Semiparametric Conditional Mixture Copula Models with Copula Selection

Zongwu Cai* Guannan Liu† Wei Long‡ Xuehong Luo§

October 16, 2023

Abstract

This study proposes a semiparametric conditional mixture copula model, that allows for unspecified functions of a covariate in both the (conditional) marginal distributions and the copula dependence and weight parameters. To estimate this model, we propose a two-step procedure. In the first step, the (conditional) marginal distributions are nonparametrically estimated using the weighted Nadaraya-Watson method. In the second step, we apply a penalized local log-likelihood function with a penalty term to simultaneously estimate the copula parameters and choose an appropriate copula model. Furthermore, we propose a test of covariate effects for time series data. We establish the large sample properties of both the penalized and unpenalized estimators based on α-mixing conditions. Monte Carlo simulations show that the proposed method performs well in selecting and estimating conditional mixture copulas under various model specifications. Finally, we apply the proposed method to investigate the dynamic patterns of dependence among four states' housing markets along the interest rate path.

Keywords: Conditional Copula; Mixture Copula; Semiparametric Estimation; Copula Selection; SCAD; EM algorithm

*Department of Economics, University of Kansas, Lawrence, KS 66045, United States.

†Department of Statistics and Data Science, School of Economics, and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China. Email: gliuecon@gmail.com.

‡Department of Economics, Tulane University, New Orleans, LA 70118, United States.

§Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.

1 Introduction

Knowledge of multivariate distributions plays a crucial role in various economic and financial applications, including portfolio selection, risk management, option pricing, and asset pricing models. The seminal work by Sklar (1959) establishes that any joint distribution with continuous margins can be decomposed into marginal distributions and a unique copula, which describes the dependence structure between variables. Copula-based models provide greater flexibility and accuracy than classical Gaussian modeling, allowing for the capture of complex dependence structures characterized by features such as fat-tails, asymmetry, and positive or negative dependence. For a comprehensive survey of copulas and their applications in economics and finance, see Patton (2012) and Fan and Patton (2014).

Sklar's theorem has been extended to conditional distributions by Patton (2006), introducing the concept of conditional copulas. Since then, the study of conditional copulas has gained significant attention, particularly for modeling time-varying dependence structures. Various approaches have been proposed in the literature. Patton (2006) considers a parametric function of lagged terms as the copula parameter, while Giacomini et al. (2009) assume a constant parameter within a time interval. Hafner and Manner (2012) and Almeida and Czado (2012) propose dynamic stochastic copula models and stochastic copula autoregressive models, respectively. In recent years, semiparametric methods have been applied to address potential misspecification issues. Acar et al. (2011) and Abegaz et al. (2012) assume that the copula parameter varies as an unknown function of a covariate. Fermanian and Lopez (2018) and Yang et al. (2021) propose a single-index copula with the parameter as an unknown link function of a univariate index.

Despite the abundance of conditional copula models in the literature, the majority of the aforementioned works utilize a single copula instead of a mixture copula. Thus, although the copula dependence parameters that reflect the level of dependence can vary with a covariate, the copula function that represents the pattern of dependence is still time-invariant. However, a covariate can affect not only the level but also the pattern of dependence. For example, international equity markets may exhibit different dependence patterns along volatility in exchange rate markets (Liu et al. (2022)).

In the literature, Liu et al. (2022) first propose an innovative semiparametric conditional mixture copula model that allows both the copula weight and dependence parameters in the mixture copula model to vary with a covariate in a nonparametric way. However, they assume that the covariate does not impact the marginal distributions and apply unconditional margins in a conditional mixture copula model. As Patton (2012) and Fermanian and Wegkamp (2012) point out, for modeling the joint conditional distribution, the information set used for the marginals and the copula must be the same. Therefore, the unconditional marginal distribution assumption in Liu et al. (2022) is fairly restrictive in practice. Furthermore, Liu et al. (2022) do not provide a data-driven method for choosing a suitable mixed copula model. To address these issues, in this paper, we allow for conditional marginal distributions in the model and estimate them by Cai (2002)'s weighted Nadaraya-Watson (WNW) conditional distribution method in the first step. Then, after plugging in the WNW estimators of the conditional marginals, we conduct parameter estimation and copula selection simultaneously for the semiparametric conditional mixture copula model under the α-mixing conditions.

Our study makes four primary contributions. First, we introduce a two-step procedure to estimate the components of the copula model and perform copula selection. In the initial step, we estimate the conditional marginal distributions through the WNW method in a nonparametric manner. In the second step, we employ a penalized local log-likelihood function with a SCAD penalty term (Fan and Li, 2001) to simultaneously estimate the parameters of the mixture copula and select an appropriate mixture copula model. The weight functions' functional norms are penalized to enable their shrinkage to zero in cases where the corresponding copulas contribute minimally. To make the methodology practical and applicable, we use a semiparametric version of the EM algorithm to estimate the weight and dependence parameters in the penalized local copula log-likelihood function. We also discuss essential practical issues, including the selection of the bandwidth and tuning parameter, and the construction of confidence intervals.

Another significant contribution of this paper is establishing the asymptotic properties of the proposed estimators in nonstandard settings. Of particular interest is the fact that the convergence rates of the conditional estimators for the marginal distributions are of order $\sqrt{Th}$, assuming a common bandwidth h for all marginals. These rates are not necessarily faster than the convergence rates of the copula parameters (including weight and dependence parameters), which renders the asymptotic properties of copula parameters complex, especially in time series data. Additionally, some weight parameters of the conditional mixture copula may be situated on the boundary of the parameter space, presenting another challenge. To address these nonstandard situations, we establish the large sample properties of the unpenalized estimator and demonstrate the consistency and oracle properties of the penalized estimator.

The third contribution is constructing a feasible test statistic for the covariate effects. Because the assumptions underlying the proposed conditional mixture copula model exclude scenarios where the copula parameters are constant and independent of the covariate, it is valuable to test the relevance of the covariate. To investigate whether the copula parameters truly exhibit functional dependencies on the covariate, we propose a feasible test statistic to assess the covariate effects within the framework of the conditional mixture copula model for weakly dependent time series data. Notably, the asymptotic variance of the test statistic mirrors that observed in situations involving independent data under the null hypothesis. Furthermore, we establish the asymptotic chi-square distribution of the test statistic under the null hypothesis.

Finally, we conduct simulations to assess the method's finite-sample performance and apply it to a real world example. Simulation results show that the proposed method performs well in selecting an appropriate copula model and in estimating the unknown copula parameters, when the true model is either an individual copula or a mixture copula. In the empirical section, we use the proposed method to study the comovement of housing markets in four states with the highest median home prices in 2020 (California, Colorado, Massachusetts, and Washington). We consider the interest rate as the covariate. The empirical results show that the levels of dependence always increase as the interest rate decreases. For the patterns of dependence, the proposed method chooses a single Gumbel copula for four pairs, and a mixture of Gumbel and Frank copulas for the other two pairs. Furthermore, for the mixture copula pairs, the weights associated with the Gumbel copula increase as the interest rate decreases. Therefore, all pairs of housing markets show significant upper-tail dependence, and both the upper-tail dependence and the overall dependence attain higher levels under a lower interest rate.

The remainder of this article is organized as follows. Section 2 introduces the semiparametric conditional mixture copula model and the two-step estimating procedure. This section also includes a test for covariate effects on the mixture copula model. Asymptotic properties are given in Section 3. Section 4 discusses practical issues, including a semiparametric EM algorithm, the bandwidth and tuning parameter selection, and the construction of confidence intervals. Section 5 reports the Monte Carlo simulation results. Section 6 presents a real data example. We conclude the paper in Section 7. Proofs of the asymptotic properties are presented in Appendix.

2 Theoretical Model and Methodology

2.1 Construction of Conditional Mixture Copula

Let $\{\boldsymbol{X}_t, Z_t\}_{t=1}^T$ be a series of random vectors with a p-dimensional time series sequence $\boldsymbol{X}_t = (X_{1t}, \cdots, X_{pt})^\top$ and a 1-dimensional covariate Z_t. Denote $F_{\boldsymbol{X}|Z}(\boldsymbol{x}_t|z_t)$ and $F_{\boldsymbol{X},Z}(\boldsymbol{x}_t, z_t)$ as the conditional joint distribution function of $\boldsymbol{X}$ evaluated at $\boldsymbol{x}_t$ given z_t and the joint distribution of $(\boldsymbol{X}_t, Z_t)$, respectively. Throughout the paper, for simplicity, we write $F_{\boldsymbol{X}|z_t}(\boldsymbol{x}_t) \equiv F_{\boldsymbol{X}|Z}(\boldsymbol{x}_t|Z = z_t)$. In this article, we aim to estimate the conditional joint distribution $F_{\boldsymbol{X}|z_t}(\boldsymbol{x}_t)$. According to the extension of Sklar's theorem in Patton (2006), it is clear that

$$F_{\boldsymbol{X}|z_t}(x_{1t}, \cdots, x_{pt}) = C\left(F_{X_1|z_t}(x_{1t}), \cdots, F_{X_p|z_t}(x_{pt})\right), \tag{1}$$

where $C(\cdot)$ is the unknown conditional copula function, and $F_{X_j|z_t}(x_{jt}) = F_{X_j|Z}(x_{jt}|z_t)$ is the marginal cumulative distribution function of X_j evaluated at x_{jt} and conditional on $Z = z_t$ for $j = 1, \cdots, p$. The extension of Sklar's theorem indicates that there exists a unique conditional copula function $C(\cdot)$ in (1) to link the conditional joint distribution function and the conditional marginal distribution functions. To avoid potential misspecification issues when using a single conditional copula model, which may significantly differ from the true conditional copula function, we propose the use of a conditional mixture copula model. This approach approximates the unknown conditional copula by a concave combination of infinite individual conditional copulas:

$$C\left(\boldsymbol{u}(z_t); \boldsymbol{\omega}(z_t), \boldsymbol{\theta}(z_t)\right) = \sum_{k=1}^{\infty} \omega_k(z_t) C_k\left(\boldsymbol{u}(z_t); \theta_k(z_t)\right),$$

where $\boldsymbol{u}(z_t) = (F_{X_1|z_t}(x_{1t}), \cdots, F_{X_p|z_t}(x_{pt}))$, $\{\omega_k(z_t)\}_{k=1}^{\infty}$ is a sequence of unknown weight parameters, and $\{C_k(\cdot;\cdot)\}_{k=1}^{\infty}$ is a set of candidate copula functions with unknown parameters $\{\theta_k(z_t)\}_{k=1}^{\infty}$. Here, $\{C_k(\cdot;\cdot)\}_{k=1}^{\infty}$ represents a set of known basis copula functions, allowing us to express $C\left(\boldsymbol{u}(z_t);\boldsymbol{\omega}(z_t),\boldsymbol{\theta}(z_t)\right)$ as a series expansion using these basis copula functions. In practical applications, we employ a finite number of d individual conditional copulas to approximate the true conditional copula:

$$C\left(\boldsymbol{u}(z_t);\boldsymbol{\omega}(z_t),\boldsymbol{\theta}(z_t)\right)=\sum_{k=1}^{d}\omega_k(z_t)C_k\left(\boldsymbol{u}(z_t);\theta_k(z_t)\right), \tag{2}$$

where $\boldsymbol{\theta}(z_t)=\left(\theta_1^\top(z_t),\cdots,\theta_d^\top(z_t)\right)^\top\in\boldsymbol{\Theta}$ is a $(p_1+\cdots+p_d)$−dimensional vector of copula dependent parameters, $\boldsymbol{\omega}(z_t)=(\omega_1(z_t),\cdots,\omega_d(z_t))^\top\in[0,1]^d$ is a vector of nonnegative weight parameters that satisfies $\sum_{k=1}^{d}\omega_k(z_t)=1$. For simplicity of presentation, we set $p_1=\cdots=p_d=1$, that is, $\boldsymbol{\theta}(z_t)$ is a $d\times 1$ vector. It is worth mentioning that the basis copulas with multi-dimensional dependence parameters are allowed to be selected in (2), and the parameter estimation method in this section is also applicable to multi-dimensional dependence parameters. The copula dependence parameters $\{\theta_k(z_t)\}_{k=1}^{d}$ measure the degree of dependence corresponding to the basis copulas $\{C_k(\cdot;\cdot)\}_{k=1}^{d}$, and weight parameters $\{\omega_k(z_t)\}_{k=1}^{d}$ characterize shape of conditional copula. Both are unknown functions of the covariate. The conditional marginal distributions $\boldsymbol{u}(z_t)$ are also unknown functions of z_t to address potential misspecification concerns.

To ensure a comprehensive approximation of the true model without excluding relevant basis copulas, we initially consider a rich set of basis copulas to capture a wide range of dependence structures. As the mixture model can theoretically utilize an infinite number of basis copulas to approximate any conditional copula function, this approach allows for greater flexibility. To achieve an efficient and optimal finite approximation, we employ the penalized local log-likelihood method described in Section 2.2. This method shrinks the weight parameters associated with "insignificant" basis copulas toward zero. The selection process of basis copulas, coupled with the nonparametric specification of copula parameters and conditional marginal distributions, endows the model in (2) with ample flexibility to capture various possible dependence structures.

