Modeling Dependence Structure of Multiple Outcomes in Causal Inference: A Bayesian Approach *

Yaohan Chen
Anhui University

Zhou Wu
Zhejiang University

Tao Zeng
Zhejiang University

Qiankun Zhou
Louisiana State University

November 6, 2025

Abstract

Inspired by the increasing attention to the distributional regression approach among econometricians, we propose a semiparametric copula-based approach for modeling the dependence structure of multiple outcomes, potentially to be applied to model heterogeneous (distributional) intervention effects with multiple response outcomes in a causal inference setting. The method we propose in this paper is simple to implement by integrating copula methods into a distributional regression framework. Simulation studies demonstrate the effectiveness of our approach, and we illustrate its application by examining how minimum wages affect the dependence between part-time and full-time employment, following Card and Kruger (1994).

Keywords: causal inference, dependence structure, treatment effects, copula, MCMC.
JEL Classification: C11, C14, C51, J00.

*Address correspondence to: Yaohan Chen, School of Big Data and Statistics, Anhui University, Hefei, China. Email: yaohan.chen@ahu.edu.cn.

1. Introduction

After decades of methodological development following Ashenfelter and Card (1985), causal inference methods, such as difference-in-differences (DiD), are now widely adopted in applied empirical research. Potential outcome setup (Robins, 1986; Imbens and Rubin, 2015) is most widely-used framework for conducting causal inference. In the traditional potential outcome framework, where a binary indicator assigns units to either the treated or control group, most causal inference methods focus on estimating the effect of the treatment on a single outcome, by separating this effect from common trends shared by both groups. Among all the methods in the literature, difference-in-differences (DiD) is a representative approach, which is essentially a linear approach for modeling the average (or aggregate) causal effect on a single outcome. However, after decades of development in causal inference methodologies, the academic community has realized that conventional methods—focusing on the mean effect (i.e., mean or aggregate effect) of the treatment on a single outcome—are somewhat limited. By contrast, a more comprehensive evaluation of a policy—one that can reveal how multiple outcomes are jointly affected by the policy—would be of greater interest (Fernández-Val et al., 2025). More importantly, within the setting featuring multiple outcomes that are potentially affected by the treatment, we might be interested not only in how each of the outcomes is affected by the treatment, but also in how the relationship between the outcomes is affected by the treatment, which we refer to as the dependence structure in this paper. Modeling multiple outcomes in causal inference has become an area of growing interest among researchers because it aligns with the natural intuition that—without a model capturing how other variables would change due to a policy change—the predictions based solely on a specific policy variable, while holding other variables constant, are likely to be misleading. As Athey (2025) suggests, modeling multiple outcomes is a crucial and hence promising future direction in causal inference. Against this backdrop, the primary objective of this paper is to develop a modeling

framework in which policy intervention effects are characterized as distributional effects, as in Gunsilius (2023) and William, Gunsilius, and Rigollet (2024), among others, contrasting with the traditional focus on mean effects. Building on this, we aim to model and estimate how the dependence structure among multiple outcomes is affected by the policy intervention—that is, the causal effects on the relationships between outcomes. However, it is worthwhile pointing out that the objective of this paper — modeling multiple outcomes — differs from the objective in the setting of modeling multivalued treatments. Briefly, in settings with multivalued treatments—where the treatment variable may assume more than two distinct values—treatment effects generally depend on the counterfactual alternatives that treated individuals would have chosen in the absence of treatment (Heckman et al., 2000). By contrast, the multiple-outcome framework analyzed in this paper is principally concerned with two objectives: (i) the effect of the treatment on each marginal outcome, and (ii) the effect of the treatment on the joint distribution of outcomes and on the dependence structure that links them.

Copula methods are widely used techniques for modeling the dependence structure of multiple outcomes and thus serve as fundamental tools for the main objectives of this paper. A copula is a joint distribution on $[0,1]^M$ (M stands for the dimension), whose marginals are uniform on $[0,1]$. When the marginals are continuous the copula is unique. With copula methods, we can split the marginals and dependence between the random variables. Specifically, on one side, we have the marginals and on the other side, we can use copula to link the marginals and model the dependence structure (Joe, 2014). Sklar's (1959) theorem lays the theoretical foundation and says that any multivariate distribution can be decomposed into its marginal distributions and a copula that ties them together. Because of their flexibility, copula methods have gained widespread popularity in both statistics and econometrics: see Nelsen (2010), Fan and Patton (2014), and Joe (2014) for nice and comprehensive reviews.

In this paper, we use Gaussian CDF as the link function that associates with the conditional transformation method borrowed from the distributional regression modeling framework (see Arellano and Bonhomme, 2017; Chen et al., 2024; Chen, Liu, and Zhang, 2025; Chernozhukov, Fernández-Val, and Luo, 2023; Chernozhukov et al., 2025; Spady and Stouli, 2025, and references therein) to establish a framework for quantitatively investigating how the dependence structure of multiple outcomes are affected by the corresponding exogenous interventions. Based on this, a Bayesian pseudo-likelihood method is established to estimate and conduct inference on the copula parameters that characterize the dependence structure. The estimation method proposed in this paper is closely related to the two-stage method for estimating parameters that characterize the dependence structure, initially proposed in Genest, Ghoudi, and Rivest (1995), which is also currently known as the method of inference function of margins (IFM) (Joe, 2005). This two-stage method is popular partly because it naturally reflects the copula principle of modeling the dependence structure separately from the marginal distributions and is straightforward for implementation. Set in this two-stage estimation framework, one readily available estimation strategy for the dependence parameters is maximum likelihood estimation method (MLE). Estimators obtained from the two-stage methods are well-behaved for continuous data and theoretically involve a minor loss of efficiency. However, even when the corresponding likelihood has an analytic form and optimization can be used to obtain an estimator, the associated large-sample theory remains comparatively complicated, making it difficult to incorporate into inference for parameters of the dependence structure in causal-inference settings — for example, when constructing confidence intervals for the estimated dependence parameters corresponding to both treated group and control group. By contrast, the advent of modern Bayesian methods, including the Markov Chain Monte Carlo (MCMC) method and variational Bayes (VB) method, facilitates the inference based on posterior analysis. Under regularity conditions, the Bernstein–von Mises (BvM hereafter) theorem

implies that the posterior is asymptotically equivalent (in total variation) to the MLE's asymptotic normal distribution. Bayesian posterior inference provides an approximation to maximum-likelihood-based inference and, under standard regularity conditions (by the BvM theorem), yields posterior credible intervals that coincide asymptotically with frequentist confidence intervals. Accordingly, the Bayesian methods we propose can obviate the need to compute complex variance estimators required for frequentist inference, while producing asymptotically equivalent frequentist results under the conditions of the BvM theorem. Relatedly, theoretical work has established semiparametric Bernstein–von Mises theorems that justify the use of Bayesian methods for estimation and inference for treatment effects in causal inference (Ray and van der Vaart, 2020; Breunig, Liu, and Yu, 2025).

