mathematics

MDPI

Article

A Copula-Tensor Neural Network Framework for High-Dimensional Causal Inference

Jong-Min Kim [1,2]

[1] Statistics and Data Science Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, USA; jongmink@morris.umn.edu
[2] EGADE Business School, Tecnológico de Monterrey, Ave. Rufino Tamayo, Rufino Tamayo, Monterrey 66269, Mexico

Abstract

Estimating conditional average treatment effects (CATEs) in high-dimensional causal inference problems remains challenging because complex nonlinear relationships, heterogeneous feature distributions, and dependence among covariates can limit the effectiveness of conventional machine learning approaches. To address this challenge, we propose a copula-enhanced neural learning framework that integrates empirical copula transformations, manifold-based feature augmentation, structured treatment–covariate interaction representations, and deep neural networks for flexible CATE estimation. The empirical copula transformation does not introduce additional dependence information; instead, it provides a rank-based feature representation that normalizes marginal distributions, reduces sensitivity to heterogeneous feature scales and extreme observations, and offers a dependence-aware representation for subsequent learning. The proposed framework is evaluated through Monte Carlo simulations under diverse data-generating mechanisms and a real-world application using the Criteo uplift dataset. The simulation study examines the contribution of individual model components through ablation experiments and compares the proposed approach with established causal learning methods. Results demonstrate that the proposed framework achieves competitive CATE estimation accuracy while providing stable policy evaluation based on Inverse Propensity Scoring (IPS) and Doubly Robust (DR) estimators. In the Criteo application, the proposed method exhibits predictive performance comparable to conventional neural-network approaches while producing more stable Doubly Robust policy value estimates. These findings suggest that copula-based feature representations combined with deep learning provide a flexible approach for heterogeneous treatment effect estimation, particularly in high-dimensional settings with complex covariate dependence. The benefits of the proposed framework depend on data characteristics, including sample size, dimension, dependence structure, and treatment assignment mechanisms.

Keywords: causal inference; copula modeling; policy value estimation; doubly robust estimators; RMSE; model stability

MSC: 62H22; 68T07; 62H20; 62F40; 62H35

Academic Editors: Josep Maria Oller and David Bernal-Casas

Received: 15 July 2026
Revised: 7 August 2026
Accepted: 18 August 2026
Published: 21 August 2026

1. Introduction

Estimating heterogeneous treatment effects is a fundamental challenge in causal inference, with important applications in precision medicine, public health, economics,

marketing, and policy optimization. Rather than focusing solely on a single average treatment effect for the entire population, modern decision-making seeks to characterize how treatment effects vary across individuals with different covariate profiles. The conditional average treatment effect (CATE) provides a useful framework for quantifying such treatment-effect heterogeneity and identifying individuals or subpopulations that are more likely to benefit from a particular treatment. Accurate CATE estimation enables personalized interventions by identifying individuals or subpopulations that are most likely to benefit from a particular treatment. In observational studies, however, estimating CATE remains challenging because treatment assignment is not randomized, confounding variables may be high-dimensional, and complex nonlinear relationships frequently exist among covariates, treatments, and outcomes.

Classical causal inference methods, including outcome regression, inverse probability weighting (IPW), and Doubly Robust (DR) estimators, provide rigorous statistical foundations under assumptions such as consistency, ignorability, and positivity [1–3]. Although these approaches possess attractive theoretical properties, their practical performance often depends on correctly specifying either the outcome regression model or the propensity score model. When covariate effects are highly nonlinear or involve complex interactions, model misspecification can substantially reduce estimation accuracy.

To address these limitations, machine learning methods have become increasingly popular for heterogeneous treatment effect estimation. Bayesian Additive Regression Trees (BARTs), Causal Forests, meta-learners, and deep neural networks provide flexible function approximation methods capable of modeling nonlinear treatment effects without requiring restrictive parametric assumptions [4–7]. These methods primarily focus on learning predictive relationships from observed covariates. However, relatively less attention has been given to how the representation of covariates themselves affects CATE estimation, particularly when predictors exhibit complex dependence structures, heterogeneous marginal distributions, or heavy-tailed behavior. In such settings, learning algorithms may become sensitive to feature scaling, extreme observations, and optimization instability, potentially affecting estimated treatment effects and downstream policy evaluation.

Recent advances in representation learning have demonstrated that transforming the covariate space can improve causal estimation. For example, manifold learning approaches have been used to extract lower-dimensional geometric representations that capture intrinsic data structure before estimating heterogeneous treatment effects [8]. While such approaches exploit geometric information, they do not explicitly incorporate dependence structures among covariates. In contrast, copula methodology provides a principled statistical framework for separating marginal distributions from dependence structures through rank-based transformations [9–11]. Because empirical copula transformations are invariant to monotone marginal transformations, they provide a rank-based representation that reduces sensitivity to heterogeneous scales and extreme observations while reflecting the dependence structure already present in the data.

The potential value of such representations for causal learning remains largely unexplored. Existing heterogeneous treatment effect estimators typically utilize either the original covariates or learned latent representations, without systematically investigating whether rank-based dependence-aware representations can improve the stability of CATE estimation and policy evaluation. Furthermore, limited empirical evidence exists regarding the individual contributions of copula transformations, manifold representations, and treatment–covariate interaction features within a unified causal learning framework.

Motivated by these observations, we propose a copula-enhanced neural learning framework for heterogeneous treatment effect estimation. The proposed approach integrates empirical copula transformations, manifold-based feature augmentation, treatment–

covariate interaction features, and deep neural networks to construct flexible representations for CATE estimation. The empirical copula transformation is not intended to introduce new dependence information; instead, it provides a robust rank-based representation by normalizing marginal distributions and reflecting the dependence structure already contained in the original covariates. Treatment–covariate interaction features are incorporated to facilitate learning heterogeneous treatment effects through a flexible neural-network architecture.

The proposed methodology is evaluated through Monte Carlo simulations under multiple data-generating scenarios with varying dependence structures, dimensional settings, and marginal distributions. Additional ablation experiments investigate the individual contributions of copula transformations, manifold features, and interaction features. Benchmark comparisons with established causal learning methods are conducted to assess the relative performance of the proposed framework. Finally, the methodology is illustrated using the Criteo uplift dataset, where predictive performance and policy evaluation are assessed using Inverse Propensity Scoring (IPS) and Doubly Robust (DR) estimators.

The primary contributions of this paper are summarized as follows.

1. We propose a copula-enhanced deep learning framework that integrates empirical copula transformations, manifold-based feature augmentation, and treatment–covariate interaction features for heterogeneous treatment effect estimation.
2. We investigate the role of rank-based covariate representations in causal learning and examine how empirical copula transformations can improve robustness to heterogeneous marginal distributions and extreme observations while providing a dependence-aware representation of high-dimensional covariates.
3. We conduct simulation studies, including ablation analyses and comparisons with established causal inference methods, to evaluate the contribution of each component of the proposed framework under diverse data-generating settings.
4. We demonstrate through a real-world uplift modeling application that the proposed framework achieves competitive policy evaluation performance and provides stable Doubly Robust estimates in high-dimensional treatment effect learning problems.

The remainder of this paper is organized as follows. Section 2 presents the proposed methodology, including the empirical copula transformation, feature construction procedure, and neural network architecture. Section 3 reports the simulation studies and benchmark comparisons. Section 4 presents the application to the Criteo uplift dataset. Section 5 concludes with a discussion of empirical findings, computational considerations, limitations of the proposed framework, and directions for future research.

2. Method

2.1. Simulation Data-Generating Mechanism

To evaluate the proposed copula-augmented neural network framework, we construct a series of observational data-generating mechanisms that represent key challenges encountered in high-dimensional causal inference, including correlated covariates, nonlinear treatment assignment, and heterogeneous treatment effects. The simulation design is motivated by applications such as precision medicine and policy evaluation, where observed characteristics may exhibit complex dependence structures and treatment decisions are influenced by multiple patient-level or contextual variables.

Throughout the simulation study, we assume that the standard identification conditions for causal inference hold: consistency, conditional exchangeability (unconfoundedness), and positivity. Under these assumptions, the conditional average treatment effect (CATE) is identifiable from the observed data.

Suppose that

$$(X_i, T_i, Y_i(0), Y_i(1)), \quad i = 1, \ldots, n,$$

are independently generated observations, where $X_i \in \mathbb{R}^p$ denotes the covariate vector, $T_i \in \{0, 1\}$ is the treatment indicator, and $Y_i(0)$ and $Y_i(1)$ represent potential outcomes under control and treatment, respectively.

2.1.1. Covariate Generation

To introduce realistic dependence among high-dimensional covariates, we generate predictors using a latent factor model,

$$X_i = AZ_i + \epsilon_i,$$

where $Z_i \in \mathbb{R}^k$ represents latent factors, $A \in \mathbb{R}^{p \times k}$ is a loading matrix, and

$$\epsilon_i \sim N(0, \sigma^2 I_p)$$

is independent Gaussian noise.

The latent factor representation creates correlated predictors because multiple observed variables share common underlying factors. This structure reflects settings such as biomedical and socioeconomic studies, where measured covariates are often driven by unobserved common mechanisms.

The strength of dependence is controlled through the loading matrix A and the noise variance σ^2. By varying these quantities, different levels of covariance dependence can be generated. This allows investigation of the robustness of competing CATE estimators under weak and strong correlation structures.

Before model fitting, continuous covariates are standardized using parameters estimated from the training sample only. The same transformation is then applied to the test sample to avoid information leakage.

2.1.2. Nonlinear Treatment Assignment Mechanism

Treatment assignment is generated according to a nonlinear propensity score model,

$$e(X_i) = P(T_i = 1 | X_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)},$$

where

$$\eta_i = 0.5X_{i1} + 0.5X_{i2} + 0.3X_{i3} + 0.2X_{i1}X_{i2}.$$

The treatment indicator is sampled from

$$T_i \sim \text{Bernoulli}(e(X_i)).$$

This mechanism introduces confounding because treatment assignment depends on observed covariates. The nonlinear interaction term creates a more challenging assignment mechanism than a purely linear propensity model. The coefficients are selected so that estimated propensity scores remain bounded away from zero and one, satisfying the positivity assumption.