To circumvent the model identification issue, we follow Cai and Wang (2014) and assume that the conditional mixture copula model is identified throughout this paper. For the identification assumption, two conditional mixture copulas $C\left(\boldsymbol{u}(z_t);\boldsymbol{\omega}(z_t),\boldsymbol{\theta}(z_t)\right)=\sum_{k=1}^{d}\omega_k(z_t)C_k\left(\boldsymbol{u}(z_t);\theta_k(z_t)\right)$ and $C^*\left(\boldsymbol{u}(z_t);\boldsymbol{\omega}^*(z_t),\boldsymbol{\theta}^*(z_t)\right)=\sum_{k=1}^{d^*}\omega_k^*(z_t)C_k^*\left(\boldsymbol{u}(z_t);\theta_k^*(z_t)\right)$ are considered to be identified if and only if $d = d^*$, and we can order the summations such that $C_k(\boldsymbol{u}(z_t); \theta_k(z_t)) = C_k^*(\boldsymbol{u}(z_t); \theta_k^*(z_t))$ and $\omega_k(z_t) = \omega_k^*(z_t)$ for all possible values of $\boldsymbol{u}(z_t)$ and z_t for $k = 1, \cdots, d$. Without loss of generality, the conditional mixture copula model under investigation is assumed to be identified.

Model (2) extends previous research in several aspects. First, unlike Liu et al. (2022), we allow for nonparametric variation in the conditional marginal distribution functions with a covariate and employ a penalized local likelihood approach with shrinkage to select the optimal conditional mixture copula from a candidate set. This extension is valuable because it retains the conditional correlation when decomposing the conditional multivariate distribution into marginal distributions (Patton, 2006). Second, Hu (2006), Cai and Wang (2014), and Liu et al. (2019) consider invariant mixture copulas, assuming known parametric forms for all marginal distributions of independent random variables. However, these assumptions may be challenging to apply in practice, as the specific form of the marginal distributions is often unknown and covariates can strongly influence the dependence structures. Third, compared to the semiparametric conditional copula models proposed by Acar et al. (2011), Abegaz et al. (2012), and Yang et al. (2021), our conditional mixture copula accommodates different copula families and enhances flexibility by integrating nonparametric conditional marginal distributions. Finally, we provide a test statistic for weakly dependent data to examine the null hypothesis of no effect from the covariate. The omnibus test by Gijbels et al. (2021) is applicable to independent data only, making it crucial to develop a suitable test for dependent data.

Notably, while the conditional mixture copula model in this study accommodates multidimensional covariates, the inclusion of a large number of covariates can lead to the curse of dimensionality. To address this issue, we recommend employing a mixture single index copula model, an extension of the single index copula proposed by Fermanian and Lopez (2018) and Yang et al. (2021). The estimation and selection procedures for such a model merit further investigation as a separate study in the future.

2.2 Estimating Steps

To estimate the unknown copula dependent and weight parameters in Model (2), we propose a two-step semiparametric method to avoid inconsistent estimates due to misspecification of the functional forms of the unknown functions $\boldsymbol{\omega}(z_t), \boldsymbol{\theta}(z_t)$ and $\boldsymbol{u}(z_t)$. In the first step, we use the WNW conditional distribution function method proposed by Cai (2002) to estimate the unknown conditional marginal distribution functions $F_{X_j|z_t}(x_{jt})$ for $j = 1, \cdots, p$. Notably, the nonparametric estimation procedure in the first step differs from that in Acar et al. (2011), Abegaz et al. (2012) and Liu et al. (2022). In the second step, we incorporate the estimated conditional marginal distributions into the mixture copula and use a penalized local log-likelihood function with a shrinkage operator to simultaneously estimate the unknown copula parameters and select an appropriate mixture copula. To accomplish the aforementioned procedures, we develop a feasible expectation-maximization (EM) algorithm for the penalized local pseudo log-likelihood function, which is presented in Section 4.1.

Step 1. In the first step, our target is to estimate the unknown conditional marginal distributions based on the available observations. We use the WNW estimator proposed by Cai (2002) to estimate $F_{X_j|z}(x_j)$. That is,

$$\widehat{F}_{X_j|z}(x_j) = \frac{\sum_{t=1}^{T} \vartheta_{jt}(z) K_{h_j}(Z_t - z) I(X_{jt} \leq x_j)}{\sum_{t=1}^{T} \vartheta_{jt}(z) K_{h_j}(Z_t - z)}, \tag{3}$$

where $K_{h_j}(\cdot) = K(\cdot/h_j)/h_j$ and $K(\cdot)$ is a kernel function and h_j is the bandwidth. The sequence of weights $\{\vartheta_{jt}(z)\}_{t=1}^{T}$ is smoothed over the covariate space with the constraint $\sum_{t=1}^{n} \vartheta_{jt}(z) = 1$. Cai (2002) shows that this constraint can be expressed as

$$\vartheta_{jt}(z) = \frac{1}{T\left[1 + \lambda_j(Z_t - z)K_{h_j}(Z_t - z)\right]} \tag{4}$$

by maximizing the sum of $\ln(\vartheta_{jt}(z))$ directly, and

$$\lambda_j = \arg\max_{\lambda_j} \frac{1}{Th_j} \sum_{t=1}^{T} \ln\left(1 + \lambda_j(Z_t - z)K_{h_j}(Z_t - z)\right). \tag{5}$$

We use the nonparametric Akaike information criterion, proposed by Cai and Tiwari (2000), to select the optimal bandwidths h_j for $j = 1, \cdots, p$ as suggested by Cai (2002). The WNW estimator $\widehat{F}_{X_j|z}(x_j)$ not only possesses desirable properties in terms of feasibility and eligibility, but also is first-order equivalent to a local linear estimator, which has ideal sampling properties both at the interior and boundary points of the support of design density (Cai, 2002). Therefore, $\widehat{F}_{X_j|z}(x_j)$ is sufficient for the purpose of conditional marginal distribution estimation.

Step 2. The second step aims to estimate the copula dependent and weight parameters via the penalized local log-likelihood function. Suppose both the $(q+1)$th derivatives of functions θ_k and ω_k exist at point z for $k = 1, \cdots, d$. Applying the Taylor expansion to $\boldsymbol{\omega}(z_t) = (\omega_1(z_t), \cdots, \omega_d(z_t))^\top$ and $\boldsymbol{\theta}(z_t) = (\theta_1(z_t), \cdots, \theta_d(z_t))^\top$ at z which is the neighborhood of data point z_t, we obtain

$$\omega_k(z_t) \approx \omega_k(z) + \omega_k^{(1)}(z)(z_t - z) + \cdots + \frac{\omega_k^{(q)}(z)}{q!}(z_t - z)^q \equiv \nu_{1,k}, \tag{6}$$

and

$$\theta_k(z_t) \approx \theta_k(z) + \theta_k^{(1)}(z)(z_t - z) + \cdots + \frac{\theta_k^{(q)}(z)}{q!}(z_t - z)^q \equiv \nu_{2,k}. \tag{7}$$

By using the estimates $\widehat{F}_{\boldsymbol{X}|z}(\boldsymbol{x}_t)$ of conditional marginal distributions in the first step, and then when z_t is close to z and $q = 1$, the local kernel-weighted pseudo penalized log-likelihood function takes the following form with a Lagrangian multiplier

$$\begin{aligned} Q(\boldsymbol{\eta}(z)) &= \sum_{t=1}^{T} \Bigg\{ \ln \left[\sum_{k=1}^{d} \left(\omega_k(z) + \omega_k^{(1)}(z)(z_t - z) \right) c_k \left(\widehat{F}_{\boldsymbol{X}|z}(\boldsymbol{x}_t); \theta_k(z) + \theta_k^{(1)}(z)(z_t - z) \right) \right] \\ &\quad \times K_h(z_t - z) \Bigg\} - \left\{ T \sum_{k=1}^{d} P_{\gamma_T}(\omega_k(z)) - \rho \left(1 - \sum_{k=1}^{d} \omega_k(z) \right) \right\} \\ &\equiv \sum_{t=1}^{T} [\ell(\boldsymbol{\eta}(z); \hat{\boldsymbol{u}}(z)) K_h(z_t - z)] - \tilde{L}(\boldsymbol{\omega}(z)) \\ &\equiv \sum_{t=1}^{T} L_t(\boldsymbol{\eta}(z)) - \tilde{L}(\boldsymbol{\omega}(z)), \end{aligned} \tag{8}$$

where $c_k(\cdot)$ is the density of $C_k(\cdot)$, and $\boldsymbol{\eta}(z) = \left(\boldsymbol{\omega}^\top(z); \boldsymbol{\theta}^\top(z); \boldsymbol{\omega}^{(1)\top}(z); \boldsymbol{\theta}^{(1)\top}(z)\right)^\top = \left(\omega_1(z), \cdots, \omega_d(z); \theta_1(z), \cdots, \theta_d(z); \omega_1^{(1)}(z), \cdots, \omega_d^{(1)}(z); \theta_1^{(1)}(z), \cdots, \theta_d^{(1)}(z)\right)^\top$. The penalty function $P_{\gamma_T}(\cdot)$ with tuning parameter γ_T accommodates various penalization methods, including the classical L_1 penalty (Tibshirani, 1996), and the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001). The SCAD penalty function is chosen because of its desirable properties (unbiasedness, sparsity and continuity). The first-order derivative $P'_{\gamma_T}(t)$ of the SCAD penalty function is given by

$$P'_{\gamma_T}(t) = \gamma_T I(t \le \gamma_T) + \frac{\max(a\gamma_T - t, 0)}{a-1} I(t > \gamma_T)$$

with indicator function $I(\cdot)$ for $a > 2$, $t > 0$ and $P_{\gamma_T}(0) = 0$. For simplicity, we assume that the penalty function used is the same across all weight parameters, and we let $a = 3.7$ for the SCAD penalty following Fan and Li (2001). The selection procedure of tuning parameter γ_T is discussed in Section 4.2. Because the penalized local maximum likelihood estimator $\hat{\boldsymbol{\eta}}(z)$ may not have a closed form, an iterative EM algorithm is introduced in Section 4.1 to find the numerical solution. Note that there are $p+1$ different bandwidths $h_1, \cdots, h_p$ and h, where $\{h_j\}_{j=1}^p$ can be selected by the nonparametric Akaike information criterion in Step 1, the bandwidth h is a component of the kernel weight of the local log-likelihood in (8), and the selection procedure of h is given in Section 4.2.

Notably, there are interior and boundary points for the penalized estimator $\hat{\boldsymbol{\omega}}(z)$. To investigate its asymptotic behavior at the boundary, we partition $\boldsymbol{\omega}(z) = (\omega_1(z), \cdots, \omega_d(z))^\top$ into $\boldsymbol{\omega}^\dagger(z)$ and $\boldsymbol{\omega}^*(z)$, where $\boldsymbol{\omega}^\dagger(z)$ is in the interior of weight parameter space, and $\boldsymbol{\omega}^*(z)$ is on the boundary of the weight parameter space. That is, the $d^\dagger \times 1$ vector $\boldsymbol{\omega}^\dagger(z)$ consists entirely of non-zero elements of $\boldsymbol{\omega}(z)$, while the $(d-d^\dagger)\times 1$ vector $\boldsymbol{\omega}^*(z)$ is exclusively composed of zero elements from $\boldsymbol{\omega}(z)$. We also denote $\left(\boldsymbol{\omega}^{(1)\dagger}(z), \boldsymbol{\theta}^\dagger(z), \boldsymbol{\theta}^{(1)\dagger}(z)\right)$ corresponding to $\boldsymbol{\omega}^\dagger(z)$, and $\left(\boldsymbol{\omega}^{(1)*}(z), \boldsymbol{\theta}^*(z), \boldsymbol{\theta}^{(1)*}(z)\right)$ corresponding to $\boldsymbol{\omega}^*(z)$. Furthermore, we denote $\boldsymbol{\eta}(z) = \left(\boldsymbol{\eta}^{\dagger\top}(z), \boldsymbol{\eta}^{*\top}(z)\right)^\top$, where $\boldsymbol{\eta}^\dagger(z) = \left(\boldsymbol{\omega}^{\dagger\top}(z), \boldsymbol{\theta}^{\dagger\top}(z), \boldsymbol{\omega}^{(1)\dagger\top}(z), \boldsymbol{\theta}^{(1)\dagger\top}(z)\right)^\top$ and $\boldsymbol{\eta}^*(z) = \left(\boldsymbol{\omega}^{*\top}(z), \boldsymbol{\theta}^{*\top}(z), \boldsymbol{\omega}^{(1)*\top}(z), \boldsymbol{\theta}^{(1)*\top}(z)\right)^\top$. It is clear that the estimates of the copula dependent parameters $\boldsymbol{\theta}^*(z)$ can be arbitrary, because their corresponding weight parameters are zero.

2.3 Test of Covariate Effects

The conditional mixture copula model in (2) allows for variation in the copula dependent and weight parameters with a covariate, which enhances the flexibility of the model. However, in some cases, empirical practitioners may assume that the conditional copula does not depend on covariates, known as the "simplifying assumption" (Derumigny and Fermanian, 2017). This assumption is commonly used in vine models. Our conditional mixture copula model excludes the scenario where the copula parameters are constant and independent of the covariate. When the covariate is irrelevant, the inclusion of nonparametric functions of the covariate becomes unnecessary. Therefore, it is crucial to test whether the copula parameters are constant and do not vary with the covariate.

