
Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

# Solving Estimating Equations With Copulas

Thomas Nagler & Thibault Vatter

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 546, 1168–1180: Theory and Methods
https://doi.org/10.1080/01621459.2023.2177545

Taylor & Francis
Taylor & Francis Group



# Solving Estimating Equations With Copulas

Thomas Nagler [a] and Thibault Vatter [b]

[a]Department of Statistics, LMU Munich, Munich, Germany; [b]Meta Platforms Inc., New York, NY

**ABSTRACT**

Thanks to their ability to capture complex dependence structures, copulas are frequently used to glue random variables into a joint model with arbitrary marginal distributions. More recently, they have been applied to solve statistical learning problems such as regression or classification. Framing such approaches as solutions of estimating equations, we generalize them in a unified framework. We can then obtain simultaneous, coherent inferences across multiple regression-like problems. We derive consistency, asymptotic normality, and validity of the bootstrap for corresponding estimators. The conditions allow for both continuous and discrete data as well as parametric, nonparametric, and semiparametric estimators of the copula and marginal distributions. The versatility of this methodology is illustrated by several theoretical examples, a simulation study, and an application to financial portfolio allocation. Supplementary materials for this article are available online.

**ARTICLE HISTORY**
Received May 2020
Accepted February 2023

**KEYWORDS**
Bootstrap; Nonparametric; Quantile; Regression; Statistical learning

## 1. Introduction

Any multivariate distribution is composed of marginal distributions and a *copula* characterizing dependence. Copula-based models combine complex dependencies with arbitrary marginal distributions and have become increasingly popular over the last two decades. They have also been applied to solve statistical learning problems like mean regression (Pitt, Chan, and Kohn 2006; Kolev and Paiva 2009; Noh, Ghouch, and Bouezmarni 2013; Cooke, Joe, and Chang 2015; Cai and Zhang 2018), quantile regression (Bouyé and Salmon 2009; Chen, Koenker, and Xiao 2009; Noh, Ghouch, and Van Keilegom 2015; Kraus and Czado 2017; Rémillard, Nasri, and Bouezmarni 2017), and classification (Elidan 2012; Han, Zhao, and Liu 2013; Nagler and Czado 2016; Carrera et al. 2019). And when parametric models fail (e.g., Dette, Van Hecke, and Volgushev 2014), semi/nonparametric approaches can be used (De Backer, El Ghouch, and Van Keilegom 2017; Schallhorn et al. 2017).

A criticism is that copula-based methods lead to overcomplicated inferential procedures and/or sub-optimal rates, as they use the joint distribution to extract features of the conditional distribution. In other words, they solve a problem that is harder than necessary. In this article, we show that this has a flip side: copula-based methods yield simultaneous and coherent inferences across arbitrary combinations of finite/infinite-dimensional and potentially constrained features of the conditional distribution.

Consider for instance a portfolio manager tasked with investing in $d$ assets conditionally on $p$ covariates representing the state of the economy. Denote by $\boldsymbol{Y} \in \mathbb{R}^d$, $\boldsymbol{\gamma} \in \mathbb{R}^d$, and $\boldsymbol{X} \in \mathbb{R}^p$ the returns on the assets, fractions of total wealth invested in each asset, and covariates. Quantitative portfolio management relies on properties of the distribution of $\boldsymbol{\gamma}^\top \boldsymbol{Y}$ conditional on $\boldsymbol{X} = \boldsymbol{x}$, like the expected return $\mu_{\boldsymbol{\gamma}}$, standard deviation $\sigma_{\boldsymbol{\gamma}}$, or a quantile $q_{\boldsymbol{\gamma}}$, as functions of $\boldsymbol{\gamma}$. Because $\mu_{\boldsymbol{\gamma}}$ and $\sigma_{\boldsymbol{\gamma}}$ are linear in the components of the conditional expectation vector and covariance matrix of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$, they only pose finite-dimensional problems. The conditional quantile $q_{\boldsymbol{\gamma}}$ is nonlinear in the portfolio weight $\boldsymbol{\gamma}$, however, and therefore, poses an infinite-dimensional problem. The approach in this article allows to construct estimators of $\{(\mu_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}, q_{\boldsymbol{\gamma}}) \colon \boldsymbol{\gamma} \in \mathbb{R}^d\}$ with asymptotics holding uniformly over portfolio weights. Conveniently, positive definiteness of the estimated covariance matrix and monotonicity of quantiles are automatically preserved.

Broadly, we propose a framework to solve a wide range of statistical learning problems using copulas. It is general enough to cover most types of regression, including mean, quantile, expectile, exponential family, or even instrumental variables and censored regression, as well as classification. Such problems can be characterized by estimating equations involving conditional expectations. Our approach builds on a key insight: conditional expectations can be replaced by weighted unconditional ones, and the weight is a ratio of measures associated with copulas.

In Section 2, we construct corresponding estimators in two steps: first estimate the copula, and then solve an approximate version of the estimating equation. Examples of compatible regression problems and estimators are given in Section 3. Given an estimated copula, all those problems can be solved simultaneously with coherent answers. We justify this approach by rigorous asymptotic theory in Section 4. We prove consistency, weak convergence, and validity of a bootstrap procedure for the proposed estimators under verifiable

**CONTACT** Thibault Vatter thibault.vatter@gmail.com Meta Platforms Inc., New York, NY.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

assumptions. Our asymptotic results are substantially more general than known results. In particular, we allow for virtually all types of regression problems, continuous and discrete variables, and for parametric, semiparametric, and nonparametric estimators in a single framework. In Section 5, we illustrate our method in simulated examples. We revisit the portfolio performance and risk management example described above using real data in Section 6. Section 7 relates our results to the literature and outlines further applications to more involved regression problems.

An implementation of the methods of this article is provided by the R package `eecop` (Nagler and Vatter 2020a). Proofs and additional results are in the supplementary material.

## 2. Copula-Based Estimating Equations

### 2.1. Estimating Equations for Regression Problems

For two random vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$, denote by $F_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{x}) = \Pr(\boldsymbol{Y} \leq \boldsymbol{y}, \boldsymbol{X} \leq \boldsymbol{x})$ their joint distribution, $F_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x}) = \Pr(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x})$ the distribution of $\boldsymbol{Y}$ conditional on $\boldsymbol{X} = \boldsymbol{x}$, $F_{Y_j}(y_j) = \Pr(Y_j \leq y_j)$ and $F_{X_j}(x_j) = \Pr(X_j \leq x_j)$ the marginal distributions.

Let $\boldsymbol{Y} \in \mathcal{Y} \subseteq \mathbb{R}^d$ be the *response* and $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ a vector of *covariates*. The response is often univariate in the context of regression or classification, but can also be vector-valued as in the asset allocation example from Section 1, or be enriched to encompass censoring indicators and instrumental variables (see Section 7).

Fix $\boldsymbol{x} \in \mathbb{R}^p$ and let the parameter of interest $\theta = \theta(\boldsymbol{x})$ be related to the conditional distribution $F_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$. Denoting the parameter space by $\Theta$ and $\theta^* \in \Theta$ the true parameter, suppose there is a family of functions $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ with the property

$$\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{x}\} = 0. \tag{1}$$

Most regression problems can be formulated that way (see Section 3.1). The set $\mathcal{G}$ is called a family of *identifying functions* and (1) the (population version of an) *estimating equation*. The name estimating equation stems from the fact that an estimator of $\theta^*$ can be constructed from solving a sample version of (1). Unconditional expectations have a canonical sample version in the sample average. Conditional expectations are more challenging, but can be replaced by unconditional ones since $\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{x}\} = \mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) w^*(\boldsymbol{Y})\}$ with

$$w^*(\boldsymbol{y}) = \frac{dF_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x})}{dF_{\boldsymbol{Y}}(\boldsymbol{y})} = \frac{dF_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{x})}{dF_{\boldsymbol{Y}}(\boldsymbol{y}) dF_{\boldsymbol{X}}(\boldsymbol{x})}, \tag{2}$$

which can be understood as a weight function that accounts for the conditioning on $\boldsymbol{X} = \boldsymbol{x}$. Hence, the estimating equation (1) can be written equivalently as

$$\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) w^*(\boldsymbol{Y})\} = 0, \tag{3}$$

which can be used to construct estimators. More generally, (3) holds for weights of the form

$$w^*(\boldsymbol{y}) = \frac{dF_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{x})}{dF_{\boldsymbol{Y}}(\boldsymbol{y}) dF_{\boldsymbol{X}}(\boldsymbol{x})} \nu(\boldsymbol{x}),$$

with $\nu$ an arbitrary function such that $0 < |\nu(\boldsymbol{x})| < \infty$. This can sometimes lead to useful simplifications as in (6).

### 2.2. Representation Using Copulas

Sklar's theorem (Sklar 1959) states that the joint distribution $F_{\boldsymbol{Z}}$ of any random vector $\boldsymbol{Z} \in \mathbb{R}^k$ can be represented as

$$F_{\boldsymbol{Z}}(\boldsymbol{z}) = C_{\boldsymbol{Z}}\{F_{Z_1}(z_1), \dots, F_{Z_k}(z_k)\}, \tag{4}$$

where the function $C_{\boldsymbol{Z}}$ is called a *copula*. The copula is a distribution function with uniform margins and unique on the ranges of $F_{Z_i}$, $i \in \{1, \dots, d\}$. Since, in what follows, it is evaluated solely on this set, potential nonuniqueness is not an issue. The weight in (2) can then be expressed as

$$w^*(\boldsymbol{y}) = \frac{dC_{\boldsymbol{Y},\boldsymbol{X}}\{F_{Y_1}(y_1), \dots, F_{Y_d}(y_d), F_{X_1}(x_1), \dots, F_{X_p}(x_p)\}}{dC_{\boldsymbol{Y}}\{F_{Y_1}(y_1), \dots, F_{Y_d}(y_d)) dC_{\boldsymbol{X}}\{F_{X_1}(x_1), \dots, F_{X_p}(x_p))\}}. \tag{5}$$

The measure $dC_{\boldsymbol{Z}}\{F_{Z_1}(z_1), \dots, F_{Z_k}(z_k)\}$ can be represented by a density with respect to the product of Lebesgue and counting measures for continuous and discrete variables, respectively. Suppose that $Z_1, \dots, Z_m$ are integer-valued and $Z_{m+1}, \dots, Z_k$ are continuous, which generalizes to categorical variables by identifying ordered categories with integers and unordered categories with binary dummy variables. Then

$$\begin{aligned} &dC_{\boldsymbol{Z}}\{F_{Z_1}(z_1), \dots, F_{Z_k}(z_k)\} \\ &= \sum_{\substack{(j_1,\dots,j_m)\in\{0,1\}^m \\ (j_{m+1},\dots,j_k)=0}} (-1)^{\sum_{r=1}^k j_r} \frac{\partial^{k-m} C_{\boldsymbol{Z}}\{F_{Z_1}(z_1 - j_1), \dots, F_{Z_k}(z_k - j_k)\}}{\partial z_{m+1} \cdots \partial z_k}. \end{aligned}$$

In many relevant cases, the expression (5) for $w^*$ can be simplified further. For instance, copula models are most commonly applied to continuous random vectors. If $\boldsymbol{Z}$ is a continuous random vector with joint density $f_{\boldsymbol{Z}}$ and marginal densities $f_{Z_k}$, $k = 1, \dots, d$, we can take derivatives in (4) to obtain

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = c_{\boldsymbol{Z}}\{F_{Z_1}(z_1), \dots, F_{Z_d}(z_d)\} \times \prod_{k=1}^d f_{Z_k}(z_k),$$

with $c_{\boldsymbol{Z}}$ the density corresponding to $C_{\boldsymbol{Z}}$. Hence, (1) is equivalent to (3) with

$$w^*(\boldsymbol{y}) = \frac{c_{\boldsymbol{Y},\boldsymbol{X}}\{F_{Y_1}(y_1), \dots, F_{Y_d}(y_d), F_{X_1}(x_1), \dots, F_{X_p}(x_p)\}}{c_{\boldsymbol{Y}}\{F_{Y_1}(y_1), \dots, F_{Y_d}(y_d))}, \tag{6}$$

where we dropped $c_{\boldsymbol{X}}$ as it does not depend on $\boldsymbol{y}$ or $\theta$. And because copulas have uniform marginals, $d = 1$ implies $c_Y \equiv 1$ and $w^*(y) = c_{Y,\boldsymbol{X}}\{F_Y(y), F_{X_1}(x_1), \dots, F_{X_p}(x_p)\}$.

