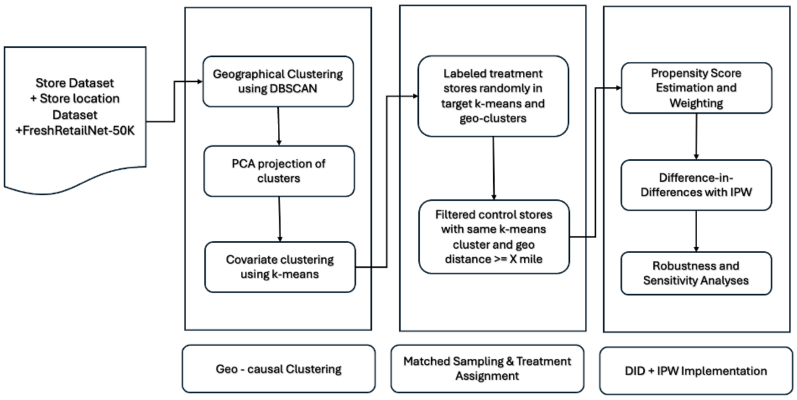
Literature Review

It is difficult to evaluate geographically targeted interventions due to spatial dependence, spillovers, and heterogeneity. Early econometric work tackled spatial correlation (Anselin, 1988; Cliff & Ord, 1981), which was then augmented using Bayesian and machine learning-oriented spatial models (Lee & Yu, 2022; Lesage & Pace, 2009; Zhang et al., 2024). Rubin's propensity score framework enhanced covariate balance (Rosenbaum & Rubin, 1983) and motivated interference-aware matching methods (Kondo, 2023; Papadogeorgou et al., 2020). Nevertheless, traditional matching does not behave well under spatial interdependence when units on the boundary between treatment and control regions interfere with one another (Eckles et al., 2023; Hudgens & Halloran, 2008). Spatial extensions for DiD have incorporated distance and spillover effects (Arefin et al., 2023; Conley, 1999; Gibbons et al., 2019; Lee et al., 2022), whereas automatic causal pipelines more frequently incorporate distance-based buffers for controlling interference (Chen & Mueller, 2023). Non-parametric clipping methods like PCA + K-Means clustering and DBSCAN increase spatial coherence and treatment- control balancing (Ester et al., 1996; Han et al., 2022; Jolliffe & Cadima, 2016;). Furthermore, adding IPW (Inverse Propensity Weighting) to DiD makes the approach more robust against confounding (Abadie, 2005; Forastiere et al., 2021; Hirano et al., 2003; Imai & Ratkovic, 2014; Robins et al., 2000). While there have been such advancements, many spatial DiD applications remain based on fixed spatial assumptions or constrained interference modelling. Classic constructions (Conley, 1999; Gibbons et al., 2019) cover geographic proximity but have no adaptive spillover adjustment, whereas hierarchical or Bayesian constructions (Lee & Yu, 2022; Lesage & Pace, 2009) estimate dependence without tractable scalability for large, realistic datasets. The presented hybrid framework combines unsupervised spatial clipping, adaptive buffering at a geographically adapted distance, and IP-weighted DiD within a common design, all at once gaining dynamic spillover adjustment, covariate balancing, and robust causal estimation.

To our knowledge, our paper is the first to combine such pieces of work within a scalable framework for geographically localised, non-randomised assessment of interventions. The techniques, when employed collectively, handle the primary spatial assessment problems, covariate imbalance, interference, and unobserved heterogeneity, yielding credible numbers that have a clear application in high-stakes business decisions for the practitioner. While demonstrated here for an omni-retail fee waiver, the framework extends to logistics, advertising, and healthcare, providing researchers and decision-makers a generalizable, flexible spatial policy evaluation tool.

Methodology

Figure 1
Step-by-Step Execution of Methodology



(Source: Developed by authors based on literature (2025))

This study estimates the causal impact of pricing interventions on store-level sales. The methodology, as described in Figure 1, consists of the following steps.

