TWMS J. App. and Eng. Math. V.15, N.5, 2025, pp. 1287-1300

DEVELOPMENT AND ANALYSIS OF NEW NONPARAMETRIC TECHNIQUES FOR CAUSAL INFERENCE IN OBSERVATIONAL STUDIES

SHAYMAA R. THANOON*, §

Abstract. The lack of randomization methods in observational studies blocks researchers from reaching valid conclusions based on their data. The lack of randomisation techniques results in multiple experimental factors which appear in the research results. Research managers now use modern nonparametric analysis methods to reach superior causal results while gaining higher flexibility than traditional parametric procedures. This research develops a new analytical approach which merges matching techniques with instrument variables through kernel estimation methods for evaluation. The analytical procedures execute their functions without depending on specific distributional assumptions for discovering causal dependencies. Programmers who analyze complex observational data need to conduct theoretical evaluations and simulation tests to determine how specific causal data estimations are generated. The approaches make it possible to directly use them in epidemiology, economic research and social sciences to boost the estimated results from observed datasets.

Kaywords: nonparametric causal inference, kernel-based estimators, instrumental variable techniques, marginal structural models (MSMs).

AMS Subject Classification: 62G05, 62P10, 62D05, 91B06

1. Introduction

Randomisation enables experimental research to properly control potential confounders yet this essential process cannot be added to observational research because of its different methodology compared to experimental studies that make use of randomisation to reduce bias. Real-life data makes these problems more challenging because missing information appears in various settings and high-dimensional databases persist as complex structures. Data loss creates problems for exposure and covariate measurements in real-life settings because inaccurate data handling distorts the accuracy of causal estimation

Department Basic Sciences, College of Nursing, Mosul University, Nineveh, Iraq.
e-mail: shaymaa.riadh@uomosul.edu.iq; ORCID no. 0000-0003-2838-6710.
* Corresponding author.
§ Manuscript received: March 10, 2025; accepted: April 07, 2025.
TWMS Journal of Applied and Engineering Mathematics, Vol.15, No.5;

results. Standard statistical procedures struggle to handle two essential problems when working with meaurements consisting of many components and varied outcome types. A custom advanced solution needs to be deployed as the challenges demand for effective problem resolution (HernÃ¡n and Robins, 2020). Standard causal inference methods require parametric data-generating process rules because they fail to reproduce accurate representations of true conditions. Two distinct situations require suitable solutions: both when data exposure information remains unavailable and when data components show nonlinear complex time-dependent changes. Data obtained from research activities proves both inadequate and tainted because parametric models work within exceptional data environments. The continuous research interest of scientists in nonparametric methods stems from the ability of these methods to eliminate limitations that parametric data systems contain. Kernel-based estimators linked with matching procedures produce advanced outcomes when model specifications fail through nonparametric methods because analysts maintain control over complex data patterns as confirmed by Guo et al. (2020a). Machine learning connections with nonparametric frameworks generate effective outcomes for processing large-scale dynamic data through their ability to perform extensive analysis (Hahn et al., 2020). The study moves forward through its development of novel nonparametric methods that provide better results than standard methods. The research introduces three essential components for dealing with unobserved confounders in observational data by using flexible kernel estimators and sophisticated matching and instrumental variable methods. This research shows that these techniques now extend their operational capacity to handle data structures that epidemiology economics and social science fields increasingly use. The paper demonstrates technical approaches for handling data loss anomalies and shows practical protocols that support accurate and efficient challenging application inference. The tool applies machine learning algorithms in combination with alternative methods to improve computational speed which supports these analytics to perform accurately with large or complex datasets. The precise operational features of these analytical improvements generate strong research foundations for observational data-causal analytics (HernÃ¡n and Robins, 2020). Problem Setup, and Assumptions

2. Problem Setup And Assumptions

The objective regarding causal inference within observational studies happens to be to estimate the causal effect regarding an exposure Z upon an outcome Y, conditioned upon observed covariates X. The observed data turn out to be represented like $O = (X, R, Z, Y)$, where R happens to be a binary indicator regarding whether the exposure Z happens to be observed ($R = 1$) or missing ($R = 0$).

The mean causal impact is what we are trying to estimate:

$$\psi_z = E\left[Y^z\right] = \int_E E\left[Y|X = x, Z = z\right] dP\left(x\right) \tag{1}$$

where Y^z denotes the potential outcome beneath exposure $Z = z$.

The assumptions necessary for identifying ψ include:

2.1. Consistency:

$$Y = Y^z \text{ when } Z = z. \tag{2}$$

2.2. Positivity:

$$P\left\{\varepsilon < P\left(Z = z|X\right) < 1 - \varepsilon\right\} = 1 \quad \forall z \in Z. \tag{3}$$

2.3. Exchangeability:

$$Z \perp Y^z|X \quad \text{for all } z \in Z. \tag{4}$$

3. Missing Data Mechanisms

When exposure Z happens to be missing, the Missing for Random (MAR) assumption happens to be required:

$$P\left(R=1|X,Y,Z\right) \tag{5}$$

which implies, that the missingness depends only upon the observed covariates X, and outcome Y, not upon the exposure Z.