### 2.3. Estimators for Copula-Based Estimating Equations

Suppose we observe an iid sequence of random vectors $(\boldsymbol{Y}_1, \boldsymbol{X}_1), \dots, (\boldsymbol{Y}_n, \boldsymbol{X}_n)$. We can use a sample version of (3) to construct estimators for the parameter $\theta^*$. To do this, all unknown quantities in the unconditional estimating equation are replaced by estimates.

Let $\widehat{w}(\boldsymbol{y}) = \widehat{w}(\boldsymbol{y}; \boldsymbol{Y}_1, \boldsymbol{X}_1, \dots, \boldsymbol{Y}_n, \boldsymbol{X}_n)$ be an estimator of $w^*$ (see, e.g., Section 3.2). Recall that $w^*$ is a ratio of measures associated with copulas. Hence, we can construct $\widehat{w}$ by plugging in estimators of the copulas and margins. This allows us to harness the rich toolbox of existing copula models and associated estimating techniques.

Then we define an estimator $\widehat{\theta} = \widehat{\theta}(\boldsymbol{Y}_1, \boldsymbol{X}_1, \ldots, \boldsymbol{Y}_n, \boldsymbol{X}_n)$ of $\theta^*$ as the solution to

$$\frac{1}{n}\sum_{i=1}^{n} g_{\widehat{\theta}}(\boldsymbol{Y}_i)\widehat{w}(\boldsymbol{Y}_i) = 0, \tag{7}$$

using the sample average as a natural estimate of the expectation in (3). For a few specific copula models, the expectation $\mathrm{E}\{g_{\widehat{\theta}}(\boldsymbol{Y})\widehat{w}(\boldsymbol{Y})\}$ might be available in closed form. Or it could be computed using numerical integration if, additionally, an estimate of the density $f_{\boldsymbol{Y}}$ is available. But the sample average provides a simpler and generally valid method.

## 3. Examples

The procedure described in Section 2.3 is quite versatile. With different choices of the identifying function $g_\theta$, we can estimate various features of the conditional distribution. In the following, we introduce popular examples covered by the theory developed in Section 4.

### 3.1. Examples of Identifying Functions

*Example 1 (Mean regression).* A classical example is $\theta^* = \mathrm{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})$ and $g_\theta(y) = y - \theta$. Given an estimator $\widehat{w}$ of the weight $w^*$, the estimating (7) has the explicit solution $\widehat{\theta} = \sum_{i=1}^n Y_i \widehat{w}(Y_i) / \sum_{i=1}^n \widehat{w}(Y_i)$. It is similar to the Nadaraya-Watson estimator for the conditional mean, albeit the weights are also functions of the response $Y$.

*Example 2 (Quantile regression).* Let $\theta_t^* = F_{Y|\boldsymbol{X}}^{-1}(t \mid \boldsymbol{X} = \boldsymbol{x})$ be the conditional $t$-quantile at level $t \in T = (0, 1)$ and consider all levels jointly. The parameter of interest is the conditional quantile function $\theta^* = \{\theta_t^* : t \in T\}$ and $\Theta$ is a space of functions from $T$ to $\mathbb{R}$. Then solving (3) with $g_{\theta,t}(y) = t - \mathbb{1}(y < \theta_t)$ identifies $\theta^*$. Here, the identifying function $g_\theta = \{g_{\theta,t} : t \in T\}$ is also indexed by the quantile level $t$.

*Example 3 (Expectile regression).* Expectiles generalize the mean of a distribution similarly as to how quantiles generalize the median (e.g., Newey and Powell 1987). They are also indexed by a level $t \in (0, 1)$, where $t = 1/2$ corresponds to mean regression. The conditional expectile is identified by (3) with $g_{\theta,t}(y) = t(y - \theta_t)\mathbb{1}(y \geq \theta_t) - (1 - t)(\theta_t - y)\mathbb{1}(y < \theta_t)$.

*Example 4 (Exponential family regression).* Suppose $f_{Y|\boldsymbol{X}=\boldsymbol{x}}$ is a one-parameter exponential family with canonical parameter $\theta$, that is $f(y; \theta) = h(y) \exp\{a(y)\theta - b(\theta)\}$ where $h$, $a$, and $b$ are known functions. Using the score equations, $\theta$ can be identified via $g_\theta(y) = a(y) - b'(\theta)$.

*Example 5 (Binary classification).* Let $Y \sim \mathrm{Bernoulli}(p)$ be a class indicator with the target being the conditional probability $\theta^* = P(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) = \mathrm{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})$. Since $F_{\boldsymbol{X}|Y}(\boldsymbol{x} \mid y) = F_{\boldsymbol{X}|Y=0}(\boldsymbol{x})\mathbb{1}(y = 0) + F_{\boldsymbol{X}|Y=1}(\boldsymbol{x})\mathbb{1}(y = 1)$, Bayes' rule leads to

$$w^*(y) = \frac{dF_{\boldsymbol{X}|Y}(\boldsymbol{x} \mid y)}{dF_{\boldsymbol{X}}(\boldsymbol{x})}$$

$$= \frac{dF_{\boldsymbol{X}|Y=0}(\boldsymbol{x})\mathbb{1}(y = 0) + dF_{\boldsymbol{X}|Y=1}(\boldsymbol{x})\mathbb{1}(y = 1)}{(1 - p)dF_{\boldsymbol{X}|Y=0}(\boldsymbol{x}) + p dF_{\boldsymbol{X}|Y=0}(\boldsymbol{x})}.$$

Replacing all quantities by estimates and solving (7) with $g_\theta(y) = y - \theta$ yields

$$\widehat{\theta} = \frac{\widehat{p}\widehat{dF}_{\boldsymbol{X}|Y=1}(\boldsymbol{x})}{(1 - \widehat{p})\widehat{dF}_{\boldsymbol{X}|Y=0}(\boldsymbol{x}) + \widehat{p}\widehat{dF}_{\boldsymbol{X}|Y=1}(\boldsymbol{x})},$$

where $\widehat{p} = n^{-1}\sum_{i=1} \mathbb{1}(Y_i = 1)$. Modeling $dF_{\boldsymbol{X}|Y=0}$ and $dF_{\boldsymbol{X}|Y=1}$ with copulas and marginal distributions, such classifiers have been used in Elidan (2012), Nagler and Czado (2016), and Carrera et al. (2019), but so far without asymptotic guarantees.

### 3.2. Estimators for the Weight Function

We now discuss simple estimators of $w^*$, focusing on continuous data for simplicity's sake.

*Weight Estimator 1 (Fully parametric estimator).* Let $\boldsymbol{\eta}_{\boldsymbol{Y}}$, $\boldsymbol{\eta}_{\boldsymbol{X}}$, and $\boldsymbol{\eta}_C$ be parameter vectors, indexing families of marginal and copula densities, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_{\boldsymbol{Y}}, \boldsymbol{\eta}_{\boldsymbol{X}}, \boldsymbol{\eta}_C)$. This defines a parametric model for the weight (6):

$$w(\boldsymbol{y}; \boldsymbol{\eta}) = \frac{\begin{aligned} c_{\boldsymbol{Y},\boldsymbol{X}}\{&F_{Y_1}(y_1; \boldsymbol{\eta}_{Y_1}), \ldots, F_{Y_d}(y_d; \boldsymbol{\eta}_{Y_d}), \\ &F_{X_1}(x_1; \boldsymbol{\eta}_{X_1}), \ldots, F_{X_d}(x_p; \boldsymbol{\eta}_{X_p}); \boldsymbol{\eta}_C\}\end{aligned}}{c_{\boldsymbol{Y}}\{F_{Y_1}(y_1; \boldsymbol{\eta}_{Y_1}), \ldots, F_{Y_d}(y_d; \boldsymbol{\eta}_{Y_d}); \boldsymbol{\eta}_C\}}.$$

If $\widehat{\boldsymbol{\eta}}$ is an estimator for the true parameter $\boldsymbol{\eta}^*$, such as a maximum-likelihood or method of moment estimator, the estimated weight is then simply $\widehat{w}(\boldsymbol{y}) = w(\boldsymbol{y}; \widehat{\boldsymbol{\eta}})$.

*Weight Estimator 2 (Semiparametric estimator).* Semiparametric models combine a parametric copula density $c_{\boldsymbol{Y},\boldsymbol{X}}(\cdot; \boldsymbol{\eta})$ with nonparametrically estimated margins. Denote by $\widehat{\boldsymbol{\eta}}$ an estimator of the true parameter $\boldsymbol{\eta}^*$. Examples are the pseudo-maximum-likelihood estimator of Genest, Ghoudi, and Rivest (1995) or the method-of-moment type estimators discussed in Tsukahara (2005). A semiparametric estimator for the weight (6) is then given by

$$\widehat{w}(\boldsymbol{y}) = \frac{c_{\boldsymbol{Y},\boldsymbol{X}}\{\widehat{F}_{Y_1}(y_1), \ldots, \widehat{F}_{Y_d}(y_d), \widehat{F}_{X_1}(x_1), \ldots, \widehat{F}_{X_p}(x_p); \widehat{\boldsymbol{\eta}}\}}{c_{\boldsymbol{Y}}\{\widehat{F}_{Y_1}(y_1), \ldots, \widehat{F}_{Y_d}(y_d); \widehat{\boldsymbol{\eta}}\}},$$

where $\widehat{F}$ denotes the empirical distribution function.

*Weight Estimator 3 (Simple kernel estimator).* Observe that taking

$$w^*(\boldsymbol{y}) = \frac{c_{\boldsymbol{Y},\boldsymbol{X}}\{F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d), F_{X_1}(x_1), \ldots, F_{X_p}(x_p)\}}{c_{\boldsymbol{Y}}\{F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)\}} \times \prod_{j=1}^{p} f_{X_k}(x_k) = \frac{f_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x})}{f_{\boldsymbol{Y}}(\boldsymbol{y})}$$

as weight function solves (3). Then a natural estimator is $\widehat{w}(\boldsymbol{y}) = \widehat{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x}) / \widehat{f}_{\boldsymbol{Y}}(\boldsymbol{y})$, where $\widehat{f}_{\boldsymbol{Y},\boldsymbol{X}}$ and $\widehat{f}_{\boldsymbol{Y}}$ are estimators of $f_{\boldsymbol{Y},\boldsymbol{X}}$ and $f_{\boldsymbol{Y}}$. As an example, consider kernel density estimators (KDEs). For some univariate probability density $K : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and bandwidth sequences $b_n, \sigma_n \to 0$, define the KDE for $f_{\boldsymbol{Y},\boldsymbol{X}}$ as

$$\widehat{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{n b_n^d \sigma_n^p} \sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{Y_{i,j} - y_j}{b_n}\right) \prod_{k=1}^{p} K\left(\frac{X_{i,k} - x_k}{\sigma_n}\right).$$

Its margin $\widehat{f}_{\boldsymbol{Y}}(\boldsymbol{y}) = \int \widehat{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{x})d\boldsymbol{x} = n^{-1}b_n^{-d}\sum_{i=1}^{n}\prod_{j=1}^{d} K\{(Y_{i,j}-y_j)/b_n\}$ is also a KDE.

## 4. Asymptotic Theory

In this section, we derive some asymptotic properties of $\widehat{\theta}$. Section 4.1 introduces required notations. For our main results in Section 4.2, we use a general framework to encompass a wide range of identifying functions and weight estimators. Then, Section 4.3 contains specialized results corresponding to the parametric, semiparametric, and nonparametric weight estimators of Section 3.2. The assumptions are stated and discussed in Appendix A. The proofs and additional verifications of the assumptions for the examples from Section 3.1 are in the supplementary material.

### 4.1. Setup and Notation

We use $\to_P$ and $o_P(a_n)$ for convergence in probability without rate and with rate $a_n$, and $\rightsquigarrow$ for weak convergence. For an arbitrary set $T$, denote the space of all bounded functions from $T$ to $\mathbb{R}$ by $\ell^\infty(T) = \{f\colon T \to \mathbb{R}, \|f\|_T < \infty\}$, with $\|f\|_T = \sup_{t\in T}|f(t)|$.

