Copula-GLMNet: Interpretable Deep Learning of Dynamic Dependency Structures

Tinghe Zhang, Yifan Huang

University of Intenrnational Business and Economics

July, 2025 CCIRM 2025

The Challenge in Dependency Modeling

Dependency modeling is a fundamental problem in statistics, finance, and risk management.

Traditional Models:

- GLMs: Great interpretability but assume linear relationships.
- Copulas: Rigorously separate marginals from dependency, but often rely on static correlations and fixed marginal forms.

Limitation: Less adaptable to complex, high-dimensional, and nonlinear data structures.

Modern Models:

- Deep Learning: High predictive power.
- Often lacks interpretability.
- Not directly designed for dependency estimation.

Limitation: A "black box" nature, which is a major drawback in regulated fields like insurance.

Our Contribution: CopulaGLMnet

We propose CopulaGLMnet, a framework that bridges this gap.

- **Integrates** the strengths of Generalized Linear Models (GLMs), deep neural networks, and copula theory.
- **Balances** statistical interpretability with the modeling flexibility of deep learning.
- **Captures** complex, nonlinear dependencies where the dependency structure itself can be learned from data.

Core Components Overview

Our framework is built on three key statistical concepts:

1. **Generalized Linear Models (GLM)**
   - Foundation: Exponential dispersion family
   - Links mean to covariates via link function
2. **Neural-Enhanced GLM**
   - Replaces linear predictor with neural networks
   - Captures non-linear relationships while maintaining interpretability
3. **Copula Theory**
   - Separates marginal distributions from dependency structure
   - Enables flexible joint distribution modeling

Next, we examine each component in detail.

1. Introduction
2. Background Concepts
   - Generalized Linear Models
   - Neural-Enhanced GLM
   - Copula Theory
3. CopulaGLMnet
4. Simulation and Results
5. Empirical Analysis
6. Conclusion

Generalized Linear Models (GLM)

Foundation: Exponential Dispersion Family (EDF)
Consider response variables Y with density:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - \kappa(\theta)}{\phi/v} + a(y;\phi/v)\right\}$$

where:

- $\theta \in \boldsymbol{\Theta}$: canonical parameter
- $\phi > 0$: dispersion parameter
- $\kappa(\theta)$: cumulant function
- $v > 0$: exposure weights

Moments:

$$\mu = \mathbb{E}[Y] = \kappa'(\theta), \quad \mathrm{Var}(Y) = \frac{\phi}{v}\kappa''(\theta)$$

Regression Structure:

$$g(\mu) = \beta_0 + \langle \boldsymbol{\beta}, \mathrm{x} \rangle$$

EDF Distributions in Insurance

Gamma Distribution: (Heavy-tailed claims)

$$f(y;\mu,\phi) = \frac{(y\phi/\mu)^{1/\phi} y^{-1}}{\Gamma(1/\phi)} \exp\left(-\frac{y\phi}{\mu}\right)$$

Inverse Gaussian Distribution: (Extreme claims)

$$f(y;\mu,\phi) = \sqrt{\frac{\phi}{2\pi y^3}} \exp\left(-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right)$$

Key Properties:

- Both distributions have positive support (suitable for claim amounts)
- Gamma: constant coefficient of variation
- Inverse Gaussian: heavier right tail, suitable for extreme values
- Both allow for flexible variance modeling via ϕ

Limitation: Linear predictor assumes $g(\mu) = \beta_0 + \langle \boldsymbol{\beta}, \mathrm{x} \rangle$

Neural-Enhanced GLM

Architecture: Replace linear predictor with neural network

Layer Transformation:

$$z^{(m)} : \mathbb{R}^{q_{m-1}} \to \mathbb{R}^{q_m}$$

$$z_j^{(m)}(x) = \phi_m \left(w_{0,j}^{(m)} + \sum_{l=1}^{q_{m-1}} w_{l,j}^{(m)} x_l \right)$$

where ϕ_m is the activation function for layer m.