There has been vast literature on applying Bayesian analysis for estimation and inference within the copula modeling framework. For instance, Pitt, Chan, and Kohn (2006) develop an estimation procedure for multivariate normal copula by modeling the marginal distributions via specified parametric families. Smith and Khaled (2012) establish a Bayesian estimation strategy for copula model with discrete margins and Smith and Klein (2021) comprehensively discuss the how the Hamiltonian Monte Carlo (HMC) method and VB methods estimation and inference of copula scalable to the high dimensions. Literature of this strand also includes the application of Bayesian in complete distributional regression as studied in Murray et al. (2013) and Klein, Kneib, and Lang (2015), among others. We employ standard MCMC with carefully chosen Gaussian random-walk proposals based on auxiliary optimization. We find that our method is easy to implement and readily extensible to settings involving more complex dependence structures, where both estimation and inference of the dependence parameters are required.

The paper proceeds as follows: Section 2 establishes the model setup and discusses how the transformation method from the distribution regression framework applies in integrating into the copula methods for modeling dependence structure of multiple outcomes in a

canonical causal inference setting. Section 3 details the Bayesian estimation and inference methods and Section 4 covers the corresponding simulations for demonstrating efficacy of the proposed method. Section 5 applies the proposed method to study how the dependence structure of part-time and full-time employment is affected by increasing the minimum wage. Section 6 concludes.

2. Model Setup

We consider a standard DiD design with 2 periods, $T \in \{0, 1\}$, and 2 groups, $G \in \{0, 1\}$ in which a binary treatment, $D \in \{0, 1\}$, assigned to the treatment group with $G = 1$ in the second period $T = 1$. There are M multiple observed outcomes collected in $\boldsymbol{Y} = \left(Y^{(1)}, \ldots, Y^{(M)}\right)^\top$. According to Imai and Li (2023), the observed outcomes are $\boldsymbol{Y} = \boldsymbol{Y}(D) = D\boldsymbol{Y}(1) + (1 - D)\boldsymbol{Y}(0)$, where $\boldsymbol{Y}(1) = \left(Y^{(1)}(1), \ldots, Y^{(M)}(1)\right)^\top$ and $\boldsymbol{Y}(0) = \left(Y^{(1)}(0), \ldots, Y^{(M)}(0)\right)^\top$ refers to the potential outcomes. Given this specification, the distribution of $\boldsymbol{Y}$ is the conditional distribution $F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid g, t)$. The treatment refers to a shared event that follows a block-adoption design. When $T = 0$, since no treatments are assigned to both groups, $\boldsymbol{Y} = \boldsymbol{Y}(0)$, and accordingly for $G = 0, T = 0$ and $G = 1, T = 0$, we can identify distributions from the observed outcomes such that

$$F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid 0, 0) = F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 0, 0), \tag{1}$$

$$F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid 1, 0) = F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 1, 0). \tag{2}$$

When $T = 1$, since treatments are only assigned to the treatment group ($G = 1$), and accordingly for $G = 0, T = 1$ and $G = 1, T = 1$, we can identify distributions from the observed outcomes such that

$$F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid 0, 1) = F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 0, 1), \tag{3}$$

$$F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid 1, 1) = F_{\boldsymbol{Y}(1)|G,T}(\boldsymbol{y} \mid 1, 1). \tag{4}$$

The distribution of potential outcomes in the block-adoption design can be further explained using Table 1.

Table 1: Distributions in Block-adoption Design

(a) Observed

	T	
G	0	1
0	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 0,0)$	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 0,1)$
1	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 1,0)$	$F_{\boldsymbol{Y}(1)\|G,T}(\boldsymbol{y} \mid 1,1)$

(b) Counterfactual

	T	
G	0	1
0	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 0,0)$	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 0,1)$
1	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 1,0)$	$F_{\boldsymbol{Y}(0)\|G,T}(\boldsymbol{y} \mid 1,1)$

Each entry in (a) above corresponds to a distribution that can be identified from the observed outcomes for each case. By contrast, each entry in (b) corresponds to the counterfactual distribution. In fact, by comparing (a) and (b), it is evident that, for the non-treated group, the distributions that can be identified from the observed outcomes are identical to the corresponding counterfactual distribution, i.e., the distributions of $\boldsymbol{Y}(0)$, which remains unidentified in the observed outcomes but serves as the target of interest: $F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 1,1)$ (the $(1,1)$ entry in (b)), the distribution of the potential outcomes under the non-treated status when $G=1$ and $T=1$. In other words, if one can identify $F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 1,1)$ under certain regular conditions, then the treatment effect can also be identified by comparing $F_{\boldsymbol{Y}(1)|G,T}(\boldsymbol{y} \mid 1,1)$ and $F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 1,1)$. In this paper, we show that we can first model the corresponding marginal univariate distributions and then link these marginal distributions via copula methods in combination with the monotonic transformation in the distributional regression to model the dependence structure captured in $F_{\boldsymbol{Y}|G,T}(\boldsymbol{y} \mid g,t)$. Specifically, we use $F_{Y^{(m)}|G,T}(y^{(m)} \mid g,t)$ to denote the univariate marginal distribution of the m-th outcome $(1 \leqslant m \leqslant M)$. Then we apply the monotonic transformation the distributional regression method as studied in Fernández-Val et al. (2025) to model each $F_{Y^{(m)}|G,T}(y^{(m)} \mid g,t)$. We proceed to detail the modeling framework in the following discussion.

For arbitrary (g,t) taking value in $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$, we follow Fernández-

Val et al. (2025) to model the distribution of the potential outcome $Y^{(m)}(0)$ under the non-treated status using the following distributional regression approach

$$F_{Y^{(m)}(0)|G,T}\left(y^{(m)} \mid g,t\right) = \Lambda\left(\alpha(y^{(m)}) + \beta(y^{(m)})t + \gamma(y^{(m)})g + \delta(y^{(m)})gt\right), \quad y^{(m)} \in \mathbb{R}, \quad (5)$$

where Λ is an invertible CDF. Λ only plays the role of a link function and hence (5) implies no restrictive parametric assumptions about the underlying distribution of $Y^{(m)}(0) \mid G,T$. In this paper, we choose Λ as the univariate standard Gaussian CDF as it naturally leads to Gaussian copula representation. When choosing the Gaussian CDF as the link function Λ, it is also referred to as the local Gaussian representation (LGR) in Chernozhukov, Fernández-Val, and Luo (2023). Given the monotonic property of CDF, functionals in (5) can be decomposed as the quantile discrepancy respectively as follows,

$$\begin{aligned}
\alpha(y^{(m)}) &= \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right) && (6)\\
\beta(y^{(m)}) &= \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)\right) - \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right) && (7)\\
\gamma(y^{(m)}) &= \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)\right) - \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right) && (8)\\
\delta(y^{(m)}) &= \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1)\right) - \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)\right) && (9)\\
&- \left[\Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)\right) - \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right)\right]. && (10)
\end{aligned}$$

We follow the extant literature and impose following assumptions for identification.