Data sources and preprocessing

We merge two large public datasets at scale to guarantee both context richness and transactional accuracy. The Store Dataset allows for store-level metadata and covariates like store_size, foot_traffic, and inventory_level for all outlets within the retailing network, which gives robust structural information necessary for comparison at baseline. The FreshRetailNet-50K dataset fills out this information by recording fine-grained, daily-level transactional sales and inventory movement, enabling accurate time-wise tracking of treatment impact. The study duration is from January 2024 through August 2024, a period that reflects enough pre- and post-intervention variation to create plausible counterfactual trends.

After pooling the inventory and sale data at store_id × day level, we create a treatment date t_0= {2024-05-10} and treat all the observations as post-treatment or pre-treatment, respectively

$$\mathrm{period}_{i,t} = \begin{cases} \text{pre if } t < t_0 \\ \text{post if } t > t_0 \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Equation (1) simply indicates that all records before the intervention are labelled as 'pre-treatment' and those after the intervention as 'post-treatment'.

The data preparation included cleaning of transaction anomalies, harmonisation of store metadata using store_id merges, truncation of incomplete daily observations (<95% coverage), and winsorizing of extreme values (top and bottom 1%) for reduction of bias. This step provided for temporal validity, stability of covariates, and robust identification of treatment exposure across stores.

Clustering for Spatial and Covariate Balance

Geographic Clustering: We apply DBSCAN with Haversine distance for unsupervised clustering of stores using latitude and longitude. The Haversine distance between two coordinates is given by:

$$d_{ij} = 2r.\arcsin\left(\sqrt{\sin^2\left(\frac{\mathrm{Dj}}{2}\right) + \cos(\mathrm{j_i})\cos(\mathrm{j_j})\sin^2\left(\frac{\mathrm{D}\lambda}{2}\right)}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

where r = 3959 miles is the Earth's radius.

Cluster validity was judged based on the silhouette measure of spatial cohesion and separation. Generally, less than 0.1 signifies poor clustering, 0.1- 0.25 signifies moderate, but understandable structure, while greater than 0.25 signifies strong, separated clusters. This range reflects plausible spatial structure for retail catchments, for which moderate overlap is desired for geographic contiguity.

Covariate Clustering: Thirteen normalised covariates (e.g., store size, inventory level, online sales share) are reduced via Principal Component Analysis (PCA), retaining components that explain at least 90% of variance. K-Means clustering is then applied, with the optimal k = 4 selected using the Elbow Method, where the sum of squared errors (SSE) is given by:

$$\mathrm{SSE}_{\mathrm{k}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} ||\mathrm{x}_{\mathrm{i}} - \mu_{\mathrm{k}}||^2 \dots\dots\dots\dots\dots\dots \text{(3)}$$

where μ_{k} denotes the centroid assigned to the observation.

It addresses RQ1 by gauging how hybrid spatial and covariate clustering facilitates both covariate balance and geographic coherence across retail outlets.

Matched sampling and treatment assignment

For validation, 50% of stores were pseudo-randomly assigned to treatment. Each treated store was matched to a control based on: (1) behavioural similarity via PCA-based K-Means clustering (cluster difference ≤ 1) and (2) geographic spillover mitigation using a dynamically chosen Haversine-based buffer within the same DBSCAN cluster. Buffers from 0–50 miles (5-mile increments) were tested, assessing matched pairs, standardised mean differences (SMDs), and control duplication rates. A 25-mile buffer achieved strong covariate balance (all SMDs < 0.1), minimised control reuse, and ensured complete coverage, producing geographically coherent matches while reducing contamination risk from proximity spillovers.