Assume that the parameter space satisfies $\Theta \subseteq \ell^\infty(T)$ for some indexing set $T$. For simplicity, assume $T$ to be a compact subset of a Euclidean space. It means that any $\theta \in \Theta$ and $g_\theta \in \mathcal{G} = \{g_\theta\colon \theta \in \Theta\}$ is indexed by $T$, that is $\theta = \{\theta_t\colon t \in T\}$ and $g_\theta = \{g_{\theta,t}\colon t \in T\}$. In particular, this holds for the estimator $\widehat{\theta}$, true parameter $\theta^*$, and the corresponding $g_{\widehat{\theta}}$ and $g_{\theta^*}$. For scalar parameters of interest, we may take $T = \{1\}$, implying that $\Theta$ is isometric to $\mathbb{R}$, and write $\theta \in \Theta \subseteq \mathbb{R}$ by slight abuse of notation. As for vectors of dimension $k$, one can use $T = \{1,\ldots,k\}$ with $\theta_t$ referring to the corresponding vector's $t$th component, translating into $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ and $\|\boldsymbol{\theta}\|_T = \max_{t=1,\ldots,K}|\theta_t|$ by the same abuse of notation. An example of the general case, where the parameter of interest is a genuine function from $T$ to $\mathbb{R}$, is quantile regression (see Example 2). Here $T = (0, 1)$ is a natural choice with $\Theta$ being a space of functions from $(0, 1)$ to $\mathbb{R}$ and $\theta \in \Theta$ being indexed by the quantile level.

Recall that $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ and $\theta^* = \theta^*(\boldsymbol{x})$ are also functions of the value conditioned upon through $\boldsymbol{X} = \boldsymbol{x}$. This is reflected in their definitions (7) and (3) through $\widehat{w}$ and $w^*$. But $\widehat{\theta}$ as a function of $\boldsymbol{x}$ cannot have a meaningful limit in many cases of practical interest. Hence, we assume $\boldsymbol{x}$ fixed for the remainder of this section and simply write $\widehat{\theta}$ and $\theta^*$.

### 4.2. Main Results

Our first result shows that $\widehat{\theta}$ is consistent, uniformly in the indexing set $T$.

*Theorem 1 (Consistency).* Under Assumptions 1–6, $\|\widehat{\theta} - \theta^*\|_T \to_P 0$ as $n \to \infty$.

If the parameter of interest is scalar, vector, or matrix valued, uniform consistency is the usual consistency, due to the problem's finite dimensional nature. But Theorem 1 is insufficient for statistical inference. For instance, to test hypotheses and construct confidence bands, an asymptotic distribution is needed. And to establish weak convergence, we need to specify the estimator $\widehat{w}$ further.

In what follows, we assume that $\widehat{w}$ is asymptotically linear in the sense that there is a sequence of functions $w_n\colon \mathbb{R}^{2d+p} \to \mathbb{R}$ and a sequence $r_n^{-1} = o(n^{1/2})$ such that

$$\widehat{w}(\boldsymbol{y}) \approx \frac{1}{n}\sum_{i=1}^{n} w_n(\boldsymbol{y}, \boldsymbol{Y}_i, \boldsymbol{X}_i) + o_P\big(n^{-1/2}r_n\big), \tag{8}$$

in a sense that is made precise by Assumption 3. This assumption is satisfied by many estimators of $w^*$ including those from Section 3.2, see Section 4.3. The rate $r_n$ allows to encompass both parametric and nonparametric estimators of $w^*$. While $r_n$ is a diverging sequence for nonparametric estimators, $r_n = 1$ gives the standard $\sqrt{n}$ rate for parametric ones.

*Theorem 2 (Weak convergence).* Under Assumptions 1–8,

$$r_n^{-1}\sqrt{n}(\widehat{\theta} - \theta^* - \beta_n) \rightsquigarrow -V_{\theta^*}^{-1}\mathbb{G} \text{ in } \ell^\infty(T),$$

where $V_{\theta^*}$ is the Fréchet derivative of the map $\theta \mapsto \mathrm{E}\big\{g_\theta(\boldsymbol{Y})w^*(\boldsymbol{Y})\big\}$ taken at $\theta^*$,

$$\beta_n = -V_{\theta^*}^{-1}\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}')\}$$

for $(\boldsymbol{Y}', \boldsymbol{X}')$ an independent copy of $(\boldsymbol{Y}, \boldsymbol{X})$, and $\mathbb{G}$ is a tight, mean-zero Gaussian with

$$\mathrm{cov}(\mathbb{G}_{t_1}, \mathbb{G}_{t_2}) = \lim_{n\to\infty} r_n^{-2}\mathrm{E}\{h_{n,t_1}(\boldsymbol{Y}', \boldsymbol{X}')h_{n,t_2}(\boldsymbol{Y}', \boldsymbol{X}')\},$$

where $h_{n,t}(\boldsymbol{y}', \boldsymbol{x}') = \mathrm{E}\{g_{\theta^*,t}(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{y}', \boldsymbol{x}') + g_{\theta^*,t}(\boldsymbol{y}')w_n(\boldsymbol{y}', \boldsymbol{Y}, \boldsymbol{X})\}$ for $(\boldsymbol{y}', \boldsymbol{x}') \in \mathcal{Y} \times \mathcal{X}$.

While this result is functional, it can be understood in the usual sense for scalar and vector valued parameters of interest. For scalars, we can take $T = \{1\}$, so $\mathbb{G}$ is mean-zero univariate Gaussian. For $k$-dimensional vectors, using $T = \{1,\ldots,k\}$, $V_{\theta^*}$ is the derivative of a map from $\mathbb{R}^k$ to $\mathbb{R}^k$, that is a $k \times k$ matrix. And $\mathbb{G}$ is a mean-zero $k$-dimensional Gaussian with covariance matrix with entries $\mathrm{cov}(\mathbb{G}_i, \mathbb{G}_j)$ for $1 \le i, j \le k$.

To better understand Theorem 2, recall that $\widehat{\theta}$ is a plug-in type estimator. Its asymptotic distribution combines the effects of two steps: replacing (i) $w^*$ with $\widehat{w}$, and (ii) a population expectation with a sample average. Since (ii) is unbiased, the bias $\beta_n$ is driven by the bias of $\widehat{w}$. This becomes obvious when writing

$$\beta_n = -V_{\theta^*}^{-1}\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}')\} = -V_{\theta^*}^{-1}\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y})b_n(\boldsymbol{Y})\},$$

with $b_n(\boldsymbol{y}) = \mathrm{E}[w_n(\boldsymbol{y}, \boldsymbol{Y}', \boldsymbol{X}')] - w^*(\boldsymbol{y})$ the first-order term in the bias of $\widehat{w}$. Hence, $\beta_n$ is proportional to a weighted average of the bias of $\widehat{w}$, where the averaging accounts for the second step. For interpreting the variance, assume that $b_n(\boldsymbol{y}) = o(1)$ such that

$$h_{n,t}(\boldsymbol{y}', \boldsymbol{x}') \approx \mathrm{E}\{g_{\theta^*,t}(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{y}', \boldsymbol{x}')\} + g_{\theta^*,t}(\boldsymbol{y}')w^*(\boldsymbol{y}').$$

Then

$$\begin{aligned}\mathrm{var}\{h_{n,t}(\boldsymbol{Y}', \boldsymbol{X}')\} \approx{}& \mathrm{var}\big[\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}') \mid \boldsymbol{Y}', \boldsymbol{X}'\}\big] \\ &+ \mathrm{var}\{g_{\theta^*,t}(\boldsymbol{Y})w^*(\boldsymbol{Y})\} + 2\big[\mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) \\ &\quad w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}') \mid \boldsymbol{Y}', \boldsymbol{X}'\}g_{\theta^*,t}(\boldsymbol{Y}')w^*(\boldsymbol{Y}')\big].\end{aligned}$$

The first and second terms, respectively, reflect the variability caused by steps (i) and (ii). And the third echoes the dependence between both steps, since the same data is used twice.

Theorem 2 allows to compute the limiting distribution for specific choices of identifying functions and weight estimators. But such computations can be rather involved in practice, especially for complex $\widehat{w}$. To remedy this issue, we propose a *bootstrap* method. The idea is to define a new estimator $\widetilde{\theta}$ based on a randomly reweighted version of the data. The distribution of $\sqrt{n}(\widehat{\theta} - \theta^*)$ is then approximated by that of $\sqrt{n}(\widetilde{\theta} - \widehat{\theta})$.

Specifically, let $\xi_1, \dots, \xi_n$ be an iid sequence of positive random variables independent of the data and satisfying $\mathrm{E}(\xi_1) = \mathrm{var}(\xi_1) = 1$, $\mathrm{E}(|\xi_1|^{2+\epsilon}) < \infty$ for some $\epsilon > 0$. For instance, $\xi_i \sim \mathrm{Exp}(1)$ is the *Bayesian bootstrap* of Rubin (1981). Define the bootstrap estimator $\widetilde{\theta} = \{\widetilde{\theta}_t : t \in T\}$ as solving

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i g_{\widetilde{\theta},t}(\boldsymbol{Y}_i)\widetilde{w}(\boldsymbol{Y}_i) = 0,$$

where the bootstrapped weight $\widetilde{w}$ is constructed so that

$$\widetilde{w}(\boldsymbol{y}) \approx \frac{1}{n}\sum_{i=1}^{n} \xi_i w_n(\boldsymbol{y}, \boldsymbol{Y}_i, \boldsymbol{X}_i) + o_p(n^{-1/2} r_n), \tag{9}$$

in a sense that is made precise by Assumption 9. In Section 4.3, we explain how such bootstrapped weights are constructed for the weight estimators (1)–(3).

Our final theorem implies that $r_n^{-1}\sqrt{n}(\widetilde{\theta} - \widehat{\theta})$ converges to the same limit as in Theorem 2. See also Bücher and Kojadinovic (2019) for equivalent formulations of bootstrap validity.

*Theorem 3 (Validity of the bootstrap).* Under Assumptions 1–9,

$$r_n^{-1}\sqrt{n}(\widehat{\theta} - \theta^* - \beta_n, \widetilde{\theta} - \widehat{\theta}) \rightsquigarrow -(V_{\theta^*}^{-1}\mathbb{G}, V_{\theta^*}^{-1}\widetilde{\mathbb{G}}) \quad \text{in } \ell^\infty(T) \times \ell^\infty(T),$$

where $\widetilde{\mathbb{G}}$ is an independent copy of $\mathbb{G}$ and $V_{\theta^*}, \mathbb{G}$ are as in Theorem 2.

*Remark 1.* The resampling technique of Efron (1979) uses dependent bootstrap weights $(\xi_1, \dots, \xi_n) \sim \mathrm{Multinomial}(n, 1/n, \dots, 1/n)$. Our formulation simplifies the asymptotic analysis and is rather natural in the context of estimating equations.

### 4.3. Examples Continued

Theorems 1–3 are general enough to cover a broad range of estimation methods and regression problems. In particular, they apply to all the examples given in Section 3. In this section, we give three corollaries corresponding to the weight estimators 1 to 3.

The parametric weight estimator 1 is defined as $\widehat{w}(\cdot) = w(\cdot; \widehat{\boldsymbol{\eta}})$, where $\widehat{\boldsymbol{\eta}}$ is the estimated model parameter. Assume that $\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = n^{-1}\sum_{i=1}^{n} \boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i) + o_P(n^{-1/2})$ for some function $\boldsymbol{\gamma}$ with $\mathrm{E}\{\boldsymbol{\gamma}(\boldsymbol{Y}, \boldsymbol{X})\} = 0$, which is satisfied for both maximum likelihood and method of moments estimators under some regularity conditions. If $w$ is sufficiently smooth, we get

$$\widehat{w}(\boldsymbol{y}) \approx w(\boldsymbol{y}; \boldsymbol{\eta}^*) + \frac{1}{n}\sum_{i=1}^{n} \nabla_{\boldsymbol{\eta}}^{\top} w(\boldsymbol{y}; \boldsymbol{\eta}^*)\boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i) + o_p(n^{-1/2}).$$

In other words, (8) holds with $r_n = 1$ and $w_n(\boldsymbol{y}, \boldsymbol{Y}_i, \boldsymbol{X}_i) = w(\boldsymbol{y}; \boldsymbol{\eta}^*) + \nabla_{\boldsymbol{\eta}}^{\top} w(\boldsymbol{y}; \boldsymbol{\eta}^*)\boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i)$. Furthermore, (9) holds with the same $r_n$ and $w_n$ if the bootstrap estimator $\widetilde{\boldsymbol{\eta}}$ satisfies $\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = n^{-1}\sum_{i=1}^{n} \xi_i \boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i) + o_P(n^{-1/2})$. For example, if $\widehat{\boldsymbol{\eta}}$ is the maximum likelihood estimator, we may take $\widetilde{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \sum_{i=1}^{n} \xi_i \log f_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\eta})$.