Assumption 1. *$\delta(y^{(m)}) = 0$ for all $1 \leqslant m \leqslant M$. Note that this $\delta(y^{(m)}) = 0$ condition can also be interpreted as the parallel trend assumption as in the conventional difference-in-difference literature.*

Assumption 2. *The potential outcomes $\boldsymbol{Y}(0)$ are contained within the support of the observed outcomes $\boldsymbol{Y}$. Since for $(G=0,T=0)$, $(G=0,T=1)$, and $(G=1,T=0)$ we have $\boldsymbol{Y} = \boldsymbol{Y}(0) = \boldsymbol{Y}(1)$, this assumption can be equivalent expressed as follows*

$$(\boldsymbol{Y}(0) \mid G=1,T=1) \subseteq (\boldsymbol{Y}(0) \mid G=0,T=1) \cup (\boldsymbol{Y}(0) \mid G=1,T=0) \cup (\boldsymbol{Y}(0) \mid G=0,T=0). \quad (11)$$

With this decomposition and the assumptions, we can show that $\delta(y^{(m)}) = 0$ can serve as a sufficient condition to identify $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1)$ from distributions that can be identified from the observed outcomes (Fernández-Val et al., 2025). In fact, $\delta(y^{(m)}) = 0$ and (5) jointly imply that,

$$\begin{aligned} & F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1) \\ = \; & \Lambda\left(\alpha(y^{(m)}) + \beta(y^{(m)}) + \gamma(y^{(m)})\right) \\ = \; & \Lambda\left[\Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)\right) + \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)\right) - \right. \\ & \left. \Lambda^{-1}\left(F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right)\right] \end{aligned} \tag{12}$$

where $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)$, $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)$, and $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)$ can be identified from the observed outcomes and estimated via nonparametric methods. Conventional DiD method mainly focus on a single response outcome, namely $M = 1$. When $M \geqslant 2$, i.e., when multiple response outcomes are exposed to treatment, we aim to model the dependence structure across outcomes and to quantify how this dependence changes under the intervention. We discuss the methods we propose for this target in the following discussion.

3. Modeling Dependence Structure

To model the dependence structure of multiple outcomes, we use Gaussian copula methods in combination with the transformation as in (5). For ease of notation, we simply use $Y^{(m)}$ to denote the observed data of the m-th potential outcomes ($Y^{(m)}(1)$ or $Y^{(m)}(0)$). Specifically, for each m, we define $z^{(m)} = \Phi_1^{-1}\left(F_{Y^{(m)}|G,T}(y^{(m)} \mid g,t)\right)$ and $\boldsymbol{z} = \left(z^{(1)}, \ldots, z^{(M)}\right)^\top$.[1] $\Phi_1(\cdot)$ denotes the CDF of the univariate standard Gaussian distribution. This transformation can be visually demonstrated in Figure 1. For the Gaussian copula structure, we use $\mathbf{R}(g,t)$ to denote the correlation matrix of the Gaussian copula function conditional

[1] For ease of notation, we suppress the dependency of $z^{(m)}$ on the (g,t) pairs.

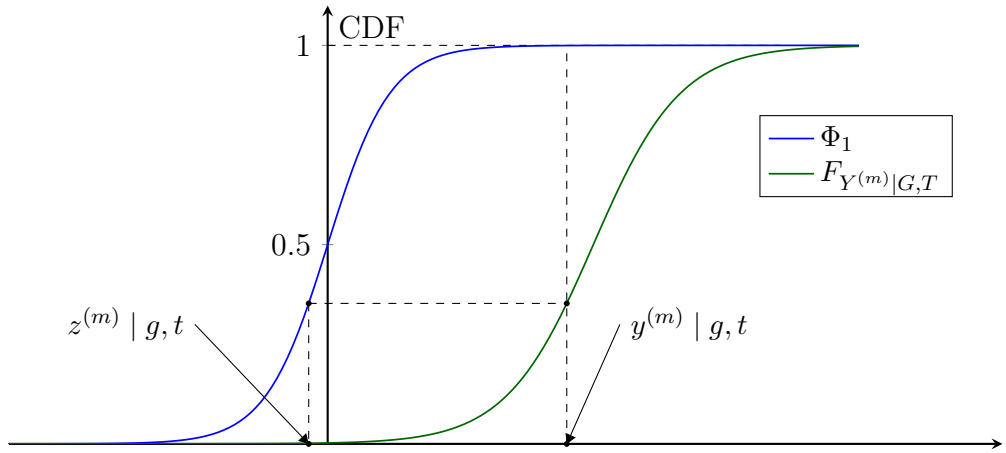


Figure 1: Standard Gaussian distribution as link function.

on $G = g$ and $T = t$. $\mathbf{R}(g,t)$ collects the measure of dependence structure of multiple outcomes, which degenerates to a scalar $\rho(g,t)$ when $M = 2$. One main target of this paper is to estimate and making inference of $\mathbf{R}(1,1)$ for the ex post observed outcomes of the treated group, denoted by $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$, and $\mathbf{R}(1,1)$ for the potential outcomes $\boldsymbol{Y}(0)$, denoted by $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$. Comparison of $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ and $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ implies treatment effects on the dependence structure of multiple outcomes.

By using the Gaussian copula as a linking method and assuming that for each m, $F_{Y^{(m)}|G,T}(y^{(m)} \mid g,t)$ is differentiable with $f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right)$ as the associated PDF, we can derive (pseudo) likelihood function in PDF form as follows,

$$\begin{aligned}
&\frac{\partial^M F_{\boldsymbol{Y}|G,T}(y^{(1)},\ldots,y^{(M)} \mid g,t)}{\partial y^{(1)},\ldots,\partial y^{(M)}} \\
&= \frac{1}{\sqrt{\det\left(\mathbf{R}(g,t)\right)}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^\top\left(\left[\mathbf{R}(g,t)\right]^{-1} - \mathbf{I}_M\right)\boldsymbol{z}\right\} \prod_{m=1}^{M} f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right). \qquad (13)
\end{aligned}$$

The detailed derivations of (13) are summarized in Appendix A.