In this section, we propose a feasible hypothesis testing procedure, which belongs to the generalized likelihood ratio test (GLRT), to test the impact of the covariate on copula parameters for the weakly dependent data. Specifically, we follow Acar et al. (2013) and test the linear function on covariate Z, which encompasses not only constant scenarios but also frequently employed parametric structures. More precisely, we assume $\boldsymbol{\Theta}_L = \{\boldsymbol{\eta}^\dagger(\cdot) : \exists \boldsymbol{\eta}_0^\dagger, \boldsymbol{\eta}_1^\dagger \in \mathbb{R}^{4d^\dagger} \text{ such that } \boldsymbol{\eta}^\dagger(Z) = \boldsymbol{\eta}_0^\dagger + \boldsymbol{\eta}_1^\dagger Z \text{ for any } Z\}$ which constitutes the collective of all linear functions within its domain. We are interested in testing $\mathbb{H}_0 : \boldsymbol{\eta}^\dagger(\cdot) \in \boldsymbol{\Theta}_L$ against $\mathbb{H}_1 : \boldsymbol{\eta}^\dagger(\cdot) \notin \boldsymbol{\Theta}_L$. Note that when $\boldsymbol{\eta}_1^\dagger$ under $\mathbb{H}_0$ is set to $\mathbf{0}$, the null hypothesis becomes a constant vector that is independent of the covariate.

Based on the estimator $\widehat{F}_{X_j}(x_{jt})$ of conditional marginal distribution for $j \in \{1, \cdots, p\}$, and the maximum likelihood estimator $\widetilde{\boldsymbol{\eta}}^\dagger(Z)$ under $\mathbb{H}_0$, we construct the GLRT statistic

$$M_T = \sum_{t=1}^{T} L_t\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t)\right) - \sum_{t=1}^{T} L_t\left(\widetilde{\boldsymbol{\eta}}^\dagger(Z_t)\right), \tag{9}$$

which is the difference between the kernel-weighted log-likelihood of the full model and the kernel-weighted log-likelihood of the linear restricted model. We reject $\mathbb{H}_0$ when the value of M_T is sufficiently large. The asymptotic distribution of the test statistic M_T is established in Theorem 3 in Section 3.

3 Asymptotic Theory

3.1 Notations

We first introduce some useful notations. For a generic $n \times 1$ vector $\mathbf{a} = (a_1, \cdots, a_n)^\top$, $\|\mathbf{a}\|_2$ denotes the Euclidean norm (L_2 norm) and is equal to $\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$. Let I_n be the $n \times n$ identity matrix and $\mathbf{1}_n$ be an $n \times 1$ vector where all the elements are equal to 1. Furthermore, we define $\boldsymbol{U}(z_t) = (F_{X_1|z_t}(X_{1t}), \cdots, F_{X_p|z_t}(X_{pt}))$, the derivative function $F^{(i)}_{X|z}(x) = (\partial/\partial z)^i F_{X|z}(x)$, and $\mu_2 = \int v^2 K(v) dv$. We use V to denote the random variable $V = (Z - z)/h$. For the sake of simplicity in presenting the derivation process, we rewrite the function $\ell(\boldsymbol{\eta}(z);\boldsymbol{u}(z))$ as

$$\ell(\boldsymbol{\eta}(z);\boldsymbol{u}(z)) = \ln \sum_{k=1}^{d} \nu_{1,k} c_k \left(F_{\boldsymbol{X}|z}(\boldsymbol{x}_t);\nu_{2,k}\right) \equiv \ell(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2;\boldsymbol{u}(z)),$$

where $\nu_{1,k}$ and $\nu_{2,k}$ are defined in (6) and (7), and $\boldsymbol{\nu}_1 = (\nu_{1,1},\cdots,\nu_{1,d})^\top$ and $\boldsymbol{\nu}_2 = (\nu_{2,1},\cdots,\nu_{2,d})^\top$. Define $\boldsymbol{\nu} = \left(\boldsymbol{\nu}_1^\top,\boldsymbol{\nu}_2^\top\right)^\top$, then the derivative functions can be represented by

$$\ell_i(\boldsymbol{\eta};\boldsymbol{u}) = \ell_i(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2;\boldsymbol{u}) = \frac{\partial^i}{\partial(\tilde{\boldsymbol{\nu}})^i}\ell(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2;\boldsymbol{u}), \quad \ell_{i,j}(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2;\boldsymbol{u}) = \frac{\partial}{\partial u_j}\ell_i(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2;\boldsymbol{u}), \tag{10}$$

where $\tilde{\boldsymbol{\nu}} = \mathbf{1}_2 \otimes \boldsymbol{\nu}$ and u_j is the jth element of vector $\boldsymbol{u}$. Further, we define a matrix function $H(x,y) = diag(x,y) \otimes \mathrm{I}_{2d^\dagger}$, where $\otimes$ is the Kronecker product. Let $\boldsymbol{\Omega}_z = E\left\{H\left(K(V),V^2K(V)\right)\left(-2\ell_2(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t))\right)\Big|Z_t = z\right\}$.

3.2 Regularity Conditions

To study the asymptotic properties of the estimators, we set the following regularity conditions:

A1. For the identified conditional copula model $C(\boldsymbol{u}(z);\boldsymbol{\omega}(z),\boldsymbol{\theta}(z))$, the true weight parameter satisfies $\omega_k(z) \in [0,1]$ and $\sum_{k=1}^{d}\omega_k(z) = 1$ for all $k \in \{1,\cdots,d\}$ and $z \in \Phi_z$, and there is an open subset $\widetilde{\boldsymbol{\Theta}}_k$ of $\boldsymbol{\Theta}_k$ that contains the true parameter $\theta_k(z)$ such that for all $\theta_k(z) \in \widetilde{\boldsymbol{\Theta}}_k$ and $\boldsymbol{u}(z)$, the density $c_k(\boldsymbol{u}(z);\theta_k(z))$ admits all third derivatives with respect to the parameter $\theta_k(z)$.

A2. The distribution F_Z of covariate Z is twice differentiable and continuous on Φ_z. The conditional marginal distribution $F_{X_j|z}(x_j)$ has continuous second order derivatives with respect to z for $j = 1,\cdots,p$.

A3. The continuous quantities $\ell_i(\boldsymbol{\eta};\boldsymbol{u})$ defined in (10) are Liptschitz for all $\boldsymbol{\eta}$ and $i \in \{0,1,2\}$ such that $E(\ell_2(\boldsymbol{\eta}(Z);\boldsymbol{U}(Z))) < 0$ and $\boldsymbol{\Omega}_z$ is invertible for true $\boldsymbol{\eta}(Z)$ and $\boldsymbol{U}(Z)$. Moreover, each Liptschitz continuous quantity is bounded by a function, and the bound function has a finite second moment.

A4. The symmetric bounded kernel function $K(\cdot)$ is twice continuously differentiable on the compact support $[-1,1]$.

A5. As $T \to \infty$, $h \to 0, h_j \to 0, Th^{2+1/\kappa} \to \infty, Th_j^{1+2/\kappa} \to \infty$ for some $\kappa > 0$ and $j \in \{1, \cdots, p\}$.

A6. The strictly stationary process $\{\boldsymbol{X}_t, Z_t\}$ is α-mixing with the mixing coefficient satisfies $\sum_{t=1}^{\infty} \alpha(t)^{\frac{\varsigma}{4+\varsigma}} < \infty$ for some positive constant ς such that

$$\int \left(v^2 K(v)\right)^{\frac{4+\varsigma}{2}} h dF_{\boldsymbol{X},Z}(\boldsymbol{x}, hv + z) < \infty. \tag{11}$$

Conditions A1 - A3 are standard for conditional copula models. They are necessary to ensure the existence of asymptotic properties and to obtain suitable solutions for the dependent parameters and weight parameters. Condition A3 requires that the partial derivatives of $L(\boldsymbol{\eta}(z))$, which is the sum of $L_t(\boldsymbol{\eta}(z))$, as shown in Equation (8), are dominated by a function with a finite second moment. However, Fisher's information matrix is not equal to $E(-\ell_2(\boldsymbol{\eta}(z); \boldsymbol{U}(z)))$. The reason is that the Fisher's information matrix is calculated based on the penalized log-likelihood function $Q(\boldsymbol{\eta}(z))$, rather than $L(\boldsymbol{\eta}(z))$. Condition A4 is common in the nonparametric literature. Condition A5 allows the conditional copula model to work with different bandwidths of the conditional marginal distribution functions, and guarantees the automatic good behavior of the WNW estimator at boundaries, as well as consistency. A similar condition can also be found in Cai (2002).

Condition A6 establishes the α-mixing condition for weakly dependent data, which is commonly found in financial econometric models of time series processes, such as ARMA and ARCH models (Pham and Tran, 1985). The finite summation of mixing coefficients in Condition A6 is critical to the asymptotic normality (Pham and Tran, 1985; Fan and Yao, 2003), and it is not contradictory to $\alpha(t) = O(t^{-(2+\kappa)})$ which is Assumption B3 in Cai (2002). In addition, the finite integration (11) in the last condition implies that the fourth moment of variable Z_t is finite, since

$$E\left(|Z_t|^{4+\varsigma}\right) = \int |hv + z|^{4+\varsigma} dF_{\boldsymbol{X},Z}(\boldsymbol{x}, hv + z)$$

for the bandwidth $h = o(1)$ specified in Condition A5, and

$$\int |hv^2|^{\frac{4+\varsigma}{2}} h dF_{\boldsymbol{X},Z}(\boldsymbol{x}, hv + z) \leq \int (v^2 K(v))^{\frac{4+\varsigma}{2}} h dF_{\boldsymbol{X},Z}(\boldsymbol{x}, hv + z) < \infty$$

follows from Condition A4. Therefore, the finite fourth moment of Z_t, Assumption A8 in Yang et al. (2021), and Assumption C7 in Liu et al. (2022), can be derived for a finite z.

3.3 Asymptotic Properties

We first establish the consistency property of the penalized local log-likelihood estimator $\hat{\boldsymbol{\eta}}(z)$ to $\boldsymbol{\eta}(z)$ in Theorem 1.

Theorem 1. *Under Conditions A1 - A6, if* $\max_{1\le k\le d}\{|\partial^2 P_{\gamma_T}(\omega_k(z))/\partial\omega_k^2(z)|\} \to 0$ *holds, then* $\hat{\boldsymbol{\eta}}(z)$ *is consistent for estimating* $\boldsymbol{\eta}(z)$.

Remark 1. *We can partition* $\boldsymbol{\eta}(z)$ *as an identified subset* $\left(\boldsymbol{\eta}^{\dagger\top}(z), \boldsymbol{\omega}^{*\top}(z)\right)^\top$ *and an unidentified subset* $\left(\boldsymbol{\theta}^{*\top}(z), \boldsymbol{\omega}^{(1)*\top}(z), \boldsymbol{\theta}^{(1)*\top}(z)\right)^\top$. *There exists a unique vector of true parameters, and the estimator is consistent in the identified subset. For the unidentified subset, the true values of parameters can be arbitrary because the corresponding weights are equal to zero.*

We next present Theorem 2, which establishes the oracle properties of the penalized local likelihood estimator.

Theorem 2. *Under Conditions A1 - A6, if* $\lim\inf_{T\to\infty}\lim\inf_{\pi_k\to\omega_k^*(z)+}\frac{P'_{\gamma_T}(\pi_k)}{\gamma_T}$ *is positive and* $\sqrt{Th}\gamma_T \to \infty$, *we have*

(a) Sparsity: $\hat{\boldsymbol{\omega}}^*(z) = \mathbf{0}$.

(b) Asymptotic normality:

$$D_T\left(\hat{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) - \boldsymbol{\Omega}_z^{-1}\boldsymbol{\Xi}_z^*\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_z^{-1}\boldsymbol{\Gamma}_z^*\boldsymbol{\Omega}_z^{-1}\right),$$

where matrix $\boldsymbol{\Gamma}_z^* = E\left\{H\left(K^2(V), V^2K^2(V)\right)\ell_1(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t))\ell_1^\top(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t))\Big|Z_t = z\right\}$ *with the random variable* V *and the* $4d^\dagger \times 4d^\dagger$ *matrix* $D_T = H\left(\sqrt{Th}, \sqrt{Th^3}\right)$, $\boldsymbol{\Xi}_z^* = E\left\{H(K(V), VK(V))\left(h^2\ell_1(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t)) + \frac{\mu_2}{2}\sum_{j=1}^p h_j^2 F_{X_j|Z_t}^{(2)}(X_j)\ell_{1,j}(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t))\right)\Big|Z_t = z\right\}$, *and* $\boldsymbol{\Omega}_z = E\left\{H\left(K(V), V^2K(V)\right)\left(-2\ell_2(\boldsymbol{\eta}^\dagger(Z_t);\boldsymbol{U}(Z_t))\right)\Big|Z_t = z\right\}$.

In Theorem 2, the sparsity result is a desired property for high-dimensional models. Thus, the proposed penalized local likelihood estimator can reduce the complexity of the mixture model by correctly estimating some weight parameters as zeros with probability one when $T \to \infty$. For the asymptotic normality result, the asymptotic bias term $\boldsymbol{\Omega}_z^{-1}\boldsymbol{\Xi}_z^*$ can be decomposed into $\boldsymbol{\Omega}_z^{-1}$, which originates essentially from the estimation of parameter function $\hat{\boldsymbol{\eta}}^\dagger(z)$, and $\boldsymbol{\Xi}_z^*$, which is derived from the estimation of both $\hat{\boldsymbol{\eta}}^\dagger(z)$ and the conditional marginal distributions. The sample versions of $\boldsymbol{\Omega}_z, \boldsymbol{\Xi}_z^*$ and $\boldsymbol{\Gamma}_z^*$ can be used to estimate the covariance, for example, $\boldsymbol{\Omega}_z$ can be estimated by $\widehat{\boldsymbol{\Omega}}_z = T^{-1}\sum_{t=1}^{T}\left\{H\left(K(V_t), V_t^2K(V_t)\right)\left(-2\ell_2(\boldsymbol{\eta}^\dagger(z);\boldsymbol{U}(z))\right)\right\}$. We also provide a feasible approach to construct the confidence intervals via bootstrap technique in Section 4.3.