Deep Representation:

$$z^{(d:1)}(x) = \left(z^{(d)} \circ \cdots \circ z^{(1)} \right)(x)$$

Output Layer:

$$g(\mu(x)) = \mathrm{NN}(x)$$

Copula Theory: Sklar's Theorem

Fundamental Theorem: Any joint distribution can be decomposed as:

$$F(y_1, y_2|x_1, x_2) = C_\theta\left(F_1(y_1|x_1), F_2(y_2|x_2)\right)$$

where:

- $F_i(y_i|x_i)$: marginal distribution functions
- $C_\theta : [0, 1]^2 \to [0, 1]$: copula function
- $\theta \in \Theta$: dependence parameter

Joint Density:

$$f(y_1, y_2|x_1, x_2) = c_\theta(F_1(y_1|x_1), F_2(y_2|x_2)) \cdot f_1(y_1|x_1) \cdot f_2(y_2|x_2)$$

where $c_\theta(u_1, u_2) = \frac{\partial^2 C_\theta(u_1, u_2)}{\partial u_1 \partial u_2}$ is the copula density.
Key Insight: Separates marginal behavior from dependency structure.

Copula Families

Gaussian Copula: (Elliptical dependence)

$$C_\rho(u_1, u_2) = \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$$

- $\rho \in (-1, 1)$: correlation parameter
- Symmetric dependence, no tail dependence ($\lambda_L = \lambda_U = 0$)

Clayton Copula: (Lower tail dependence)

$$C_\theta(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$$

- $\theta \in (-1, \infty) \setminus \{0\}$: dependence parameter
- Lower tail dependence: $\lambda_L = 2^{-1/\theta}$

Frank Copula: (Symmetric, no tail dependence)

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln\left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right]$$

Model 1: Baseline GLM with Static Copula Dependency

Marginal Parameter Estimation:

- Simple GLM transformation for marginal means:

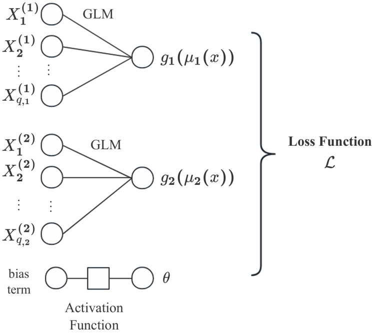
$$\mu_i = \exp(\omega_\mu \mathrm{x}^{(\mathrm{i})} + b_\mu), \quad i = 1, 2$$

Copula Parameter Estimation:

- Static parameter independent of covariates:

$$\theta = \phi(\omega_\theta + 1 \cdot \omega_{0,\theta})$$

- ϕ: activation function (copula-specific)
- Acts as neural network bias unit



Key Features:

- Comparable to traditional copula models
- Interpretable baseline

Model 2: Trainable Marginal Estimation with Static Copula Dependency

Enhanced Marginal Networks:

- Multi-layer feedforward networks for marginals:

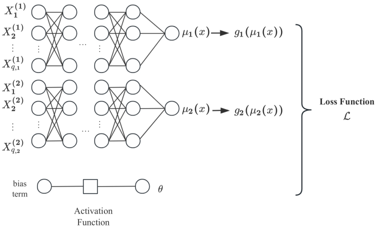
$$\mu_i = \exp(\mathrm{z}_\mu^{(L)}(\mathrm{x}^{(\mathrm{i})})), \quad i = 1, 2$$

- Layer-wise transformation:

$$\mathrm{z}_\mu^{(m)}(\mathrm{x}^{(\mathrm{i})}) = \phi_m(\omega_\mu^{(m)} \mathrm{z}_\mu^{(m-1)}(\mathrm{x}^{(\mathrm{i})}) + b_\mu^{(m)})$$

Copula Parameter:

- Remains static as in Model 1



Key Features:

- Flexible marginal estimation
- Nonlinear feature interactions
- Fixed dependency structure

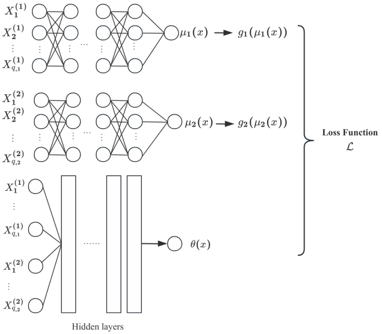
Model 3: Fully Trainable Marginal and Copula Parameter Estimation

Integrated Architecture:

- Independent subnetworks for θ estimation
- Concatenated input: $\mathrm{x} = [\mathrm{x}^{(1)}, \mathrm{x}^{(2)}]$
- Multi-layer transformation:

$$H^{(m)} : \mathbb{R}^{q_{m-1}} \to \mathbb{R}^{q_m}$$

$$h_i^{(m)}(\mathrm{x}) = \phi_m \left(\omega_{0,i}^{(m)} + \sum_{k=1}^{q_{m-1}} \omega_{k,i}^{(m)} \cdot \mathrm{x}_k \right)$$