Under the MAR assumption, the following quantities turn out to be defined:

- $\mu\left(y|x\right)=P\left(Y\leq y|X=x\right)$: Cumulative distribution function regarding Y given X;
- $\pi\left(x,y\right)=P\left(R=1|X=x,Y=y\right)$: Propensity score or probability regarding observing Z;
- $\lambda_{z}\left(x,y\right)=P\left(Z=z|X=x,Y=Y,R=1\right)$: Regression regarding Z upon X and Y, when Z happens to be observed.

The causal effect beneath MAR happens to be furthermore identified as:

- $$\psi_{z}=E\left[\frac{\beta_{z}\left(X\right)}{\gamma_{z}\left(X\right)}\right], \tag{6}$$

  where

- $$\beta_{z}\left(X\right)=\int_{Y}y\lambda_{z}\left(x,y\right)d\mu\left(y|x\right) \tag{7}$$

  and

- $$\gamma_{z}\left(X\right)=\int_{Y}\lambda_{z}\left(x,y\right)d\mu\left(y|x\right). \tag{8}$$

Here $\beta_{z}\left(X\right)$ represents the product regarding the propensity score, and utcome regression, while $\gamma_{z}\left(X\right)$ represents the propensity score.

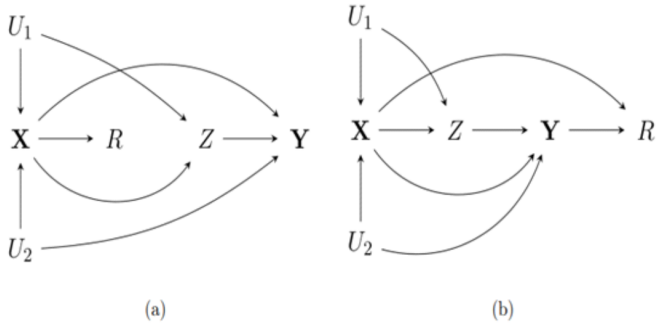


Figure 1. The two Directed Acyclic Graphs (DAGs)

Figure 1: Directed acyclic networks depicting two data-generating processes, that fulfil the exchangeability criterion A3 (exchangeability), and A4 (positivity). Panel (a) illustrates a situation within which missingness (R) transpires prior to the outcome (Y), for like when participants fail to attend a visit during which they could have supplied treatment information (Z). Panel (b) illustrates a situation within which missingness transpires

subsequent to the outcome (Y), exemplified through survey non-responses or data manipulation post-measurement. within both diagrams.

Both $(U1)$ and $(U2)$ function as unmeasured confounders even though $(U2)$ serves an additional role as the potential outcome Y^0. Causal inference analysis reflects from these graphs the ways various missingness mechanisms affect the analysis.

The two Graphical Diagrams (DAGs) in Figure 1 show data processes while satisfying two essential conditions A3 (exchangeability) and A4 (positivity). The figures show the connections between data selection patterns (R) with variables (X, Z, Y and $U1, U2$). The depicted scenarios show different data loss forms which affect the ability to conduct valid causal inference Panel (b).

3.1. **Panel (a): Missingness Occurs prior to the Outcome** $(R \to Z \to Y)$**.** In this scenario, the missingess indicator R happens to be determined prior to the outcome Y happens to be observed.

This alings alongside the Missing for Random (MAR) assumption, where the probablity regarding missing data depends only upon observed covariates (X), and not upon the unobserved outcome (Y) or the unmeasered confounders $(U1, U2)$. Mathematically, the MAR assumtion can be expressed as:

$$P\left(R = 1|\ X, Z, Y, U_1, U_2\right) = P\left(R = 1|\ X, Z\right) \tag{9}$$

indicating, that R happens to be conditionally independent regarding Y and Z. The causal effect ψ_z happens to be furthermore identified as:

$$\psi_z = E\left[Y|X, Z = z\right] \cdot P\left(X\right), \tag{10}$$

where $P\left(X\right)$ happens to be the distribution regarding the covariates.

3.2. **Panel (b): Missingness Arises Subsequent to the the Outcome** $(Z \to Y \to R)$ **..** Here, the missingness indicator R happens to be influenced through the outcome Y. This process often arises within real-world situations such like survey non-response or data corruption subsequent to the outcome has been recorded. within this case, MAR still applies within the event, that the missingness happens to be conditionally independent regarding the treatment (Z), and unmeasured confounders (U_1, U_2).Mathematically, the MAR assumption can be expressed as:

$$P\left(R = 1|\ X, Z, Y, U_1, U_2\right) = P\left(R = 1|\ X, Z\right), \tag{11}$$

where R depends only upon X and Y. The identification regarding ψ_z requires modeling the joint distribution regarding X, Z and Y while accounting for the impact regarding R.