Given (13) and the i.i.d. assumptions with sample size denoted by n, we obtain pseudo

log-likelihood function by taking logarithm of (13),

$$L = \underbrace{\sum_{i=1}^{n}\left\{-\frac{1}{2}\ln\left(\det\left(\mathbf{R}(g,t)\right)\right)-\frac{1}{2}\boldsymbol{z}_i^{\top}\left(\left[\mathbf{R}(g,t)\right]^{-1}-\mathbf{I}_M\right)\boldsymbol{z}_i\right\}}_{\text{part I}}+ \underbrace{\sum_{i=1}^{n}\left\{\sum_{m=1}^{M}\ln f_{Y^{(m)}|G,T}\left(y_i^{(m)}\mid g,t\right)\right\}}_{\text{part II}}. \tag{14}$$

Remark 1. *It should be emphasized that the sample size varies across different* (g,t)*-pairs. In our model setup, we denote the sample sizes as* $n_{0,0}$*,* $n_{1,0}$*,* $n_{0,1}$*, and* $n_{1,1}$ *for all possible* (g,t)*-pairs. Typically, the notation* n *refers to* $n(1)=n_{1,1}$ *in the observed universe or to* $n(0)=n_{0,0}+n_{1,0}+n_{0,1}$ *in the counterfactual universe.*

Decomposition as in (14) is informative for designing estimation and inference strategy. This pseudo log-likelihood function contains two parts and only part I contain parameters modeling the dependence structure $\mathbf{R}(g,t)$. The two-stage estimation strategy proposed in Genest, Ghoudi, and Rivest (1995) models the marginal distributions nonparametrically using empirical distribution functions in the first stage; in the second stage the original data are transformed, via a link function and the empirical marginal distributions, into pseudo-data $\boldsymbol{z}_i$, and then parameters that model the dependence structure are estimated via Part I. This estimation strategy is also referred to as the inference for margin (IFM) method. Genest, Ghoudi, and Rivest (1995) and Joe (2005) show that under some regular conditions, the copula dependence parameters, namely the $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ and $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ in our model, can be consistently estimated using the IFM estimation procedure with maximum likelihood estimation (MLE) method adopted in the second stage. Although the MLE within the IFM procedure can provide consistent estimates, inference is more involved: the asymptotic distribution typically depends on the Fisher information, which often must be computed numerically (Genest, Ghoudi, and Rivest, 1995; Joe, 2005), and the variance estimator required for constructing test statistics, and for that reason constructing confidence intervals in the frequentist setting is more complicated in practice. One alternative

for making inference with MLE in IFM is bootstrap inference, which is a broadly adopted inference strategy from the Frequentists' perspectives, see Fernández-Val et al. (2025).

We retain the two-stage procedure but propose to perform Bayesian posterior analysis of $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ and $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ in the second stage, which we refer to as Bayesian IFM. In the first stage of Bayesian IFM, we follow the convention in semiparametric literature (Hoff, 2007) to scale the empirical CDF as

$$\hat{F}_{Y^{(m)}|G,T}(y^{(m)} \mid g,t) = \frac{n_{g,t}}{n_{g,t}+1}\frac{1}{n_{g,t}}\sum_{i=1}^{n_{g,t}} \mathbf{1}\left\{Y_i^{(m)} \leqslant y^{(m)}\right\}, \tag{15}$$

where $\mathbf{1}\{\cdot\}$ denotes an indicator function and $n_{g,t} = \sum_{i=1}^{n} \mathbf{1}\{G_i = g\}\mathbf{1}\{T_i = t\}$. This scaling is to ensure the computational stability. With the scaled marginal empirical CDFs, we can estimate marginal distribution of $Y^{(m)}(1)$ and $Y^{(m)}(0)$ of the treated group after the treatment interventions. This is summarized as follows:

- The estimation of marginal distribution of $Y^{(m)}(1)$ after the treatment interventions,

$$\hat{F}_{Y^{(m)}(1)|G,T}(y^{(m)} \mid 1,1) = \frac{n_{1,1}}{n_{1,1}+1}\frac{1}{n_{1,1}}\sum_{i=1}^{n_{1,1}} \mathbf{1}\left\{Y_i^{(m)} \leqslant y^{(m)}\right\}.$$

- The estimation of marginal distribution of $Y^{(m)}(0)$ after the treatment interventions,

$$\begin{aligned}
&\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1) \\
&= \Lambda\left[\Lambda^{-1}\left(\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)\right) + \Lambda^{-1}\left(\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)\right) - \right. \\
&\qquad \left. \Lambda^{-1}\left(\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)\right)\right],
\end{aligned}$$

  where $\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)$, $\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)$, and $\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)$ are the scaled empirical CDF using (15).

Then, in the second stage we first transform data into pseudo data for each $1 \leqslant m \leqslant M$, using $\hat{F}_{Y^{(m)}(1)|G,T}(y^{(m)} \mid 1,1)$ and $\hat{F}_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1)$, such that ,

$$\begin{aligned}
\hat{z}_i^{(m)}(1) &= \Phi_1^{-1}\left(\hat{F}_{Y^{(m)}(1)|G,T}(Y_i^{(m)} \mid 1,1)\right), \\
\hat{z}_i^{(m)}(0) &= \Phi_1^{-1}\left(\hat{F}_{Y^{(m)}(0)|G,T}(Y_i^{(m)} \mid 1,1)\right).
\end{aligned}$$

For clarity, we collect all the transformed pseudo data into vector, that is

$$\hat{\boldsymbol{z}}_i(1) = \left(\hat{z}_i^{(1)}(1), \ldots, \hat{z}_i^{(M)}(1)\right)^\top, \quad \hat{\boldsymbol{z}}_i(0) = \left(\hat{z}_i^{(1)}(0), \ldots, \hat{z}_i^{(M)}(0)\right)^\top. \tag{16}$$

Next in the second stage, with $\hat{\boldsymbol{z}}_i(1)$ and $\hat{\boldsymbol{z}}_i(0)$, we estimate $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ and $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ via MCMC respectively. The pseudo log-likelihood function, as shown in (14), suggests that the key to modeling the dependence structure of multiple outcomes lies in parameterizing the correlation matrix $\mathbf{R}(g,t)$. In contrast to the methods in Murray et al. (2013) and Klein, Kneib, and Lang (2015) that parameterize the covariance matrix through Cholesky decomposition, we use the method in Archakov and Hansen (2021) and Hansen and Luo (2025) for parameterizing the correlation matrix $\mathbf{R}(g,t)$. The main advantage of the method in Archakov and Hansen (2021) is that it establishes a one-to-one correspondence between a nonsingular correlation matrix and an unrestricted vector of matching dimension, which facilitates the construction of joint likelihood based on (14), which is necessary for posterior analysis. This parameterization has gained popularity in recent literature for its flexibility and strong theoretical grounding, see the application in Chen, Fei, and Yu (2025) for modeling a multivariate stochastic volatility model.