*Corollary 1 (Fully parametric estimator).* Under Assumption 1, 2, 8, and 10, the Weight estimator 1 satisfies the conditions of Theorems 1–3. The bias and covariance are given by $\beta_n = 0$ and $\mathrm{cov}(\mathbb{G}_{t_1}, \mathbb{G}_{t_1}) = \mathrm{cov}(Z_{t_1}, Z_{t_2})$, where

$$Z_t = g_{\theta^*,t}(\boldsymbol{Y})w(\boldsymbol{Y}; \boldsymbol{\eta}^*) + \nabla_{\boldsymbol{\eta}^*}^{\top}\mathrm{E}\{g_{\theta^*,t}(\boldsymbol{Y})w(\boldsymbol{Y}; \boldsymbol{\eta}^*)\}\boldsymbol{\gamma}(\boldsymbol{Y}, \boldsymbol{X}).$$

A similar result was obtained by Rémillard, Nasri, and Bouezmarni (2017) for the weak convergence of conditional quantile estimators. For the semiparametric weight estimator 2, we can proceed similarly, but the resulting variance is larger due to nonparametric margins estimation.

*Corollary 2 (Semiparametric estimator).* Under Assumption 1, 2, 8, and 11, the weight estimator 2 satisfies the conditions of Theorem 1–3. The bias and covariance are given by $\beta_n = 0$ and $\mathrm{cov}(\mathbb{G}_{t_1}, \mathbb{G}_{t_1}) = \mathrm{cov}(Z_{t_1}, Z_{t_2})$, where

$$Z_t = g_{\theta^*,t}(\boldsymbol{Y})w(\boldsymbol{Y}; \boldsymbol{\eta}^*) + \nabla_{\boldsymbol{\eta}^*}^{\top}\mathrm{E}\{g_{\theta^*,t}(\boldsymbol{Y})w(\boldsymbol{Y}; \boldsymbol{\eta}^*)\}\boldsymbol{\gamma}(\boldsymbol{Y}, \boldsymbol{X}) + \lambda(\boldsymbol{Y}, \boldsymbol{X}),$$

and $\lambda(\boldsymbol{Y}, \boldsymbol{X})$ defined in (11).

Asymptotic normality of this estimator for the specific cases of mean and quantile regression was previously established by Noh, Ghouch, and Bouezmarni (2013), Noh, Ghouch, and Van Keilegom (2015). Rémillard, Nasri, and Bouezmarni (2017) extended the latter to weak convergence uniformly in the quantile level. Although Noh, Ghouch, and Bouezmarni (2013), Noh, Ghouch, and Van Keilegom (2015), and Rémillard, Nasri, and Bouezmarni (2017) are framed differently, they operate under regularity conditions similar to ours, and lead to the same expression for the asymptotic (co)variances. Our corollary extend them by allowing for almost arbitrary regression problems with potentially multivariate responses. A further generalization that allows for discrete variables can be obtained similarly from our main theorems.

Lastly, we can verify the conditions of the theorems for the nonparametric weight estimator (3) and its bootstrapped version $\widetilde{w}(\boldsymbol{y}) = \widetilde{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x}) / \int \widetilde{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x})d\boldsymbol{x}$, where

$$\widetilde{f}_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x}) = \frac{1}{nb_n^d\sigma_n^p}\sum_{i=1}^{n} \xi_i \prod_{j=1}^{d} K\left(\frac{Y_{i,j} - y_j}{b_n}\right)\prod_{k=1}^{p} K\left(\frac{X_{i,k} - x_k}{\sigma_n}\right).$$

*Corollary 3 (Simple kernel estimator).* Under Assumption 1, 2, 8, and 12 with $r_n = \sigma_n^{-p/2}$, the weight estimator 3 satisfies the conditions of Theorem 1–3. The bias satisfies

$$\beta_n = -\frac{\sigma_n^2 \int s^2 K(s)ds}{2} V_{\theta^*}^{-1} \sum_{k=1}^{p} \left[\partial_{x_k}^2 \psi(\boldsymbol{x}) \times f_{\boldsymbol{X}}(\boldsymbol{x}) + 2\partial_{x_k}\psi(\boldsymbol{x}) \times \partial_{x_k} f_{\boldsymbol{X}}(\boldsymbol{x})\right] + o(s_n),$$

where $s_n = n^{-1/2}\sigma_n^{-p/2} + \sigma_n^2$ and $\psi(\boldsymbol{x}) = \mathrm{E}\{g_{\theta^*}(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{x}\}$. If, additionally, $g_{\theta^*,t}$ is continuous almost everywhere for all $t \in T$, we have

$$\mathrm{cov}(\mathbb{G}_{t_1}, \mathbb{G}_{t_1}) = \left\{\int K(s)^2 ds\right\}^p \mathrm{E}\{g_{\theta^*,t_1}(\boldsymbol{Y}) g_{\theta^*,t_2}(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{x}\} f_X(\boldsymbol{x}).$$

This estimator is biased, which is expected from kernel based methods. In the case of mean-regression (i.e., $g_\theta(y) = y - \theta$), one can verify that the bias and variances are asymptotically equivalent to those of the Nadaraya-Watson estimator (e.g., Fan and Gijbels 1996, Theorem 3.1). The simple kernel estimator therefore has no advantages over traditional kernel methods and should only be seen as an illustrative example.

Benefits of the copula-based approach can be expected when imposing more structure on the copula model. For example, Nagler and Czado (2016) showed that simplified vine copulas evade the curse of dimensionality prevalent in nonparametric estimation. Similar effects can be expected with other hierarchical models, like nested Archimedean copulas (see e.g., Okhrin, Okhrin, and Schmid 2013) or the aggregation copula model of Côté and Genest (2015). This is confirmed empirically in the following section.

## 5. Numerical Validation

The proposed methodology subsumes a range of problems and estimation methods too wide to be exhaustively covered here. Instead, we validate our theoretical results in simple settings. We also provide a brief comparison to some benchmark methods.

We simulate from a linear Gaussian model $Y = \boldsymbol{\beta X} + Z$, with $\boldsymbol{X} \in \mathbb{R}^p$ a vector of iid $X \sim N(0, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}_{i,j} = \mathbb{1}(i = j) + 0.3\mathbb{1}(i \neq j)$, and $Z \sim N(0, 1)$ independent of $X$. We set $\boldsymbol{\beta} = (1, \dots, 1)^\top/\sqrt{p}$ to ensure that the signal-to-noise ratio is unaffected by the number of covariates $p$. The main advantage of this example is that the copula of $(Y, \boldsymbol{X})$ is known to be Gaussian. As such, we can use the maximum likelihood estimation as baseline, and compare it to our method when the densities are estimated (semi/non)parametrically.

We use one parametric, one semiparametric, and two nonparametric estimators of $\widehat{w}$. The first, *Gaussian + Gaussian*, is fully parametric and constructed by fitting Gaussian marginal distributions and a Gaussian copula using maximum-likelihood at each step. The second, KDE + Gaussian, is semiparametric and uses kernel estimators for the marginal distributions along with the Gaussian copula. The last two are fully nonparametric, but differ in the copula estimator. The method KDE + KDE uses weight estimator 3 and KDE + Kernel Vine uses a nonparametric vine estimator (`tll2` in Nagler, Schellhase, and Czado 2017).

These estimators are provided in the `R` package `eecop` (Nagler and Vatter 2020a), providing routines to fit and predict conditional expectiles and quantiles using copulas. Vine-related functionality is powered by `rvinecopulib` (Nagler and Vatter 2020b), the Gaussian copula by `copula` (Hofert et al. 2020), and kernel estimators for the margins by `kde1d` (Nagler and Vatter 2019). The scripts to reproduce the results are in the supplementary materials.

In Section 5.1, we look at the accuracy and convergence rate of estimators, in Section 5.2 we study the coverage of bootstrapped confidence intervals. In both cases, we consider both expectile and quantile regression at levels $t = 0.5$, that is mean and median regression, respectively, and $t = 0.95$. Because the results are qualitatively similar across estimation targets, we only report numbers averaged across targets. Full results and additional simulations with discrete covariates can be found in the supplementary materials.

### 5.1. Estimation Accuracy

We evaluate each estimator's accuracy using its empirical risk. In other words, we calculate $\left\{50^{-1}\sum_{i=1}^{50} \epsilon_i^2 \,|t - \mathbb{1}(\epsilon_i \leq 0)|\right\}^{1/2}$ for expectiles and $50^{-1}\sum_{i=1}^{50} \epsilon_i\{t - \mathbb{1}(\epsilon_i \leq 0)\}$ for quantiles, where $\epsilon_i = Y_i - \widehat{\theta}_{\boldsymbol{X}_i,t}$, on an independent test sample $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \dots, (\boldsymbol{X}_{50}, \boldsymbol{Y}_{50})$. All results are based on 100 replications. The x-axis of Figure 1 contains the training sample size $n$, the y-axis the risk; both have logarithmic scale.

There are two main observations. First, all parametric and semiparametric methods appear to converge at $\sqrt{n}$ rate, as seen from the $-1/2$ slopes. As expected from Corollaries 1 and 2, the MLE is slightly more efficient than the parametric copula-based

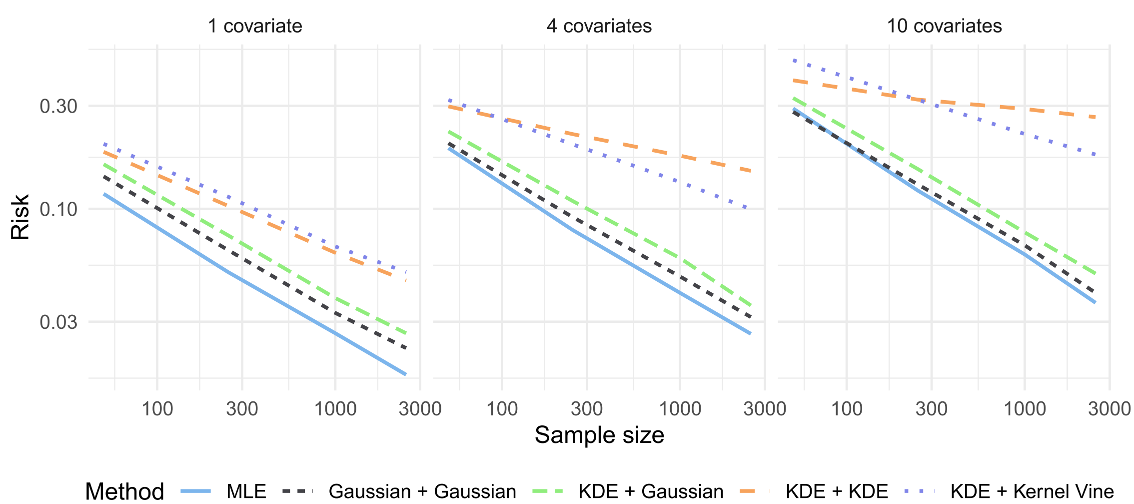


**Figure 1.** Average risk for expectile and quantile regression in a linear Gaussian model. Both axes have logarithmic scale.

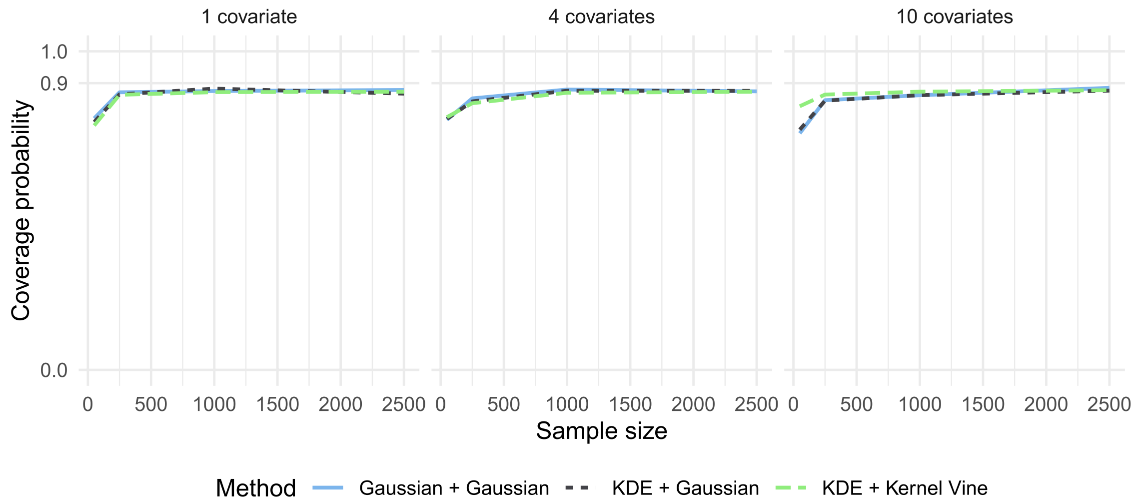


**Figure 2.** Coverage frequencies of bootstrapped 90%-confidence intervals.

method, which itself is slighlty more efficient than the semiparametric copula-based method. Second, the convergence rate of the KDE + KDE estimator deteriorates with the number of covariates, in line with Corollary 3. The KDE + Kernel Vine estimator achieves a better rate of convergence, largely unaffected by the dimension, as suggested by our discussion following Corollary 3. This illustrates the potential advantage of structural copula models for nonparametric regression.