2.1.3. Outcome Model

The baseline outcome under control is generated according to

$$\mu(X_i) = X_{i1} + 0.5X_{i2} + 0.5\sin(X_{i3}) + 0.2X_{i4}^2.$$

The heterogeneous treatment effect is defined as

$$\tau(X_i) = 1 + 0.5X_{i1}X_{i3} + 0.5X_{i5}.$$

This specification introduces nonlinear and interaction-based treatment effect heterogeneity, creating a setting where flexible models are required to recover the underlying CATE function.

The potential outcomes are generated as

$$Y_i(0) = \mu(X_i) + \varepsilon_i,$$

and

$$Y_i(1) = \mu(X_i) + \tau(X_i) + \varepsilon_i,$$

where

$$\varepsilon_i \sim N(0,1)$$

is independent noise.

The observed outcome follows the standard potential outcome formulation,

$$Y_i = Y_i(0) + T_i\{Y_i(1) - Y_i(0)\}.$$

Because the true treatment effect function is known,

$$\tau(X_i) = E[Y_i(1) - Y_i(0)|X_i],$$

the simulation allows direct evaluation of CATE estimation accuracy through comparison between the estimated and true treatment effects.

2.1.4. Motivation for the Simulation Design

The simulation design combines several challenging characteristics of high-dimensional observational studies.

First, correlated covariates generated from a latent factor model allow us to investigate whether dependence-aware feature representations provide benefits when predictors contain complex correlation patterns.

Second, nonlinear treatment assignment creates realistic confounding while maintaining causal identifiability.

Third, nonlinear and interaction-based treatment effect functions generate heterogeneous responses across individuals, providing a challenging benchmark for evaluating individualized treatment effect estimation.

Importantly, the simulation study is not intended to demonstrate that copula transformations create additional dependence information. Rather, it evaluates whether a copula-based marginal representation can improve numerical stability and learning performance when covariates have heterogeneous marginal distributions and complex dependence structures.

2.2. Manifold Feature Augmentation

High-dimensional observational data often contain complex nonlinear structures arising from latent biological, behavioral, or socioeconomic mechanisms. Although the observed covariates are represented in a high-dimensional ambient space, the effective dimension of the underlying data-generating process may be substantially lower. Consequently, low-dimensional representations that preserve important geometric relationships among observations may provide useful additional information for heterogeneous treatment effect estimation.

In the proposed framework, manifold learning is used as a feature augmentation procedure rather than as a replacement for the original covariates. The objective is to construct additional representations that summarize nonlinear relationships among observations, which may assist the neural network in learning complex treatment assignment mechanisms and outcome heterogeneity.

Let

$$X_i \in \mathbb{R}^p$$

denote the original covariate vector. A manifold learning algorithm defines a mapping

$$\phi : \mathbb{R}^p \to \mathbb{R}^d,$$

where $d \ll p$ represents the dimension of the learned embedding. The resulting manifold coordinates are denoted by

$$Z_i^M = \phi(X_i).$$

The augmented representation is then constructed as

$$X_i^A = [X_i, Z_i^M],$$

where the original covariates are retained together with the learned manifold features.

2.2.1. Manifold Learning Methods

Several nonlinear dimension-reduction methods can be used to obtain the manifold representation. In this study, we consider three commonly used approaches:

- **Uniform Manifold Approximation and Projection (UMAP):** UMAP constructs a low-dimensional representation by preserving local neighborhood relationships while maintaining aspects of global structure.
- **t-distributed Stochastic Neighbor Embedding (t-SNE):** t-SNE emphasizes preservation of local similarity and is particularly effective for identifying cluster-like structures.
- **Isometric Mapping (Isomap):** Isomap preserves approximate geodesic distances and captures global nonlinear geometry.

These methods are not assumed to provide complementary information in all settings. Instead, they represent alternative strategies for extracting low-dimensional geometric features from high-dimensional covariates. Their relative contribution is evaluated through ablation experiments comparing the original covariates, individual manifold representations, and combined feature sets.

The manifold representation used in the proposed framework is therefore interpreted as an auxiliary feature construction step. The main contribution of the proposed method does not rely on introducing a new manifold learning algorithm, but rather on integrating dependence-aware preprocessing, interaction features, and neural outcome modeling within a unified causal learning framework.

2.2.2. Training and Test Data Embedding

To prevent information leakage, all preprocessing operations are performed using the training data only. Specifically, the manifold transformation is estimated from the training sample,

$$\hat{\phi}_{train},$$

and the resulting mapping is applied to the test observations whenever an out-of-sample transformation is available.

For UMAP and Isomap, the fitted embedding transformation can be used to obtain representations for new observations. Therefore, the test manifold coordinates are calculated as

$$Z^{M}_{i,test} = \hat{\phi}_{train}(X_{i,test}).$$

Because standard t-SNE is primarily designed for visualization and does not provide a natural out-of-sample mapping, it requires special consideration. Accordingly, t-SNE representations are used only when an appropriate out-of-sample extension is available; otherwise, t-SNE is excluded from the predictive pipeline and retained only as an exploratory visualization tool.

This procedure ensures that no information from the test sample is used during feature construction, model training, or hyperparameter selection.

2.2.3. Integration into the Proposed Framework

The purpose of manifold feature augmentation is to provide the neural network with additional nonlinear summaries of the covariate structure. The original covariates preserve individual-level information, while manifold coordinates provide compressed representations of latent geometric relationships.

The augmented features are subsequently processed through the empirical copula transformation and treatment-interaction construction steps. Therefore, the manifold component should be viewed as a representation-learning module that may improve prediction in settings with nonlinear covariate structures, rather than as an independent causal estimation method.

2.3. Empirical Copula Transformation

High-dimensional observational data frequently contain covariates with heterogeneous marginal distributions, including different scales, skewed distributions, and heavy-tailed behavior. Such characteristics may affect the optimization and numerical stability of machine learning models, particularly when flexible neural networks are used to estimate heterogeneous treatment effects.

To obtain a common marginal representation while retaining rank-based dependence information, we apply a nonparametric empirical copula transformation. The purpose of this transformation is not to create additional dependence information, but rather to separate marginal behavior from the dependence structure and provide a more homogeneous input representation for subsequent neural network learning.

2.3.1. Empirical Copula Mapping

Let X_j denote the jth covariate. The empirical marginal distribution function is defined as

$$\widehat{F}_j(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_{ij} \leq x).$$

The corresponding empirical copula pseudo-observation is obtained as

$$\widetilde{X}_{ij} = \frac{\text{rank}(X_{ij})}{n+1}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, p.$$

The transformed variable satisfies

$$0 < \widetilde{X}_{ij} < 1,$$

and represents the relative position of an observation within the empirical marginal distribution of each covariate.

2.3.2. Connection with Copula Theory

According to Sklar's theorem [9,10], a multivariate distribution function can be represented as

$$F(x_1, \ldots, x_p) = C\{F_1(x_1), \ldots, F_p(x_p)\},$$

where F_j denotes the marginal distribution of the jth variable and C denotes the copula function describing the dependence structure among variables.

The empirical copula transformation replaces the unknown marginal distributions with their empirical counterparts,

$$\widehat{F}_j(X_{ij}),$$

which are estimated through normalized ranks. Therefore, the transformed variables have approximately uniform marginal distributions while maintaining the ordering relationships among observations.

It is important to emphasize that the rank transformation itself does not introduce new dependence information. Instead, it provides a scale-invariant representation in which the effects of marginal distributions are reduced, allowing subsequent learning algorithms to focus on nonlinear associations between variables.

2.3.3. Statistical Properties

The empirical copula transformation has several useful properties.

First, it is invariant to strictly monotone transformations of individual covariates. Therefore, variables measured on different scales or with different marginal distributions can be represented in a comparable manner.

Second, each transformed marginal approximately follows a uniform distribution on $(0, 1)$, avoiding the need to specify parametric marginal distributions.

Third, because the transformation is rank-based, it is less sensitive to extreme observations than methods relying directly on raw numerical values.

Finally, the transformation preserves rank-based dependence measures, including Spearman's rank correlation and Kendall's tau. Thus, important ordering relationships among covariates remain available to the downstream learning algorithm.

2.3.4. Training and Test Transformation

To avoid information leakage, empirical marginal distributions are estimated using the training sample only. Specifically, for each covariate X_j, we estimate

$$\widehat{F}_{j,train},$$

using the training observations. The transformation of a test observation is then obtained by applying the training empirical distribution,

$$\widetilde{X}_{ij,test} = \widehat{F}_{j,train}(X_{ij,test}).$$

Consequently, no information from the test sample is used in constructing the copula-based features.

2.3.5. Integration into the Proposed Framework

Within the proposed framework, the empirical copula transformation serves as a nonparametric marginal normalization step before neural network modeling. It allows covariates and manifold features with heterogeneous distributions to be represented on a common scale while retaining rank-based dependence patterns.

The transformed features are subsequently combined with treatment indicators and treatment–covariate interaction terms. The neural network then learns the relationship between these representations and observed outcomes.

Therefore, the potential advantage of the copula-based representation is attributed to improved robustness and numerical conditioning under complex marginal distributions, rather than the creation of additional dependence information.

2.4. Treatment–Covariate Interaction Feature Construction

The conditional average treatment effect is defined as

$$\tau(X) = E\{Y(1) - Y(0) \mid X\},$$

which characterizes how treatment effects vary across individuals with different covariate profiles. A central challenge in CATE estimation is that treatment effect heterogeneity arises through interactions between treatment assignment and subject characteristics.

In conventional regression models, treatment heterogeneity is commonly represented through treatment–covariate interaction terms. Following this principle, we explicitly augment the model input with treatment interaction features to facilitate learning heterogeneous treatment effects using a neural network.