For verification, 50% of the stores were pseudo-randomly designated for treatment. A control was matched with a treated store on: (1) behavioural similarity based on PCA-based K-Means clustering (differing clusters ≤ 1) and (2) adjustment for geographic spillover using a dynamically chosen Haversine-based buffer within the same DBSCAN cluster. We checked buffers 0-50 miles apart (in 5-mile increments), testing on matched pairs, standardised mean differences, and rates of duplication for controls.

The adaptive buffer approach balances empirical optimisation with theoretical reasoning grounded in retail catchment dynamics and spatial decay principles, where customer influence typically declines sharply beyond 20-30 miles. Within this range, proximity captures local market effects without introducing spillovers from adjacent territories. The 25-mile buffer thus represents a stable equilibrium, wide enough to ensure sufficient control availability yet narrow enough to maintain treatment isolation. This configuration achieved strong covariate balance (all SMDs < 0.1), minimised control reuse, and ensured complete geographic coverage.

It addresses RQ2 through exploration of the capacity of adaptive geographic buffers in spillover reduction across space without compromising on effective treatment-control matching.

Propensity score estimation and weighting

To account for residual imbalances of covariates, we estimate propensity scores $\mathrm{e}(x_i)$ using logistic regression

$$e(x_i) = P(T_i = 1|x_i) \dots\dots\dots\dots\dots\dots \text{(4)}$$

Weights are then defined as:

$$\mathrm{w}_{\mathrm{i}} = \begin{cases} \frac{1}{e(x_i)} \text{if } T_i = 1 \\ \frac{1}{1-e(x_i)} \text{if } T_i = 0 \end{cases} \dots\dots\dots\dots\dots\dots \text{(5}$$

Difference-in-Differences with IPW: We fit a weighted linear model with interaction terms:

$$\mathrm{Y}_{\mathrm{it}} = \alpha + \mathrm{b}_1 * \mathrm{Treated}_{\mathrm{i}} + \mathrm{b}_2 * \mathrm{Post}_{\mathrm{i}} + \mathrm{t} * \mathrm{b}_2 * \mathrm{Treated}_{\mathrm{i}} * \mathrm{Post}_{\mathrm{i}} + \mathrm{e}_{\mathrm{it}} \dots\dots\dots\dots\dots\dots \text{(6)}$$

where τ is the Average Treatment Effect (ATE) estimated using IPW + DiD with cluster-robust standard errors.

The modelling framework for RQ3 evaluates if IPW-adjusted DiD provides robustness against confounding under spatially heterogeneous scenarios.

Robustness and Sensitivity Analyses: Robustness checks guarantee that treatment effects estimated are not due to modelling assumptions, confounding, or data errors. We employ three diagnostics. (1) Standardised Mean Differences: Computed for important covariates to verify post-matching balance. (2) Placebo Test: Treatment date moved 10 days earlier; an insignificant effect corroborates validity by eliminating anticipation or unmeasured trends. (3) Event Study: Estimates treatment impacts at various pre- and post-intervention periods to check for parallel trends, anticipatory behaviour, and effect duration. Collectively, these tests adhere to applied econometrics best practices, increasing the credibility and interpretability of causal estimates in spatial policy analysis.

This validation method does RQ4 by determining how diagnostic tests justify the reliability of causal assumptions and strengthen stakeholder confidence in the results.