### 5.2. Bootstrap Confidence Intervals

We now consider the same setup, but compute 90% confidence intervals using 500 bootstrapped replicates. Because nonparametric estimators are generally biased, we have to undersmooth them to make the bias asymptotically negligible. In preliminary experiments, we discovered a small bias between the estimated regression target and its bootstrap replicates nevertheless. More details and a justification are given in the supplementary material.

The method KDE + KDE is omitted in these experiments because it is both heavily biased and computationally too demanding. In Figure 2, we observe that the coverage probabilities are all close to the target level of 90%. Coverage generally improved for larger sample sizes, but is insufficient for very small samples with $n = 50$. Additional results for all sample sizes and the uncorrected bootstrap are in the supplementary materials.

### 5.3. Comparison to other Methods

We conclude with a comparison of the KDE + Kernel Vine estimator from the previous section, Copula in the following, to three quantile regression benchmarks: linear quantile regression (*LQR*, Koenker 2022), support vector machines (*SVM*, Karatzoglou et al. 2004), and gradient boosted trees (*GBM*, Greenwell et al. 2020). All implementations are used with default parameters.

We simulate from two models with $d = 10$ covariates. The first is the Gaussian from the previous section with marginal distributions of the response and covariates

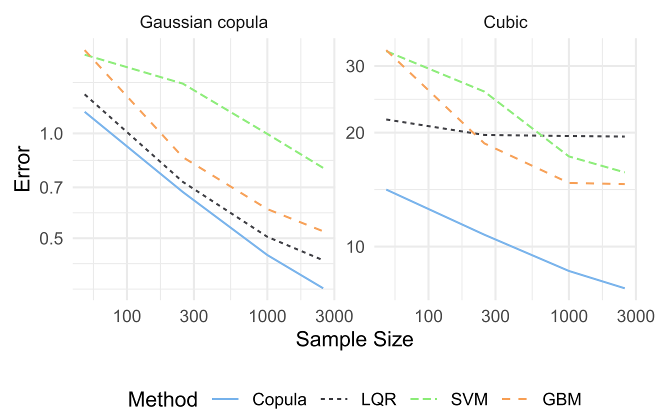


**Figure 3.** Absolute error of the conditional quantiles.

**Table 1.** Fraction of test points with quantile crossing.

| gap | Copula | LQR | SVM | GBM |
|---|---|---|---|---|
| 0.05 | 0% | 52% | 42% | 99% |
| 0.10 | 0% | 37% | 16% | 75% |
| 0.30 | 0% | 27% | 1% | 22% |

transformed, respectively to lognormal and exponential. The second is a cubic additive model with three active covariates, namely $Y = (X_1 + X_2 + X_3)^3 + \epsilon$, where $X_1, \dots, X_{10}$ and $\epsilon$ as before.

Using a test sample of size 250, we compare the estimators in two ways. First, we estimate conditional quantiles at levels 0.5 and 0.95, and report the average absolute distance from the true conditional quantiles. Second, we estimate conditional quantiles at all levels in $\{0.05, 0.1, \dots, 0.9, 0.95\}$, and report the fraction of observations where a $\tau_1$-quantile and a $\tau_2$-quantile with a certain gap $\tau_1 - \tau_2$ cross. All results are based on 100 replications.

Figure 3 indicates that our method outperforms the benchmarks for both models. This was expected from the Gaussian, which amounts to a copula regression, but not necessarily from the cubic. Table 1 further shows that our method satisfy the monotonicity constraints of conditional quantiles, which is not the case for the benchmarks. GBM quantiles almost always cross when the gap is 0.05, on 22% of the samples even at a gap as large as 0.3. SVM and LQR do only slightly better. Depending on context, such behavior may be prohibitive. Quantiles predicted by our method never cross.

## 6. Application: Quantitative Asset Allocation

We revisit the application mentioned in Section 1 and illustrate how our method can help blend (subjective) economic views in an otherwise quantitative asset allocation. Suppose that an investor faces the problem of shifting the fraction of her total wealth across various categories to take advantage of evolving market conditions. Such categories might be asset classes, such as bonds, stocks, and commodities, or industry sectors in a stock portfolio. She might also need to evaluate a range of performance and risk measures for different portfolios under economic scenarios like "the GDP will grow by 0.5%" or "the unemployment rate will decrease by 1%," over a given horizon.

Denote by $\boldsymbol{Y} \in \mathbb{R}^d$, $\boldsymbol{\gamma} \in \mathbb{R}^d$, and $\boldsymbol{X} \in \mathbb{R}^p$ the returns on the different categories, the fractions of total wealth invested in each of them, and the covariates. As measures of portfolio performance and risk, we consider the conditional mean, standard deviation or *volatility*, and 95% *Value-at-Risk* (VaR, see, e.g., McNeil, Frey, and Embrechts 2015), defined, respectively, as $\mu^*_{\boldsymbol{\gamma}} = \mathrm{E}(\boldsymbol{\gamma}^\top \boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})$, $\sigma^*_{\boldsymbol{\gamma}} = \mathrm{var}(\boldsymbol{\gamma}^\top \boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})^{1/2}$, and $q^*_{\boldsymbol{\gamma}} = F^{-1}_{-\boldsymbol{\gamma}^\top \boldsymbol{Y}|\boldsymbol{X}}(0.95 \mid \boldsymbol{x})$.

Let the indexing set be $T = \{1, 2, 3\} \times \Gamma$, where $\Gamma = \{\boldsymbol{\gamma}\colon \sum_j \gamma_j = 1, \max_j |\gamma_j| \leq M\}$ for some $M$ representing a constraint on how large a position in a single category can get. The parameter space and identifying functions are $\Theta = \{\theta_t \colon t \in T\} = \{\theta_{(j,\boldsymbol{\gamma})} \colon j \in \{1, 2, 3\}, \boldsymbol{\gamma} \in \Gamma\}$ and $\mathcal{G} = \{g_{\theta,t} \colon t \in T\} = \{g_{\theta,(j,\boldsymbol{\gamma})} \colon j \in \{1, 2, 3\}, \boldsymbol{\gamma} \in \Gamma\}$, where

$$\theta_t = \theta_{(j,\boldsymbol{\gamma})} = \begin{cases} \mu_{\boldsymbol{\gamma}}, & j = 1 \\ \sigma_{\boldsymbol{\gamma}}, & j = 2 \\ q_{\boldsymbol{\gamma}}, & j = 3 \end{cases}$$

$$g_{\theta,t} = g_{\theta,(j,\boldsymbol{\gamma})} = \begin{cases} \boldsymbol{\gamma}^\top \boldsymbol{y} - \mu_{\boldsymbol{\gamma}}, & j = 1 \\ (\boldsymbol{\gamma}^\top \boldsymbol{y} - \mu_{\boldsymbol{\gamma}})(\boldsymbol{\gamma}^\top \boldsymbol{y} - \mu_{\boldsymbol{\gamma}})^\top - \sigma^2_{\boldsymbol{\gamma}}, & j = 2 \\ 0.95 - \mathbb{1}(-\boldsymbol{\gamma}^\top \boldsymbol{y} < q_{\boldsymbol{\gamma}}), & j = 3 \end{cases}.$$

With $\{\boldsymbol{e}_i\}_{i=1}^d$ the standard basis of $\mathbb{R}^d$, $t = (1, \boldsymbol{e}_i)$ and $t = (2, \boldsymbol{e}_i)$ correspond, respectively, to the conditional expected return and variance of asset $i$, and $t = (2, \boldsymbol{e}_i + \boldsymbol{e}_j)$ to the conditional covariance between assets $i$ and $j$. This formulation thus also identifies the vector of expected returns $\boldsymbol{\mu}^* = \mathrm{E}(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})$ and the covariance matrix $\boldsymbol{\Sigma}^* = \mathrm{cov}(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x})$.

Given an estimated weight $\widehat{w}$, $\widehat{q}_{\boldsymbol{\gamma}}$ has to be solved for numerically, but there are closed form solutions for the conditional mean vector and covariance matrix estimators, that is

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n \boldsymbol{Y}_i \widehat{w}(\boldsymbol{Y}_i)}{\sum_{i=1}^n \widehat{w}(\boldsymbol{Y}_i)}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{\sum_{i=1}^n (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}})^\top \widehat{w}(\boldsymbol{Y}_i)}{\sum_{i=1}^n \widehat{w}(\boldsymbol{Y}_i)},$$

leading to $\widehat{\mu}_{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^\top \widehat{\boldsymbol{\mu}}$ and $\widehat{\sigma}_{\boldsymbol{\gamma}} = \sqrt{\boldsymbol{\gamma}^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{\gamma}}$.

If for all $\boldsymbol{\gamma} \in \Gamma$, $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = (\mu_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}, q_{\boldsymbol{\gamma}})$ lies in a compact subset of $\mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}$, the collection of identifying functions $\mathcal{G}$ satisfies the conditions of our theorems. Thus, our results in Section 4 imply that the estimators $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}} = (\widehat{\mu}_{\boldsymbol{\gamma}}, \widehat{\sigma}_{\boldsymbol{\gamma}}, \widehat{q}_{\boldsymbol{\gamma}})$ are consistent, converge weakly, and that their bootstrap versions yield valid inferences, uniformly in $\boldsymbol{\gamma} \in \Gamma$. To obtain uniform confidence bands over portfolios for the volatility and VaR, we compute the statistics $\widehat{\sigma}_b = \sup_{\boldsymbol{\gamma} \in \Gamma} \left|\widehat{\sigma}_{b,\boldsymbol{\gamma}} - \widehat{\sigma}_{\boldsymbol{\gamma}}\right|$ and $\widehat{q}_b = \sup_{\boldsymbol{\gamma} \in \Gamma} \left|\widehat{q}_{b,\boldsymbol{\gamma}} - \widehat{q}_{\boldsymbol{\gamma}}\right|$ for bootstrap samples $b = 1, \ldots, B$. Uniform $\alpha$-confidence bands are then given by $\boldsymbol{\gamma} \mapsto \widehat{\sigma}_{\boldsymbol{\gamma}} \pm \widehat{s}_{\sigma,\alpha}$ and $\boldsymbol{\gamma} \mapsto \widehat{q}_{\boldsymbol{\gamma}} \pm \widehat{s}_{q,\alpha}$, with $\widehat{s}_{\sigma,\alpha}$ and $\widehat{s}_{q,\alpha}$, respectively, the $\alpha$-quantiles of $\widehat{\sigma}_b$ and $\widehat{q}_b$. Scripts to reproduce the following analysis with the `eecop` (Nagler and Vatter 2020a) package are in the supplementary material.

### *6.1. The Data*

For $\boldsymbol{Y}$, we use value-weighted returns on five industry portfolios. For $\boldsymbol{X}$, we use the real gross domestic product and the seasonally adjusted unemployment rate. With yearly data covering 1947–2019, we have a total of 72 observations. The average yearly returns on stocks are in the 13%–17% range, but with large variations over time. While the GDP has been growing steadily at around 3% per year, the evolution of the unemployment has varied widely. Further, the returns on all industry sectors are positively correlated among each other and the growth in GDP, and negatively correlated with the growth in unemployment. Because the auto-correlations of all variables and their squares are statistically indistinguishable from zero, we treat the data as iid. Plots of time-series, auto-correlation functions, cross-correlations and summary statistics, along with additional details on the data sources, are in the supplementary material.

To study the impact of the predictors in the quantitative allocation scheme, we create three scenarios for the change in GDP and unemployment: good economy (+4.35/−11.52), median economy (+3.22/−2.48), poor economy (+1.59/+5.44). The good and bad scenarios are obtained using the 75th and 25th percentile for the growth in GDP unemployment, and conversely for the unemployment.

### *6.2. Results*

To derive predictions for each scenario, we estimate all margins by kernel estimators with plug-in bandwidths and fit a parametric vine copula model for the copula density. Predicted conditional means, standard deviations, and 95% VaRs are given in Table 2. We see that, the better the economic outlook, the higher the expected returns, and the lower the risk.