2.4.1. Interaction Feature Representation

Let

$$X_i^C$$

denote the covariate representation after preprocessing, including the original covariates, manifold features, and empirical copula-transformed variables. For each observation, we construct the augmented representation

$$\mathcal{X}_i = [X_i^C, T_i, X_i^C \odot T_i],$$

where $\odot$ denotes element-wise multiplication.

The resulting representation contains three components:

- The covariate information X_i^C;
- The treatment indicator T_i;
- The treatment–covariate interaction features $X_i^C \odot T_i$.

For a covariate representation of dimension p^*, the augmented input has dimension

$$2p^* + 1.$$

This construction provides the neural network with explicit access to treatment effect modifiers rather than requiring all interaction structures to be learned implicitly through hidden layers.

2.4.2. Relationship to the S-Learner Framework

The proposed model follows the S-learner paradigm, in which a single outcome model is fitted using treatment as an additional input variable,

$$E(Y \mid X, T).$$

A conventional S-learner can theoretically learn treatment heterogeneity through nonlinear interactions within the neural network. However, when treatment effects depend

on complex interactions between many covariates, explicitly including treatment–covariate products provides an inductive bias that encourages the model to focus on effect modification.

Specifically, the proposed representation extends the standard S-learner input

$$[X, T]$$

to

$$[X, T, X \odot T],$$

thereby providing a direct representation of first-order treatment interactions.

The neural network subsequently learns nonlinear transformations of these features, allowing higher-order interactions to be captured through the hidden layers.

2.4.3. Counterfactual Outcome Estimation

After model training, counterfactual outcomes are estimated by evaluating the same fitted model under alternative treatment assignments.

For a fixed covariate vector X_i^C, define

$$\mathcal{X}_i^{(1)} = [X_i^C, 1, X_i^C],$$

and

$$\mathcal{X}_i^{(0)} = [X_i^C, 0, 0].$$

The estimated potential outcomes are

$$\widehat{Y}_i(1) = f_\theta(\mathcal{X}_i^{(1)}),$$

and

$$\widehat{Y}_i(0) = f_\theta(\mathcal{X}_i^{(0)}).$$

The estimated CATE is therefore

$$\widehat{\tau}(X_i) = \widehat{Y}_i(1) - \widehat{Y}_i(0).$$

This procedure follows the standard S-learner counterfactual prediction strategy while using an enriched representation designed to improve learning of heterogeneous treatment effects.

2.4.4. Treatment–Covariate Interaction in the Proposed Framework

The treatment–covariate interaction representation integrates treatment information with the transformed feature space created by the previous preprocessing stages. Specifically:

- Manifold augmentation provides nonlinear geometric summaries;
- Empirical copula transformation provides marginally normalized rank-based representations;
- Interaction construction explicitly introduces treatment effect modifiers.

The final neural network combines these components to estimate the conditional outcome function and generate counterfactual predictions.

Importantly, the contribution of this component is not a new tensor operation or tensor decomposition. Rather, it is an explicit interaction-based feature representation within a deep S-learner framework, designed to improve the learning of heterogeneous treatment effects.

2.5. Neural Network Outcome Model

The final component of the proposed framework is a neural network outcome model that estimates the conditional response function

$$E(Y \mid X, T).$$

The neural network is implemented as a flexible S-learner, where treatment assignment is included as an input variable and a single predictive model is trained using both treated and control observations.

Let

$$\mathcal{X}_i$$

denote the treatment–covariate interaction representation constructed in the previous subsection. The neural network estimates

$$\widehat{Y}_i = f_\theta(\mathcal{X}_i),$$

where f_θ represents a multilayer perceptron parameterized by weights and biases θ.

The role of the neural network is not to introduce a new deep learning architecture, but rather to provide a flexible nonlinear function approximator for integrating the copula-based features, manifold representations, and treatment interaction variables.

2.5.1. Network Architecture

The neural network consists of multiple fully connected hidden layers with rectified linear unit (ReLU) activation functions.

The forward propagation through the network is defined as

$$h_1 = \mathrm{ReLU}(W_1 \mathcal{X}_i + b_1),$$

$$h_2 = \mathrm{ReLU}(W_2 h_1 + b_2),$$

$$h_3 = \mathrm{ReLU}(W_3 h_2 + b_3),$$

and the final outcome prediction is

$$\widehat{Y}_i = W_4 h_3 + b_4.$$

Here, W_l and b_l represent the weight matrix and bias vector of layer l, respectively.

The network architecture is intentionally kept identical between the proposed copula-based model and the baseline neural network model. Consequently, any performance differences can be attributed to the input representation rather than differences in model capacity.

2.5.2. Training Objective

The network parameters are estimated by minimizing the empirical squared loss,

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - f_\theta(\mathcal{X}_i)]^2.$$

This objective corresponds to maximum likelihood estimation under the working Gaussian outcome model,

$$Y_i = f_\theta(\mathcal{X}_i) + \varepsilon_i,$$

where

$$\varepsilon_i \sim N(0, \sigma^2).$$

Optimization is performed using the Adam algorithm with mini-batch stochastic gradient descent. Dropout regularization is applied during training to reduce overfitting in high-dimensional settings.

2.5.3. Counterfactual Prediction

After training, the fitted outcome model is evaluated under both possible treatment assignments.

For each individual with covariate representation X_i^C,

$$\mathcal{X}_i^{(1)} = [X_i^C, 1, X_i^C],$$

and

$$\mathcal{X}_i^{(0)} = [X_i^C, 0, 0].$$

The predicted potential outcomes are

$$\widehat{Y}_i(1) = f_\theta(\mathcal{X}_i^{(1)}),$$

and

$$\widehat{Y}_i(0) = f_\theta(\mathcal{X}_i^{(0)}).$$

The estimated treatment effect is obtained by

$$\widehat{\tau}(X_i) = \widehat{Y}_i(1) - \widehat{Y}_i(0).$$

This procedure follows the standard S-learner framework. The proposed method differs from a conventional S-learner only through the enriched feature representation used as network input.

2.5.4. Computational Complexity

The computational cost of the proposed framework consists of three main components: feature preprocessing, interaction construction, and neural network optimization.

For empirical copula transformation, computing marginal ranks requires

$$O(np \log n)$$

operations, where n is the sample size and p is the number of variables.

The treatment–covariate interaction construction requires linear computation,

$$O(np).$$

For a neural network with P trainable parameters, E training epochs, and mini-batch optimization, the approximate computational complexity is

$$O(EnP).$$

Therefore, the dominant computational cost generally arises from neural network training rather than copula transformation.

When manifold learning is included, its computational burden depends on the specific algorithm and embedding dimension. In particular, nonlinear manifold methods may be-

come expensive for very large datasets. Thus, manifold augmentation should be considered as a trade-off between additional representation capacity and computational scalability.

2.5.5. Role in the Proposed Framework

The neural network serves as the final integration module of the proposed framework. The complete representation-learning process consists of the following:

1. Manifold augmentation to obtain nonlinear geometric features;
2. Empirical copula transformation to provide marginally normalized rank-based representations;
3. Treatment–covariate interaction construction to explicitly represent effect modifiers;
4. Neural network outcome modeling to estimate observed and counterfactual responses.

Thus, the methodological contribution lies in the integration of these components for high-dimensional CATE estimation rather than in introducing a new neural network architecture.

2.6. Conditional Average Treatment Effect Estimation

The primary objective of the proposed framework is to estimate the conditional average treatment effect (CATE),

$$\tau(X) = E\{Y(1) - Y(0) \mid X\},$$

where $Y(1)$ and $Y(0)$ denote the potential outcomes under treatment and control, respectively [1]. The CATE characterizes how the expected treatment effect varies across individuals with different covariate profiles.

Unlike the average treatment effect (ATE), which summarizes the treatment effect over the entire population, the CATE provides a conditional description of treatment heterogeneity. Such information can support subgroup analysis and treatment prioritization when combined with additional clinical or policy considerations.

2.6.1. Identification Assumptions

Identification of the CATE relies on the standard assumptions of causal inference:

1. **Consistency:** The observed outcome corresponds to the potential outcome associated with the received treatment,
$$Y = Y(T).$$
2. **Conditional exchangeability:**
$$(Y(1), Y(0)) \perp T \mid X.$$
This assumption implies that all variables affecting both treatment assignment and outcomes are sufficiently captured by the observed covariates.
3. **Positivity:**
$$0 < P(T = 1 \mid X) < 1.$$
This ensures that individuals with similar covariate characteristics have a positive probability of receiving either treatment.

Under these assumptions,

$$\tau(X) = E(Y \mid X, T = 1) - E(Y \mid X, T = 0).$$

Therefore, estimation of the CATE can be formulated as estimating the conditional outcome function under alternative treatment assignments.

2.6.2. Outcome Regression and Counterfactual Prediction

The proposed framework estimates the conditional outcome function

$$\widehat{f}_\theta(X, T) \approx E(Y \mid X, T)$$

using the neural network S-learner introduced previously.

For an individual i, predicted potential outcomes are obtained by evaluating the fitted model under both treatment conditions:

$$\widehat{Y}_i(1) = f_\theta(\mathcal{X}_i^{(1)}),$$

and

$$\widehat{Y}_i(0) = f_\theta(\mathcal{X}_i^{(0)}).$$

The estimated CATE is then

$$\widehat{\tau}(X_i) = \widehat{Y}_i(1) - \widehat{Y}_i(0).$$

Because both potential outcomes cannot be simultaneously observed for the same individual, the quality of CATE estimation depends on the validity of the identification assumptions, the informativeness of the observed covariates, and the predictive accuracy of the outcome model.

2.6.3. Relation to the S-Learner Framework

The proposed approach belongs to the class of S-learners, where treatment assignment is included as an input variable in a single outcome prediction model. The advantage of this strategy is that treated and control observations contribute jointly to learning the conditional outcome surface.

The proposed framework extends the standard S-learner by enriching the input representation with:

- Manifold-based features capturing nonlinear covariate geometry;
- Empirical copula features providing marginally normalized rank-based representations;
- Explicit treatment–covariate interaction features.

The neural network then learns a flexible approximation to the conditional outcome function using this enriched representation.