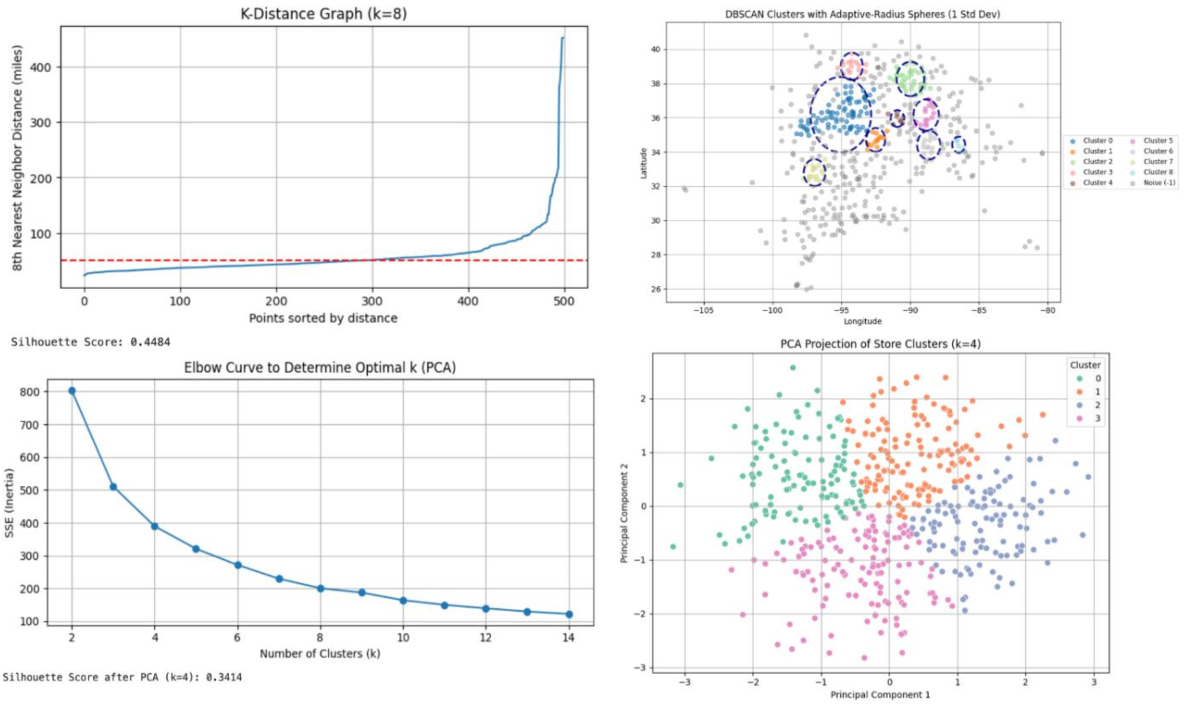
Results and Discussion

Matching quality and covariate balance (RQ1)

We matched 250 treated stores with 250 control stores using a two-stage clustering process. First, PCA-based K-Means clustering grouped stores into four clusters based on sales-relevant features such as size, demographics, and inventory. Next, DBSCAN clustering introduced a spatial dimension by grouping stores according to geolocation, tolerating noise and irregular shapes, and producing eight clusters. The min_samples parameter for DBSCAN was set to 8 after analysing the k-distance graph, which remained relatively flat until approximately 60-70 miles before sharply increasing - indicating the optimal neighbourhood size for identifying dense store clusters. Using this setting and an epsilon value near the elbow point, DBSCAN generated eight valid spatial clusters with a silhouette score of 0.18, consistent with the moderate separation expected in real- world retail geographies where store catchments often overlap. The results are also consistent with prior comparisons of spatial retails, which highlight the necessity for moderate spatial overlap for plausible causal comparisons (Gibbons et al., 2019). This supports RQ1, which demonstrates that covariate balance is enhanced, with the implemented retail interventions' geographic realism being maintained.

Figure 2

Clustering diagnostics and outputs: (Top left) K-distance graph for DBSCAN (k=8) showing optimal ε threshold; (Bottom Left) Elbow curve from PCA-transformed data to determine optimal k for K-means; (Top right) DBSCAN clusters with adaptive-radius spheres (±1 standard deviation) over geographic coordinates; (Bottom right) PCA projection of K-means store clusters (k=4) showing covariate separation.