Next, to establish the asymptotic variance of the test statistic, we define

$$\begin{aligned}\mathcal{M}(Z_t, Z_s) \;=\;& \ell_1^\top\left(\boldsymbol{\eta}^\dagger(Z_t), \boldsymbol{U}(Z_t)\right)E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^\dagger(Z), \boldsymbol{U}(Z)))\right]\times\\ & H\left(h, (Z_t - Z_s)^2\right)\ell_1\left(\boldsymbol{\eta}^\dagger(Z_s), \boldsymbol{U}(Z_s)\right)K^2\left(\frac{Z_t - Z_s}{h}\right)\left(1 + \frac{1}{\sqrt{Th^5}}\right),\end{aligned}$$

where $t \neq s$, and

$$\begin{aligned}\mu_T \;=\;& \mathbf{1}_{4d^\dagger}^\top E\left[\ell_1\left(\boldsymbol{\eta}^\dagger(Z), \boldsymbol{U}(Z)\right)\ell_1^\top\left(\boldsymbol{\eta}^\dagger(Z), \boldsymbol{U}(Z)\right)\right]H(h, 0)\times\\ & E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^\dagger(Z), \boldsymbol{U}(Z)))\right]\mathbf{1}_{4d^\dagger}K^2(0)\left(1 + \frac{1}{\sqrt{Th^5}}\right).\end{aligned}$$

Let $\left\{\widetilde{Z}_t\right\}_{t=1}^T$ be a sequence of independent and identically distributed variables that share the same marginal distribution as that of $\{Z_t\}_{t=1}^T$.

Theorem 3. *Under Conditions A1 - A6, we have, for the GLRT statistic in (9),*

$$r_T M_T \xrightarrow{d} \chi^2(r_T\mu_T)$$

under $\mathbb{H}_0$, where $r_T = T^2\mu_T/E\left(\mathcal{M}^2(\widetilde{Z}_1, \widetilde{Z}_2)\right)$ as $T \to \infty$.

Theorem 3 shows the asymptotic chi-square distribution of the GLRT statistic. Based on the maximum likelihood estimators derived from the restricted and unrestricted models, $\mathbb{H}_0$ can be rejected if the test statistic exceeds a critical value by the theorem above.

To obtain a feasible test statistic, our primary task is to derive a consistent estimator for the quantity $r_T\mu_T$. Given that $\widetilde{Z}_t$ shares the same distribution as Z_t, we follow Gijbels et al. (2021) and consider the following consistent estimators

$$\begin{aligned}\widehat{\mathcal{M}}(\widetilde{Z}_t, \widetilde{Z}_s) &= \ell_1^\top \left(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) \left[\sum_{t=1}^{T} H(K(V_t), (V_t)^2 K(V_t))(-2\ell_2(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)))\right]^{-1} \times \\ & \quad H\left(h, (Z_t - Z_s)^2\right) \ell_1 \left(\hat{\boldsymbol{\eta}}^\dagger(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) K^2\left(V_t - V_s\right) \left(T + \sqrt{\frac{T}{h^5}}\right),\end{aligned}$$

where $V_t = \frac{Z_t - z}{h}$, and

$$\begin{aligned}\widehat{\mu}_T &= \mathbf{1}_{4d^\dagger}^\top \left[\sum_{t=1}^{T} \ell_1 \left(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) \ell_1^\top \left(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right)\right] H\left(h, 0\right) \times \\ & \quad \left[\sum_{t=1}^{T} H(K(V_t), (V_t)^2 K(V_t))(-2\ell_2(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)))\right]^{-1} \mathbf{1}_{4d^\dagger} K^2\left(0\right) \left(1 + \frac{1}{\sqrt{Th^5}}\right),\end{aligned}$$

$$\widehat{r}_T = T^2 \widehat{\mu}_T \left[\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} \widehat{\mathcal{M}}^2\left(\widetilde{Z}_t, \widetilde{Z}_s\right)\right]^{-1}.$$

4 Practical Issues

4.1 EM Algorithm

Of significant interest is the application of the conditional mixture copula model to practical cases and obtaining precise estimates. However, the penalized local maximum likelihood estimator for the local kernel-weighted pseudo penalized log-likelihood function (8) cannot be explicitly expressed in mathematical terms. Hence, we adopt a numerical optimization method, the EM algorithm proposed by Dempster et al. (1977), to solve this computational issue and obtain the numerical maximum likelihood estimates. The EM algorithm decomposes the process of finding the maximum likelihood estimator into two steps: first, for each $k \in \{1, \cdots, d\}$, the expectation step computes and updates $\hat{\omega}_k(z)$ for each candidate copula, and second, the maximization step maximizes the local kernel-weighted pseudo penalized log-likelihood function to estimate the remaining parameters $\omega_k^{(1)}(z)$, and $\widetilde{\boldsymbol{\theta}}_k(z) = \left(\theta_k(z), \theta_k^{(1)}(z)\right)^\top$.

We take the first derivative of $Q(\boldsymbol{\eta}(z))$ with respect to $\omega_k(z)$ and let it equal zero:

$$\sum_{t=1}^{T}\left\{\frac{\left(\omega_k(z)+\omega_k^{(1)}(z)(z_t-z)\right)c_k\left(\widehat{F}_{\boldsymbol{X}|z}(\boldsymbol{x}_t);\theta_k(z)+\theta_k^{(1)}(z)(z_t-z)\right)K_h(z_t-z)}{\sum_{k=1}^{d}\left[\left(\omega_k(z)+\omega_k^{(1)}(z)(z_t-z)\right)c_k\left(\widehat{F}_{\boldsymbol{X}|z}(\boldsymbol{x}_t);\theta_k(z)+\theta_k^{(1)}(z)(z_t-z)\right)\right]}\right\}$$
$$=TP_{\gamma_T}^{'}\left(\omega_k(z)\right)-\rho. \tag{12}$$

For the sake of brevity, we use $\boldsymbol{B}_k$ to denote the left hand side of Equation (12). By multiplying by $\omega_k(z)$ on both sides of (12), we have

$$\omega_k(z)\boldsymbol{B}_k=TP_{\gamma_T}^{'}\left(\omega_k(z)\right)\omega_k(z)-\rho\omega_k(z). \tag{13}$$

Then, taking summation on both sides for all k, we obtain

$$\rho=\sum_{k=1}^{d}\omega_k(z)\left[TP_{\gamma_T}^{'}\left(\omega_k(z)\right)-\boldsymbol{B}_k\right]. \tag{14}$$

At the expectation step, letting $\hat{\boldsymbol{\eta}}^{[0]}(z)$ be the initial values, we can compute the corresponding $\rho^{[0]}$ by Equation (14) at a given point z. Then, we can update $\hat{\omega}_k^{[0]}(z)$ to

$$\hat{\omega}_k^{[1]}(z)=\frac{\hat{\omega}_k^{[0]}(z)}{\rho^{[0]}}\left(TP_{\gamma_T}^{'}\left(\hat{\omega}_k^{[0]}(z)\right)-\boldsymbol{B}_k^{[0]}\right)$$

by plugging $\rho^{[0]}$ back into Equation (13) for $k=1,\cdots,d$.

At the maximization step, we aim to estimate the remaining parameters in the conditional mixture copula model (2). With $\hat{\omega}_k^{[1]}(z)$ updated from the expectation step above for $k=1,\cdots,d$, we can update $\left(\hat{\omega}_k^{(1)}(z)\right)^{[0]}$ to

$$\left(\hat{\omega}_k^{(1)}(z)\right)^{[1]}=\left(\hat{\omega}_k^{(1)}(z)\right)^{[0]}-\frac{Q_{\omega_k^{(1)}(z)}^{'}\left(\hat{\omega}_k^{[1]}(z),\left(\hat{\omega}_k^{(1)}(z)\right)^{[0]},\hat{\tilde{\boldsymbol{\theta}}}_k^{[0]}(z)\right)}{Q_{\omega_k^{(1)}(z)}^{''}\left(\hat{\omega}_k^{[1]}(z),\left(\hat{\omega}_k^{(1)}(z)\right)^{[0]},\hat{\tilde{\boldsymbol{\theta}}}_k^{[0]}(z)\right)} \tag{15}$$

by the Newton-Raphson algorithm, where $Q_{\omega_k^{(1)}(z)}^{'}(\cdot)$ and $Q_{\omega_k^{(1)}(z)}^{''}(\cdot)$ are the first and second derivative functions of $Q(\cdot)$ with respect to $\omega_k^{(1)}(z)$ respectively, and $\hat{\tilde{\boldsymbol{\theta}}}_k^{[0]}(z)=\left(\hat{\theta}_k^{[0]}(z),\left(\hat{\theta}_k^{(1)}(z)\right)^{[0]}\right)^{\top}$. Due to the presence of several basis copulas, the second derivative of $Q(\boldsymbol{\eta}(z))$ with respect to $\widetilde{\boldsymbol{\theta}}_k(z)$ may not have an explicit expression. Hence, we adopt the gradient ascent method to find the numerical solution

$$\hat{\widetilde{\boldsymbol{\theta}}}_k^{[1]}(z) = \hat{\widetilde{\boldsymbol{\theta}}}_k^{[0]}(z) + s_k \cdot \bigtriangledown Q_{\widetilde{\boldsymbol{\theta}}_k(z)} \left(\hat{\omega}_k^{[1]}(z), \left(\hat{\omega}_k^{(1)}(z) \right)^{[1]}, \hat{\widetilde{\boldsymbol{\theta}}}_k^{[0]}(z) \right), \tag{16}$$

where s_k is the learning rate (step size), and $\bigtriangledown Q_{\widetilde{\boldsymbol{\theta}}_k(z)}$ is a gradient vector of $Q(\boldsymbol{\eta}(z))$ with respect to $\widetilde{\boldsymbol{\theta}}_k(z)$. Note that we can use backtracking line search to adaptively choose s_k.

4.2 Selection of Bandwidth and Tuning Parameter

The selection of the bandwidth h and the tuning parameter γ_T is crucial in the estimation of copula parameters as the bandwidth h plays a key role in balancing the trade-off between the bias and the variance of nonparametric estimators, while the tuning parameter γ_T regulates the weight of the penalty term. However, traditional methods such as multifold cross-validation, information criterion, and plug-in methods may not perform well for weakly dependent data. To address this issue, we recommend using the forward leave-one-out cross-validation method, proposed by Yang et al. (2022), to select the optimal bandwidth h^* and suitable tuning parameter γ_T^* simultaneously.

Specifically, we define $\hat{\boldsymbol{\eta}}(z_t; h, \gamma_T)$ as the estimates for the penalized conditional mixture copula model with known bandwidth h and tuning parameter γ_T. We use the data $\{x_{1t}, \cdots, x_{pt}, z_t : t < t^*\}$ to construct the estimates $\hat{\boldsymbol{\eta}}(z_{t^*}; h, \gamma_T)$ and $\hat{\boldsymbol{u}}(z_{t^*})$ at the corresponding sample point $\{x_{1t^*}, \cdots, x_{pt^*}, z_{t^*}\}$ for each $t_0 + 1 \le t^* \le T$, where t_0 is the minimum window size used to estimate $\hat{\boldsymbol{\eta}}(z_{t_0+1}; h, \gamma_T)$. By employing the forward recursive approach, the sequential estimators $\{\hat{\boldsymbol{\eta}}(z_{t^*}; h, \gamma_T)\}_{t^*=t_0+1}^{T}$ can be constructed. Therefore, the forward leave-one-out cross-validation estimators of the optimal bandwidth and tuning parameter can be obtained by maximizing the objective function

$$(h^*, \gamma_T^*) = \arg\max_{(h,\gamma_T)} \sum_{t^*=t_0+1}^{T} \left(\ell\left(\hat{\boldsymbol{\eta}}(z_{t^*}; h, \gamma_T); \hat{\boldsymbol{u}}(z_{t^*}) \right) \Big| h, \gamma_T \right). \tag{17}$$

In practice, we can choose the initial bandwidth $h = cT^{-1/5}$ for some constant c by the rule of thumb when maximizing the objective functions above, and Condition A5 is still satisfied as long as $\kappa > \frac{1}{3}$.

4.3 Construction of Confidence Intervals

As economic or financial series data are time series, the traditional independent and identically distributed bootstrap technique may not yield accurate results in practice. To address this issue, we adopt the block bootstrap method in this article to construct the pointwise confidence intervals of the copula weight and dependent parameters for serially dependent data. This method preserves the original time series dependence within each block by partitioning the series data into several blocks (Politis and Romano, 1994). The block bootstrap method has also been employed in previous studies, such as Patton (2013) and Yang et al. (2022).

We first describe a stationary bootstrap resampling method to generate sample series from observations. Specifically, we define the block $B_{t,b} = \{\boldsymbol{y}_t, \boldsymbol{y}_{t+1}, \cdots, \boldsymbol{y}_{t+b-1}\}$, where $\{\boldsymbol{y}_t\}_{t=1}^T \equiv \{x_{1t}, \cdots, x_{pt}, z_t\}_{t=1}^T$ is a strictly stationary and weakly dependent time series. In the case $j > T$, $\boldsymbol{y}_j$ is defined as $\boldsymbol{y}_t$, where $t = j (\mathrm{mod}\ T)$ and $\boldsymbol{y}_0 = \boldsymbol{y}_T$. To construct the blocks, let $\{\varpi_k\}_{k=1}^{\infty}$ be a sequence of i.i.d. random variables from the geometric distribution with probability $p = T^{-1/3}$. Independent of both $\{\boldsymbol{y}_t\}_{t=1}^T$ and ϖ_k, let $\{\varpi_k'\}_{k=1}^{\infty}$ be a sequence of i.i.d. random variables from the discrete uniform distribution on $\{1, \cdots, T\}$.