For each scenario, we compute the *efficient frontier*, that is a set of portfolios where $\boldsymbol{\gamma}$ can be expressed as a linear combination of the minimum variance and market portfolios. Background details on asset allocation can be found in the supplementary materials. In the top panel of Figure 4, we show the frontiers as a lineplot with uniform 75% bands as the dashed lines, and the corresponding quantities for the individual assets as a scatterplot. In the bottom panel, the portfolio weights are displayed as a function of the expected return. We can make the following observations:

- The upper half of all the parabolas yields higher returns for the same risk compared with the lower half, rational investors would only choose such portfolios.

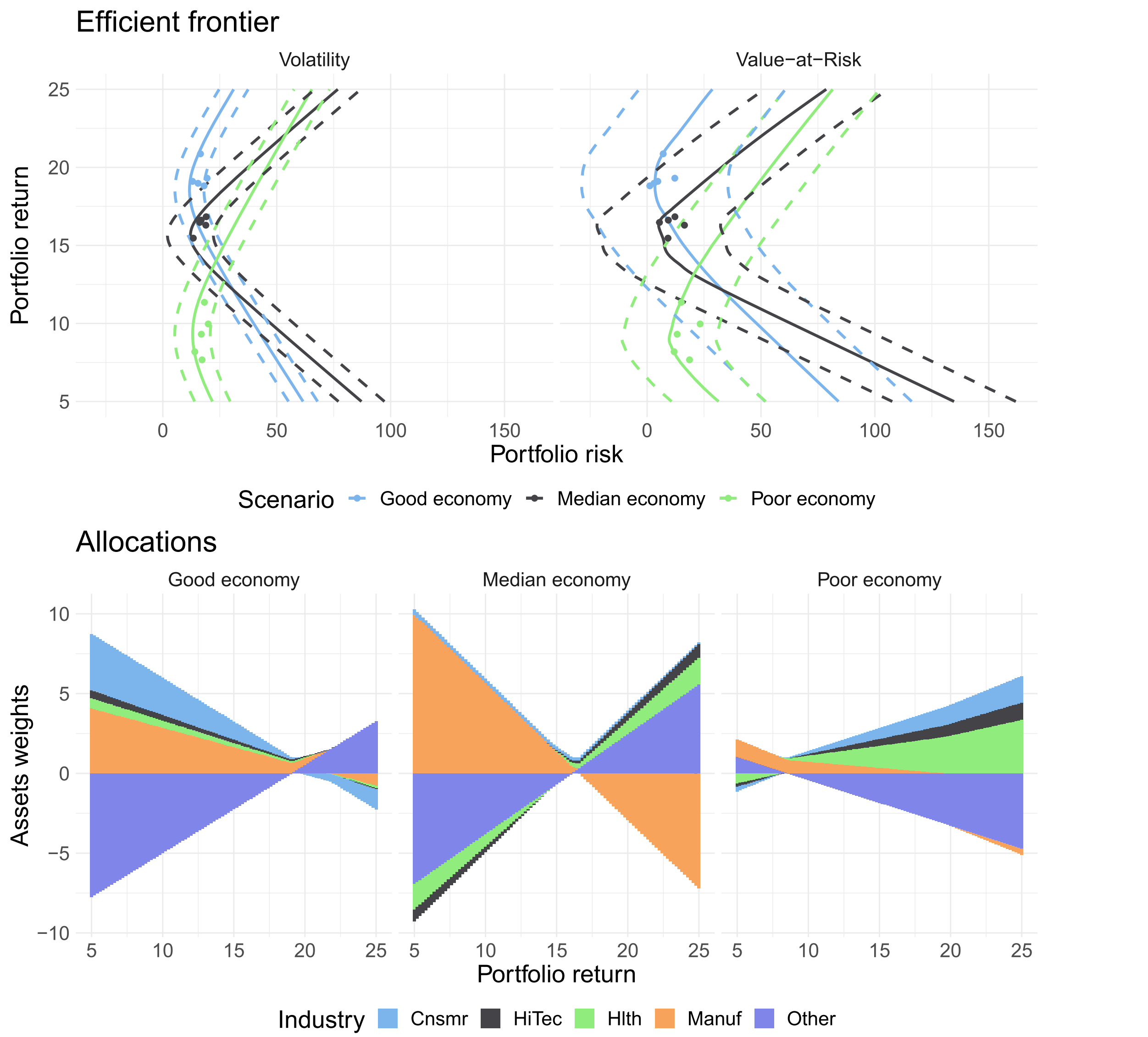


**Figure 4.** Efficient frontiers with uniform 75%-confidence bands and scatterplots for the individual assets (top panel), and corresponding portfolio weights (bottom panel).

**Table 2.** Predicted means ($\mu$), standard deviations ($\sigma$), and VaRs ($q$) for each scenario.

| | Cnsmr | | | Manuf | | | HiTec | | | Hlth | | | Other | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scenario | $\mu$ | $\sigma$ | $q$ | $\mu$ | $\sigma$ | $q$ | $\mu$ | $\sigma$ | $q$ | $\mu$ | $\sigma$ | $q$ | $\mu$ | $\sigma$ | $q$ |
| Good economy | 17.7 | 15.0 | 3.7 | 17.5 | 13.1 | 7.2 | 17.5 | 18.2 | 11.8 | 17.9 | 18.6 | 12.1 | 18.4 | 16.5 | 9.2 |
| Median economy | 16.1 | 15.9 | 5.4 | 14.7 | 13.4 | 11.1 | 15.6 | 18.9 | 16.6 | 16.2 | 18.3 | 12.2 | 15.2 | 16.7 | 9.8 |
| Poor economy | 9.1 | 16.6 | 13.2 | 8.0 | 14.2 | 12.8 | 9.5 | 20.1 | 25.0 | 11.1 | 17.6 | 17.3 | 7.1 | 17.0 | 18.8 |

- Similarly, since the scatterpoints corresponding to the individual assets are "inside" the parabolas, rational investors would prefer diversified portfolios from the upper half.
- Better economic outlooks generally imply higher expected returns at a given risk level.
- The weights of efficient portfolios with returns between that of minimum variance and market portfolios are generally reasonable, that is mostly included in $[-1, 1]$. And targeting returns outside this range generally requires large long and short positions, accompanied by quick an increase in portfolio risk.

Nonetheless, the uncertainties are fairly large, due to the small sample size. In higher frequency data (e.g., monthly) however, serial dependence can no longer be ignored. An extension of our results to this setting is discussed briefly in the following section.

## 7. Discussion

Section 4 provides an umbrella theory for solutions to copula-based estimating equations, and has close connections to several recent results. Noh, Ghouch, and Bouezmarni (2013) shows consistency and asymptotic normality for mean regression in a semiparametric model. Noh, Ghouch, and Van Keilegom (2015) and De Backer, El Ghouch, and Van Keilegom (2017) derive similar results for quantile regression, the latter for semiparametric copula densities. Rémillard, Nasri, and Bouezmarni (2017) establish weak convergence of parameteric and semiparametric estimators and validity of a parametric bootstrap procedure for the conditional quantile function indexed by the level. Note that Noh, Ghouch, and Van Keilegom (2015) replace the iid-assumption by a mixing condition, which we do not cover. But our results can likely be extended to stationary sequences under more stringent conditions using the techniques in Dehling, Mikosch, and Sörensen (2002). Nonetheless, we generalize and extend the above in several ways:

- The focus of previous research is on mean and quantile regression with univariate response. We allow for large classes of potentially multivariate identifying functions. This opens new possibilities for copula-based solutions to other regression problems.
- Previous results cover only parametric or semiparametric copula estimators. Our theory allows for parametric, semiparametric, and nonparametric methods.
- Weak convergence of $\widehat{\theta}$ as a process indexed by some potentially dense set $T$ is established. It can be used to derive simultaneous and coherent inferences across arbitrary combinations of finite/infinite-dimensional features of the conditional distribution.
- We propose and validate a bootstrapping scheme. It is applicable whenever obtaining the asymptotic distribution in closed form is inconvenient or infeasible.
- The theory also applies to $M$-estimators, defined as maximizers of a criterion function, when the criterion is differentiable (see, e.g., Kosorok 2007, sec. 2.2.6).
- Our results apply to both continuous and discrete data. Admittedly, the practical applicability for discrete data is limited by a lack of available software. While experimental features for discrete variables exist in the `rvinecopulib` package (Nagler and Vatter 2020b), they still need to mature. But the theoretical foundation is set and we hope that the computational limitations are overcome in the near future.

Being broadly applicable, our main results are somewhat abstract. But it is often straightforward to specialize them given a specific identifying function and weight estimator, as we do in Corollaries 1–3. We close our discussion by outlining applications to additional regression problems covered by our theory in Section 4.

First, suppose the goal is to characterize the relationship between a response $Y_1$ and a treatment $Y_2$ using an instrument $Y_3$, conditionally on a set of exogenous covariates $\boldsymbol{X}$. Specifically, assume as in Newey and Powell (2003) that $Y_1 = \boldsymbol{b}(Y_2)^\top \boldsymbol{\theta}(\boldsymbol{X}) + Z$, where $\boldsymbol{b}$ is known vector of basis functions, and $Z$ is a zero-mean error term. When the treatment is endogenous, that is $Y_2 \not\perp\!\!\!\perp Z$, identifying $\boldsymbol{\theta}$ requires an instrument $Y_3$ satisfying $E(Z \mid \boldsymbol{X} = \boldsymbol{x}, Y_3 = y_3) = 0$ for all $\boldsymbol{x}$ and $y_3$. In other words, one has to solve $\mathrm{E}\big(Y_1 | \boldsymbol{X} = \boldsymbol{x}, Y_3 = y_3\big) = E\big\{\boldsymbol{b}(Y_2)^\top \boldsymbol{\theta}(\boldsymbol{X}) | \boldsymbol{X} = \boldsymbol{x}, Y_3 = y_3\big\}$ for all $\boldsymbol{x}$ and $y_3$. Then the collection $\mathcal{G} = \{g_{\boldsymbol{\theta},t} : \boldsymbol{\theta} \in \Theta, t \in T\}$ with $g_{\boldsymbol{\theta},t}(\boldsymbol{y}) = b_t(y_3)\big\{y_1 - \boldsymbol{b}(y_2)^\top \boldsymbol{\theta}\big\}$ identifies the parameter $\boldsymbol{\theta}(\boldsymbol{x})$ for given $\boldsymbol{x}$. In the supplementary materials, we show that Assumptions 2 and 8 are satisfied, and it is then left to check Assumptions 3–7, and 9.

Further, our results apply to problems with censored or missing responses by tweaking the weight function. For example, suppose we only observe the right-censored version $\bar{S} = \min(S, Z) \in \mathbb{R}$ of a survival time $S$, along with a censoring indicator $\Delta = \mathbb{1}(S \leq Z)$. For any identifying function $\phi_\theta$, $\mathrm{E}\big\{\phi_\theta(S) \mid \boldsymbol{X} = \boldsymbol{x}\big\} = 0$ if and only if $E\big\{\phi_\theta(\bar{S})\Delta \bar{w}^*(\bar{S})\zeta(\bar{S})\big\} = 0$, where $\zeta(t) = 1/\{1 - F_{Z|\boldsymbol{X}}(t \mid \boldsymbol{x})\}$ and $\bar{w}^*$ is defined as in (2), but replacing $F_{Y,\boldsymbol{X}}$ and $F_Y$ by $F_{\bar{S},\boldsymbol{X}}$ and $F_{\bar{S}}$. This technique was used in De Backer, El Ghouch, and Van Keilegom (2017) to allow for right censoring in copula-based quantile regression, and it fits the setup of Section 4. With $\boldsymbol{Y} = (\bar{S}, \Delta)$, $g_\theta(\boldsymbol{y}) = \phi_\theta(y_1)y_2$, and $w^*(\boldsymbol{y}) = \bar{w}^*(y_1)\zeta(y_1)$, then $E\big\{\phi_\theta(\bar{S})\Delta \bar{w}^*(\bar{S})\zeta(\bar{S})\big\} = E\big\{g_\theta(\boldsymbol{Y})w^*(\boldsymbol{Y})\big\}$ and it remains to check the assumptions of Theorems 1–3. The above is only one instance of a larger class of methods based on *inverse probability weighting*. Other weight functionals $\zeta$ can be used similarly to account for other forms of censoring and missingness (see, e.g., Robins, Rotnitzky, and Zhao 1994; Wooldridge 2007; Han, Wang, and Song 2016). Following an earlier version of this article that appeared online, this idea was already picked by Hamori, Motegi, and Zhang (2020), albeit in a restricted semiparametric context.