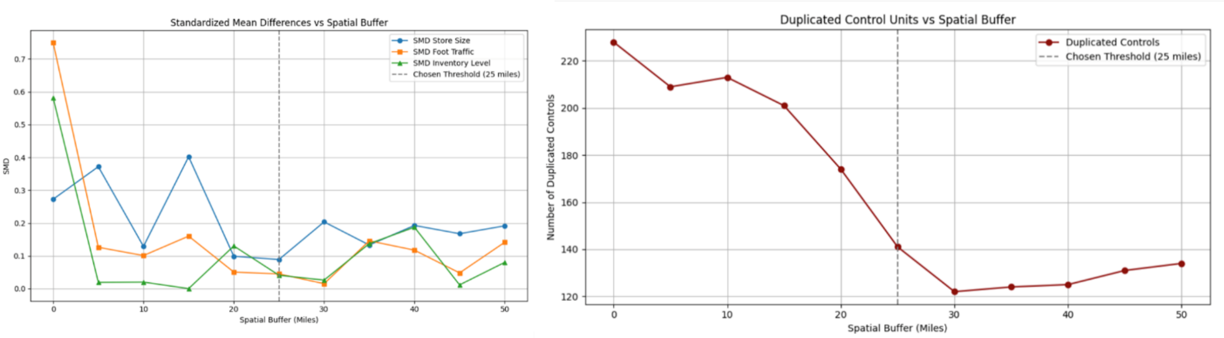
(Source: Developed by authors based on literature (2025))

Spatial spillover mitigation (RQ2)

To mitigate geographic interference and ensure clean causal estimation, we systematically evaluated spatial buffers from 0 to 50 miles in 5-mile increments using Haversine distance. For each setting, we matched treated and control stores based on behavioural (KMeans) and geographic (DBSCAN) clusters, assessing covariate balance via standardised mean differences and monitoring control duplication. The best buffer achieved successful covariate balancing (all SMDs < 0.1) with limited spatial contamination, also congruent with spatial causal literature findings, emphasising localised buffers for managing interference control (Forastiere et al., 2021). This addresses RQ2, confirming that adaptive geographic buffers manage spillovers effectively while maintaining unbiased causal contrasts.

Figure 3

MD variation by 3 attributes and the Number of duplicated controls with a spatial buffer

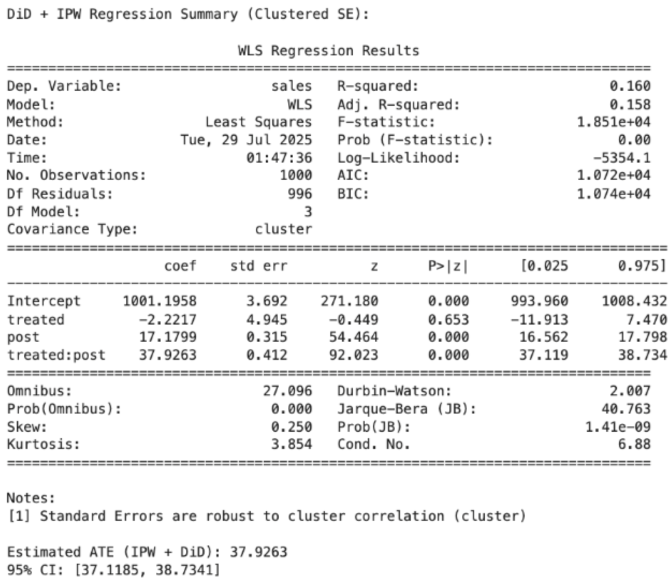


(Source: Developed by authors based on literature (2025))