A pseudo time series $\{\boldsymbol{y}_t^*\}_{t=1}^T$ can be generated through the following process. Sample a sequence of blocks of random length by the prescription $\left\{B_{\varpi_k', \varpi_k}\right\}_{k=1}^{\infty}$. The first ϖ_1 observations in the pseudo time series $\{\boldsymbol{y}_t^*\}_{t=1}^T$ are determined by the first block B_{ϖ_1', ϖ_1} of observations $\boldsymbol{y}_{\varpi_1'}, \cdots, \boldsymbol{y}_{\varpi_1' + \varpi_1 - 1}$, the next ϖ_2 observations in the pseudo time series are the observations in the second sampled block B_{ϖ_2', ϖ_2}, namely $\boldsymbol{y}_{\varpi_2'}, \cdots, \boldsymbol{y}_{\varpi_2' + \varpi_2 - 1}$. This process is repeated until T observations have been generated for the pseudo time series.

Our target in this section is to construct confidence intervals. Therefore, we provide some details of the practical implementation based on the block bootstrap method below.

Step 1. Generate a pseudo sample series $\{\boldsymbol{y}_t^*\}_{t=1}^T$ from the original observations $\{\boldsymbol{y}_t\}_{t=1}^T$ by using the stationary bootstrap technique described above.

Step 2. Calculate the corresponding $\hat{\boldsymbol{\eta}}^*(z)$ at grid point z with the WNW estimates $\widehat{F}_{X_j|z}^*(x_{jt})$ for $j \in \{1, \cdots, p\}$.

Step 3. Get R values of $\hat{\boldsymbol{\eta}}^*(z)$ at each z if we repeat the first two steps R times with a large integer R (says, $R = 1000$). Denote the $\frac{\alpha}{2}$-th and $\left(1 - \frac{\alpha}{2}\right)$-th percentiles of the sequence $\{\hat{\boldsymbol{\eta}}^*(z)\}$ by $q_{\frac{\alpha}{2}}$ and $q_{1-\frac{\alpha}{2}}$. Then, the interval $\left(q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}\right)$ is the empirical $100(1-\alpha)\%$ confidence interval for $\hat{\boldsymbol{\eta}}(z)$.

5 Numerical Studies

In this section, we investigate the finite-sample performance of our estimation and model selection procedures through Monte Carlo simulations. For convenience, we define a mixture copula consisting of the Clayton, Gumbel, and Frank copulas. They are widely used in empirical studies because they can describe different dependence structures. Specifically, the Clayton copula exhibits strong lower tail dependence, and can capture cases such as two markets that are likely to crash simultaneously. The Gumbel copula shows strong upper tail dependence and can be an appropriate model when two markets are likely to boom together. The Frank copula exhibits symmetric tail dependence.

The working mixture copula model is then formulated as

$$\begin{aligned} C_{z_t}(u_1,u_2) = \omega_{Cl}(z_t)C_{Cl}\left(u_1(z_t),u_2(z_t);\theta_{Cl}(z_t)\right) + \\ \omega_{Gu}(z_t)C_{Gu}(u_1(z_t),u_2(z_t);\theta_{Gu}(z_t)) + \\ \omega_{Fr}(z_t)C_{Fr}(u_1(z_t),u_2(z_t);\theta_{Fr}(z_t)), \end{aligned} \tag{18}$$

where $C_{Cl}(\cdot)$, $C_{Gu}(\cdot)$, and $C_{Fr}(\cdot)$ denote the Clayton, Gumbel, and Frank copulas, respectively. Similar to Abegaz et al. (2012), we generate the covariate z from the truncated normal distribution within $[-2,2]$ with mean 0 and variance 9, and then consider three different types of the copula dependent parameter function $\theta(z)$:

Type 1: $\theta(z) = 10 - 1.5z^2$;

Type 2: $\theta(z) = 10 - 0.02z^2 + 0.4z^3$;

Type 3: $\theta(z) = 3 + z + 2\exp(-2z^2)$.

For the two conditional marginal distributions, given a specific level of z, we assume u_1 and u_2 respectively denote the cumulative distribution functions of $X_{1t} \sim N(\exp(z/2),1)$ and $X_{2t} \sim N(\exp(z/2),2)$. For each generated sample, we calculate the estimates $\hat{\theta}$ at 101 equally-spaced grid points $z_i = -1.95 + 0.039i$ for $i \in \{0,1,...,100\}$. We use local linear fitting with the Epanechnikov kernel. Each simulation is repeated $M = 1000$ times with sample size $T \in \{200,400,800\}$.

We start with a scenario in which data are generated from a single copula. That is, the true model is an individual copula selected from the three candidates. For each individual copula used to generate data, we assume that the function of the copula dependent parameter θ follows one of the three types listed above. Then, we fit Equation (18) to the generated data and investigate the performance of the proposed method. To assess the performance of our method, we report the percentage that each copula is correctly (incorrectly) selected and the mean squared errors (MSEs) of the copula dependent parameter estimates over the 101 grid points, which is defined as

$$MSE(\hat{\theta}) = \frac{1}{M}\frac{1}{101}\sum_{j=1}^{M}\sum_{i=1}^{101}\left(\hat{\theta}_j(z_i) - \theta(z_i)\right)^2.$$

Table 1 documents the MSEs of copula dependent parameter estimates by the proposed method for the three functional forms of $\theta(z)$ as the sample size T increases from 200 to 800, and the accurate (inaccurate) rate of copula selection. We draw three observations from Table 1. First, the MSEs of the copula dependent parameter estimates remarkably decrease as the sample size increases. Second, the accuracy rate of copula selection is high in all three panels and robust to the different combinations of sample size and functional form of $\theta(z)$. Third, as the sample size increases, the inaccurate rate of copula selection decreases, and becomes negligible when $T = 800$. In Figure 1, we provide visual evidence of the performance of the proposed method by sketching the paths of the estimated copula dependent parameters along the covariate z when $T = 400$. For ease of comparison, we transform the copula dependent parameters into Kendall's τ. In each panel, the black solid line denotes the actual path of Kendall's τ, which respectively follows Types 1 - 3 of $\theta(z)$, and the other two dashed curves denote the mean (red) and median (blue) of the copula dependent parameter estimates at the 101 grid points from the 1000 simulations. The two green dotted curves connect the 5% and 95% percentiles of the copula dependent parameter estimates at the 101 grid points. As seen therein, both the mean and median curves closely track the actual path in all three types of $\theta(z)$, indicating a reasonably good performance of the proposed method in estimating the parameter of the selected copula model.

Next, we examine the performance of the proposed method when the true model is a mixture of two copulas. Specifically, for the three-component mixture copula model (18), we assume two candidate copulas' weights respectively equal to $(1 + z)^2/29 + 0.3$ and $1 - ((1 + z)^2/29 + 0.3)$, while the remaining copula's weight uniformly equals zero. For the two component copulas with nonzero weights, we further assume that their copula dependent parameter functions $\theta(z)$ follow different patterns determined by the three models discussed above. Table 2 presents the MSEs of the copula dependent parameter estimates and the accurate (inaccurate) rates of copula selection by the proposed method. For instance, Panel (a) corresponds to the case where the true model is a combination of the Clayton and Gumbel copulas. The results indicate that the MSEs of the two copulas' dependent parameter estimates significantly decrease as the sample size T increases from 200 to 800. Moreover, the likelihood of selecting Gumbel gradually increases as T increases, while the likelihood of inaccurately selecting Frank substantially decreases. Similar patterns can be observed in Panels (b) and (c) when the true models are combinations of Clayton and Frank, and Gumbel and Frank, respectively. As in the single copula scenario, we also sketch the paths of the estimated dependent parameters of the selected copulas. To conserve space, here, we display only the three combinations of dependent parameter functions when the true model is a combination of the Clayton and Gumbel copulas. For example, in Figure 2(a), the left panel shows the true path (Type 1), the mean, and the median of the estimated paths by the proposed method for Clayton, while the right panel exhibits the corresponding results for Gumbel. Figure 2 shows that the copula dependent parameters of the Clayton-Gumbel mixture can be reasonably well estimated by the proposed method, as the mean and median paths closely track the actual paths of the dependent parameter functions. Furthermore, results in Table 3 indicate that the MSEs of the corresponding weight estimates decrease in all cases when T increases. In Figure 3, we additionally draw the actual, mean, and median paths of weight parameters of the Clayton-Gumbel mixture over the covariate Z. Both the mean and median paths of the weight estimates track the true paths closely, implying good performance of the proposed method in estimating the weights of the Clayton-Gumbel combination.

6 An Empirical Illustration

In this section, we implement the proposed method to explore the comovement of the housing markets in four states in the United States: California (CA), Colorado (CO), Massachusetts (MA), and Washington (WA). These states were chosen because they had the highest median home prices in 2020, according to information released by Zillow.[1] We obtain the four states' quarterly housing price indices (HPIs) from 1975:Q1 to 2022:Q4 from the Federal Housing Finance Agency. Figure 4 presents the HPI trajectories for these states. The HPIs in all four states exhibit similar patterns over the sample period. Housing prices showed a significant increase after 2000 but declined considerably following the subprime mortgage crisis in 2007. Thanks to the Federal Reserve's quantitative easing policy, housing prices resumed their upward trend after 2011 and saw a sharp increase at the onset of the COVID pandemic in 2020.

Several economic factors have the potential to impact the housing market. For simplicity, we examine only the influence of the interest rate in this study. The interest rate can directly affect mortgage rates and purchasing costs. Additionally, the interest rate is closely linked with the business cycle and serves as an effective monetary tool for the Federal Reserve to adjust the macroeconomy. In this study, we use the quarterly Federal Funds Effective Rate, retrieved from FRED (`https://fred.stlouisfed.org/`), to proxy the interest rate, and sketch its trajectory in Figure 5.

We compute the quarterly growth rates of the four states' HPIs and then present some summary statistics. As shown in Table 4, the upper panel reveals that the housing prices in all four states experienced positive average growth rates over the past few decades, although CA and WA exhibited larger fluctuations than the other two states. Moreover, the growth rates in CA and WA are left-skewed, while those in CO and MA are right-skewed. The Jarque-Bera test results indicate that the null hypothesis of normality of the growth rates is rejected for all states except MA, whose kurtosis is close to 3. Panel (b) illustrates the linear (Pearson) correlation coefficients across the four housing markets. The results reveal that, among the six pairs, the correlation between CA and WA is the highest, which is not surprising given the cluster of high-tech companies in both states.

Furthermore, to estimate the conditional marginal distribution of the four series, we use the WNW estimator formulated in (3), and then fit the estimated conditional marginals to the conditional mixture copula model outlined in (18). Due to space constraints, here we present only the copula parameter estimates of the conditional mixture copula model.[2] Among the six pairs of local housing markets, the proposed method

[1] `https://www.businessinsider.com/average-home-prices-in-every-state-washington-dc-2019-6`

[2] We conduct the hypothesis test proposed in Section 2.3 and find that the null hypothesis of constant copula parameters could be rejected at the 5% level in all six pairs. The test results, as well as the estimates of the marginal distributions, are not tabulated but are available upon request.

chooses a single Gumbel copula for CA-CO, CA-WA, CO-WA, and MA-WA. The estimated paths of Kendall's τ for the single Gumbel copula are shown in Figure 6, with the shaded area representing the 95% confidence interval. As seen therein, the degree of dependence decreases as the interest rate increases. This pattern is particularly evident in the CA-WA pair, where Kendall's τ drops significantly from 0.8 to approximately 0.05 as the interest rate increases. This finding indicates a strong comovement between the two western coast states when the Fed conducted several rounds of quantitative easing over the past decade.

For the remaining two pairs, CA-MA and CO-MA, the proposed method yields a mixture of Gumbel and Frank copulas. For CA-MA, Panel (a) in Figure 7 suggests that the degree of dependence increases remarkably when the interest rate decreases, indicating a higher probability of upward comovement between the two housing markets. The weight associated with the Gumbel copula, displayed by Panel (c), also increases as the interest rate decreases. On the other hand, although the parameter of the Frank copula fluctuates along the interest rate, it briefly falls within the range between 0 and 0.2, and the associated weight declines when the interest rate decreases, as displayed by Panels (b) and (d). For CO-MA, we can observe a similar pattern in which the HPIs between the two housing markets are more likely to increase simultaneously when the interest rate is low. The degree of dependence is also high for a low interest rate. We further consider the 4-dimensional mixture copula which contains the four series simultaneously. Results in Figure 9 show that the Gumbel and Frank copulas are selected, and the four markets exhibit strong comovement when the interest rate is low, and the weight parameter of the Gumbel copula is also relatively larger than that of the Frank copula.

7 Conclusion

In this article, we propose a semiparametric conditional mixture copula model, estimating both conditional marginal distributions and copula parameters as functions of a covariate. We use a two-step procedure: first, estimate the conditional marginal distributions using a nonparametric method, and then, simultaneously estimate the model and select the copula through a penalized local log-likelihood function. We provide a generalized likelihood ratio test statistic for the covariate effects. We establish the asymptotic properties of the estimators and discuss algorithmic and selection issues. Monte Carlo simulations demonstrate the method's effectiveness, and we apply it to study interest rate effects on housing market dependence. Importantly, the conditional mixture copula model in this study allows for multidimensional covariates. However, researchers may encounter the curse of dimensionality when too many covariates are involved. In such cases, we suggest using a mixture single index copula model that extends the single index copula model proposed in Fermanian and Lopez (2018) and Yang et al. (2021). The estimation and selection for a mixture single index copula model warrant future investigation.