2.6.4. Interpretation of Estimated Treatment Effects

The estimated quantity

$$\widehat{\tau}(X_i)$$

represents the predicted difference between the expected outcomes under treatment and control for individuals with similar observed characteristics.

However, estimated CATE values should not automatically be interpreted as guaranteed optimal treatment recommendations. Practical decisions may require additional considerations, including treatment costs, safety constraints, uncertainty quantification, competing outcomes, and domain-specific knowledge.

Therefore, in this study, CATE estimation is evaluated as a statistical learning problem, while treatment assignment policies are assessed separately through policy evaluation metrics.

2.6.5. Relation to Policy Evaluation

The primary target of the proposed methodology is estimation of heterogeneous treatment effects. In addition to CATE accuracy, we evaluate whether the estimated effects can support effective treatment prioritization using policy evaluation measures.

Given an estimated CATE function, a treatment policy may be defined as

$$\pi(X) = I\{\widehat{\tau}(X) > 0\}.$$

This policy assigns treatment to individuals predicted to have positive treatment effects.

The quality of this policy is evaluated separately using inverse propensity scoring (IPS) and Doubly Robust (DR) estimators. These policy metrics evaluate the consequences of treatment assignment decisions and are not used as components of the proposed CATE estimation procedure.

2.6.6. Sources of Estimation Uncertainty

CATE estimation involves several sources of uncertainty, including finite sample variation, model approximation error, and limited overlap between treated and untreated populations.

In particular, when the positivity assumption is weak, counterfactual predictions may be unreliable because the model must extrapolate beyond the observed treatment distribution.

Consequently, the proposed framework should be interpreted as a flexible representation-learning approach for CATE estimation under identifiable observational settings, rather than a universally optimal decision rule.

2.7. Evaluation Metrics

The proposed framework is evaluated from two complementary perspectives:

1. Accuracy of CATE estimation when the true treatment effect function is known (simulation studies);
2. Effectiveness of treatment prioritization when individual treatment effects are not directly observable (real-world applications).

Because individual causal effects cannot be observed simultaneously under both treatment conditions, direct evaluation of CATE accuracy is only possible in simulation settings where the data-generating mechanism is known. For observational datasets such as the Criteo uplift dataset, evaluation relies on policy-based measures rather than comparisons with unknown true CATE values.

2.7.1. CATE Estimation Accuracy in Simulation Studies

In simulation experiments, the true conditional average treatment effect

$$\tau(X_i) = E\{Y_i(1) - Y_i(0) \mid X_i\}$$

is available by construction. Therefore, the estimated treatment effect

$$\widehat{\tau}(X_i)$$

can be directly compared with the known target.

2.7.2. Bias

Bias measures systematic deviation of the estimated CATE from the true CATE:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}[\widehat{\tau}(X_i) - \tau(X_i)].$$

A value close to zero indicates that the estimator does not systematically overestimate or underestimate treatment effects.

2.7.3. Root Mean Squared Error

The root mean squared error (RMSE) measures overall estimation accuracy:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\widehat{\tau}(X_i) - \tau(X_i)]^2}.$$

RMSE reflects both systematic error and variability in recovering treatment effect heterogeneity. Smaller values indicate more accurate estimation of the true CATE function.

2.7.4. Distributional Assessment Across Replications

Because simulation results are obtained through repeated Monte Carlo experiments, we report both average performance and variability across replications.

For a metric M, the Monte Carlo average is calculated as

$$\overline{M} = \frac{1}{R}\sum_{r=1}^{R} M^{(r)},$$

where R denotes the number of simulation replications.

In addition to mean values, dispersion measures such as interquartile ranges are reported to evaluate estimator stability.

2.7.5. Evaluation in Real Observational Data

For real observational datasets, the true individual treatment effects are not available. Consequently, Bias and RMSE relative to a true CATE function cannot be computed.

If pseudo-outcomes are constructed using inverse propensity weighting or other methods, the resulting quantities measure prediction error relative to an estimated surrogate target rather than the true causal effect. Such metrics should therefore be interpreted cautiously and are not referred to as ordinary CATE bias or RMSE.

Instead, real-data experiments focus on treatment policy evaluation using Inverse Propensity Scoring and Doubly Robust estimators, which assess whether the estimated treatment effects lead to improved treatment prioritization.

2.8. Policy Evaluation

Although accurate CATE estimation is the primary objective of the proposed framework, practical applications often require translating estimated treatment effects into treatment assignment policies. A treatment policy specifies which individuals should receive treatment based on their observed characteristics.

Therefore, in addition to evaluating CATE estimation accuracy in simulation studies, we assess whether the estimated treatment effects lead to effective treatment prioritization through policy evaluation metrics.

Importantly, policy evaluation addresses a different question from CATE estimation. CATE estimation evaluates whether the conditional treatment effect function is accurately recovered, whereas policy evaluation evaluates the expected outcome resulting from applying a decision rule based on the estimated effects.

2.8.1. Treatment Policy

Given an estimated CATE function

$$\widehat{\tau}(X),$$

we define a treatment policy as

$$\pi(X) = I\{\hat{\tau}(X) > 0\}.$$

Under this policy, individuals predicted to have positive treatment effects are assigned treatment, while individuals predicted to have non-positive effects remain untreated.

This rule represents a simple treatment prioritization strategy. More complex policies incorporating treatment cost, risk constraints, or resource limitations can be considered in future extensions.

2.8.2. Propensity Score Estimation

Both Inverse Propensity Scoring (IPS) and Doubly Robust (DR) estimators require an estimate of the propensity score,

$$e(X) = P(T = 1 \mid X).$$

The propensity score model estimates the probability of receiving treatment conditional on observed covariates.

In the simulation study, propensity scores are estimated using logistic regression:

$$\hat{e}(X) = P(T = 1 \mid X; \hat{\gamma}),$$

where $\hat{\gamma}$ denotes estimated regression coefficients.

The propensity score model is used only for policy evaluation and is not included in the training procedure of the proposed CATE estimator. This separation ensures that evaluation of treatment policies is independent of the representation-learning process used by the proposed framework.

2.8.3. Inverse Propensity Score Estimator

The inverse propensity score (IPS) estimator evaluates the expected outcome under policy π by reweighting observed outcomes according to the probability of receiving the observed treatment:

$$\hat{V}_{\mathrm{IPS}}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \frac{I\{T_i = \pi(X_i)\} Y_i}{T_i \hat{e}(X_i) + (1 - T_i)(1 - \hat{e}(X_i))}.$$

IPS provides an unbiased estimate when the propensity score model is correctly specified. However, the estimator can exhibit high variance when estimated propensity scores approach zero or one.

2.8.4. Doubly Robust Estimator

To improve stability, we also consider the Doubly Robust (DR) estimator, which combines outcome regression and propensity score weighting.

The DR policy value estimator is

$$\hat{V}_{\mathrm{DR}}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\mu}_{\pi(X_i)}(X_i) + \frac{I\{T_i = \pi(X_i)\}}{T_i \hat{e}(X_i) + (1 - T_i)(1 - \hat{e}(X_i))} \{ Y_i - \hat{\mu}_{T_i}(X_i) \} \right],$$

where

$$\hat{\mu}_t(X) = \hat{E}(Y \mid X, T = t)$$

is obtained from the fitted neural network outcome model.

The DR estimator remains consistent if either the propensity score model or the outcome regression model is correctly specified. This property often leads to greater robustness than IPS alone, particularly in observational settings where treatment assignment mechanisms may be complex.

2.8.5. Practical Interpretation

Policy evaluation metrics provide a measure of whether estimated treatment effects translate into useful decision rules.

For example, in personalized treatment settings, a model with improved DR policy value suggests that its estimated treatment prioritization rule may lead to better expected outcomes compared with alternative strategies.

However, higher policy value does not necessarily imply uniformly accurate individual treatment effect estimation. A model may achieve good policy performance by correctly ranking individuals according to treatment benefit even if the numerical CATE estimates are imperfect.

Therefore, IPS and DR are interpreted as complementary evaluation criteria rather than replacements for direct CATE accuracy assessment in simulation studies.

2.9. Monte Carlo Experimental Design

We conduct Monte Carlo simulation experiments to evaluate the finite-sample performance of the proposed copula-augmented neural network framework under controlled causal settings.

The simulation study is designed to address the following three main questions:

1. Does the proposed copula-based representation improve CATE estimation compared with the same neural network without copula transformation?
2. How does the proposed framework compare with established causal machine learning approaches?
3. Under what data characteristics, including dimensionality and dependence structure, does the proposed representation provide the greatest benefit?

Each simulation replication generates an independent dataset according to the data-generating mechanism described in Section 2.1. Because the true CATE function is known, direct evaluation of treatment effect estimation accuracy is possible.

2.9.1. Simulation Replications

For each experimental setting, we perform R independent Monte Carlo replications.

For replication r,

$$\{(X_i, T_i, Y_i)\}_{i=1}^{n}$$

is generated independently. Each dataset is randomly divided into training and test samples.

The training data are used exclusively for the following:

- Estimating preprocessing transformations;
- Fitting prediction models;
- Selecting model parameters.

The test data are used only for final evaluation of CATE estimation accuracy and policy performance.

2.9.2. Data Splitting

For each replication, observations are randomly partitioned into training and test sets using a fixed proportion.

All preprocessing procedures are fitted using training observations only:

- Standardization parameters;
- Empirical distribution functions for copula transformation;
- Manifold representations.

The learned transformations are then applied to the independent test sample.

This procedure prevents information leakage and provides an unbiased assessment of out-of-sample performance.

2.9.3. Compared Methods

To evaluate the contribution of the proposed representation, we compare the following approaches.

- **Proposed Copula-Augmented Neural S-Learner:**
  The proposed framework uses manifold augmentation, empirical copula transformation, treatment interaction features, and neural network outcome modeling.
- **Non-Copula Neural S-Learner:**
  The same neural network architecture and interaction representation without the empirical copula transformation.
- **Random Forest-Based CATE Estimator:**
  A causal forest approach that estimates heterogeneous treatment effects using tree-based partitioning.
- **X-Learner:**
  A meta-learning approach that estimates treatment effects through separate outcome models and treatment effect modeling.
- **Bayesian Additive Regression Trees (BART):**
  A Bayesian nonparametric method capable of capturing nonlinear outcome relationships and treatment effect heterogeneity.