Copula Parameter Output:

$$\theta = \phi_{\text{out}} \left(\omega_{0,\theta} + H^{(d)}(\mathrm{x})^\top \omega_\theta \right)$$

Key Features:

- End-to-end trainable
- Covariate-dependent θ, Dynamic dependency

Model Architecture Summary

Component	Model 1	Model 2	Model 3
Marginal μ_1, μ_2	GLM	Neural Network	Neural Network
Copula θ	Static	Static	Neural Network
Covariate Dependency	Marginals only	Marginals only	Full Integration
Complexity	Low	Medium	High
Interpretability	High	Medium	Medium
Flexibility	Low	Medium	High

Progressive Enhancement:

- Model 1 $\rightarrow$ Model 2: Enhanced marginal flexibility
- Model 2 $\rightarrow$ Model 3: Dynamic dependency structure
- Each model builds upon the previous architecture

Experimental Setup: Three Simulation Scenarios

We assessed our CopulaGLMnet models using synthetic datasets of increasing complexity, designed to test their ability to capture varying dependency structures and marginal distributions. Each dataset featured covariates from uniform distributions, response variables modeled with Gamma distributions, and dependencies controlled by Gaussian copulas with adjustable parameters.

Case I: Constant Dependence

Objective: Evaluate model performance under a simple, homogeneous dependency.

Data Generation:

- **Covariates:** $X_1, X_2, X_3 \sim \text{Uniform}(0,1)$, and a categorical variable $t \in \{1,2,3\}$.
- **Copula:** Gaussian copula with a fixed parameter ($\theta = 0.4$) for all observations.
- **Marginal Means:** Linear functions:

$$\mu_1 = \exp(6 + 0.5X_{1i} + 2X_{2i} - X_{3i})$$

$$\mu_2 = \exp(5 + X_{1i} + 1.5X_{2i} + 0.5X_{3i})$$

- **Dispersion:** $\phi_1 = 0.5$, $\phi_2 = 1.25$.

Case II: Linear Marginals with Varying Dependence

Objective: Assess the model's capacity to identify structured heterogeneity in dependence.

Data Generation:

- **Covariates:** Same as Case I.
- **Copula:** Gaussian copula with a varying parameter θ across three distinct groups, guided by t:

$$(U_{i1}, U_{i2}) \sim \begin{cases} \text{Gaussian Copula}(\theta_1), & t = 1, \quad \theta_1 = 0.3, \\ \text{Gaussian Copula}(\theta_2), & t = 2, \quad \theta_2 = -0.1, \\ \text{Gaussian Copula}(\theta_3), & t = 3, \quad \theta_3 = -0.5. \end{cases}$$

- **Marginal Means:** Linear functions, identical to Case I.
- **Dispersion:** $\phi_1 = 0.5$, $\phi_2 = 1.25$.

Discussion: This case is relevant to actuarial science, where dependence can shift across subpopulations.

Case III: Nonlinear Marginals with Varying Dependence

Objective: Challenge the model with both complex marginals and dynamic dependencies, simulating a more realistic scenario.

Data Generation:

- **Covariates:** Same as Case I.
- **Copula:** Varying parameters across groups, identical to Case II.
- **Marginal Means:** Nonlinear functions:

$$\mu_1 = \exp(6 + 0.5X_{1i} + 2X_{2i}^{\frac{1}{2}} - X_{3i}^{\frac{2}{3}})$$

$$\mu_2 = \exp(5 + X_{1i} + 1.5X_{2i}^{\frac{1}{3}} + 0.5X_{3i}^{2})$$

- **Dispersion:** $\phi_1 = 0.5$, $\phi_2 = 1.25$.

Results Analysis: Model Performance Comparison

Out-of-Sample Results (Gaussian Copula, Gamma, Gamma)

Case	Model	Loss	Sum of RMSE
Case I	Model 1	15.447	1965.530
	Model 2	15.246	1723.996
	Model 3	15.222	1701.785
Case II	Model 1	15.318	1918.822
	Model 2	15.158	1710.709
	Model 3	15.137	1696.088
Case III	Model 1	15.661	2137.198
	Model 2	15.557	2044.070
	Model 3	15.539	2010.086

Finding:

- In all three scenarios, Model 3 consistently shows the best performance, followed by Model 2 and then Model 1.
- This demonstrates the superiority of jointly training marginal distributions and copula dependency parameters.

Discussion:

- **Case I (Simple Data):** Model 3's flexibility allows it to capture subtle non-linearities and better adapt to random noise in the data, leading to an improved fit even in this simple setting.
- **Case III (Complex Data):** The performance gap between Model 2 and Model 3 is not as large as the gap between Model 1 and Model 2.