Given the structure of Gaussian copula we adopt, it would be of greater interests to conduct pairwise estimation, i.e. the case when $M = 2$. For this scenario, $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ degenerates to a 2×2 matrix

$$\begin{pmatrix} 1 & \rho_{\boldsymbol{Y}(1)}(1,1) \\ \rho_{\boldsymbol{Y}(1)}(1,1) & 1 \end{pmatrix}, \tag{17}$$

and $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ degenerates to a 2×2 matrix

$$\begin{pmatrix} 1 & \rho_{\boldsymbol{Y}(0)}(1,1) \\ \rho_{\boldsymbol{Y}(0)}(1,1) & 1 \end{pmatrix}, \tag{18}$$

then it is equivalently to estimate $\rho_{\boldsymbol{Y}(1)}(1,1)$ and $\rho_{\boldsymbol{Y}(0)}(1,1)$ using the Bayesian IFM method pairwise to arbitrary pair of multiple outcomes.[2] For this reason, we mainly focus on es-

[2]It is a well-known result that if a random vector is multivariate Gaussian, then any subvector (i.e., any

timating and making inference of (17) and (18) in the following discussion. To conduct Bayesian analysis, we choose a uniform prior for $\rho_{\boldsymbol{Y}(1)}(1,1)$ and $\rho_{\boldsymbol{Y}(0)}(1,1)$ with an independent structure, such that the corresponding density function is given by

$$\pi\left(\rho_{\boldsymbol{Y}(1)}(1,1)\right) = \mathbf{1}\left\{\rho_{\boldsymbol{Y}(1)}(1,1) \in (-1,1)\right\}, \quad \pi\left(\rho_{\boldsymbol{Y}(0)}(1,1)\right) = \mathbf{1}\left\{\rho_{\boldsymbol{Y}(0)}(1,1) \in (-1,1)\right\}.$$

As pointed in Archakov and Hansen (2021), the parameterization of 2×2 correlation matrix based on the matrix-logarithm is equivalent to applying the Fisher transformation on the off-diagonal elements. Accordingly,

$$\tilde{\rho}_{\boldsymbol{Y}(1)}(1,1) = \frac{1}{2}\ln\left(\frac{1+\rho_{\boldsymbol{Y}(1)}(1,1)}{1-\rho_{\boldsymbol{Y}(1)}(1,1)}\right) \in \mathbb{R}, \quad \tilde{\rho}_{\boldsymbol{Y}(0)}(1,1) = \frac{1}{2}\ln\left(\frac{1+\rho_{\boldsymbol{Y}(1)}(1,1)}{1-\rho_{\boldsymbol{Y}(0)}(1,1)}\right) \in \mathbb{R}.$$

Given the parameterization, the posterior of $\tilde{\rho}_{\boldsymbol{Y}(1)}(1,1)$ is

$$p\left(\tilde{\rho}_{\boldsymbol{Y}(1)}(1,1) \mid \{\hat{\boldsymbol{z}}_i(1)\}_{i=1}^{n(1)}\right) \propto \prod_{i=1}^{n(1)} p\left(\hat{\boldsymbol{z}}_i(1) \mid \tilde{\rho}_{\boldsymbol{Y}(1)}(1,1)\right) \times \left[1-\tanh^2(\tilde{\rho}_{\boldsymbol{Y}(1)}(1,1))\right], \quad (19)$$

with $n(1) = n_{1,1}$, and the posterior of $\tilde{\rho}_{\boldsymbol{Y}(0)}(1,1)$ is

$$p\left(\tilde{\rho}_{\boldsymbol{Y}(0)}(1,1) \mid \{\hat{\boldsymbol{z}}_i(0)\}_{i=1}^{n(0)}\right) \propto \prod_{i=1}^{n(0)} p\left(\hat{\boldsymbol{z}}_i(0) \mid \tilde{\rho}_{\boldsymbol{Y}(0)}(1,1)\right) \times \left[1-\tanh^2(\tilde{\rho}_{\boldsymbol{Y}(0)}(1,1))\right], \quad (20)$$

with $n(0) = n_{1,0} + n_{0,1} + n_{0,0}$. The $p\left(\hat{\boldsymbol{z}}_i(1) \mid \tilde{\rho}_{\boldsymbol{Y}(1)}(1,1)\right)$ and $p\left(\hat{\boldsymbol{z}}_i(0) \mid \tilde{\rho}_{\boldsymbol{Y}(0)}(1,1)\right)$ can be easily calculated via part I of (14), and we use MCMC to make posterior draws of $\tilde{\rho}_{\boldsymbol{Y}(1)}(1,1)$ and $\tilde{\rho}_{\boldsymbol{Y}(0)}(1,1)$, and accordingly the posterior draws of $\rho_{\boldsymbol{Y}(1)}(1,1)$ and $\rho_{\boldsymbol{Y}(0)}(1,1)$. To prevent poor mixing, we run an auxiliary optimization to identify the posterior mode and set the Gaussian Metropolis–Hastings random-walk proposal covariance to the negative inverse Hessian evaluated at the mode (Schorfheide, 2000). Estimation and inference of $\rho_{\boldsymbol{Y}(1)}(1,1)$ and $\rho_{\boldsymbol{Y}(0)}(1,1)$, which is our main target, can then be obtained via posterior sampling. The main advantage of Bayesian IFM is that it obviates the need to compute complex variance estimators for statistical inference, while retaining the key asymptotic properties of the maximum likelihood estimator, as we have mentioned previously. We demonstrate the effectiveness of MCMC sampling in Section 4.

selection of components) is also Gaussian. This result lays the foundation for estimating $\mathbf{R}_{\boldsymbol{Y}(1)}(1,1)$ and $\mathbf{R}_{\boldsymbol{Y}(0)}(1,1)$ in a pairwise manner, i.e., estimating $\rho_{\boldsymbol{Y}(1)}(1,1)$ and $\rho_{\boldsymbol{Y}(0)}(1,1)$.

4. Simulation

As we have discussed in the estimation procedure previously, the key is to obtain estimation of $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,0)$, $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,1)$, and $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 0,0)$ for each m and then identify $F_{Y^{(m)}(0)|G,T}(y^{(m)} \mid 1,1)$ using (5). For illustration purpose, we demonstrate the estimated dependence structure of $F_{\boldsymbol{Y}(1)|G,T}(\boldsymbol{y} \mid 1,1)$ (distribution of the treated group, observed) and $F_{\boldsymbol{Y}(0)|G,T}(\boldsymbol{y} \mid 1,1)$ (distribution of the non-treated group, counterfactual). That is, in the Monte Carlo experiment, we are interested in uncovering the ex post discrepancy in the dependence structure of multiple outcomes between the treated group and the control group after the treatment assignments.