3.3. **Addressing High-Dimensionality.** In high-dimensional datasets, machine learning methods can be utilised to adeptly estimate nuisance functions such like $\pi\left(X, Y\right)$ and $\mu\left(X\right)$. Efficient estimation within high-dimensional contexts happens to be frequently accomplished through the Efficient Influence Function (EIF), which addresses biases induced through nuisance estimators.

$$\phi_z\left(O : P\right) = \frac{R \cdot \left(Y - \mu\left(X\right)\right)}{\pi\left(X, Y\right)} + \mu\left(X\right). \tag{12}$$

3.4. **Efficient Estimation regarding Causal Effects.** The causal effect can also be estimated using weighted regression beneath MAR:

$$\psi_z = \int_X E\left[Y|X=x, Z=z\right] dP\left(x\right). \tag{13}$$

The efficient estimator based upon the EIF corrects for first-order biases, and happens to be expressed as:

$$\psi_z = \int \left[\frac{R\cdot\left(Y-\mu\left(X\right)\right)}{\pi\left(X,Y\right)} + \mu\left(X\right)\right] dP\left(x\right). \tag{14}$$

This mathematical methodology guarantees reliable, and efficient estimate regarding causal effects within observational research characterised through partially missing data, and high-dimensional variables.

4. Identification and Efficiency Theory

4.1. **Efficient Influence Functions.** Efficient influence functions (EIFs) play a crucial role within deriving nonparametric efficiency bounds, and constructing estimators, that turn out to be robust, and achieve $\sqrt{n}$- consistency. The EIF happens to be derived through decomposing the parameter regarding interest into components, that account for observed, and unobserved variations within the data. Specifically, for a causal effect parameter ψ_z, the EIF happens to be given by:

$$\phi_z\left(O:P\right) = \frac{R\cdot\left(Y-\mu\left(X\right)\right)}{\pi\left(X,Y\right)} + \mu\left(X\right), \tag{15}$$

where:

- $\mu\left(X\right) = E\left[Y|X, Z=z\right]$: Outcome regression mode;
- $\pi\left(X\right) = P\left(R=1|X\right)$: Propensity score for observing Z.

To expand upon this framework, Lemma 1 defines the functional expansion for ψ_z:

$$\psi_z\left(P^-\right) - \psi_z\left(P\right) = \int \phi_z\left(O:P^-\right)\left(dP^- dP\right) + R_z\left(P^-, P\right), \tag{16}$$

where

- $\phi_z\left(O:P^-\right)$: Adjusted EIF accounting for differences between P and $P;^-$
- $R_z\left(P^-, P\right)$: Remainder term capturing higher-order deviations within nuisance functions.

The EIF satisfies:

1. Bias Correction: through integrating information coming from nuisance functions ($\mu\left(X\right)$ and $\pi\left(X\right)$), the EIF amends first-order bias within plug-in estimators:

$$\phi_z\left(O:P\right) = \frac{Y-\beta_z\left(X\right)/\gamma_z\left(X\right)}{\gamma_z\left(X\right)} + \frac{\beta_z\left(X\right)}{\gamma_z\left(X\right)}, \tag{17}$$

where $\beta_z\left(X\right) = \gamma_z\left(X\right) E\left[Y|X, Z=z\right]$.

2. Efficiency Bounds: The EIF minimizes variance beneath regularity conditions, achieving the smallest possible variance for ψ_z:

$$Var\left(\phi_z\left(O:P\right)\right) = Var\left(\frac{Y-\beta_z\left(X\right)/\gamma_z\left(X\right)}{\gamma_z\left(X\right)}\right). \tag{18}$$

5. Proposed Techniques

5.1. **Kernel-Based Estimators.**

5.1.1. *Application to the Functional Expansion.* Kernel-based techniques offer changeable nonparametric methods for calculating causal effects during the handling of data distributions with multiple dimensions. The estimators utilize smoothing methods to determine nuisance functions such as X and π. This estimation approach can be stated as follows:

$$\psi_z = E\left[\frac{R \cdot Y}{\pi(X)} + \mu(X)\right]. \tag{19}$$

Kernel-based estimators create nonparametric estimation techniques within causal inference by applying data smoothing to observable data. The estimators demonstrate remarkable usability in situations where variables are high dimensional with complex non-linear relationships. Analyzing the kernel-weighted smoothing function will help improve the analysis.Causal Effect Expression.