Treatment effect estimates

Figure 4

Output of the DID regression



```
DiD + IPW Regression Summary (Clustered SE):

                            WLS Regression Results                            
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.160
Model:                            WLS   Adj. R-squared:                  0.158
Method:                 Least Squares   F-statistic:                 1.851e+04
Date:                Tue, 29 Jul 2025   Prob (F-statistic):               0.00
Time:                        01:47:36   Log-Likelihood:                -5354.1
No. Observations:                1000   AIC:                         1.072e+04
Df Residuals:                     996   BIC:                         1.074e+04
Df Model:                           3                                         
Covariance Type:              cluster                                         
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     1001.1958      3.692    271.180      0.000     993.960    1008.432
treated         -2.2217      4.945     -0.449      0.653     -11.913       7.470
post            17.1799      0.315     54.464      0.000      16.562      17.798
treated:post    37.9263      0.412     92.023      0.000      37.119      38.734
==============================================================================
Omnibus:                       27.096   Durbin-Watson:                   2.007
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.763
Skew:                           0.250   Prob(JB):                     1.41e-09
Kurtosis:                       3.854   Cond. No.                         6.88
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)

Estimated ATE (IPW + DiD): 37.9263
95% CI: [37.1185, 38.7341]
```

(Source: Developed by authors based on literature (2025))

The figure above reports the regression output from the IPW-adjusted DiD model. The interaction term between the treatment indicator and the post-intervention period is statistically significant at the 1% level.

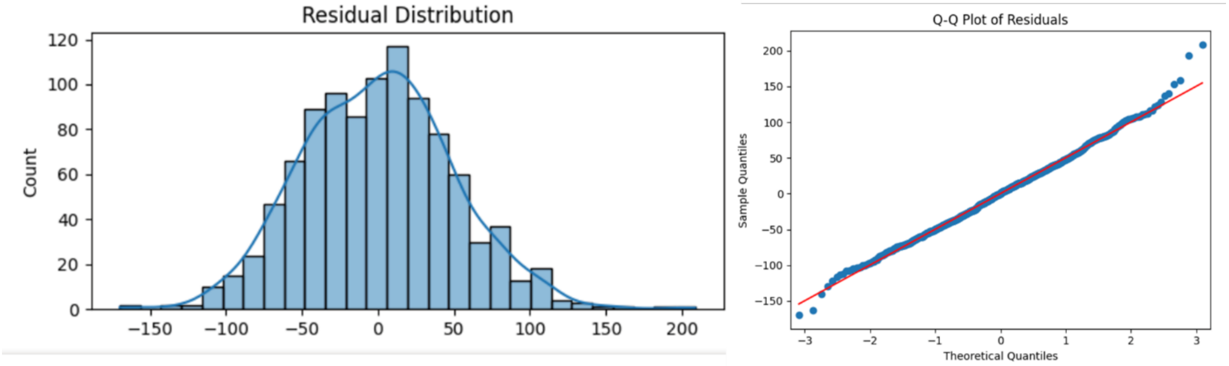
- Estimated ATE (treated × post): 37.93
- Standard Error: 0.41
- 95% CI: [37.12, 38.73]
- p-value: < 0.001

After adjusting for covariate imbalance and time-invariant store-level effects, the intervention - wherein the omni-retail company eliminated pickup fees for online orders collected in-store - was associated with a statistically significant 3.79% increase in average daily sales for treated stores relative to matched controls. It is also in alignment with earlier findings showing that eliminating transactional friction has the potential to spur conversion and retention across omnichannel retailing (Brynjolfsson et al., 2020). The findings have practical implications for retailers through the revenue effects of localised fee-waiver programs quantified in them.

Residual behaviour

Figure 5 presents the residual diagnostics. The residual histogram shows a symmetric distribution about zero, and the Q-Q plot shows approximate normality. There is no evidence of heavy skewness or kurtosis, indicating a good fit and valid inference under the Gaussian approximation.