References

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *Journal of Multivariate Analysis*, 110(C):43–73.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67(2):445–453.

Acar, E. F., Craiu, R. V., and Yao, F. (2013). Statistical testing of covariate effects in conditional copula models. *Electronic Journal of Statistics*, 7:2822–2850.

Almeida, C. and Czado, C. (2012). Efficient bayesian inference for stochastic time-varying copula models. *Computational Statistics & Data Analysis*, 56(6):1511–1527.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory*, 18(1):169–192.

Cai, Z. and Tiwari, R. C. (2000). Application of a local linear autoregressive model to bod time series. *Environmetrics*, 11(3):341–350.

Cai, Z. and Wang, X. (2014). Selection of mixed copula model via penalized likelihood. *Journal of the American Statistical Association*, 109.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Derumigny, A. and Fermanian, J.-D. (2017). About tests of the "simplifying" assumption for conditional copulas. *Dependence Modeling*, 5(1):154–197.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. Chapman & Hall, London.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer, New York.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics*, 10(3):245–271.

Fan, Y. and Patton, A. J. (2014). Copulas in econometrics. *Annu. Rev. Econ.*, 6(1):179–200.

Fermanian, J.-D. and Lopez, O. (2018). Single-index copulas. *Journal of Multivariate Analysis*, 165:27–55.

Fermanian, J.-D. and Wegkamp, M. H. (2012). Time-dependent copulas. *Journal of Multivariate Analysis*, 110:19–29.

Giacomini, E., Härdle, W., and Spokoiny, V. (2009). Inhomogeneous dependence modeling with time-varying copulae. *Journal of Business & Economic Statistics*, 27(2):224–234.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2021). Omnibus test for covariate effects in conditional copula models. *Journal of Multivariate Analysis*, 186:104804.

Hafner, C. M. and Manner, H. (2012). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics*, 27(2):269–295.

Hu, L. (2006). Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics*, 16:717–729.

Liu, G., Long, W., Yang, B., and Cai, Z. (2022). Semiparametric estimation and model selection for conditional mixture copula models. *Scandinavian Journal of Statistics*, 49(1):287–330.

Liu, G., Long, W., Zhang, X., and Li, Q. (2019). Detecting financial data dependence structure by averaging mixture copulas. *Econometric Theory*, 35(4):777–815.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18.

Patton, A. (2013). Copula methods for forecasting multivariate time series. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 899–960. Elsevier.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19(2):297–303.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Yang, B., Cai, Z., Hafner, C. M., and Liu, G. (2022). Time-varying mixture copula models with copula selection. *Statistica Sinica*, 32:1049–1077.

Yang, B., Hafner, C. M., Liu, G., and Long, W. (2021). Semiparametric estimation and variable selection for single-index copula models. *Journal of Applied Econometrics*, 36(7):962–988.

Table 1: MSEs of copula dependent parameter estimates and accurate (inaccurate) rates of selection when the true model is a single conditional copula

		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Panel (a): True copula: Clayton										
Type 1	MSE	0.896	-	-	0.619	-	-	0.488	-	-
	Rate	1.000	(0.000)	(0.023)	1.000	(0.000)	(0.011)	1.000	(0.000)	(0.000)
Type 2	MSE	0.351	-	-	0.257	-	-	0.189	-	-
	Rate	1.000	(0.000)	(0.019)	1.000	(0.000)	(0.008)	1.000	(0.000)	(0.000)
Type 3	MSE	0.493	-	-	0.369	-	-	0.255	-	-
	Rate	1.000	(0.000)	(0.012)	1.000	(0.000)	(0.003)	1.000	(0.000)	(0.000)
Panel (b): True copula: Gumbel										
Type 1	MSE	-	0.154	-	-	0.069	-	-	0.046	-
	Rate	(0.000)	1.000	(0.000)	(0.000)	1.000	(0.000)	(0.000)	1.000	(0.000)
Type 2	MSE	-	0.281	-	-	0.209	-	-	0.131	-
	Rate	(0.000)	1.000	(0.000)	(0.000)	1.000	(0.000)	(0.000)	1.000	(0.000)
Type 3	MSE	-	0.186	-	-	0.143	-	-	0.103	-
	Rate	(0.000)	1.000	(0.071)	(0.000)	1.000	(0.000)	(0.000)	1.000	(0.000)
Panel (c): True copula: Frank										
Type 1	MSE	-	-	0.494	-	-	0.376	-	-	0.235
	Rate	(0.000)	(0.012)	0.998	(0.000)	(0.000)	1.000	(0.000)	(0.000)	1.000
Type 2	MSE	-	-	0.803	-	-	0.641	-	-	0.457
	Rate	(0.000)	(0.005)	0.997	(0.000)	(0.000)	1.000	(0.000)	(0.000)	1.000
Type 3	MSE	-	-	0.374	-	-	0.224	-	-	0.104
	Rate	(0.000)	(0.006)	1.000	(0.000)	(0.000)	1.000	(0.000)	(0.000)	1.000

Notes. This table presents the MSEs of copula dependent parameter estimates and accurate (inaccurate) selection rates of the proposed conditional mixture copula method when the true model is a single Clayton (panel (a)), Gumbel (panel (b)), and Frank (panel (c)) copula. Values in parentheses denote the inaccurate selection rate. Each simulation is repeated 1000 times.

Table 2: MSEs of copula dependent parameter estimates and accurate (inaccurate) rates of selection when the true model is a combination of two conditional copulas

		Panel (a): True Model: Clayton + Gumbel								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	4.087	3.248	-	3.120	2.689		2.386	1.981	-
	Rate	1.000	0.944	(0.176)	1.000	0.961	(0.103)	1.000	0.985	(0.062)
Type 1 + Type 3	MSE	3.205	3.117	-	2.575	2.310		1.933	1.702	-
	Rate	1.000	0.917	(0.113)	1.000	0.948	(0.076)	1.000	0.972	(0.051)
Type 2 + Type 3	MSE	4.636	4.107	-	3.425	2.804		2.545	2.112	-
	Rate	1.000	0.936	(0.158)	1.000	0.955	(0.117)	1.000	1.000	(0.083)
		Panel (b): True Model: Clayton + Frank								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	3.719	-	3.662	2.541	-	2.324	1.917	-	1.785
	Rate	1.000	(0.025)	0.943	1.000	(0.000)	0.955	1.000	(0.000)	0.992
Type 1 + Type 3	MSE	3.248	-	3.358	2.388	-	2.439	1.733	-	1.954
	Rate	1.000	(0.005)	0.908	1.000	(0.001)	0.941	1.000	(0.000)	0.983
Type 2 + Type 3	MSE	3.529	-	3.144	2.602	-	2.286	2.115	-	1.521
	Rate	1.000	(0.011)	0.921	1.000	(0.004)	0.962	1.000	(0.001)	0.976
		Panel (c): True Model: Gumbel + Frank								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	-	0.785	0.824	-	0.503	0.481	-	0.249	0.203
	Rate	(0.005)	0.985	0.913	(0.002)	1.000	0.992	(0.000)	1.000	0.996
Type 1 + Type 3	MSE	-	0.846	0.793	-	0.614	0.477	-	0.220	0.211
	Rate	(0.004)	0.973	0.884	(0.001)	0.993	0.952	(0.000)	0.998	0.973
Type 2 + Type 3	MSE	-	0.803	0.732	-	0.553	0.392	-	0.268	0.275
	Rate	(0.004)	0.977	0.944	(0.000)	0.980	0.985	(0.000)	0.997	0.991

Notes. This table presents the MSEs of copula dependent parameter estimates and accurate (inaccurate) selection rates of the proposed conditional mixture copula method when the true model is a combination of two conditional copulas: Clayton and Gumbel (panel (a)), Clayton and Frank (panel (b)), and Gumbel and Frank (panel (c)). Values in parentheses denote the inaccurate selection rate. Each simulation is repeated 1000 times.

Table 3: MSEs of copula weight parameter estimates when the true model is a combination of two conditional copulas

		Panel (a): True Model: Clayton + Gumbel								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	0.013	0.035	-	0.009	0.023		0.004	0.014	-
Type 1 + Type 3	MSE	0.036	0.022	-	0.022	0.012		0.012	0.007	-
Type 2 + Type 3	MSE	0.028	0.030	-	0.016	0.019		0.010	0.010	-
		Panel (b): True Model: Clayton + Frank								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	0.011	-	0.019	0.006	-	0.010	0.004	-	0.006
Type 1 + Type 3	MSE	0.021	-	0.018	0.012	-	0.012	0.007	-	0.003
Type 2 + Type 3	MSE	0.024	-	0.019	0.011	-	0.008	0.006	-	0.005
		Panel (c): True Model: Gumbel + Frank								
		$T = 200$			$T = 400$			$T = 800$		
		Clayton	Gumbel	Frank	Clayton	Gumbel	Frank	Clayton	Gumbel	Frank
Type 1 + Type 2	MSE	-	0.010	0.015	-	0.008	0.006	-	0.003	0.003
Type 1 + Type 3	MSE	-	0.012	0.011	-	0.007	0.006	-	0.003	0.004
Type 2 + Type 3	MSE	-	0.014	0.014	-	0.007	0.009	-	0.004	0.006

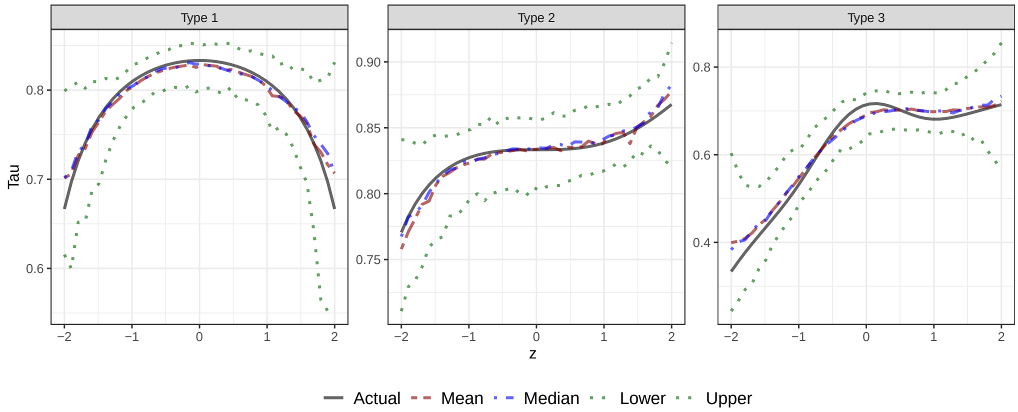
Notes. This table presents the MSEs of copula weight parameter estimates of the proposed conditional mixture copula method when the true model is a combination of two conditional copulas: Clayton and Gumbel (panel (a)), Clayton and Frank (panel (b)), and Gumbel and Frank (panel (c)). Each simulation is repeated 1000 times.

Table 4: Summary Statistics

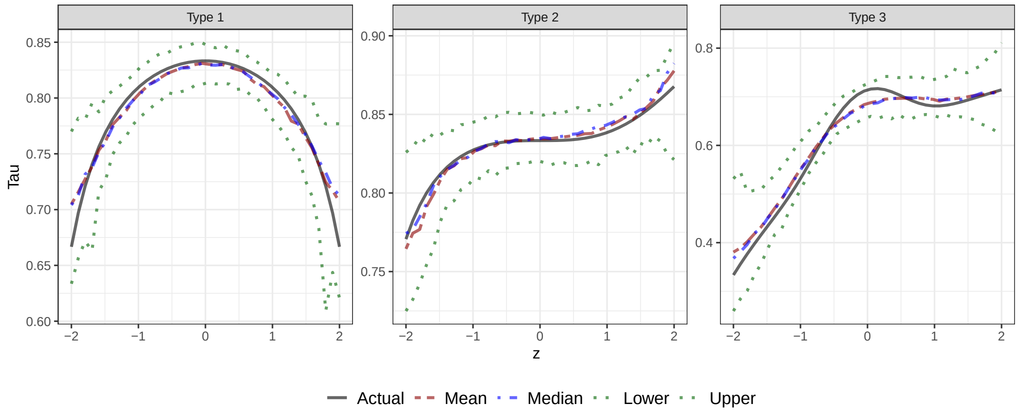
	CA	CO	MA	WA	Interest Rate (Z)
Panel (a): Summary statistics					
Mean	0.0166	0.0145	0.0149	0.0162	4.6676
Median	0.0186	0.0129	0.0123	0.0132	4.8185
Min	-0.1132	-0.0361	-0.0448	-0.1644	0.0588
Max	0.1050	0.0876	0.0758	0.1041	17.7869
Std. Dev	0.0283	0.0195	0.0215	0.0258	3.9869
Skewness	-0.6393	0.7515	0.3187	-1.3384	0.9074
Kurtosis	6.2791	4.9715	3.1530	15.2772	3.6921
Jarque-Bera	98.5810***	48.9081***	3.4196	1256.6102***	30.0223***
Panel (b): Linear correlation coefficients					
	CA	CO	MA	WA	
CA	1.0000	0.2768	0.4170	0.7247	
CO	0.2768	1.0000	0.3370	0.3494	
MA	0.4170	0.3370	1.0000	0.1645	
WA	0.7247	0.3494	0.1645	1.0000	

Notes. The sample period is between 1975:Q1 and 2022:Q4. Panel (a) presents the summary statistics of the quarterly growth of HPIs in CA, CO, MA, and WA, and the quarterly Federal Funds Effective rate. Jarque-Bera denotes the statistic of the Jarque-Bera test with the null hypothesis of normality. *** indicates rejection of the null at 1%. Panel (b) presents the linear correlation coefficients among the four housing markets' quarterly growth rate.