All competing methods are evaluated using identical training and test samples within each replication.

2.9.4. Ablation Experiments

To identify the contribution of individual components of the proposed framework, we perform ablation experiments.

The following representations are compared:

- Raw covariates;
- Standardized covariates;
- Empirical rank-uniform transformed covariates;
- Copula-transformed covariates with treatment interactions;
- Copula transformation combined with manifold augmentation.

These comparisons distinguish the contribution of marginal normalization, manifold representation, and explicit treatment interaction features.

2.9.5. Simulation Scenarios

To evaluate robustness under different data conditions, we consider multiple simulation scenarios by varying:

- **Sample size:**

$$n \in \{1000, 3000, 5000\}$$

to examine finite-sample behavior.

- **Covariate dimension:**
$$p \in \{20, 50, 100\}$$
to evaluate performance in increasingly high-dimensional settings.
- **Dependence strength:**
Different values of the latent factor noise variance are used to generate weak, moderate, and strong correlations among covariates.
- **Marginal distributions:**
Covariates are generated under Gaussian, skewed, and heavy-tailed settings to evaluate robustness of marginal normalization.

These scenarios allow assessment of whether the proposed representation is particularly beneficial when covariates exhibit heterogeneous distributions and complex dependence structures.

2.9.6. Estimation Procedure

For each replication, the following procedure is performed:

1. Generate observational data according to the specified data-generating mechanism.
2. Split observations into training and test samples.
3. Estimate preprocessing transformations using training data only.
4. Construct feature representations for each competing method.
5. Fit the corresponding CATE estimator.
6. Predict counterfactual outcomes for test observations.
7. Compute Bias, RMSE, and policy evaluation metrics.

2.9.7. Summary of Simulation Results

For each performance metric M, the Monte Carlo average is calculated as

$$\overline{M} = \frac{1}{R} \sum_{r=1}^{R} M^{(r)}.$$

We report both average performance and variability across replications. In addition to mean values, confidence intervals or interquartile ranges are provided to characterize estimator stability.

Statistical comparisons between methods are conducted using paired comparisons across identical simulation replications. This design reduces variability due to random data generation and provides a more reliable assessment of relative performance.

2.10. Implementation Details

All computational experiments are conducted using the `R` programming environment. The proposed framework combines statistical preprocessing, manifold representation learning, and deep neural network modeling.

Data preprocessing and manipulation are performed using `data.table`, `dplyr`, and related statistical computing packages. Neural network models are implemented using the `keras` interface with TensorFlow as the computational backend. Manifold learning representations are obtained using appropriate implementations of UMAP, Isomap, and related nonlinear embedding algorithms. Causal benchmark methods are implemented using established causal machine learning packages, including implementations of causal forests and meta-learning approaches.

All compared methods are evaluated under identical simulation replications, training/test splits, and computational environments to ensure a fair comparison.

2.10.1. Preprocessing Procedure

To prevent information leakage, all preprocessing operations are estimated using the training sample only.

The preprocessing pipeline consists of the following steps:

1. Continuous covariates are standardized using the training sample mean and standard deviation:
$$X_{ij}^{*} = \frac{X_{ij} - \bar{X}_{j,train}}{s_{j,train}}.$$
The same transformation is applied to test observations.
2. Manifold representations are estimated using the standardized training covariates:
$$\widehat{\phi}_{train}(X_{train}),$$
and applied to test observations using the corresponding out-of-sample mapping when available.
3. For the copula-based model, empirical marginal distributions are estimated from the training data:
$$\widehat{F}_{j,train}.$$
The copula transformation for new observations is then obtained as
$$\widetilde{X}_{ij} = \widehat{F}_{j,train}(X_{ij}).$$
4. Treatment–covariate interaction features are constructed after the previous transformations.

This procedure ensures that neither test outcomes nor test covariate distributions are used during model training.

2.10.2. Neural Network Configuration

The proposed neural network is implemented as a fully connected multilayer perceptron representing an S-learner.

The network architecture consists of three hidden layers with
$$256, \quad 128, \quad 64$$
neurons, respectively.

Each hidden layer uses a rectified linear unit (ReLU) activation function, followed by dropout regularization with probability 0.3.

The final output layer produces the predicted conditional outcome
$$\widehat{Y} = f_{\theta}(X, T).$$

The copula-based model and the non-copula baseline use exactly the same neural network architecture, optimization procedure, and training settings. The only difference is the feature representation supplied to the network.

2.10.3. Training Procedure

The model parameters are estimated by minimizing the mean squared error loss,

$$\mathcal{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - f_\theta(\mathcal{X}_i))^2.$$

Optimization is performed using the Adam optimizer with a learning rate

$$0.001.$$

Training uses mini-batch stochastic gradient descent with batch size 64.

The number of epochs is fixed across competing models to ensure a consistent comparison. Early stopping may be incorporated based on validation loss to avoid unnecessary computation and overfitting.

All random seeds are fixed for data generation, parameter initialization, and training procedures to improve computational reproducibility.

2.10.4. Hyperparameter Selection

The network architecture and training parameters are selected to balance flexibility and computational efficiency in repeated Monte Carlo experiments.

The objective of the simulation study is not to optimize a particular neural network architecture, but rather to evaluate the contribution of the proposed representation. Therefore, identical hyperparameter settings are applied to all neural network-based methods.

A more extensive hyperparameter optimization study may further improve predictive performance but is outside the primary scope of this work.

2.10.5. Computational Complexity and Scalability

The computational cost of the proposed framework depends on sample size, covariate dimension, manifold dimension, and neural network complexity.

The major components are:

- **Standardization:**
  $$O(np)$$
  where n is the sample size, and p is the number of covariates.
- **Empirical copula transformation:**
  Ranking each covariate requires approximately
  $$O(np\log n)$$
  operations.
- **Interaction feature construction:**
  Treatment–covariate interactions require
  $$O(np).$$
- **Neural network training:**
  For P trainable parameters, E epochs, and n observations,
  $$O(EnP).$$

Among these components, neural network optimization generally dominates the computational cost. For very high-dimensional settings, dimensionality reduction or feature screening may be considered before applying the proposed framework.

The manifold-learning component introduces additional computational burden that depends on the selected algorithm. Therefore, in large-scale applications, careful selection of the embedding method and dimension is important to balance representation quality and computational feasibility.

2.10.6. Computational Reproducibility

For reproducibility, all simulation experiments use fixed random seeds and follow identical procedures across replications.

The complete workflow consists of the following:

1. Data generation;
2. Training/test splitting;
3. Training-only preprocessing;
4. Feature construction;
5. Model training;
6. Counterfactual prediction;
7. Evaluation of CATE accuracy and policy value.

The reported simulation results are obtained by aggregating performance metrics across independent Monte Carlo replications.

3. Simulation Data Analysis

We conduct a comprehensive Monte Carlo simulation study to evaluate the performance of the proposed Copula-Tensor Neural Network estimator for conditional average treatment effect (CATE) estimation under high-dimensional observational settings. The simulation framework is constructed such that the true treatment effect function is known, allowing direct assessment of CATE estimation accuracy.

The simulation experiments are designed to examine several challenges commonly encountered in modern causal machine learning: (i) high-dimensional correlated covariates, (ii) nonlinear treatment assignment and confounding, (iii) nonlinear heterogeneous treatment effects, and (iv) scalability with increasing feature dimensions. In addition to evaluating the contribution of the copula representation, we compare the proposed approach with several established CATE estimators, including neural-network S-learning without copula transformation, Causal Forests, X-Learners, and Bayesian Additive Regression Trees (BART).

All preprocessing procedures, including feature standardization, empirical copula transformation, and feature augmentation, are estimated using the training data only and subsequently applied to the test data. This procedure avoids information leakage between training and evaluation samples.

3.1. Scenario 1: Latent Factor Dependence with Nonlinear Treatment Effects

3.1.1. Covariate Generation

We generate $n = 1000$ observations with $p = 30$ observed covariates. To introduce realistic dependence among high-dimensional predictors, we assume a latent factor structure:

$$Z_i \sim N(0, I_3), \tag{1}$$

where $Z_i \in \mathbb{R}^3$ represents three latent factors. The observed covariates are generated according to

$$X_i = Z_i W + \epsilon_i, \tag{2}$$

where $W \in \mathbb{R}^{3\times p}$ is a random loading matrix and

$$\epsilon_{ij} \sim N(0, 0.88^2) \tag{3}$$

is independent measurement noise.

Because multiple observed covariates share common latent factors, this construction produces strong dependence among predictors while maintaining a high-dimensional feature space. The number of latent factors remains fixed while the observed dimension increases, allowing evaluation of model performance under increasing feature complexity. Before model estimation, the generated covariates are standardized using training-sample statistics only.

3.1.2. Treatment Assignment

The binary treatment variable T_i is generated from a nonlinear propensity score model:

$$e(X_i) = P(T_i = 1|X_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}, \tag{4}$$

where

$$\eta_i = 0.5X_{i1} - 0.5X_{i2} + 0.3X_{i3} + 0.2X_{i1}X_{i2}. \tag{5}$$

The treatment indicator is sampled according to

$$T_i \sim \text{Bernoulli}(e(X_i)). \tag{6}$$

This treatment assignment mechanism introduces confounding because treatment selection depends on observed covariates and includes nonlinear interaction effects.

3.1.3. Treatment Effect Function

The true conditional average treatment effect is defined as

$$\tau(X_i) = 1 + 0.5\sin(X_{i1}) + 0.3X_{i2}^2 + 0.5X_{i3}. \tag{7}$$

This specification generates heterogeneous treatment effects through periodic, nonlinear polynomial, and linear components, providing a challenging environment for CATE estimation.