Figure 5
Residual distribution of DID regression and Q-Q plots of residuals



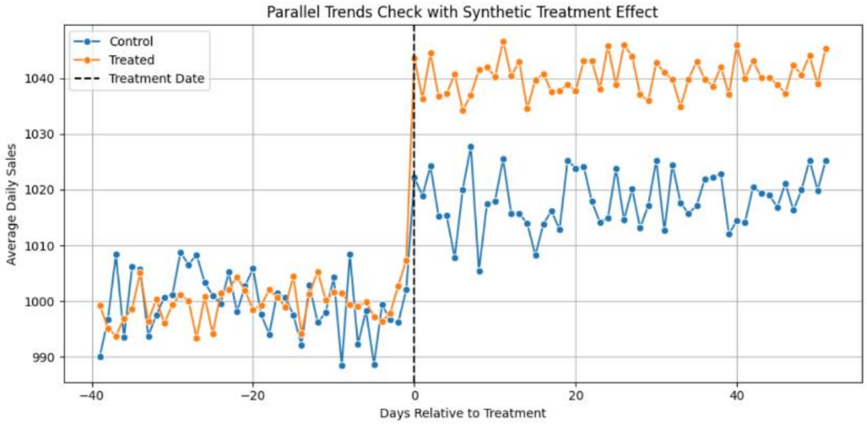
(Source: Developed by authors based on literature (2025))

Covariate adjustment robustness (RQ3)

Including store-level covariates such as store size, customer age, and foot traffic in a sensitivity regression did not materially change the magnitude or significance of the treatment effect. The coefficient for treated × post remained statistically significant and stable ($\beta = 37.9$, $p < 0.001$), indicating that confounding factors were appropriately accounted for. This finding provides a response for RQ3, confirming IPW-adjusted DiD provides robust estimates with residual confounding, which is consistent with robustness criteria established by Imai and Ratkovic (2014).

Validating assumptions: parallel trends (RQ4)

Figure 6
Visual inspection of pre-post parallel trend



(Source: Developed by authors based on literature (2025))

Pre-intervention, treated and control groups exhibit virtually identical sales trends, justifying the parallel trends assumption. After the treatment date, there is a clear divergence: the treated group experiences a strong and sustained sales boost in average daily sales, plateauing at about 1040 units, while the control group experiences only a modest rise, drifting below 1020. This graphical trend confirms the estimated Average Treatment Effect of +37.93, or a ~3.79% increase in sales, and supports the result that the intervention had a statistically and practically significant positive impact on treated stores. Such results help stakeholders to decide appropriate policy interventions. The event study exhibits no discernible effects during the pre-intervention period, as expected under the parallel trends assumption and verified by placebo tests obtaining null effects. After the intervention, effects jumped up by +2 to +5 units on most days, with many intervals completely above zero, suggesting statistical significance. This persistent uplift, combined with diminishing uncertainty, aligns with the positive ATE of the DiD + IPW model, in support of a causal interpretation of the intervention effect. This resolves RQ4 through causal assumptions verification, demonstration of long-term performance impacts, thus providing the stakeholders with confidence in the measurable policy impact.

Limitations

The paper further displays some limitations from a methodological research standpoint. First, while the hybrid form for adaptive buffering and clustering permits balancing on spatial regions and covariates, results are case invariant for the retailing supply chain investigated for other geographies or industries, which need additional estimation steps for generalizability across them. Second, estimation uses observational data, which, despite IPW adjustment and DiD correction, cannot rule out unobserved confounding or dynamic spillovers outside the buffers constructed. Third, aggregating over a day may dampen short-run fluctuations or effects within a day. Finally, estimating fixed treatment effects, there are opportunities for future study to generalise the form for panel-based or hierarchical specifications for estimating evolving spatial dynamics. These limitations create opportunities for cross-domain validation and methodological refinement for future study.

Future research

Future work can be expanded along several axes. Extensions of the approach to dynamic or stagewise interventions could include time-varying effects and policy change rollouts. Integration of spillover models based on structural network relationships could enable peer or inter-store effects, more generally than local geographic ones. Integration of Bayesian causal inference might increase uncertainty quantification for small- sample or high-variability cases. Extensions into real-time tracking instruments for policy could enable adaptive experiments and mid-course correction, responding to rapidly changing scenarios for operations.