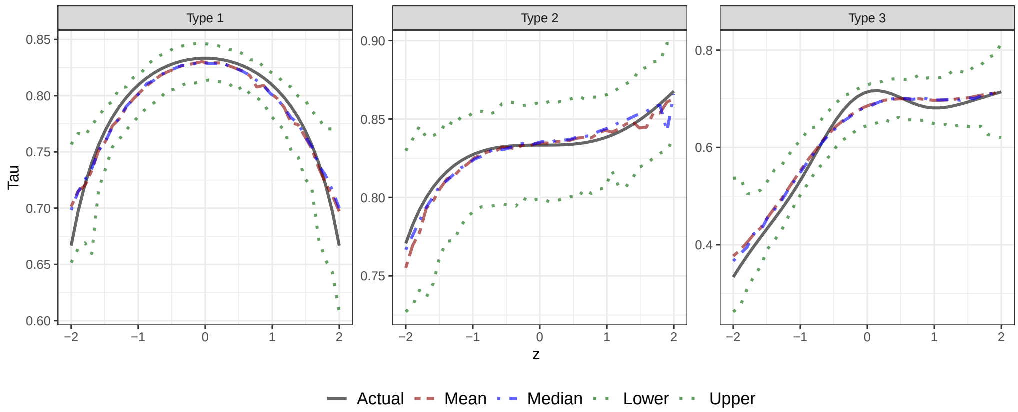
Figure 1: Estimated paths for copula dependent parameters (Kendall's τ) when the true model is a conditional individual copula



(a) True copula: Clayton



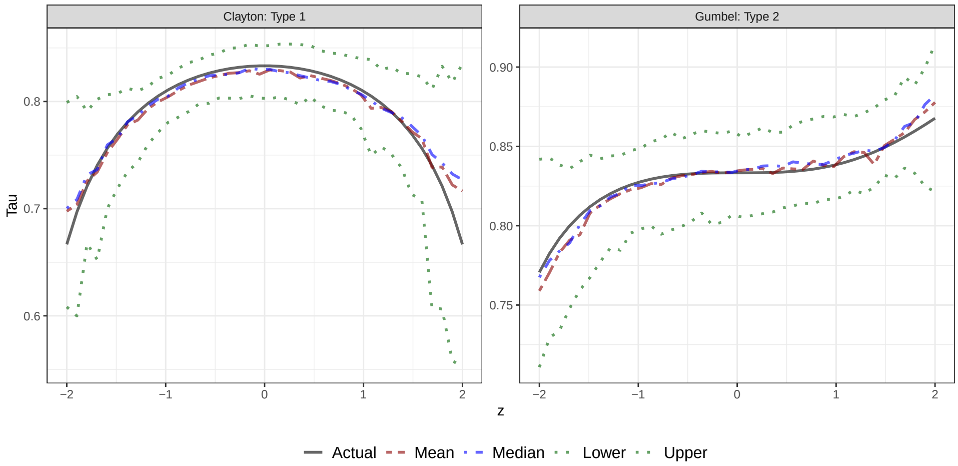
(b) True copula: Gumbel



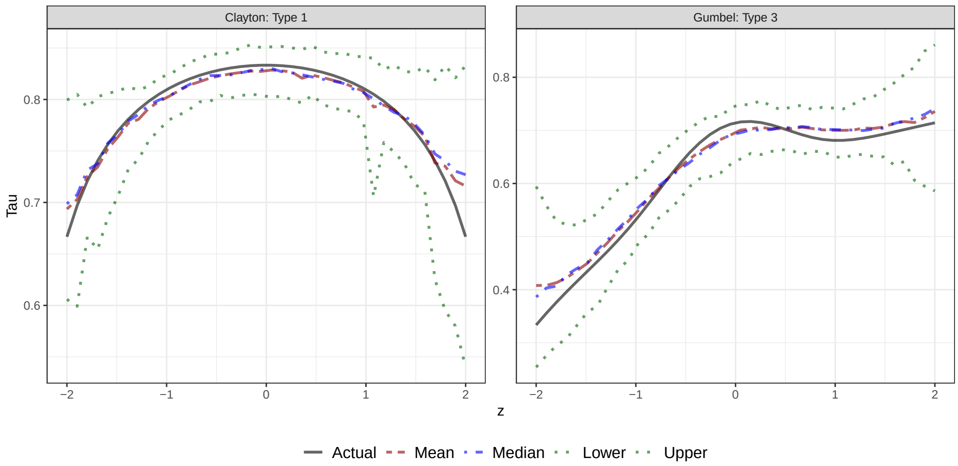
(c) True copula: Frank

Notes. This figure displays the estimated paths of copula dependent parameters (Kendall's τ) when the true mode is an individual conditional Clayton copula (panel (a)), Gumbel copula (panel (b)), and Frank copula (panel (c)), respectively. Type 1: $\theta(z) = 10 - 1.5z^2$. Type 2: $\theta(z) = 10 - 0.02z^2 + 0.4z^3$. Type 3: $\theta(z) = 3 + z + 2\exp(-2z^2)$. The sample size $T = 400$. Each simulation is repeated 1000 times.

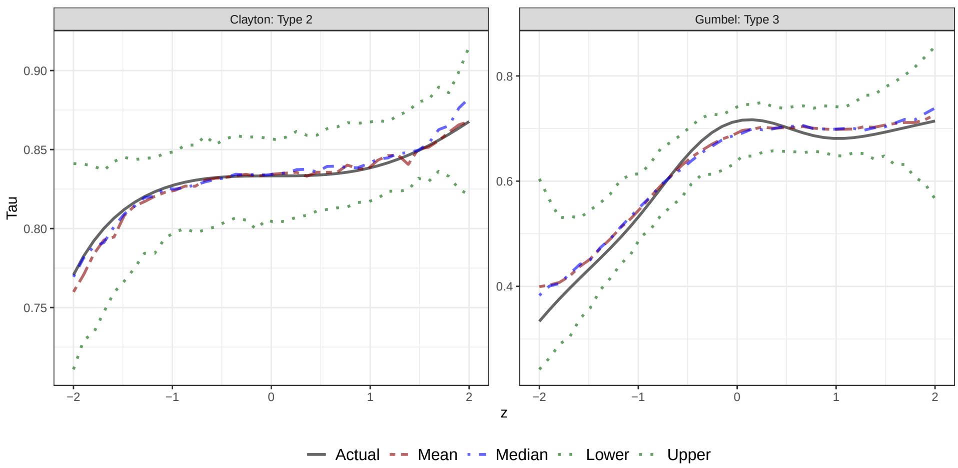
Figure 2: Estimated paths for copula dependent parameters (Kendall's τ) when the true model is a combination of Clayton and Gumbel copulas



(a) True copula: Clayton (Type 1) + Gumbel (Type 2)



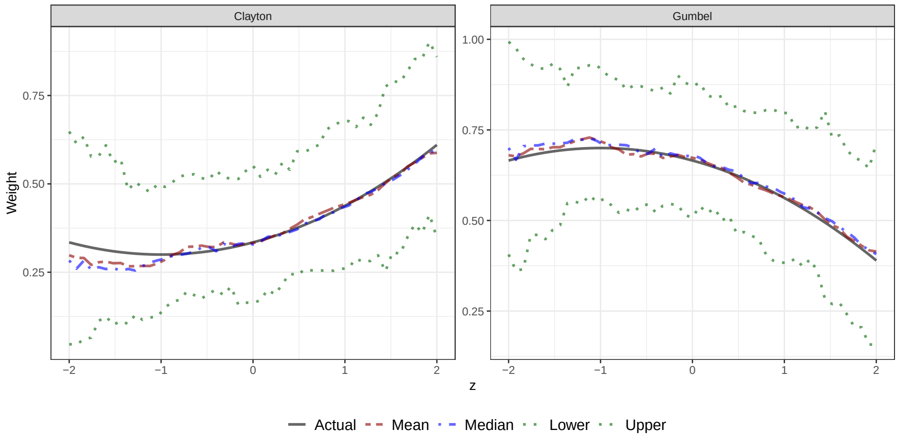
(b) True copula: Clayton (Type 1) + Gumbel (Type 3)



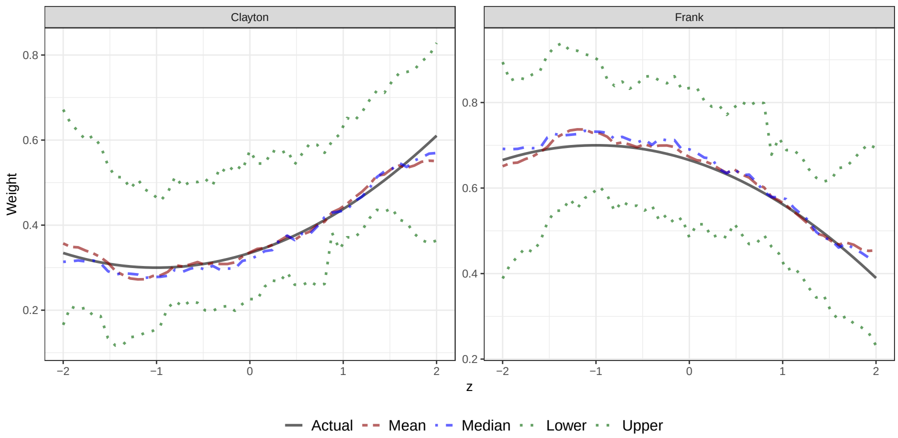
(c) True copula: Clayton (Type 2) + Gumbel (Type 3)

Notes. This figure displays the estimated paths of copula dependent parameters (Kendall's τ) when the true mode is a combination of Clayton and Gumbel copulas with different functional forms of $\theta(z)$. Type 1: $\theta(z) = 10 - 1.5z^2$. Type 2: $\theta(z) = 10 - 0.02z^2 + 0.4z^3$. Type 3: $\theta(z) = 3 + z + 2\exp(-2z^2)$. The sample size $T = 400$. Each simulation is repeated 1000 times.

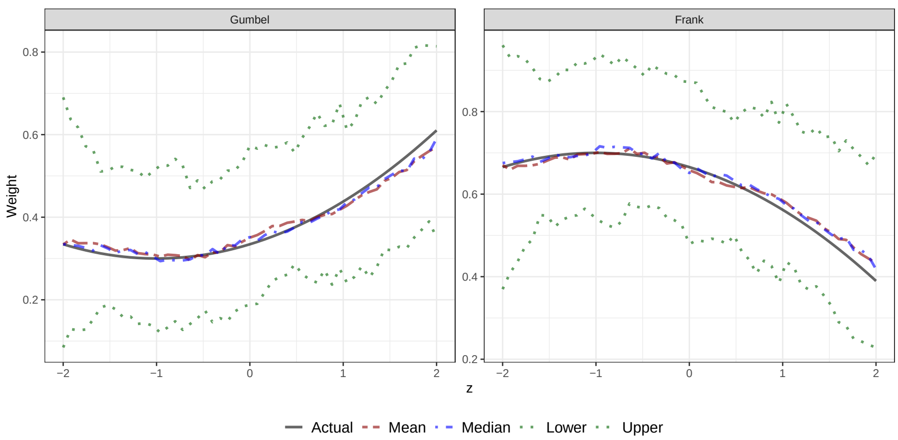
Figure 3: Estimated paths for weights (ω) when the true model is a combination of Clayton and Gumbel copulas



(a) True copula: Clayton (Type 1) + Gumbel (Type 2)



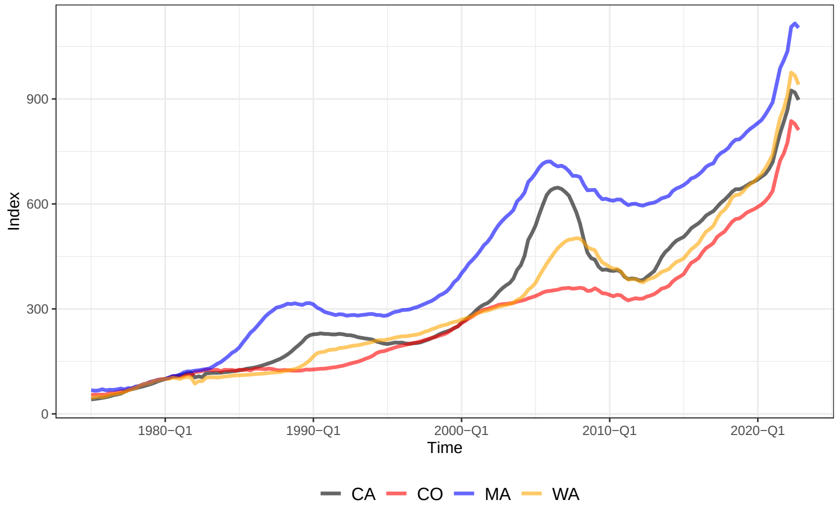
(b) True copula: Clayton (Type 1) + Gumbel (Type 3)



(c) True copula: Clayton (Type 2) + Gumbel (Type 3)

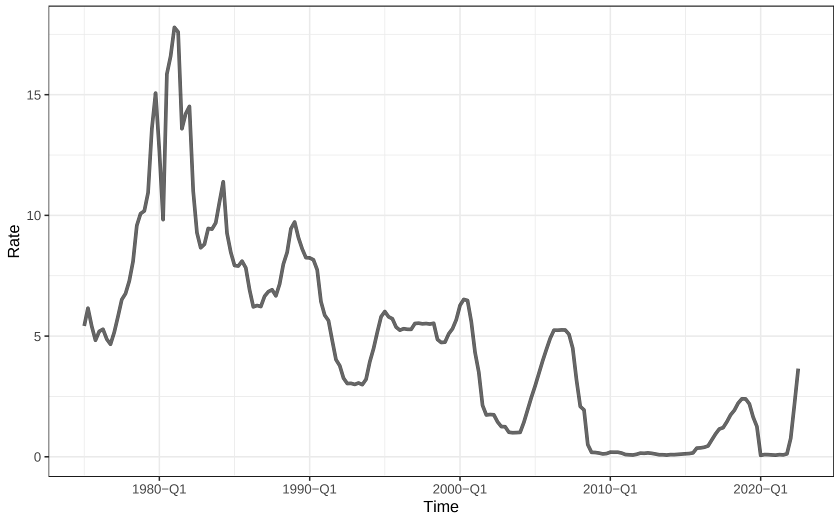
Notes. This figure displays the estimated paths of weights when the true mode is a combination of Clayton and Gumbel copulas with different functional forms of $\theta(z)$. Type 1: $\theta(z) = 10 - 1.5z^2$. Type 2: $\theta(z) = 10 - 0.02z^2 + 0.4z^3$. Type 3: $\theta(z) = 3 + z + 2\exp(-2z^2)$. The sample size $T = 400$. Each simulation is repeated 1000 times.