3.1.4. Outcome Generation

The potential outcomes are generated as

$$Y_i(0) = \mu(X_i) + \varepsilon_i, \tag{8}$$

$$Y_i(1) = \mu(X_i) + \tau(X_i) + \varepsilon_i, \tag{9}$$

where the baseline outcome function is

$$\mu(X_i) = 0.5X_{i1} + 0.5X_{i2} + 0.5\sin(X_{i3}), \tag{10}$$

and

$$\varepsilon_i \sim N(0,1) \tag{11}$$

is independent noise.

The observed outcome is obtained through the standard potential outcome framework:

$$Y_i = Y_i(0) + T_i\{Y_i(1) - Y_i(0)\}. \tag{12}$$

Under this construction,

$$E[Y_i(1) - Y_i(0)|X_i] = \tau(X_i), \tag{13}$$

which provides the ground truth for evaluating estimated CATE functions.

3.1.5. Simulation Characteristics

This scenario incorporates several challenges commonly encountered in observational causal inference:

- High-dimensional correlated covariates generated from a latent factor structure;
- Nonlinear treatment assignment producing covariate-dependent confounding;
- Nonlinear heterogeneous treatment effects;
- Continuous outcomes with stochastic noise.

3.1.6. Benchmark Methods

To evaluate the contribution of the proposed copula-based representation, we compare the proposed model with several widely used CATE estimation methods:

- **Neural Network S-Learner (NN-S):** A neural network estimator using the original covariates and treatment indicator without copula transformation.
- **Causal Forest (CF):** A nonparametric tree-based estimator designed specifically for heterogeneous treatment effect estimation.
- **X-Learner:** A meta-learning approach that estimates individual treatment effects using separate outcome models for treated and control groups.
- **Bayesian Additive Regression Trees (BART):** A Bayesian nonparametric regression framework commonly applied in causal inference problems.
- **Proposed Copula-Tensor Neural Network:** The proposed framework combining empirical copula representation, treatment–covariate interaction features, and neural network estimation.

3.1.7. Evaluation Metrics

CATE estimation accuracy is evaluated using the known true treatment effect function. For a test sample of size n_{test}, the estimation bias is defined as

$$\text{Bias} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (\hat{\tau}(X_i) - \tau(X_i)). \tag{14}$$

The root mean squared error (RMSE) is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (\hat{\tau}(X_i) - \tau(X_i))^2}. \tag{15}$$

In addition to estimation accuracy, treatment policies induced by the estimated CATE functions are evaluated using Inverse Propensity Scoring (IPS) and Doubly Robust (DR) policy value estimators.

3.1.8. Monte Carlo Experimental Design

The simulation experiment is repeated for 500 Monte Carlo replications. In each replication, observations are randomly divided into training and test sets using a 70–30% split.

All preprocessing operations, including standardization, empirical copula transformation, and feature augmentation, are fitted using only the training sample. The learned transformations are then applied to the independent test sample.

For each replication, competing models are trained using identical training data, treatment assignments, and outcome realizations. The resulting CATE estimates are evaluated using Bias, RMSE, IPS, and DR policy evaluation metrics.

Table 1 summarizes the average performance and variability of all competing methods across Monte Carlo replications.

Table 1. Monte Carlo simulation results for CATE estimation and policy evaluation. The table summarizes the distribution of performance metrics across Monte Carlo replications for the proposed Copula-Tensor Neural Network (Copula-TNN) and the neural network S-learner without copula transformation.

Model	Metric	Min	Q1	Median	Q3	Max	Mean	SD	IQR
Copula-TNN	Bias	−1.03	0.201	0.493	0.783	1.81	0.498	0.424	0.582
NN-S Learner	Bias	−0.218	0.343	0.483	0.629	1.30	0.490	0.214	0.285
Copula-TNN	DR	0.270	1.92	2.12	2.32	4.16	2.12	0.310	0.393
NN-S Learner	DR	−0.039	1.88	2.04	2.21	4.44	2.05	0.270	0.329
Copula-TNN	IPS	0.334	1.57	1.69	1.84	3.85	1.73	0.288	0.266
NN-S Learner	IPS	0.246	1.53	1.65	1.79	3.64	1.69	0.289	0.264
Copula-TNN	RMSE	1.38	1.82	1.92	2.03	2.74	1.93	0.168	0.213
NN-S Learner	RMSE	1.37	1.78	1.91	2.04	2.66	1.91	0.189	0.265

3.2. Simulation Results

Table 1 summarizes the Monte Carlo simulation results for CATE estimation accuracy and policy evaluation. The results indicate that the proposed copula-based representation provides comparable CATE estimation performance while achieving moderate improvements in policy evaluation metrics.

For pointwise CATE estimation, the proposed Copula-TNN and the neural network S-learner without copula transformation achieve similar RMSE values (1.93 and 1.91, respectively). This indicates that the empirical copula transformation does not substantially reduce prediction error under the latent-factor dependence structure considered in this simulation. However, the results also demonstrate that the copula representation does not introduce additional prediction degradation, suggesting that it provides a stable dependence-aware feature transformation for high-dimensional causal learning.

The estimated bias values are also comparable between the two approaches. The average bias is 0.498 for the Copula-TNN and 0.490 for the NN-S Learner. These results suggest that the empirical copula transformation alone does not remove finite-sample estimation bias. Instead, its potential contribution lies in representing nonlinear dependence structures among covariates and improving the ranking of individuals according to their estimated treatment effects.

The largest differences are observed in policy evaluation metrics. The average Doubly Robust (DR) value increases from 2.05 for the NN-S Learner to 2.12 for the Copula-TNN, while the average Inverse Propensity Scoring (IPS) value increases from 1.69 to 1.73. Although these improvements are moderate, they suggest that the copula-based

representation may provide more effective treatment-effect ranking, which is important for individualized treatment selection and policy optimization.

The distributional comparisons further support this observation. The Copula-TNN achieves slightly higher median and upper-quartile values for both DR and IPS metrics, indicating that the improvement is not caused solely by a small number of favorable Monte Carlo replications. Instead, the copula representation provides relatively consistent improvements in policy evaluation performance across repeated experiments.

However, the Copula-TNN also exhibits increased variability in some estimation metrics. For example, the interquartile range of the bias distribution is larger for the Copula-TNN (0.582) compared with the NN-S Learner (0.285). This result suggests a finite-sample trade-off: while dependence-aware representations may improve treatment-effect ranking, they may also introduce additional estimation variability when the sample size is limited.

Overall, these findings suggest that the proposed copula-based framework should not be interpreted as universally improving all aspects of CATE estimation. Rather, its main contribution is providing a flexible dependence-aware representation that can improve downstream policy evaluation under complex covariate dependence. The effectiveness of the proposed approach is further investigated under alternative simulation settings with varying sample sizes, feature dimensions, dependence strengths, and marginal distributions.

Figure 1 illustrates the distribution of RMSE, bias, IPS, and DR metrics across Monte Carlo replications. The boxplots summarize the central tendency, variability, and robustness of the competing approaches.

For RMSE, both models exhibit similar median performance, indicating that the empirical copula transformation does not substantially change pointwise CATE estimation accuracy under the considered latent-factor dependence structure. The NN-S Learner shows slightly smaller dispersion, suggesting more stable finite-sample estimation under the baseline feature representation.

For bias, both approaches demonstrate comparable average estimation behavior. However, the Copula-TNN exhibits a wider distribution with more extreme replications, indicating increased variability in finite samples. This result suggests that while the copula-based representation provides a flexible dependence-aware feature transformation, it may also introduce additional estimation variability when the available sample size is limited.

For IPS, the two approaches exhibit similar distributions, although the NN-S Learner shows slightly tighter concentration. For the DR estimator, the Copula-TNN achieves somewhat higher upper-tail performance, but this improvement is accompanied by increased dispersion across Monte Carlo replications.

The variability comparison provides additional insight into the behavior of the two approaches. The Copula-TNN achieves comparable or slightly lower variability for RMSE, similar variability for IPS, and higher variability for bias and DR metrics. Therefore, the empirical copula representation appears to provide benefits primarily for treatment-effect ranking and policy evaluation rather than for reducing pointwise CATE estimation error.

Overall, the simulation results demonstrate a trade-off between representation flexibility and estimation stability. The Copula-TNN improves average policy value measured by IPS and DR while maintaining comparable CATE estimation accuracy. However, these improvements are not uniform across all evaluation criteria and may depend on sample size, covariate dimension, and dependence structure. Further simulation studies considering different dimensional settings, dependence strengths, and marginal distributions are therefore necessary to characterize the conditions under which the proposed framework provides advantages.