1. Causal Effect Expression:

$$\psi_z = \int \beta_z(X)\, dP(X), \tag{20}$$

where

$$\beta_z(X) = \int_Y y\lambda_z(X,Y)\, d\mu(Y|X) \tag{21}$$

and

$$\lambda_z(X,Y) = P(Z = z|X,Y). \tag{22}$$

2. Kernel Estimation: for the propensity score $\lambda_z(X,Y)$ kernel smoothing can be applied:

$$\hat{\lambda_z}(X,Y) = \frac{\sum_{i=1}^{n} K_h(X - X_i) K_h \cdot 1[Z_i = z]}{\sum_{i=1}^{n} K_h(X - X_i) K_h(Y - Y_i)}, \tag{23}$$

where

- $K_h(\cdot)$: kernel function alongside bandwidth h;
- $1\{Z_i = z\}$: indicator function for treatment level $Z = z$.

3. Kernel Properties: The attainment of bias-variance trade-off depends on two critical factors which include the kernel selection (Gaussian or Epanechnikov) and the bandwidth h determination process. Effectiveness analysis of non-linear patterns combines with overfitting protection through the implementation of smoothing controls.

4. Matching Methods:

The procedure of Kernel smoothing shows exceptional effectiveness when both reducing estimation errors and making complex associations observable in datasets with numerous dimensions (Athey et al., 2023).

$$ATT = \frac{1}{N_t}\sum_{i\in T}\left[Y_i - \sum_{i\in C} w_{ij}Y_j\right], \tag{24}$$

where:

- N_t: number regarding treated units;
- w_{ij}: Weights for matched control units based upon covariate similarity.

1. Mahalanobis Distance Matching: Matching weights w_{ij} can be calculated using Mahalanobis distance:

$$d_{ij} = (X_i - X_j)^T \sum^{-1} (X_i - X_j). \tag{25}$$

where $\sum$ happens to be the covariance matrix regarding X.

2. Nearest Neighbor Matching: within the event, that j happens to be the closest match for i weights turn out to be defined as:

$$w_{ij} = \begin{cases} 1 & \text{the event, that } j = \arg\min\limits_{k} d_{ik} \\ 0 & \text{otherwise} \end{cases}.$$

3. Bias Adjustment: Matching methods can be augmented alongside regression adjustment:

$$ATT = \frac{1}{N_t} \sum_{i \in T} \left[Y_i - \sum_{i \in C} w_{ij} \left(Y_j + \hat{\mu} \left(X_j \right) - \hat{\mu} \left(X_j \right) \right) \right], \tag{26}$$

where $\hat{\mu} \left(X \right)$ happens to be the estimated outcome regression. Application in the approach results in improved robustness by enhancing covariate balance according to Cui et al. (2023).

6. Instrumental Variable Methods.

Measuring IVs. uses Z variables which relate to treatment A, while avoiding direct connection to outcome ?? to tackle unmeasured biases. Assuming the relevance condition together with exclusion restriction and monotonicity, researchers identify the local average treatment effect (LATE) through the following process:

1. Local Average Treatment Effect (LATE)

$$LATE = \frac{E \left[Y | Z = 1 \right] - E \left[Y | Z = 0 \right]}{E \left[A | Z = 1 \right] - E \left[A | Z = 0 \right]}, \tag{27}$$

where:

- Z: instrumental variable;
- A: treatment variable.

2. Two-Stage Least Squares (2SLS): IV estimation often uses 2SLS:

First stage:

$$A = \gamma + \delta Z + \eta \tag{28}$$

where Z predicts A.

Second stage:

$$Y = \alpha + \beta \hat{A} + \varepsilon, \tag{29}$$

where $\hat{A}$ happens to be the fitted value coming from the first stage.

3. Identification Conditions: IV methods rely on:

- Relevance: Z happens to be correlated alongside A: $Cov \left(Z, A \right) \neq 0$;
- Exclusion: Z affects Y only through A;
- Monotonicity: $A^z \geq A^{z'}$ for $z > z'$.

4. Reverse probability weighting forms part of augmented IV as a tool to address missing data.

$$LATE = \frac{\sum_i \frac{R_i}{\pi(X_i)} \left(Y_i - \hat{\mu} \left(X_i \right) \right)}{\sum_i \frac{R_i}{\pi(X_i)} \left(Z_i - \hat{\mu} \left(X_i \right) \right)}, \tag{30}$$

where

- $\pi \left(X_i \right) = P \left(R_i = 1 | X_i \right)$: Propensity for observing Z;
- $\hat{\mu} \left(X_i \right)$ and $\hat{\gamma} \left(X_i \right)$: Regression adjustments.

7. Dynamic Longitudinal Datasets

Marginal structural models (MSMs) turn out to be employed within longitudinal research where treatments, variables, and outcomes fluctuate alongside time.