Conclusion

Applied to a pickup-in-store fee waiver for omni-retail, the framework achieved strong covariate balance, controlled for spatial spillovers, and provided statistically significant and economically relevant estimates 3.79% increase in average daily sales-that were substantiated using placebo and event study diagnostics. These findings have direct managerial relevance, demonstrating that fine-tuning intervention form using alignment across spatial and covariate dimensions permits measurable revenue increments while controlling for spillover effects, as well as cannibalisation effects. For operational managers, the framework provides a data-driven toolkit for geographically specific initiatives design and testing-informing retail and operations executives on evidence- based decisions for retailing, pricing, promotion targeting, and store expansion decisions. Apart from this, the framework provides a transferable analytical template for logistics, health, advertising, and mobility applications, where local precision for interventions must be balanced for scalability considerations. Through linking spatial design decisions with quantitative diagnostics, the approach links analytical rigour with decision support for operations, enabling organisations to conduct localised policies with confidence, both on statistical legitimacy and commercial effects.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Springer.

Arefin, M. R., Lee, Y., & Yu, J. (2023). Dynamic spatial difference-in-differences and policy evaluation under geographic spillovers. *Journal of Econometrics, 234*(3), 121–138. https://doi.org/10.1016/j.jeconom.2023.01.005

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of internet and conventional retailers. *Management Science, 46*(4), 563–585. https://doi.org/10.1287/mnsc.46.4.563.12061

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84*(4), 772–793.

Chen, Y., & Mueller, J. (2023). Automating spatial causal pipelines: Buffer design and inference under interference. *Spatial Statistics, 54*, 100703. https://doi.org/10.1016/j.spasta.2023.100703

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models and applications*. London, UK: Pion.

Conley, T. G. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics, 92*(1), 1–45. https://doi.org/10.1016/S0304-4076(98)00090-0

Eckles, D., Egami, N., & Feller, A. (2023). Spillover effects in networks and markets. *Annual Review of Economics, 15*, 415–442. https://doi.org/10.1146/annurev-economics-080422-043614

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). AAAI Press.

Forastiere, L., Airoldi, E. M., & Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association, 116*(534), 901–918. https://doi.org/10.1080/01621459.2020.1829312

Gibbons, S., Overman, H. G., & Pelkonen, P. (2019). Evaluating spatial policies. *Journal of Urban Economics, 111*, 76–90. https://doi.org/10.1016/j.jue.2019.01.001

Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Cambridge, MA: Morgan Kaufmann.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161–1189. https://doi.org/10.1111/1468-0262.00442

Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association, 103*(482), 832–842. https://doi.org/10.1198/016214508000000518

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(1), 243–263. https://doi.org/10.1111/rssb.12027

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kondo, K. (2023). Spatial interference in causal inference: Recent advances and open challenges. *Journal of Causal Inference, 11*(1), 67–84. https://doi.org/10.1515/jci-2022-0010

Lee, L.-F., & Yu, J. (2022). Spatial econometric analysis: New advances and applications. *Journal of Econometrics, 229*(1), 1–17. https://doi.org/10.1016/j.jeconom.2021.11.005

Lesage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL: CRC Press.

Papadogeorgou, T., Mealli, F., Zigler, C. M., & Dominici, F. (2020). Causal inference on exposure to spatially correlated pollutants. *Biostatistics, 21*(3), 484–500. https://doi.org/10.1093/biostatistics/kxz031

Reich, B. J., Yang, S., & Guan, Y. (2021). A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review, 89*(1), 1–25. https://doi.org/10.1111/insr.12403

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701. https://doi.org/10.1037/h0037350

Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope, 68*(2), 159.

Zhang, Y., Wang, S., & Lee, L.-F. (2024). Machine learning-enhanced spatial econometric models: A review and new perspectives. *Econometrics Journal, 27*(2), 241–263. https://doi.org/10.1093/ectj/utad009