We consider the data-generating process (DGP) as follows. We first generate $X_i^{(1)}$ and $X_i^{(2)}$ independently from uniform distribution on $[0,1]$ and then make $G_i = \mathbf{1}\{X_i > 0.5\}$ and $T_i = \mathbf{1}\{X_i > 0.5\}$. By this construction, we have $D_i = G_i T_i$ as the indicator specifying whether the i-th unit is exposed to the treatment after the implementation of treatment. For each unit i, we assume there are $M = 2$ observed outcomes, that is $\boldsymbol{Y}_i = \left(Y_i^{(1)}, Y_i^{(2)}\right)^\top$. Specifically, we have

$$Y_i^{(1)} = D_i \mu_{\text{treated}}^{(1)} + (1 - D_i)\mu_{\text{control}}^{(1)} + \varepsilon_i^{(1)}(D_i), \tag{21}$$

$$Y_i^{(2)} = D_i \mu_{\text{treated}}^{(2)} + (1 - D_i)\mu_{\text{control}}^{(2)} + \varepsilon_i^{(2)}(D_i) \tag{22}$$

where $\boldsymbol{\varepsilon}_i(D_i) = \left(\varepsilon_i^{(1)}(D_i), \varepsilon_i^{(2)}(D_i)\right)^\top$, $\boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(D_i))$. For the variance-covariance matrix $\boldsymbol{\Sigma}(D_i)$, we specify it as follows,

$$\boldsymbol{\Sigma}(D_i) = \begin{pmatrix} \Sigma_{11}(D_i) & \Sigma_{12}(D_i) \\ \Sigma_{21}(D_i) & \Sigma_{22}(D_i) \end{pmatrix} \tag{23}$$

and

$$\begin{aligned} \Sigma_{11}(D_i) &= \left(D_i \sigma_{\text{treated}}^{(1)} + (1 - D_i)\sigma_{\text{control}}^{(1)}\right)^2, \\ \Sigma_{12}(D_i) = \Sigma_{21}(D_i) &= D_i \rho_{\text{treated}} \sigma_{\text{treated}}^{(1)} \sigma_{\text{treated}}^{(2)} + (1 - D_i)\rho_{\text{control}} \sigma_{\text{control}}^{(1)} \sigma_{\text{control}}^{(2)}, \\ \Sigma_{22}(D_i) &= \left(D_i \sigma_{\text{treated}}^{(2)} + (1 - D_i)\sigma_{\text{control}}^{(2)}\right)^2. \end{aligned}$$

For the treated group we specify $\mu^{(1)}_{\text{treated}} = 3.10$, $\mu^{(2)}_{\text{treated}} = 0.18$, $\sigma^{(1)}_{\text{treated}} = 0.16$, $\sigma^{(2)}_{\text{treated}} = 0.19$, $\rho_{\text{treated}} = -0.55$; while for the controlled group, we specify $\mu^{(1)}_{\text{control}} = 3.87$, $\mu^{(2)}_{\text{control}} = 6.36$, $\sigma^{(1)}_{\text{control}} = 0.48$, $\sigma^{(2)}_{\text{control}} = 0.20$, $\rho_{\text{control}} = 0.42$. We use the posterior mean of the posterior sampling generated from Bayesian IFM as the estimation of ρ_{treated} and ρ_{control}.

In Table 2, we summarize the posterior mean estimation results for ρ_{treated} and ρ_{control}, denoted by $\bar{\rho}_{\text{treated}}$ and $\bar{\rho}_{\text{control}}$ respectively, with different sample size specifications ($n \equiv n(1)+n(0) = 500, 1000, 5000, 10000$) across 1000 Monte Carlo simulations. In each MCMC sampling procedure, we run $110,000$ iterations, discarding the initial $10,000$ draws (burn-in draws), and store remained sampling for every 10 draws (thinning draws). Namely, for each exercise of the Monte Carlo simulation, the posterior mean is based on a total of $10,000$ posterior samples. Our estimation procedure is efficiently implemented in a hybrid manner using both `R` and `C++`, supported by `Rcpp` (Eddelbuettel, 2013). We provide an R package, `multdr`, that implements all the main procedures. Since the MCMC sampling procedure, along with the auxiliary optimization, is efficiently programmed in `C++`, it can easily handle more demanding situations as the total number of MCMC iterations increases.

Table 2

	$n = 500$		$n = 1000$		$n = 5000$		$n = 10000$	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\bar{\rho}_{\text{treated}}$	-0.5500	0.0616	-0.5526	0.0445	-0.5511	0.0199	-0.5500	0.0142
$\bar{\rho}_{\text{control}}$	0.3072	0.0692	0.3578	0.0509	0.4077	0.0217	0.4141	0.0142

As suggested by Table 2, $\bar{\rho}_{\text{treated}}$ and $\bar{\rho}_{\text{control}}$ approaches the true value in probability as the sample size increases, which is consistent with the theory.

5. Empirical Application

In this section, we illustrate our approach by applying it to the data of Card and Krueger (1994), hereafter CK. The work of CK is influential in labor economics, not only for its empirical findings challenging the conventional view that minimum wage hikes harm employment but also for motivating subsequent research on causal inference methodologies, such as difference-in-differences (DiD) and synthetic controls. The main focus of CK is the policy implemented in April 1992, when New Jersey increased its minimum wage from the federal level of 4.25 to 5.05 per hour. There are two corresponding outcomes associated with this policy intervention: the number of full-time and part-time employees, respectively. By employing a difference-in-differences (DiD) approach, comparing outcomes before and after wage hikes across affected and unaffected states or regions, the authors of CK find that the increase in the minimum wage does not necessarily lead to a decrease in employment. As mentioned in the original work of CK, the authors focus on the aggregate outcome—namely, total employment measured as the full-time workers plus 0.5 times the number of part-time workers. In contrast, our approach centers on decomposing the aggregate outcome by explicitly modeling the dependence structure between full-time and part-time employment, with particular attention to how this dependence may be affected by the policy intervention. Since in this setting there are two outcomes, i.e. $M = 2$, we use ρ_{treated} ($\rho_{\boldsymbol{Y}(1)}(1,1)$ in (17)) and ρ_{control} ($\rho_{\boldsymbol{Y}(0)}(1,1)$ in (18)) to denote the dependence structure parameters. We estimate that, using our method, $\bar{\rho}_{\text{treated}} = -0.176$ and $\bar{\rho}_{\text{counterfactual}} = -0.1293$. In the MCMC sampling procedure, we run $110,000$ iterations, discarding the initial $10,000$ draws (burn-in draws), and store remained sampling for every 10 draws (thinning draws). $\bar{\rho}_{\text{treated}} = -0.176$ and $\bar{\rho}_{\text{counterfactual}} = -0.1293$ are calculated as the sample average using the retained $10,000$ posterior samples.