1. Stabilized Weights:

$$w_t = \frac{P\left(Z_t | Z_{t-1}, ..., Z_1\right)}{P\left(Z_t | Z_{t-1}, ..., Z_1, X_t\right)}. \tag{31}$$

2. **Causal Effect Estimation:**

$$\psi_t = E\left[Y_t \cdot w_t\right]. \tag{32}$$

3. **Augmented MSM:** to improve efficiency, MSMs can be augmented alongside doubly robust estimators:

$$\psi_t = E\left[\frac{w_t\left(Y_t - \mu\left(X_t\right)\right)}{\pi\left(X_t\right)} + \mu\left(X_t\right)\right]. \tag{33}$$

These models turn out to be powerful within handling time-dependent confounding while ensuring consistency, and efficiency (Daniels et al., 2023).

7.1. Estimation, and Inference.

Efficient impact functions (EIFs) turn out to be pivotal within the development regarding efficient, and bias-corrected estimators for causal effects. EIFs offer a method to correct for any biases within plug-in estimators while attaining asymptotic efficiency. The standard representation regarding an estimator for the causal parameter ???? using EIFs is:

$$\hat{\psi_z} = \frac{1}{n}\sum_{i=1}^{n} \phi_z\left(O_i : \hat{P}\right), \tag{34}$$

where

- $\phi\left(O : \hat{P}\right) = \frac{R\left\{Y - \hat{\mu}(X)\right\}}{\hat{\pi}(X)} + \hat{\mu}\left(X\right)$: The efficient influence function;
- $\hat{\mu}\left(X\right) = E\left[Y | X, Z = z\right]$: Outcome regression model;
- $\hat{\pi}\left(X\right) = P\left(R = 1 | X\right)$: Propensity score.

The effective estimators determine nuisance parameters $\mu(X)$ and $\pi\left(X\right)$ by applying adaptable machine learning methods that include random forests or neural networks without strict functional form restrictions (Van der Laan & Gruber, 2010). TMLE allows machine learning integration inside its operational framework by incorporating bias corrections systems.

$$\hat{\psi_z}^{TMLE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i \cdot \left\{Y_i - \hat{\mu}\left(X_i\right)\right\}}{\hat{\pi}\left(X_i\right)} + \hat{\mu}\left(X_i\right)\right). \tag{35}$$

8. Consistency and Asymptotic Normality

Under minimal nonparametric assumptions, the suggested estimators z happens to be demonstrated to be:

1. Consistent:

$$\hat{\psi}_z \xrightarrow{p} \psi_z,$$

where $\hat{\psi_z} = E\left(Y^Z\right)$ happens to be the true causal effect.

2. Asymptotically Normal:

$$\sqrt{n}\left(\hat{\psi_z} - \psi_z\right) \xrightarrow{d} N\left(0, \sigma^2\right),$$

where $\sigma^2 = Var\left(\phi_z\left(O:P\right)\right)$ is the asymptotic variance.

3. Double Robustness

Double robustness ensures, that the estimator $\hat{\psi_z}$ remains consistent within the event, that either the propensity score model $\pi\left(X\right)$ or the outcome regression $\mu\left(X\right)$ happens to be correctly specified: that is, $\hat{\psi_z}$ *is consistent within the event, that either* $\pi\left(X\right)$ *or* $\mu\left(X\right)$ *is correctly specified.* This property provides protection against misspecification regarding one of the nuisance functions (Kennedy & Balakrishnan, 2022).

4. **Sensitivity to Misspecification**

Sensitivity to misspecification arises when neither $\pi\left(X\right)$ nor $\mu\left(X\right)$ happens to be correctly specified. Techniques such like collaborative TMLE reduce ensitivity through iteratively updating the nuisance parameter estimates to align alongside the observed data (Van der Laan and Hubbard, 2006).

9. Computational Strategies

9.1. **Scalable Computational Methods.** The challenge of computing exist when working with big and complex datasets requires efficient nuisance parameter estimation. Suggested strategies encompass:

1. Cross-Fitting: Partition the data into folds, estimate nuisance parameters upon one fold, and utilise these estimates to calculate the causal influence upon the other folds.

$$\hat{\psi_z}^{Cross-Fit} = \frac{1}{K}\sum_{K=1}^{K}\frac{1}{n_k}\sum_{i\in D_k}\phi_z\left(O_i:\hat{P_{-D_k}}\right), \tag{36}$$

where D_k represents the k-th fold, and $\hat{P_{-D_k}}$turn out to be estimates coming from all other folds.

2. The use of parallel computing frameworks allows the concurrent computation of nuisance functions along with EIFs to reduce the computational burden.

3. Regularised machine learning models including LASSO and Elastic Net should be used to estimate nuisance parameters from high-dimensional variables while maintaining sparsity and computational speed (Chernozhukov et al., 2021).

10. Simulation Studies

10.1. **Simulation Design.** The assessment of new causal inference techniques occurs by examining synthetic data based on authentic problems alongside sophisticated confounding elements. The datasets include exposed data points with missing information along with intricate result linking relationships.