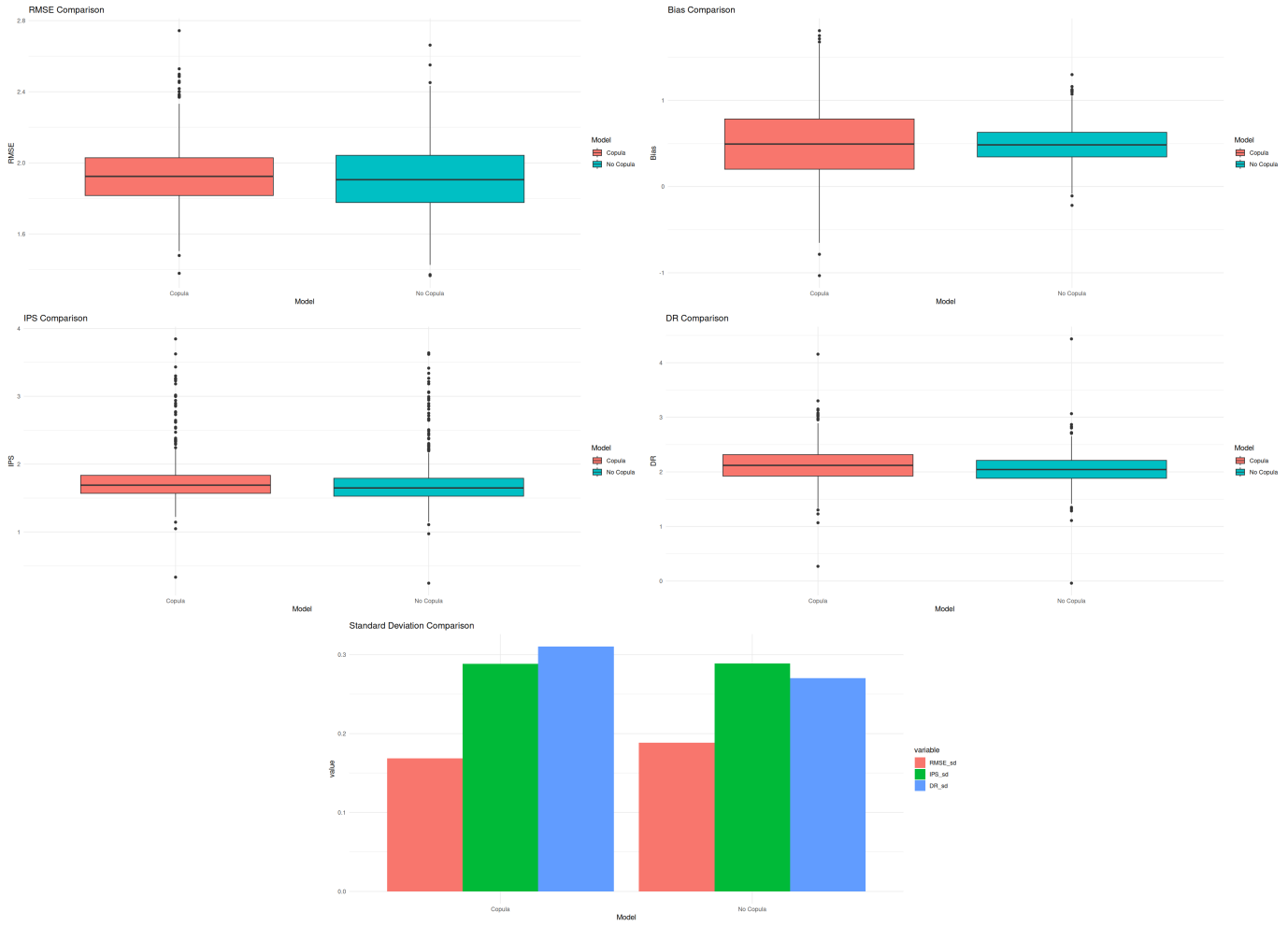

Figure 1. Distribution of CATE estimation error, bias, and policy evaluation metrics across Monte Carlo replications. The figure compares the proposed Copula-Tensor Neural Network (Copula-TNN) and neural network S-learner without copula transformation (NN-S Learner) under simulated high-dimensional dependence settings.

4. Real Data Analysis

We evaluate the proposed framework using the *Criteo Uplift Dataset* [12], a widely used benchmark dataset for heterogeneous treatment effect estimation and uplift modeling in digital advertising. The dataset is based on a randomized advertising experiment, where users are randomly assigned to either an advertising treatment group or a control group. The objective is to estimate heterogeneous treatment effects and identify users who are most likely to benefit from advertising exposure.

The dataset contains a binary treatment indicator, a binary conversion outcome, and high-dimensional user-level features describing browsing behavior and contextual characteristics. These properties make the dataset suitable for evaluating causal machine learning methods designed to capture nonlinear treatment heterogeneity in large-scale applications.

4.1. Data Description and Preprocessing

We randomly select n = 50,000 observations from the original dataset to facilitate repeated model training while preserving the heterogeneous structure of the benchmark data. The resulting dataset is represented as

$$\{(X_i, T_i, Y_i)\}_{i=1}^{50{,}000}.$$

The binary conversion outcome is defined as

$$Y_i = \texttt{conversion}_i,$$

where $Y_i = 1$ indicates that user i completed a conversion after the advertising intervention.

The treatment indicator is defined as

$$T_i = \texttt{treatment}_i,$$

where $T_i = 1$ represents exposure to advertising and $T_i = 0$ represents the control condition.

The covariate matrix is constructed by excluding the treatment and outcome variables:

$$X = \{\text{all features except } (\texttt{conversion}, \texttt{treatment})\}.$$

After preprocessing, the feature matrix is expressed as

$$X \in \mathbb{R}^{n \times p'},$$

where p' denotes the number of retained predictors after removing near-constant variables.

Variables satisfying

$$\mathrm{Var}(X_j) \leq 10^{-6}$$

are removed. Continuous variables are standardized using the mean and standard deviation estimated from the training sample only. Categorical variables are encoded using the numerical representations provided in the benchmark dataset.

Therefore, for each train-test split, preprocessing parameters are learned exclusively from the training data and subsequently applied to the independent test data to avoid information leakage.

4.2. Feature Representation

High-dimensional user characteristics may exhibit nonlinear dependence and complex geometric structures. To investigate whether nonlinear feature representations improve causal estimation, we augment the original covariates using manifold learning approaches.

The following representation methods are considered:

- Uniform Manifold Approximation and Projection (UMAP);
- t-distributed Stochastic Neighbor Embedding (t-SNE);
- Isometric Mapping (Isomap).

Each embedding is learned using the training sample only. The resulting low-dimensional representations are incorporated into the original feature space:

$$X^* = [X, \phi_{\mathrm{UMAP}}(X), \phi_{\mathrm{tSNE}}(X), \phi_{\mathrm{ISO}}(X)].$$

For prediction on test observations, the corresponding out-of-sample embedding procedures are used. This prevents information from the test set being used during feature construction.

For the proposed copula-based framework, the empirical copula transformation is subsequently applied to the augmented representation X^*. The transformed representation is then combined with treatment indicators and treatment–covariate interactions for CATE estimation.

4.3. Model Specification

The neural network architecture follows the same configuration used in the simulation experiments. The model consists of three fully connected hidden layers with 256, 128, and 64 neurons, respectively. Each hidden layer uses ReLU activation, and dropout regularization with rate 0.3 is applied to reduce overfitting.

The network is trained using the Adam optimizer with a learning rate of 0.001. The binary outcome prediction loss is defined as

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Although the response variable is binary, the regression-based objective is used because the primary goal is estimation of heterogeneous treatment effects through counterfactual prediction rather than classification accuracy.

To ensure a fair comparison, the Copula-TNN and NN-S Learner use identical network architectures, optimization procedures, and training settings. The only difference is the inclusion of the empirical copula-based representation.

4.4. Repeated Evaluation and Policy Assessment

To evaluate robustness, we repeat the analysis using 100 independent train-test splits. In each replication, the dataset is randomly divided into training and testing samples; preprocessing transformations are estimated from the training data, and the fitted models are used to generate counterfactual predictions:

$$\hat{\tau}(X_i) = \hat{Y}_i(1) - \hat{Y}_i(0).$$

Because the individual treatment effect is not directly observed in real-world applications, conventional CATE bias and RMSE cannot be computed exactly. Therefore, evaluation focuses primarily on policy-oriented metrics, including Inverse Propensity Scoring (IPS) and Doubly Robust (DR) policy value estimators.

When pseudo-outcome-based error measures are reported, they are interpreted as relative evaluation criteria rather than exact measures of individual causal effect estimation error.

Table 2 summarizes the distribution of performance metrics across repeated experiments, including average values, variability, and quantile-based measures.

The repeated evaluation provides insight into the stability of the proposed copula-based representation under large-scale high-dimensional treatment effect estimation settings.

Table 2. Real-data evaluation results for the proposed Copula-Tensor Neural Network (Copula-TNN) and neural network S-learner without copula transformation (NN-S Learner) on the Criteo uplift dataset.

Model	Metric	Min	Q1	Median	Q3	Max	Mean	SD	IQR
Copula-TNN	Bias	−0.0103	−0.00334	−0.00135	0.000737	0.00787	−0.00137	0.00334	0.00408
NN-S Learner	Bias	−0.0113	−0.00323	−0.00128	0.00120	0.00852	−0.000893	0.00339	0.00443
Copula-TNN	DR	−0.00790	−0.000247	0.00172	0.00441	0.00979	0.00196	0.00352	0.00466
NN-S Learner	DR	−0.0204	−0.00416	0.000345	0.00420	0.0224	0.000634	0.00792	0.00836
Copula-TNN	IPS	0.000534	0.00227	0.00263	0.00311	0.00419	0.00263	0.000741	0.000842
NN-S Learner	IPS	0.00116	0.00220	0.00269	0.00308	0.00491	0.00269	0.000643	0.000883
Copula-TNN	RMSE	0.0459	0.0681	0.0864	0.0984	0.172	0.0864	0.0251	0.0303
NN-S Learner	RMSE	0.0461	0.0685	0.0867	0.0998	0.172	0.0864	0.0253	0.0313

4.5. Real Data Results

Table 2 summarizes the performance of the Copula-TNN and NN-S Learner across repeated train-test evaluations on the Criteo uplift dataset. Since individual treatment effects are not directly observed in this benchmark dataset, the reported error measures should be interpreted as relative evaluation criteria based on the adopted pseudo-outcome framework rather than exact measures of individual causal effect estimation accuracy.

The two models exhibit nearly identical predictive performance, with mean RMSE values of approximately 0.0864. The similarity in median values and interquartile ranges indicates that the empirical copula transformation does not substantially change outcome prediction performance in this real-data setting. This result is consistent with the simulation findings, suggesting that the primary contribution of the copula representation is not necessarily improved pointwise prediction accuracy.

The pseudo-outcome-based bias measures are also similar between the two approaches. The Copula-TNN produces a mean bias of -0.00137, compared with -0.000893 for the NN-S Learner. The differences are small relative to the overall variability, indicating that the copula transformation does not substantially alter systematic estimation behavior under this benchmark setting.

The most notable difference is observed in the Doubly Robust (DR) policy evaluation metric. The Copula-TNN achieves a higher mean DR value (0.00196) compared with the NN-S Learner (0.000634). In addition, the variability of the DR estimator is reduced, with the standard deviation decreasing from 0.00792 to 0.00352 and the interquartile range decreasing from 0.00836 to 0.00466. These results suggest that the copula-based representation may provide a more stable feature representation for policy value estimation.