Figure 2 summarizes the main information of this estimation. This estimation result suggests that the minimum-wage increase magnified the negative dependence between

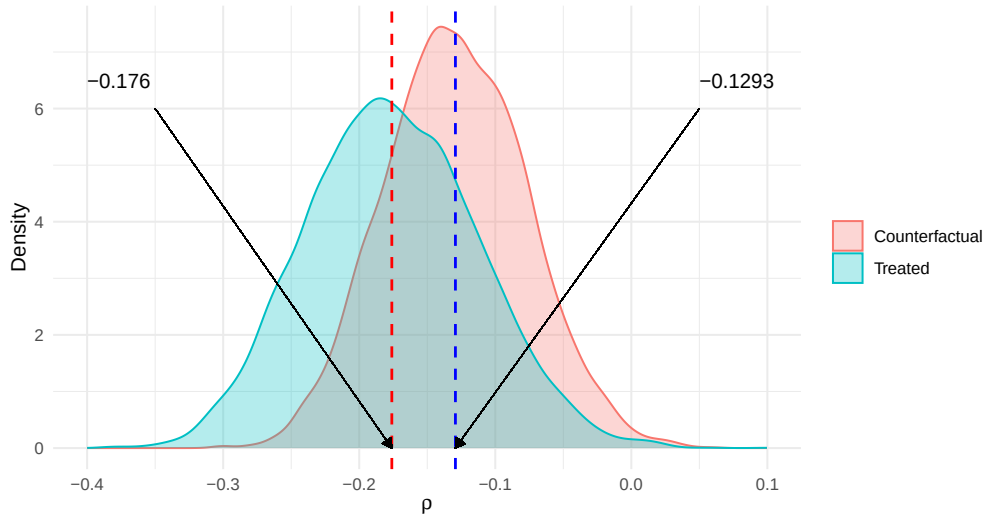
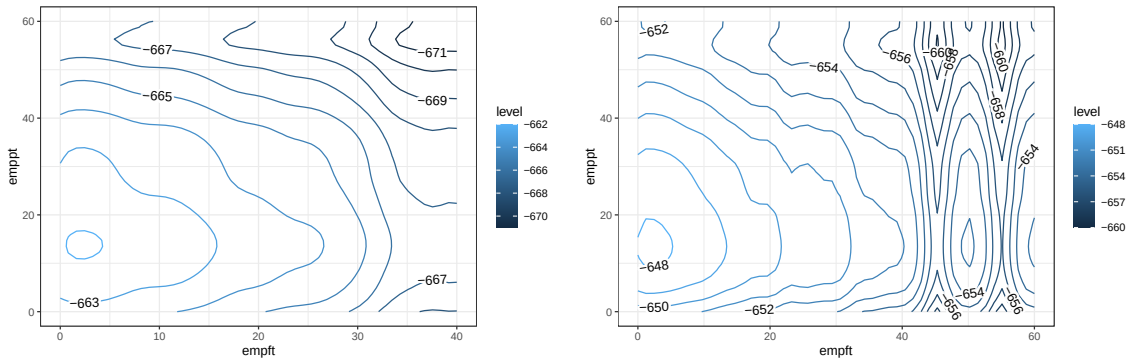


Figure 2: Posterior Distribution of ρ_{treated} and ρ_{control}.



(a) Treated **(b)** Counterfactual

Figure 3: Contour of Distribution

full-time and part-time employment compared with the counterfactual without the intervention. The slightly stronger negative correlation following the minimum-wage increase suggests a substitution effect between full-time and part-time employment, which may partly explain CK's finding that employment did not decline as expected. As a by-product, our approach can provide counterfactual joint distributions of multiple outcomes via copula link while keeping the flexibility of marginals. We demonstrate the contour (logarithm of (13)) of the joint distribution of full-time and part-time employment both in the pres-

ence of a policy intervention, in Figure 3(a), and in the counterfactual absence of such a minimum-wage policy, in Figure 3(b), respectively.

6. Conclusion

Building on recent developments in distributional regression, we develop a Bayesian approach for modeling dependence among multiple outcomes in a standard causal inference setting by integrating copula methods into the distributional regression framework. Our approach retains the key structure and advantages of distributional regression while providing more flexible Bayesian estimation and posterior inference for the effects of policy interventions on the dependence structure among multiple outcomes.

The modeling framework and methods we present in this paper are easy to implement with carefully chosen Metropolis–Hastings proposals. Inference based on posterior sampling obviates the need to compute complex variance estimators or to rely on bootstrap methods, making the approach more flexible. To demonstrate our approach, we revisit the study of Card and Krueger (1994), extending it to multiple outcomes — full-time and part-time employment — and modeling their dependence in a standard difference-in-differences (DiD) framework. We find that the minimum-wage increase slightly amplifies the substitution effect between full-time and part-time employment, relative to no policy intervention.

Our analysis can be readily extended to settings that incorporate covariates to model covariate-driven heterogeneity — for example, within the synthetic difference-in-differences framework of Arkhangelsky et al. (2021). Additionally, the model and method we propose currently handle continuous marginals, and therefore another meaningful extension would be accommodating mixed marginals including both continuous and discrete data. These extensions demand more advanced linking methodologies, which are of greater interests, and are left for future work.

References

Archakov, Ilya, and Peter Reinhard Hansen (2021): "A new parametrization of correlation matrices," *Econometrica*, 89 (4), 1699–1715.

Arellano, Manuel, and Stéphane Bonhomme (2017): "Quantile selection models with an application to understanding changes in wage inequality," *Econometrica*, 85 (1), 1–28.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021): "Synthetic difference-in-differences," *American Economic Review*, 111 (12), 4088–4118.

Ashenfelter, Orley, and David Card (1985): "Using the longitudinal structure of earnings to estimate the effect of training programs," *The Review of Economics and Statistics*, 67 (4), 648–660.

Athey, Susan (2025): "Presidential address: The economist as designer in the innovation process for socially impactful digital products," *American Economic Review*, 115 (4), 1059–99.

Breunig, Christoph, Ruixuan Liu, and Zhengfei Yu (2025): "Double robust bayesian inference on average treatment effects," *Econometrica*, 93 (2), 539–568.

Card, David, and Alan B. Krueger (1994): "Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania," *American Economic Review*, 84 (4), 772–793.

Chen, Han, Yijie Fei, and Jun Yu (2025): "Multivariate stochastic volatility models based on generalized fisher transformation," *Journal of Econometrics*, 251, 106041.

Chen, Songnian, Nianqing Liu, and Hanghui Zhang (2025): "Distribution regression with censored selection," *Journal of Econometrics*, 251, 106030.

Chen, Songnian, Nianqing Liu, Hanghui Zhang, and Yahong Zhou (2024): "Estimation of wage inequality in the uk by quantile regression with censored selection," *Journal of Econometrics*, 105733.

Chernozhukov, Victor, Iván Fernández-Val, Sukjin Han, and Kaspar Wüthrich (2025): "Estimating causal effects of discrete and continuous treatments with binary instruments," Manuscript.