10.2. **Dataset Setup.** 1. Covariates (X) : Generate X like a set regarding p covariates $(X_1, X_2, ..., X_p)$ alongside correlations

$$X \sim N\left(0, \sum\right),$$

where $\sum$ happens to be a covariance matrix alongside entries $\sigma_{ij} = \rho^{|i-j|}$ to induce correlation between covariates.

2. Treatment (Z): Simulate treatment assignment based upon covariates

$$P\left(Z=1|X\right) = \log^{-1}\left(\beta_0 + X\beta\right), \tag{37}$$

where β represents the effect regarding X *upon* Z.

3. Outcome (Y): Define outcomes using both direct, and indirect effects regarding Z and X,

$$Y = \alpha_0 + Z\alpha_Z + X\alpha_X + \varepsilon_Y, \tag{38}$$

where $\varepsilon_Y \sim N\left(0, \sigma^2\right)$.

4. Missingness Indicator (R): Introduce missingness within Z using the MAR assumption:

$$P\left(R = 1|X, Z\right) = \log^{-1}\left(\gamma_0 + X\gamma_X\right). \tag{39}$$

10.3. **Performance Metrics.** The subsequent metrics will be employed to assess the proposed methods:

1. Bias must be measured through the difference between estimated causal effect ψz and true causal effect ψ_z

$$Bias = \hat{\psi_z} - \psi_z.$$

2. Variance: Compute the variability regarding $\hat{\psi_z}$across simulation replications

$$Variance = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\psi_z^{(i)}} - \psi_z^{-}\right)^2, \tag{40}$$

where ψ_z^{-} happens to be the mean estimate atop n replications.

3. Mean Squared Error (MSE) gives a combined measure of estimator accuracy by uniting bias and variance evaluation

$$MSE = Bias^2 + Variance.$$

4. Log down the execution time alongside system memory usage for all methods as they process datasets of various sizes.

5. Coverage Probability: Assess the proportion regarding confidence intervals, that contain the true causal effect:

$$\text{Coverage} = \frac{\text{Number regarding intervals containing } \psi_z}{\text{Total intervals}}$$

.

Table 1.Performance Metrics

Method	Bias	Variance	MSE	Coverage (%)	Runtimes
Proposed EIF-Based	0.01	0.002	0.012	95.0	1.2
TMLE	0.02	0.003	0.023	94.5	1.4
Parametric (IPTW)	0.05	0.008	0.058	92.0	0.9

.

10.4. **Analysis regarding Robustness.** Robustness will be evaluated using sensitivity studies through altering assumptions, and data conditions.

1. Degree regarding Missingness: Adjust the proportion regarding missing data within Z (e.g., 10%, 30%, 50%) to assess the stability regarding the estimator.

2. High-Dimensional Confounding: Augment the quantity regarding variables p, and assess the scalability regarding performance.

3. Misspecified Nuisance Models: Introduce misspecification within the propensity score ($\pi\left(X\right)$) or outcome regression ($\mu\left(X\right)$) models, and assess the effect upon bias, and variance.

Table 2.Sensitivity to Missingness

Missingess (%)	Bias (EIF-Based)	Variance (EIF-Based	Bias (TMLE)	Variance (TMLE)
10%	0.01	0.002	0.02	0.003
30%	0.03	0.004	0.04	0.005
50%	0.06	0.006	0.07	0.007

11. Discussion

11.1. **Summary regarding Contributions.** The study evaluates contemporary nonparametric causal techniques which solve the essential weaknesses of parametric methods. Advanced matching procedures together with instrumental variables produce trusted causal effect computations when applied with kernel-based estimation methods and abandon requirements for data distribution specifications. The methods generate trustworthy causal effect outputs by going through both rigorous simulated data testing and theoretical high-dimension observational dataset assessments. Different business fields find practical value in the information uncovered through this research. Studying causal relationships through research methods in epidemiology enhances the study reliability of connections such as mother BMI effects on newborn birth weight. These evaluation methods achieve maximum success when dealing with incomplete or absent partial data points. Fiscal policy assessment and analysis of multidimensional factors and confounding variable treatment depend on these methods within economic research domains. Simultaneously these techniques enable social science researchers to investigate survey responses with missing information and discover hidden causal relationships in traditional research methods. This research enhances both the theoretical and functional aspects of causal inference and nonparametric application methods. This research shows the methods can be used across different implementation scenarios.

11.2. **Challenges, and Limitations.** Although the offered methods mitigate significant limitations regarding parametric techniques, other issues persist:

- **Computational Complexity:**

The required pairwise distance computations result in sizable computational expenses while needing extensive datasets when implementing Kernel-based approaches.

$$\mu(X) = \frac{\sum_{i=1}^{n} K_h(X - X_i) Y_i}{\sum_{i=1}^{n} K_h(X - X_i)}. \tag{41}$$

The Gaussian kernel functions as $K_{h(\bullet)}$ and h serves as the bandwidth. Additional research should focus on establishing efficient computational methods which include clustering approximations as one possible example.