Results: Learned Dependence Structures

To verify if Model 3 correctly identifies and predicts the group-specific dependency parameters for the test set, based on the categorical variable t.

Discussion:

- The plot visualizes the distribution of predicted θ values for samples in the test set, grouped by the true categorical variable t.
- The horizontal red dashed lines represent the true copula parameter values used to generate the data: 0.3 for $t = 1$, -0.1 for $t = 2$, and -0.5 for $t = 3$.
- As demonstrated by the clusters of predicted points and their mean values (blue diamonds), the model correctly learns to predict distinct θ values for each group.
- This shows Model 3's ability to successfully capture the designed structured heterogeneity in the dependency parameter, a key feature of the CopulaGLMnet framework.

Robustness Analysis: Performance Under Mismatched Assumptions

To test the framework's robustness by examining model performance under incorrect distribution assumptions.

Model	Gaussian + Gamma + Gamma (True Assumptions)		Frank Copula Assumption		Inv. Gaussian Marginal Assumption	
	Loss	RMSE	Loss	RMSE	Loss	RMSE
Model 1	15.661	2137.198	15.657	2131.171	16.315	2132.707
Model 2	15.557	2044.070	15.553	2004.374	16.211	2025.690
Model 3	15.539	2010.086	15.464	2053.174	16.191	2109.693

Discussion:

- The overall ranking of models remains consistent, showing the framework's robustness.
- **Frank Copula Test:** The Frank assumption yields lower loss values for all models. This unexpected result shows the issue of *copula non-identifiability*. The existence of alternative, non-Gaussian copula models that achieve a similar or even lower loss suggests that Model 3 may have found an equally valid, if not better, dependency structure within the Frank family, even if it is not the true underlying model.
- **Inverse Gaussian Marginal Test:** All models show an increase in loss. This indicates the importance of a correct marginal distribution assumption, as a fundamental mismatch in data structure cannot be fully compensated even by highly flexible models.

Empirical Analysis: Background

Modeling Motor Insurance Premium Dependencies

Data Source:

- French motor insurance portfolio from `CASdatasets` (`fremotor1prem0304b`)
- $n = 25,878$ policy records with complete premium and driver/vehicle information

Target Variables:

- **Y1 (Own Damage Premium)**: `PremDamAll`
- **Y2 (Third-Party Liability Premium)**: `PremTPLM + PremTPLV`

Covariates:

- **Categorical:** Gender, Marital Status, Vehicle Class, Usage, Region, etc.
- **Numerical:** Driver Age, Bonus-Malus, License Number, Vehicle Age

Empirical Analysis: Real-World Application

Out-of-Sample Performance on Motor Insurance Data

The empirical results on real-world motor insurance data corroborate the findings from the simulation studies. Model 3, with its fully trainable architecture, consistently achieves better out-of-sample performance.

Out-of-Sample Performance on Empirical Data (Case 1)

Model	Outsample 1	Outsample 2
Model 1	108.4517	8.13756
Model 2	110.3977	8.022532
Model 3	107.4961	6.576198

Discussion:

- Model 3 achieves the lowest out-of-sample error. Jointly training both marginal distributions and the dependence parameter achieves high predictive efficiency.

Conclusion

- **Contribution:** We proposed CopulaGLMnet, a novel framework integrates traditional statistical models and modern deep learning approaches to model complex dependencies.
- **Performance:** Our simulation and empirical results show that jointly training both marginal distribution and copula dependence parameters can improve model fit and generalization performance.
- **Robustness:** CopulaGLMnet exhibits superior robustness, maintaining a consistent performance ranking even when model assumptions are mismatched with the true data structure.
- **Empirical Evidence:** The findings from simulation experiments are corroborated by an empirical analysis on real-world motor insurance data, showing the framework's practical utility and effectiveness in actuarial applications.

Implications and Future Work

Future Research Directions

- **Explainability Analysis:** Explore how to further enhance model interpretability using tools like Partial Dependence Plots (PDPs) to reveal how covariates adaptively influence the dependency parameters. Other explainable AI tools like SHAP could also be integrated to provide deeper insights into model behavior.
- **Architectural Enhancements:** Investigate alternative architectures, such as LocalGLMnet, to better balance flexibility and interpretability.
- **Regularization and Constraints:** Study the use of regularized architectures or structured parameter constraints to improve model stability and prevent potential overfitting in highly complex settings.

Thank You