Chernozhukov, Victor, Iván Fernández-Val, and Siyi Luo (2023): "Distribution regression with sample selection and UK wage decomposition," CeMMAP working papers 09/23, Institute for Fiscal Studies.

Eddelbuettel, Dirk (2013): *Seamless R and C++ Integration with Rcpp*. Use R! Series. Springer.

Fan, Yanqin, and Andrew J. Patton (2014): "Copulas in econometrics," *Annual Review of Economics*, 6 (1), 179–200.

Fernández-Val, Iván, Jonas Meier, Aico van Vuuren, and Francis Vella (2025): "Distribution regression difference-in-differences," Working paper, Boston University, Swiss National Bank, University of Groningen, and Georgetown University.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995): "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82 (3), 543–552.

Gunsilius, F. F. (2023): "Distributional synthetic controls," *Econometrica*, 91 (3), 1105–1117.

Hansen, Peter Reinhard, and Yiyao Luo (2025): "Robust estimation of realized correlation: New insight about intraday fluctuations in market betas," Manuscript.

Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo (2000): "Substitution and dropout bias in social experiments: A study of an influential social experiment*," *The Quarterly Journal of Economics*, 115 (2), 651–694.

Hoff, Peter D. (2007): "Extending the rank likelihood for semiparametric copula estimation," *The Annals of Applied Statistics*, 1 (1), 265–283.

Imai, Kosuke, and Michael Lingzhi Li (2023): "Experimental evaluation of individualized treatment rules," *Journal of the American Statistical Association*, 118 (541), 242–256.

Imbens, Guido W., and Donald B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge, Cambridge University Press.

Joe, Harry (2005): "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94 (2), 401–419.

——— (2014): *Dependence Modeling with Copulas.* Chapman and Hall/CRC.

Klein, Nadja, Thomas Kneib, and Stefan Lang (2015): "Bayesian generalized additive models for location, scale, and shape for zero-inflated and overdispersed count data," *Journal of the American Statistical Association*, 110 (509), 405–419.

Murray, Jared S., David B. Dunson, Lawrence Carin, and Joseph E. Lucas and (2013): "Bayesian gaussian copula factor models for mixed data," *Journal of the American Statistical Association*, 108 (502), 656–665.

Nelsen, Roger B. (2010): *An Introduction to Copulas.* Springer Publishing Company, Incorporated.

Pitt, Michael, David Chan, and Robert Kohn (2006): "Efficient bayesian inference for gaussian copula regression models," *Biometrika*, 93 (3), 537–554.

Ray, Kolyan, and Aad van der Vaart (2020): "Semiparametric bayesian causal inference," *The Annals of Statistics*, 48 (5).

Robins, James (1986): "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical Modelling*, 7 (9), 1393–1512.

Schorfheide, Frank (2000): "Loss function-based evaluation of dsge models," *Journal of Applied Econometrics*, 15 (6), 645–670.

Smith, Michael S., and Mohamad A. Khaled (2012): "Estimation of copula models with discrete margins via bayesian data augmentation," *Journal of the American Statistical Association*, 107 (497), 290–303.

Smith, Michael Stanley, and Nadja Klein (2021): "Bayesian inference for regression copulas," *Journal of Business & Economic Statistics*, 39 (3), 712–728.

Spady, Richard H., and Sami Stouli (2025): "Gaussian transforms modeling and the estimation of distributional regression functions," *Econometrica*, 93 (5), 1885–1913.

William, Torous, Florian Gunsilius, and Philippe Rigollet (2024): "An optimal transport approach to estimating causal effects via nonlinear difference-in-differences," *Journal of Causal Inference*, 12 (1), 1–26.

Appendix

A. Derivation of the pseudo-likelihood

Let $\Phi_M(\cdot; \mathbf{R}(g,t))$ denote the cumulative distribution function (CDF) of an M-dimensional multivariate Gaussian distribution with zero mean and $\mathbf{R}(g,t)$ as the covariance matrix, while $\Phi_1(\cdot)$ denotes the CDF of the univariate standard Gaussian distribution. Additionally, we denote $\phi_M(\cdot; \mathbf{R}(g,t))$ as the probability density function (PDF) of $\Phi_M(\cdot; \mathbf{R}(g,t))$ and $\phi_1(\cdot)$ as the PDF of $\Phi_1(\cdot)$. With this notation the characterization of joint distribution, we derive the pseudo-likelihood as

$$
\begin{aligned}
&\frac{\partial^M F_{\boldsymbol{Y}|G,T}(y^{(1)}, \ldots, y^{(M)} \mid g,t)}{\partial y^{(1)}, \ldots, \partial y^{(M)}} \\
&= \frac{\phi_M\left(\Phi_1^{-1}\left(F_{Y^{(1)}|G,T}\left(y^{(1)} \mid g,t\right)\right), \ldots, \Phi_1^{-1}\left(F_{Y^{(M)}|G,T}\left(y^{(M)} \mid g,t\right)\right); \mathbf{R}(g,t)\right)}{\prod_{m=1}^{M} \phi_1\left(\Phi_1^{-1}\left(F_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right)\right)\right)} \prod_{m=1}^{M} f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right) \\
&= \frac{\phi_M\left(z^{(1)}, \ldots, z^{(M)}; \mathbf{R}(g,t)\right)}{\prod_{m=1}^{M} \phi_1\left(z^{(m)}\right)} \prod_{m=1}^{M} f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right) \\
&= \det\left(\mathbf{R}(g,t)\right)^{-1/2} \frac{\exp\left\{-\frac{1}{2}\boldsymbol{z}^\top [\mathbf{R}(g,t)]^{-1} \boldsymbol{z}\right\}}{\exp\left(-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}\right)} \prod_{m=1}^{M} f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right) \\
&= \frac{1}{\sqrt{\det\left(\mathbf{R}(g,t)\right)}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^\top \left([\mathbf{R}(g,t)]^{-1} - \mathbf{I}_M\right) \boldsymbol{z}\right\} \prod_{m=1}^{M} f_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right). \qquad \text{(A.1)}
\end{aligned}
$$

If for each $1 \leqslant m \leqslant M$, $F_{Y^{(m)}|G,T}\left(y^{(m)} \mid g,t\right)$ is the just the CDF of standard univariate Gaussian distribution, then the likelihood function will degenerates to density function of multivariate normal distribution with variance-covariance as $\mathbf{R}\left(g,t\right)$.

B. MCMC Diagnosis

We report MCMC diagnostic plots only for the empirical applications, as the corresponding check plots for the simulations are quite similar. These diagnosis plots suggest that the MCMC procedure generates a mean-stationary posterior samples both for ρ_{treated} and $\rho_{\text{counterfactual}}$.