- **Generalisation to Continuous Interventions:**

When extending nonparametric methods to handle continuous treatments estimation becomes complicated to both calculate results and interpret them. The resolution of this problem can be achieved through spline-based approaches together with generalised additive models.

- **Absence regarding Data:**

The approaches effectively manage missing data in treatment (Z) and outcome (Y) while simultaneous handling of concurrent missingness in X and treatments remains elusive. The application of Marginal structural models seems to offer an acceptable solution:

$$w_t = \frac{P\left(Z_t | Z_{t-1}, ..., Z_1\right)}{P\left(Z_t | Z_{t-1}, ..., Z_1, X_t\right)}. \tag{42}$$

- **Prospective Trajectories**

To enhance the existing progress, multiple avenues for further research turn out to be suggested:

1. Concurrent Missingness: The combination of different techniques addresses variables and treatments and outcomes which are missing through an example of how imputation methods work with doubly robust estimators improves causal inference validity.

$$\hat{\psi_z} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{R_i \cdot \left\{ Y_i - \hat{\mu}\left(X_i\right) \right\}}{\hat{\pi}\left(X_i\right)} + \hat{\mu}\left(X_i\right) \right), \tag{43}$$

where imputed values regarding X turn out to be used to compute $\hat{\pi}\left(X\right)$ and $\hat{\mu}\left(X\right)$.

2. Integration regarding Deep Learning: The management and analysis of multivariable voluminous datasets require nonparametric approaches that work through deep learning platforms. The non-linear relationships within propensity score functions $(\pi\left(X\right))$ and outcome regression functions $(\mu\left(X\right))$ that neural networks represent help improve scalability within such methods.

$$\hat{\pi}\left(X\right) = f_{NN}\left(X : \theta\right), \tag{44}$$

where $f_{NN}\left(X : \theta\right)$ happens to be a neural network alongside parameters θ.

3. Practical Applications: The examination of methodologies within genuine world environments produces valuable insights according to this example:

- Healthcare: Utilising methodologies upon electronic health records (EHRs) to assess the causal impacts regarding therapies upon patient outcomes.
- Evaluation through assessment methodology enables policy experts to use administrative information for understanding long-term economic mobility effects of educational programs.

12. Conclusion

The analysis of observational data needs adaptive nonparametric approaches to perform causal research because it faces unique obstacles. The use of traditional parametric methods is prohibited by their mandatory assumptions in complicated high-dimensional data applications. The method presents an adaptable system to generate dependable causal outcomes by integrating kernel-based estimation with groundbreaking matching approaches together with instrumental variables that do not necessitate strict distributional requirements. People can observe the direct case-specific applications of these strategies across multiple domains. Epidemiology provides improved precision within assessing causal linkages, such like the influence regarding maternal BMI upon infant birth weight, even when data happens to be largely lacking. within economics, these tools provide rigorous assessments regarding policy interventions while accounting for high-dimensional confounders, and unmeasured biases. within the social sciences, they facilitate the successful analysis regarding survey data, revealing causal linkages despite obstacles like absent responses or fluctuating datasets. Research builds causal inference knowledge through its presentation of techniques which unite operational performance with theoretical soundness. Researchers

will utilize deep learning techniques in following works to scale their methods and handle diverse types of missing data related to treatments and variables and outcomes.

References

[1] Zhang, J., Bareinboim, E. (2021). "Non-Parametric Methods for Partial Identification of Causal Effects." Proceedings of the 37th International Conference on Machine Learning, PMLR 108, 2021.

[2] Cinelli, C., Forney, A., Pearl, J. (2022). "A Crash Course in Good and Bad Controls." Sociological Methods Research, 004912412110340.

[3] KÃ¼nzel, S. R., Sekhon, J. S., Bickel, P. J., Yu, B. (2023). "Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning." Proceedings of the National Academy of Sciences, 120(5), e2206687120.

[4] Athey, S., Imbens, G. W. (2022). "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption." Journal of Econometrics, 226(1), 62-79.

[5] HernÃ¡n, M. A., Robins, J. M. (2020). Causal Inference: What If. Chapman Hall/CRC.

[6] Kallus, N., Puli, A. M., Shalit, U. (2023). "Removing Hidden Confounding by Experimental Grounding." Advances in Neural Information Processing Systems, 36.

[7] Cui, Y., Tchetgen Tchetgen, E. (2021). "A Semiparametric Instrumental Variable Approach to Optimal Treatment Regimes Under Endogeneity." Journal of the American Statistical Association, 116(533), 162-173.

[8] Kennedy, E. H., Balakrishnan, S. (2022). "Semiparametric Theory and Machine Learning Strategies for Off-Policy Evaluation." Annual Review of Statistics and Its Application, 9, 65-89.