## Appendix A: Assumptions

In general, we assume that the parameter space satisfies $\Theta \subseteq \ell^\infty(T)$ for some indexing set $T$. For simplicity, we consider $T$ to be a compact subset of a Euclidean space. We also use $(\boldsymbol{Y}', \boldsymbol{X}')$ to denote an independent copy of $(\boldsymbol{Y}, \boldsymbol{X})$.

### *A.1. Assumptions for the Main Results*

*Assumption 1.* The parameter space $\Theta$ is compact and contains an interior point $\theta^* \in \Theta$ such that $E\{g_{\theta^*}(\boldsymbol{Y})w^*(\boldsymbol{Y})\} = 0$ and, for any $\epsilon > 0$, $\inf_{|\theta_t - \theta_t^*| > \epsilon, t \in T} |E\{g_{\theta,t}(\boldsymbol{Y})w^*(\boldsymbol{Y})\}| > 0$.

*Assumption 2.* The class $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ is Euclidean[1] for some envelope function $G$.

*Assumption 3.* There exists a sequence of functions $w_n : \mathbb{R}^{2d+p} \to \mathbb{R}$ and a sequence $r_n = o(n^{1/2})$ with $r_n \geq 1$, such that $\max_{1 \leq i \leq n} \big|\widehat{w}(\boldsymbol{Y}_i) - \frac{1}{n}\sum_{j=1}^n w_n(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{X}_j)\big| = o_P\big(n^{-1/2} r_n\big)$.

*Assumption 4.* $\sup_{\theta \in \Theta} \|\mathrm{E}\{g_\theta(\boldsymbol{Y})w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}')\} - \mathrm{E}\{g_\theta(\boldsymbol{Y})w^*(\boldsymbol{Y})\}\|_T = o(1)$.

*Assumption 5.* $G$, $w_n$, and $r_n$ are such that

(i) $\mathrm{E}\{|G(\boldsymbol{Y})|\} < \infty$,

(ii) $\mathrm{E}\{G(\boldsymbol{Y})|w_n(\boldsymbol{Y}, \boldsymbol{Y}, \boldsymbol{X})|\} = o(n^{1/2} r_n)$,

(iii) $\mathrm{E}\{G(\boldsymbol{Y})^2 w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}')^2\} = o(n r_n^2)$.

[1] A formal definition is given in the supplementary material.

*Assumption 6.* With $(\boldsymbol{y}', \boldsymbol{x}')$ an arbitrary point in $\mathcal{Y} \times \mathcal{X}$, the functions

$$h_{n,\theta,t}(\boldsymbol{y}', \boldsymbol{x}') = \mathrm{E}\{g_{\theta,t}(\boldsymbol{Y}) w_n(\boldsymbol{Y}, \boldsymbol{y}', \boldsymbol{x}') + g_{\theta,t}(\boldsymbol{y}') w_n(\boldsymbol{y}', \boldsymbol{Y}, \boldsymbol{X})\},$$
$$H_n(\boldsymbol{y}', \boldsymbol{x}') = \mathrm{E}\{G(\boldsymbol{Y}) |w_n(\boldsymbol{Y}, \boldsymbol{y}', \boldsymbol{x}')| + G(\boldsymbol{y}') |w_n(\boldsymbol{y}', \boldsymbol{Y}, \boldsymbol{X})|\},$$

satisfy

(i) $\sup_{\|\theta_1-\theta_2\|_T + \|t_1-t_2\| < \delta_n} \mathrm{E}\{|h_{n,\theta_1,t_1}(\boldsymbol{Y}', \boldsymbol{X}') - h_{n,\theta_2,t_2}(\boldsymbol{Y}', \boldsymbol{X}')|^2\} = o(r_n^2)$ for every $\delta_n \to 0$,

(ii) $\mathrm{E}\{r_n^{-2} H_n(\boldsymbol{Y}', \boldsymbol{X}')^2\} = O(1)$,

(iii) $\mathrm{E}\{r_n^{-2} H_n(\boldsymbol{Y}', \boldsymbol{X}')^2 \mathbb{1}_{r_n^{-1} H_n(\boldsymbol{Y}', \boldsymbol{X}') \geq \eta\sqrt{n}}\} = o(1)$ for every $\eta > 0$.

*Assumption 7.* $\|\mathrm{E}[\{g_{\theta_n}(\boldsymbol{Y}) - g_{\theta^*}(\boldsymbol{Y})\}\{w_n(\boldsymbol{Y}, \boldsymbol{Y}', \boldsymbol{X}') - w^*(\boldsymbol{Y})\}]\|_T = o(\|\theta_n - \theta^*\|_T)$ for every $\theta_n \to \theta^*$.

*Assumption 8.* The map $\theta \mapsto \mathrm{E}\{g_\theta(\boldsymbol{Y}) w^*(\boldsymbol{Y})\}$ from $\ell^\infty(T)$ to $\ell^\infty(T)$ is Fréchet differentiable in a neighborhood of $\theta^*$ and the derivative $V_\theta \colon \ell^\infty(T) \mapsto \ell^\infty(T)$ is invertible at $\theta^*$. That is, for $\theta$ in a neighborhood of $\theta^*$, $V_\theta$ is a bounded linear operator such that

$$\|\mathrm{E}\{g_{\theta_n}(\boldsymbol{Y}) w^*(\boldsymbol{Y})\} - \mathrm{E}\{g_\theta(\boldsymbol{Y}) w^*(\boldsymbol{Y})\} - V_\theta(\theta_n - \theta)\|_T = o(\|\theta_n - \theta\|_T), \quad \forall \theta_n \to \theta.$$

And the inverse is the map $V_{\theta^*}^{-1}$ such that $V_{\theta^*}^{-1} V_{\theta^*}$ is the identity.

*Assumption 9.* There are iid random variables $\xi_i, i = 1, \dots, n$ independent of $(\boldsymbol{Y}_i, \boldsymbol{X}_i)_{i=1}^n$ with $\mathrm{E}(\xi_1) = \mathrm{var}(\xi_1) = 1$, and $\mathrm{E}\{|\xi_1|^{2+\epsilon}\} < \infty$ for some $\epsilon < 0$, such that for the same $w_n$, $r_n$ as in Assumption 3, $\max_{1 \leq i \leq n} \left|\widetilde{w}(\boldsymbol{Y}_i) - \frac{1}{n}\sum_{j=1}^n \xi_i w_n(\boldsymbol{Y}_i, \boldsymbol{Y}_j, \boldsymbol{X}_j)\right| = o_P\left(n^{-1/2} r_n\right)$.

A brief discussion of the assumptions is in order. Assumption 1 ensures identifiability of the parameter of interest $\theta^*$. Assumption 2 limits the complexity of the class of identifying functions $\mathcal{G} = \{g_\theta \colon \theta \in \Theta\}$. Importantly, this complexity is disentangled from the weight estimator. Euclidean classes (Nolan and Pollard 1987) generalize Vapnik-Cervonenkis classes of real-valued functions and are also called VC-type classes by some authors (e.g., Giné and Koltchinskii 2006). A formal definition and several convenient properties of these classes are in the supplementary material, where we also verify Assumption 2 for all examples from Section 3.1. Assumption 3 allows us to expand $\widehat{w}$ as a sample average and a uniformly negligible remainder. Since we only evaluate $\widehat{w}$ on $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ in the empirical estimating (7), the expansion only needs to be valid at these points. Assumption 9 is an analogous condition for $\widetilde{w}$, the bootstrap version of $\widehat{w}$. Assumption 4 ensures consistency of $\widehat{w}$ to $w^*$ in a weak sense. In particular, we require neither uniform nor pointwise consistency, although, with Assumption 5, either can be shown to be sufficient.

The remaining assumptions deal with joint regularity of the class of identifying functions $\mathcal{G}$ and the weight estimator $\widehat{w}$ through the sequence $w_n$. Assumptions 5, 6ii, and 6iii are moment conditions. In the latter two, the randomness from either estimating $w^*$ or replacing the population mean by its empirical counterpart has been averaged out; see also the discussion following Theorem 2. Assumption 6i is a form of stochastic equicontinuity in both $\theta$ and $t$. Note that Assumptions 6i–6iii are common in the context of function classes changing with $n$ (see e.g., van der Vaart and Wellner 1996, sec. 2.11.3). Assumption 8 ensures sufficient smoothness of the map $\theta \mapsto \mathrm{E}\{g_\theta(\boldsymbol{Y}) w^*(\boldsymbol{Y})\}$ and is verified in the supplementary material for the examples from Section 3.1. It could be weakened to Hadamard differentiability at the cost of slightly more tedious proofs. Assumption 7 ensures that replacing $w^*$ by $\widehat{w}$ has a negligible effect on the smoothness in Assumption 8.

When specifying $\widehat{w}$ further, the assumptions can often be disentangled into separate conditions on the identifying functions and weight estimator, see Assumptions 10–12.

## A.2. Assumptions for the Corollaries

*Assumption 10.* Denote respectively by $w^*(\boldsymbol{y}) = w(\boldsymbol{y}; \boldsymbol{\eta}^*)$ and $\widehat{w}(\boldsymbol{y}) = w(\boldsymbol{y}; \widehat{\boldsymbol{\eta}})$ the true and estimated weight functions.

(i) One has $\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i) + o_P(n^{-1/2})$ and $\widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = \frac{1}{n}\sum_{i=1}^n \xi_i \boldsymbol{\gamma}(\boldsymbol{Y}_i, \boldsymbol{X}_i) + o_P(n^{-1/2})$, for a function $\boldsymbol{\gamma}$ with $\mathrm{E}\{\boldsymbol{\gamma}(\boldsymbol{Y}, \boldsymbol{X})\} = 0$ and $\mathrm{E}\{\|\boldsymbol{\gamma}(\boldsymbol{Y}, \boldsymbol{X})\|_2^{2+\epsilon}\} < \infty$ and iid positive random variables $\xi_1, \dots, \xi_n$ independent of $(\boldsymbol{Y}_i, \boldsymbol{X}_i)_{i=1}^n$ with $\mathrm{E}(\xi_1) = \mathrm{var}(\xi_1) = 1$ and $\mathrm{E}\{|\xi_1|^{2+\epsilon}\} < \infty$ for some $\epsilon < 0$.

(ii) The function $w(\boldsymbol{y}; \boldsymbol{\eta}^*)$ is twice continuously differentiable in $\boldsymbol{\eta}^*$ with derivatives uniformly bounded in $\boldsymbol{y} \in \mathcal{Y}$.

(iii) One has $\mathrm{E}\{w(\boldsymbol{Y}; \boldsymbol{\eta}^*)^2 + \|\nabla_{\boldsymbol{\eta}} w(\boldsymbol{Y}; \boldsymbol{\eta}^*)\|_2^2\} < \infty$.

(iv) For some $\epsilon > 0$, one has $\mathrm{E}\{G(\boldsymbol{Y})^{2+\epsilon} \mid \boldsymbol{X} = \boldsymbol{x}\} < \infty$ and $\mathrm{E}\{G(\boldsymbol{Y})^{2+\epsilon}\}$.

(v) For every $\delta_n \to 0$ and with $\|\cdot\|$ the Euclidean norm, one has

$$\sup_{\|\theta_1-\theta_2\|_T + \|t_1-t_2\| < \delta_n} \mathrm{E}\big\{|g_{\theta_1,t_1}(\boldsymbol{Y}) - g_{\theta_2,t_2}(\boldsymbol{Y})|^2 \mid \boldsymbol{X} = \boldsymbol{x}\big\} = o(1),$$
$$\sup_{\|\theta_1-\theta_2\|_T + \|t_1-t_2\| < \delta_n} \big\|\mathrm{E}\big\{|g_{\theta_1,t_1}(\boldsymbol{Y}) - g_{\theta_2,t_2}(\boldsymbol{Y})|^2\big\}\big\|_2 = o(1).$$

*Assumption 11.* Write $\boldsymbol{F}_{\boldsymbol{Y}}(\boldsymbol{y}) = (F_{Y_1}(y_1), \dots, F_{Y_d}(y_d))$ and similarly for $\boldsymbol{F}_{\boldsymbol{X}}$, $\widehat{\boldsymbol{F}}_{\boldsymbol{Y}}$, and $\widehat{\boldsymbol{F}}_{\boldsymbol{X}}$. Define $\omega(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\eta}) = c_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\eta}) / c_{\boldsymbol{Y}}(\boldsymbol{u}; \boldsymbol{\eta})$, $w(\boldsymbol{y}; \boldsymbol{\eta}) = \omega\{\boldsymbol{F}_{\boldsymbol{Y}}(\boldsymbol{y}), \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x}); \boldsymbol{\eta}\}$, and

$$\lambda(\boldsymbol{Y}, \boldsymbol{X}) = \sum_{j=1}^d \int g_{\theta^*}(\boldsymbol{y}) \frac{\partial \omega\{\boldsymbol{F}_{\boldsymbol{Y}}(\boldsymbol{y}), \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x}); \boldsymbol{\eta}^*\}}{\partial F_{Y_j}(y_j)} \times \{\mathbb{1}(Y_j \leq y_j) - F_{Y_j}(y_j)\} f_{\boldsymbol{Y}}(\boldsymbol{y}) d\boldsymbol{y} + \sum_{k=1}^p \frac{\partial \mathrm{E}\big\{g_{\theta^*}(\boldsymbol{Y}) \mid \boldsymbol{X} = \boldsymbol{x}\big\}}{\partial x_k} \times \frac{\{\mathbb{1}(X_k \leq x_k) - F_{X_k}(x_k)\}}{f_{X_k}(x_k)}.$$

Further, for any $Z \in \{Y_1, \dots, Y_d, X_1, \dots, X_p\}$, write $\widetilde{F}_Z(z) = n^{-1}\sum_{i=1}^n \xi_i \mathbb{1}(Z_i \leq z)$ for the bootstrapped empirical distribution.