For the Inverse Propensity Scoring (IPS) estimator, both approaches exhibit very similar performance. The mean IPS values are nearly identical (0.00263 for Copula-TNN and 0.00269 for NN-S Learner), indicating that the copula transformation has limited impact on IPS-based evaluation in this application.

Overall, the real-data results suggest that the main advantage of the proposed copula-based framework is associated with the stability of Doubly Robust policy evaluation rather than substantial improvement in predictive accuracy. The results support the view that copula-based feature representations may be useful in high-dimensional treatment effect problems where reliable policy evaluation is important.

Figure 2 illustrates the distribution of performance metrics across repeated train-test evaluations. The RMSE and bias distributions are highly similar between the Copula-TNN and NN-S Learner, confirming comparable predictive performance and systematic estimation behavior.

The IPS distributions are also nearly identical, with only minor differences in dispersion. In contrast, the DR estimator exhibits a more concentrated distribution for the Copula-TNN, suggesting improved stability of Doubly Robust policy evaluation.

The variability comparison further supports this observation. The Copula-TNN achieves comparable variability for RMSE and IPS while showing reduced dispersion in DR estimates. These results indicate that the empirical copula representation does not substantially change predictive accuracy but may provide a more stable representation for policy evaluation in high-dimensional treatment effect applications.

Overall, the Criteo analysis demonstrates that both approaches achieve similar central performance, while the copula-based framework provides potential benefits in stabilizing policy evaluation metrics. These findings should be interpreted within the context of the benchmark dataset, and further studies with additional datasets are needed to characterize the settings where dependence-aware representations provide consistent advantages.

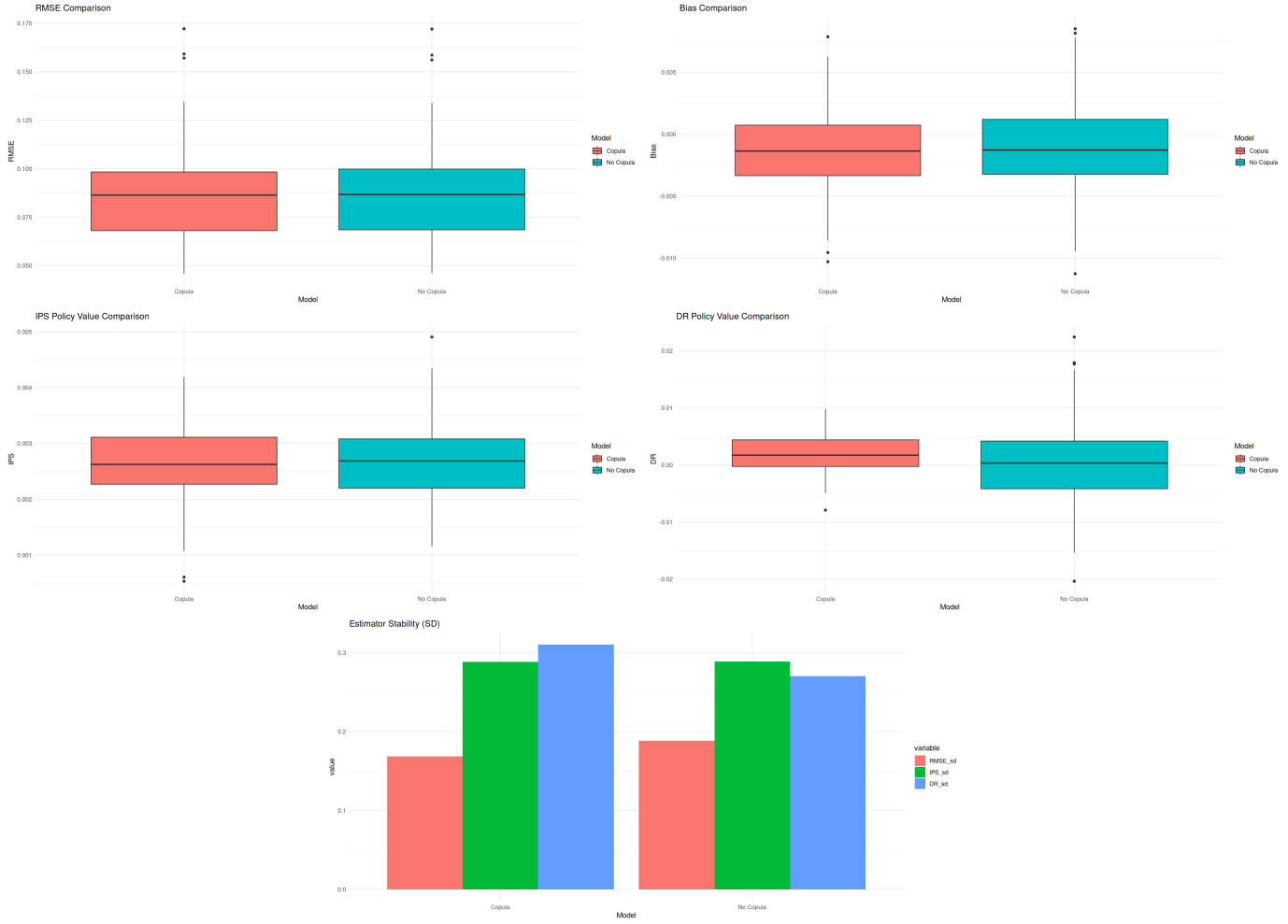

Figure 2. Distribution of pseudo-outcome-based error measures, bias, and policy evaluation metrics (IPS and DR) for the Copula-TNN and NN-S Learner across repeated evaluations on the Criteo uplift dataset.

5. Conclusions

In this paper, we proposed a copula-based feature representation framework for heterogeneous treatment effect estimation and policy evaluation in high-dimensional settings. The proposed approach integrates empirical copula transformations, structured treatment–covariate interaction representations, and deep neural networks to construct a flexible dependence-aware feature space for modeling nonlinear treatment response relationships.

Through Monte Carlo simulations, we evaluated the proposed framework under high-dimensional covariate dependence, nonlinear treatment assignment, and heterogeneous treatment effects. The results demonstrate that the proposed Copula-Tensor Neural Network achieves comparable CATE estimation accuracy to the baseline neural network approach in terms of RMSE and bias. However, the copula-based representation provides moderate improvements in downstream policy evaluation metrics, particularly for Doubly Robust (DR) policy value estimation. These findings suggest that dependence-aware representations may improve the ranking of individuals according to estimated treatment effects, which is important for individualized treatment decisions and policy optimization.

The real-data analysis using the Criteo uplift dataset further demonstrates that the copula-based and non-copula models achieve similar predictive performance. The main difference is observed in the variability of the DR estimator, where the copula-based framework produces more concentrated policy value estimates. Since individual treatment effects are not directly observed in real-world applications, these results should be inter-

preted primarily as evidence of improved policy evaluation stability rather than definitive improvement in true CATE estimation accuracy.

Overall, the proposed framework provides a flexible approach for incorporating dependence-aware representations into deep causal learning models. The results highlight an important trade-off between representation flexibility and estimation stability. While copula-based transformations may improve the robustness of policy evaluation under complex covariate structures, they do not uniformly improve every CATE estimation criterion. The effectiveness of the framework depends on characteristics of the data, including sample size, covariate dimension, dependence structure, and treatment assignment mechanism.

Several limitations should be acknowledged. First, the proposed framework introduces additional computational complexity due to the combination of copula transformation, manifold-based feature augmentation, and deep neural network optimization. Second, the current methodology does not provide formal theoretical guarantees regarding estimator consistency, asymptotic efficiency, or uncertainty quantification. Third, the scalability of the approach for extremely high-dimensional settings requires further investigation.

Future research directions include developing adaptive copula selection procedures, integrating more flexible dependence models with deep generative representations, and establishing theoretical properties of copula-based causal estimators. Additional studies involving larger-scale datasets, uncertainty quantification, and broader comparisons with state-of-the-art CATE methods will be important for further validating the practical utility of the proposed framework.

All R code used in this study is publicly available for reproducibility at: https://github.com/kjonomi/Rcode/blob/main/Tensor (accessed on 17 August 2026).

Funding: This research received no external funding.

Data Availability Statement: The publicly available datasets used in this study, including the Criteo-uplift dataset, can be accessed from their respective sources: [criteo-uplift, accessed on 18 December 2025. https://huggingface.co/datasets/criteo/criteo-uplift/blob/main/criteo-research-uplift-v2.1.csv.gz].

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **1974**, *66*, 688–701. [CrossRef]
2. Rosenbaum, P.R.; Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **1983**, *70*, 41–55. [CrossRef]
3. Bang, H.; Robins, J.M. Doubly robust estimation in missing data and causal inference models. *Biometrics* **2005**, *61*, 962–973. [CrossRef] [PubMed]
4. Hill, J.L. Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Stat.* **2011**, *20*, 217–240. [CrossRef]
5. Athey, S.; Imbens, G. Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7353–7360. [CrossRef] [PubMed]
6. Wager, S.; Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **2018**, *113*, 1228–1242. [CrossRef]
7. Shalit, U.; Johansson, F.D.; Sontag, D. Estimating individual treatment effect: Generalization bounds and algorithms. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3076–3085.
8. Kim, J.-M. Manifold causal conditional deep networks for heterogeneous treatment effect estimation and policy evaluation. *Mathematics* **2026**, *14*, 738. [CrossRef]
9. Sklar, A. *Fonctions de Répartition à n Dimensions et Leurs Marges*; L'Institut de Statistique de L'Université de Paris: Paris, France, 1959; Volume 8, pp. 229–231.
10. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006.

11. Joe, H. *Dependence Modeling with Copulas*; CRC Press: Boca Raton, FL, USA, 2014.
12. Diemert, E.; Betlei, A.; Renaudin, C.; Amini, M.-R. A large scale benchmark for uplift modeling. In *Proceedings of the AdKDD and TargetAd Workshop*; KDD: London, UK, 2018.