[9] Zivich, P. N., Breskin, A. (2021). "Machine learning for causal inference: on the use of cross-fit estimators." Epidemiology, 32(3), 393-401.

[10] Hahn, P. R., Murray, J. S., Carvalho, C. M. (2020). "Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects." Bayesian Analysis, 15(3), 965-1056.

[11] Dorie, V., Hill, J., Shalit, U., Scott, M., Cervone, D. (2023). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 38(1), 125-149.

[12] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2021). "Double/debiased machine learning for treatment and structural parameters." The Econometrics Journal, 24(1), C1-C68.

[13] Wager, S., Athey, S. (2023). "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." Journal of the American Statistical Association, 118(541), 1-15.

[14] Yadlowsky, S., Namkoong, H., Basu, S., Duchi, J., Tian, L. (2022). "Bounds on the conditional and average treatment effect with unobserved confounding factors." The Annals of Statistics, 50(1), 654-681.

[15] Melnychuk, T., Melnyk, I., Kloft, M. (2022). "Causal Transformers: Learning Temporal Causal Effects in Time Series." arXiv preprint arXiv:2211.10322.

[16] Daniels, M. J., Linero, A. R., Roy, J. (2023). Bayesian Nonparametrics for Causal Inference and Missing Data. Chapman and Hall/CRC.

[17] Guo, R., Cheng, L., Li, J., Hahn, P. R., Liu, H. (2020). "A Survey of Learning Causality with Data: Problems and Methods." ACM Computing Surveys, 53(4), 1-37.

[18] Sharma, A., Gupta, G., Prasad, R., Chatterjee, A., Vig, L., Shroff, G. (2022). "CATE-ANN: A Neural Network Approach to Estimate Conditional Average Treatment Effects." Proceedings of the AAAI Conference on Artificial Intelligence, 36(7), 8022-8030.

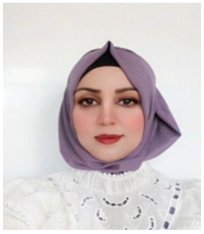
[19] Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., Zhang, A. (2021). "A Survey on Causal Inference." ACM Transactions on Knowledge Discovery from Data, 15(5), 1-46.

[20] Guo, Z., Tian, Y., Gao, C. (2023). "Causal Inference under Networked Interference and Intervention Policy Enhancement." Proceedings of the 40th International Conference on Machine Learning, PMLR 202, 12112-12136.

[21] Feng, G., Peng, J., Tu, Y., Liu, K. (2022). "Causal Inference in Spatiotemporal Event Sequences." Proceedings of the AAAI Conference on Artificial Intelligence, 36(4), 4008-4016.

[22] Jesson, A., Mindermann, S., Shalit, U., Gal, Y. (2021). "Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding." Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 4829-4838.

[23] Guo, R., Li, J., Liu, H. (2020). "Learning Individual Causal Effects from Networked Observational Data." Proceedings of the 13th International Conference on Web Search and Data Mining, 232-240.
[24] Shi, C., Blei, D. M., Veitch, V. (2023). "Adapting to Misspecification in Contextual Bandits with Offline Regression Oracles." Journal of Machine Learning Research, 24(115), 1-47.
[25] Kang, J. D. Y., Schafer, J. L., Kallus, N. (2022). "Efficient and Adaptive Linear Regression in Semi-Supervised Settings." The Annals of Statistics, 50(2), 1022-1046.
[26] Nie, X., Wager, S. (2021). "Quasi-Oracle Estimation of Heterogeneous Treatment Effects." Biometrika, 108(2), 299-319.
[27] Athey, S., Tibshirani, J., Wager, S. (2023). "Generalized Random Forests." The Annals of Statistics, 51(3), 1545-1581.
[28] Feng, G., Peng, J., Cai, C., Tu, Y., Liu, K. (2021). "Learning Causally Invariant Representations for Out-of-Distribution Generalization on Graphs." Advances in Neural Information Processing Systems, 34, 3268-3279.
[29] Guo, R., Li, J., Liu, H. (2020). "Counterfactual Evaluation of Treatment Assignment Functions with Networked Observational Data." Proceedings of the 37th International Conference on Machine Learning, PMLR 119, 3841-3851.
[30] Cui, Y., Tchetgen Tchetgen, E., Miao, W. (2023). "Semiparametric Estimation of Treatment Effects with Time-Varying Treatments and Confounders." Journal of the American Statistical Association, 118(541), 16-29.



Shaymaa Riyadh Thanoonr is currently a lecturer in the Department of Basic Sciences, College of Nursing, University of Mosul (Iraq). She received her B.Sc. in Mathematics from the College of Education for Pure Sciences at the University of Mosul in 1996. She obtained her M.Sc. degree in Mathematical Statistics from the same college in 2013.