(i) Assumptions 10i, 10iv, and 10v hold.

(ii) The function $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\eta}) \mapsto \omega(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\eta})$ has two continuous derivatives on the domain $[0,1]^d \times \{\boldsymbol{v} \colon \|\boldsymbol{v} - \boldsymbol{F}_{\boldsymbol{X}}(\boldsymbol{x})\| < \epsilon\} \times \{\boldsymbol{\eta} \colon \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\| < \epsilon\}$ for some $\epsilon > 0$.

(iii) One has $\mathrm{E}\{\omega(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\eta}^*)^2 + \|\nabla_{(\boldsymbol{\eta}, \boldsymbol{Y}, \boldsymbol{X})} \omega(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\eta}^*)\|_2^2\} < \infty$.

*Assumption 12.*

(i) $K$ is a symmetric, bounded probability density function on $[-1, 1]$.

(ii) One has $b_n, \sigma_n \to 0$, $b_n^2 = o(n^{-1/2}\sigma_n^{-p/2})$, $\sigma_n^{-p/2} b_n^d / \ln n \to \infty$, and $n b_n^d \sigma_n^p / \ln n \to \infty$.

(iii) The densities $f_{\boldsymbol{X},\boldsymbol{Y}}$ and $f_{\boldsymbol{Y}}$ have uniformly bounded and continuous derivatives up to the third order and $\sup_{\boldsymbol{y} \in \mathcal{Y}} f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) < \infty$.

(iv) One has $\lim_{\delta \to 0} \sup_{\boldsymbol{y} \in \mathcal{Y}} \sup_{\|\boldsymbol{y}'-\boldsymbol{y}\| \leq \delta} \sup_{\|\boldsymbol{x}'-\boldsymbol{x}\| \leq \delta} \left|\frac{f_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}', \boldsymbol{x}')}{f_{\boldsymbol{Y},\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x})} - 1\right| = 0.$

(v) For some $\epsilon, \delta > 0$, one has $\mathrm{E}\{\sup_{|\boldsymbol{a}| < \delta} G(\boldsymbol{Y} + \boldsymbol{a})^{2+\epsilon} \mid \boldsymbol{X} = \boldsymbol{x}\} < \infty$ and $\mathrm{E}\{G(\boldsymbol{Y})^{2+\epsilon}\}$.

(vi) For every $\theta_n \to \theta^*$ and $\delta_n \to 0$ and with $\|\cdot\|$ the Euclidean norm, one has

$$\left\| \mathrm{E}\left\{ |g_{\theta_n}(\boldsymbol{Y}) - g_{\theta^*}(\boldsymbol{Y})| \mid \boldsymbol{X} = \boldsymbol{x} \right\} \right\|_T = O(\|\theta_n - \theta^*\|_T),$$
$$\sup_{\|\theta_1 - \theta_2\|_T + \|t_1 - t_2\| < \delta_n} \mathrm{E}\left\{ |g_{\theta_1,t_1}(\boldsymbol{Y}) - g_{\theta_2,t_2}(\boldsymbol{Y})|^2 \mid \boldsymbol{X} = \boldsymbol{x} \right\} = o(1).$$

In all three cases, the specialized assumptions Assumptions 10–12 are standard regularity conditions on the estimation method, as well as smoothness and moment conditions for the identifying function. This is in contrast to the general conditions Assumptions 3–7, 9 and Equation (9), where the estimation method and identifying function are often intertwined. Unfortunately, to the best of our knowledge, this disentanglement is not feasible in the general setting of Theorems 1–3 without strengthening the conditions.

## Supplementary Materials

Online supplementary material includes all mathematical proofs and auxiliary results, additional figures and results for the simulation studies in Section 5, and code and data to reproduce all results from Sections 5 and 6.


## Acknowledgments

The authors thank Johannes Wiesel for the proof of Lemma 11 in the supplementary material. We are grateful to the Associate Editor and three referees for helpful comments. Part of the research was conducted while Thomas Nagler was at Delft University of Technology and Thibault Vatter was at Columbia University.

## Funding

Thomas Nagler was partially supported by a Spinoza grant awarded by the Netherlands Organisation of Scientific Research (NWO) and the German Federal Ministry of Education and Research (BMBF) under grant no. 01IS18036A. Thibault Vatter was partially supported by the Swiss National Science Foundation (grant 174709).



## ORCID

Thomas Nagler http://orcid.org/0000-0003-1855-0046
Thibault Vatter http://orcid.org/0000-0001-9212-0218


## References


Bouyé, E., and Salmon, M. (2009), "Dynamic Copula Quantile Regressions and Tail Area Dynamic Dependence in Forex Markets," *The European Journal of Finance*, 15, 721–750. [1168]

Bücher, A., and Kojadinovic, I. (2019), "A Note on Conditional versus Joint Unconditional Weak Convergence in Bootstrap Consistency Results," *Journal of Theoretical Probability*, 32, 1145–1165. [1172]

Cai, T. T., and Zhang, L. (2018), "High-Dimensional Gaussian Copula Regression: Adaptive Estimation and Statistical Inference," *Statistica Sinica*, 28, 963–993. [1168]

Carrera, D., Bandeira, L., Santana, R., and Lozano, J. A. (2019), "Detection of Sand Dunes on Mars Using a Regular Vine-based Classification Approach," *Knowledge-Based Systems*, 163, 858–874. [1168,1170]

Chen, X., Koenker, R., and Xiao, Z. (2009), "Copula-based Nonlinear Quantile Autoregression," *The Econometrics Journal*, 12, 50–67. [1168]

Cooke, R. M., Joe, H., and Chang, B. (2015), "Vine Regression," *Resources for the Future - Discussion Paper*, 15, 15–52. [1168]

Côté, M.-P., and Genest, C. (2015), "A Copula-based Risk Aggregation Model," *The Canadian Journal of Statistics*, 43, 60–81. [1173]

De Backer, M., El Ghouch, A., and Van Keilegom, I. (2017), "Semiparametric Copula Quantile Regression for Complete or Censored Data," *Electronic Journal of Statistics*, 11, 1660–1698. [1168,1177]

Dehling, H., Mikosch, T., and Sörensen, M. (2002), *Empirical Process Techniques for Dependent Data*, Boston, MA: Springer. [1177]

Dette, H., Van Hecke, R., and Volgushev, S. (2014), "Some Comments on Copula-Based Regression," *Journal of the American Statistical Association*, 109, 1319–1324. [1168]

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26. [1172]

Elidan, G. (2012), "Copula Network Classifiers (CNCs)," in *Artificial Intelligence and Statistics (AISTATS) Proceedings*, pp. 346–354. [1168,1170]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, Boca Raton, FL: Chapman & Hall/CRC. [1173]

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552. [1170]

Giné, E., and Koltchinskii, V. (2006), "Concentration Inequalities and Asymptotic Results for Ratio Type Empirical Processes," *The Annals of Probability*, 34, 1143–1216. [1178]

Greenwell, B., Boehmke, B., Cunningham, J., and Developers, G. (2020), *gbm: Generalized Boosted Regression Models*, R package version 2.1.8. [1174]

Hamori, S., Motegi, K., and Zhang, Z. (2020), "Copula-based Regression Models with Data Missing at Random," *Journal of Multivariate Analysis*, 180, 104654. [1177]

Han, F., Zhao, T., and Liu, H. (2013), "Coda: High Dimensional Copula Discriminant Analysis," *Journal of Machine Learning Research*, 14, 629–671. [1168]

Han, P., Wang, L., and Song, P. X.-K. (2016), "Doubly Robust and Locally Efficient Estimation with Missing Outcomes," *Statistica Sinica*, 26, 691–719. [1177]

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020), *copula: Multivariate Dependence with Copulas*, R package version 0.999-20. [1173]

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004), "kernlab – an S4 Package for Kernel Methods in R," *Journal of Statistical Software*, 11, 1–20. [1174]

Koenker, R. (2022), *quantreg: Quantile Regression*. R package version 5.94. [1174]

Kolev, N., and Paiva, D. (2009), "Copula-based Regression Models: A Survey," *Journal of Statistical Planning and Inference*, 139, 3847–3856. [1168]

Kosorok, M. R. (2007), *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics, New York: Springer. [1177]

Kraus, D., and Czado, C. (2017), "D-vine Copula based Quantile Regression," *Computational Statistics & Data Analysis*, 110, 1–18. [1168]

McNeil, A. J., Frey, R., and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools – revised edition*, Princeton, NJ: Princeton university press. [1175]

Nagler, T., and Czado, C. (2016), "Evading the Curse of Dimensionality in Nonparametric Density Estimation with Simplified Vine Copulas," *Journal of Multivariate Analysis*, 151, 69–89. [1168,1170,1173]

Nagler, T., Schellhase, C., and Czado, C. (2017), "Nonparametric Estimation of Simplified Vine Copula Models: Comparison of Methods," *Dependence Modeling*, 5, 99–120. [1173]

Nagler, T., and Vatter, T. (2019), *kde1d: Univariate Kernel Density Estimation*, R package version 1.0.2. [1173]

——— (2020a), *eecop: an R Package to Solve Estimating Equations with Copulas*, R package version 0.0.1, Available at *https://github.com/tnagler/eecop*. [1169,1173,1175]

——— (2020b), *rvinecopulib: High Performance Algorithms for Vine Copula Modeling in R*, R package version 0.5.2.1.0. [1173,1177]

Newey, W. K., and Powell, J. L. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847. [1170]

——— (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [1177]

Noh, H., Ghouch, A. E., and Bouezmarni, T. (2013), "Copula-Based Regression Estimation and Inference," *Journal of the American Statistical Association*, 108, 676–688. [1168,1172,1176]

Noh, H., Ghouch, A. E., and Van Keilegom, I. (2015), "Semiparametric Conditional Quantile Estimation Through Copula-Based Multivariate

Models," *Journal of Business and Economic Statistics*, 33, 167–178. [1168,1172,1176,1177]

Nolan, D., and Pollard, D. (1987), "U-Processes: Rates of Convergence," *The Annals of Statistics*, 15, 780–799. [1178]

Okhrin, O., Okhrin, Y., and Schmid, W. (2013), "On the Structure and Estimation of Hierarchical Archimedean Copulas," *Journal of Econometrics*, 173, 189–204. [1173]

Pitt, M., Chan, D., and Kohn, R. (2006), "Efficient Bayesian Inference for Gaussian Copula Regression Models," *Biometrika*, 93, 537–554. [1168]

Rémillard, B., Nasri, B., and Bouezmarni, T. (2017), "On Copula-based Conditional Quantile Estimators," *Statistics & Probability Letters*, 128, 14–20. [1168,1172,1177]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression coefficients When Some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866. [1177]

Rubin, D. B. (1981), "The Bayesian Bootstrap," *The Annals of Statistics*, 9, 130–134. [1172]

Schallhorn, N., Kraus, D., Nagler, T., and Czado, C. (2017), "D-vine Quantile Regression with Discrete Variables," arXiv preprint arXiv:1705.08310. [1168]

Sklar, A. (1959), "Fonctions de répartition à *n* dimensions et leurs marges," *Publications de L'Institut de Statistique de L'Université de Paris*, 8, 229–231. [1169]

Tsukahara, H. (2005), "Semiparametric Estimation in Copula Models," *Canadian Journal of Statistics*, 33, 357–375. [1170]

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer. [1178]

Wooldridge, J. M. (2007), "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, 1281–1301. [1177]