Figure 4: Housing price indices in CA, CO, MA, and WA: 1975:Q1 - 2022:Q4



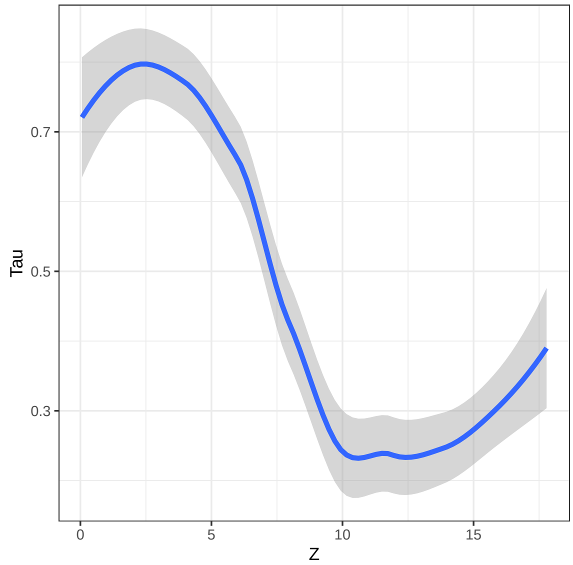
Notes. This figure displays the quarterly housing price indices in four states of the United States: California (CA), Colorado (CO), Massachusetts (MA), and Washington (WA). The sample period is from 1975:Q1 to 2022:Q4. The data can be retrieved from the Federal Housing Finance Agency (FHFA).

Figure 5: Effective federal funds rate: 1975:Q1 - 2022:Q4

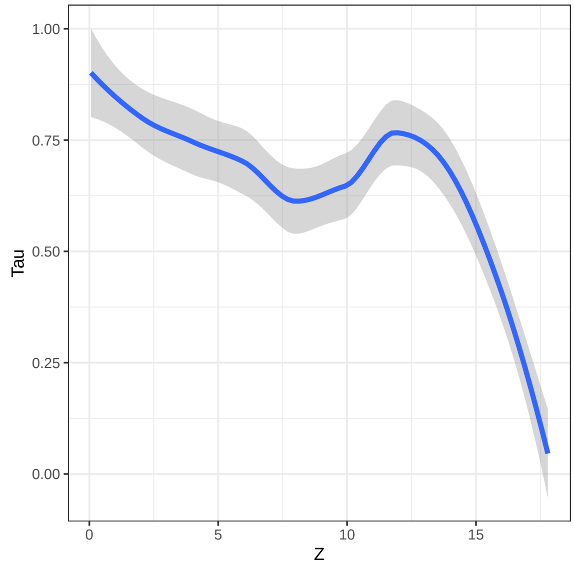


Notes. This figure displays the quarterly federal funds effective rate, or the interest rate depository institutions charge each other for overnight loans of funds. The sample period is from 1975:Q1 to 2022:Q4. The data can be downloaded from `https://fred.stlouisfed.org/`.

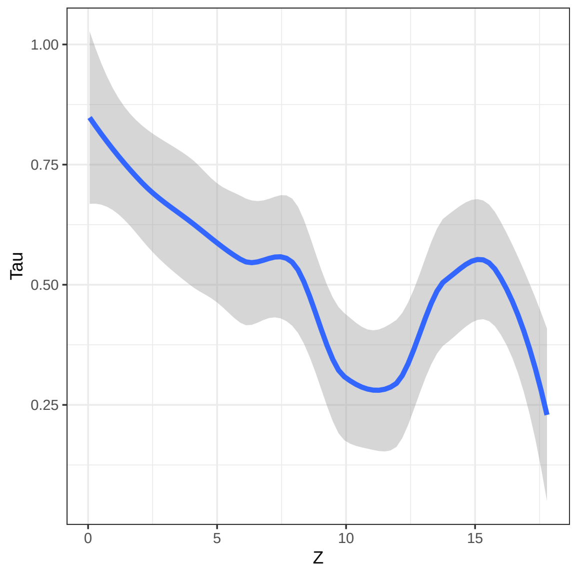
Figure 6: The estimated paths for Kendall's τs of Gumbel for CA-CO, CA-WA, CO-WA, and MA-WA



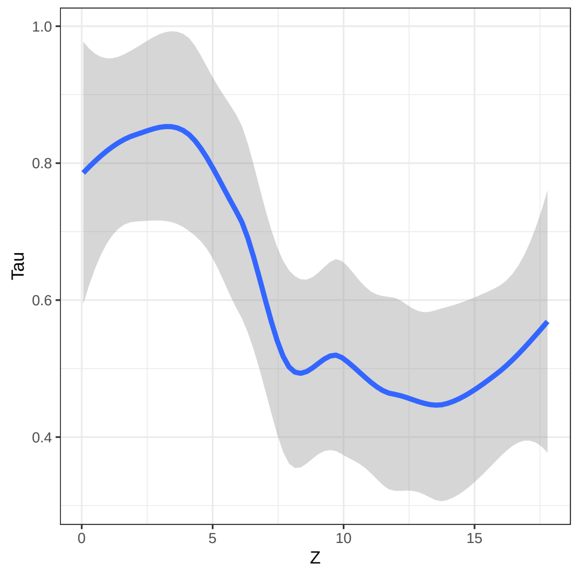
(a) CA-CO: Gumbel's τ



(b) CA-WA: Gumbel's τ



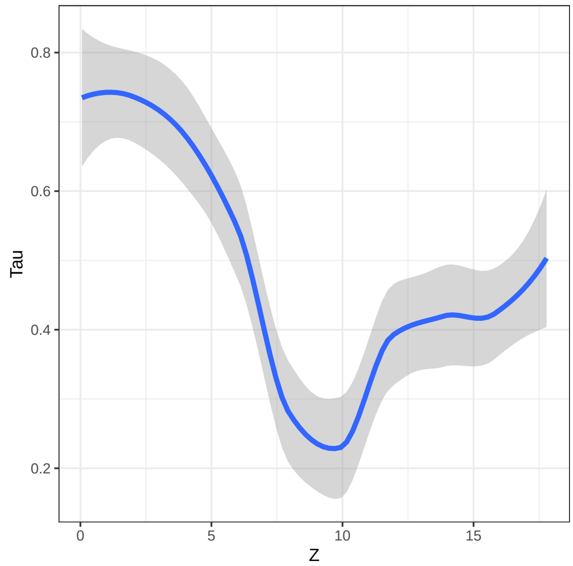
(c) CO-WA: Gumbel's τ



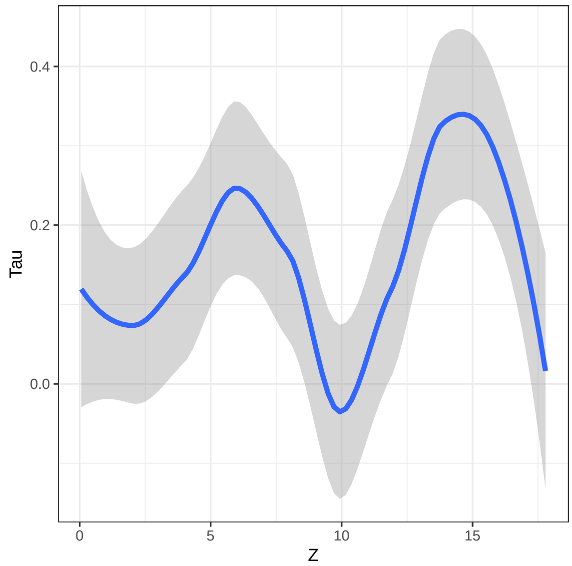
(d) MA-WA: Gumbel's τ

Notes. This figure displays the estimated paths for Kendall's τs of the selected Gumbel copula for CA-CO, CA-WA, CO-WA, and MA-WA. The shaded area represents the 95% confidence interval. The data are at quarterly frequency and span from 1975:Q1 to 2022:Q4.

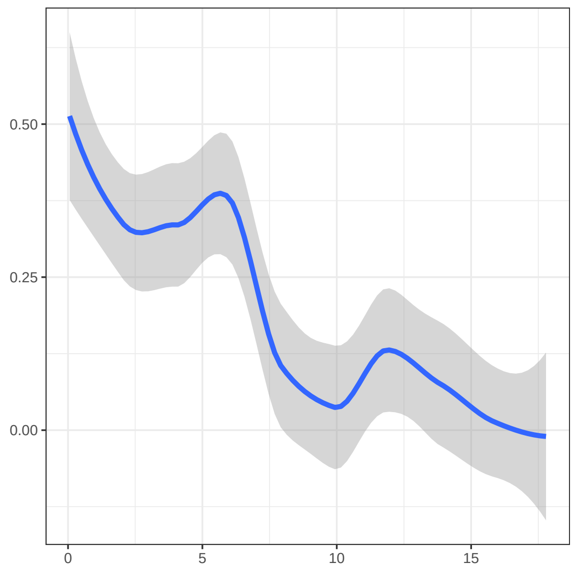
Figure 7: The estimated paths for Kendall's τs and weights of Gumbel and Frank for CA-MA



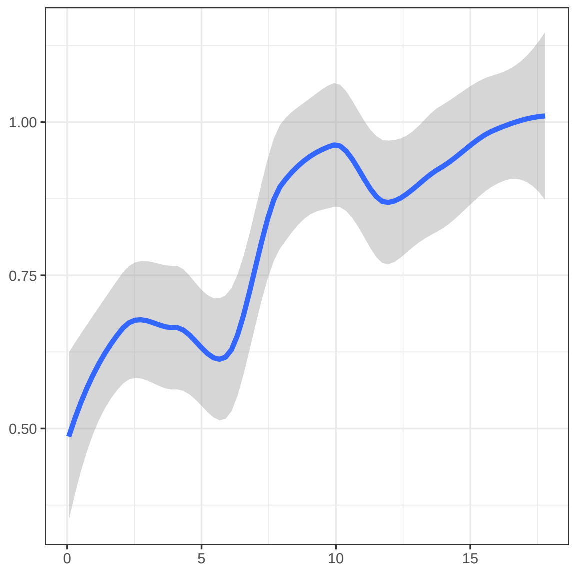
(a) CA-MA: Gumbel's τ



(b) CA-MA: Frank's τ



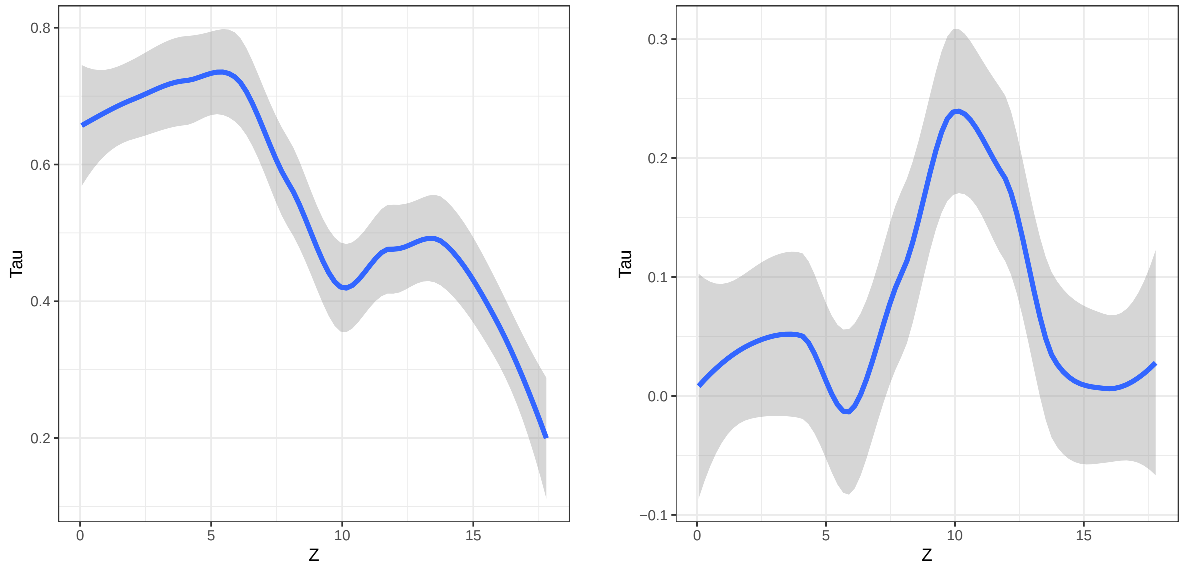
(c) CA-MA: Gumbel's weight



(d) CA-MA: Frank's weight

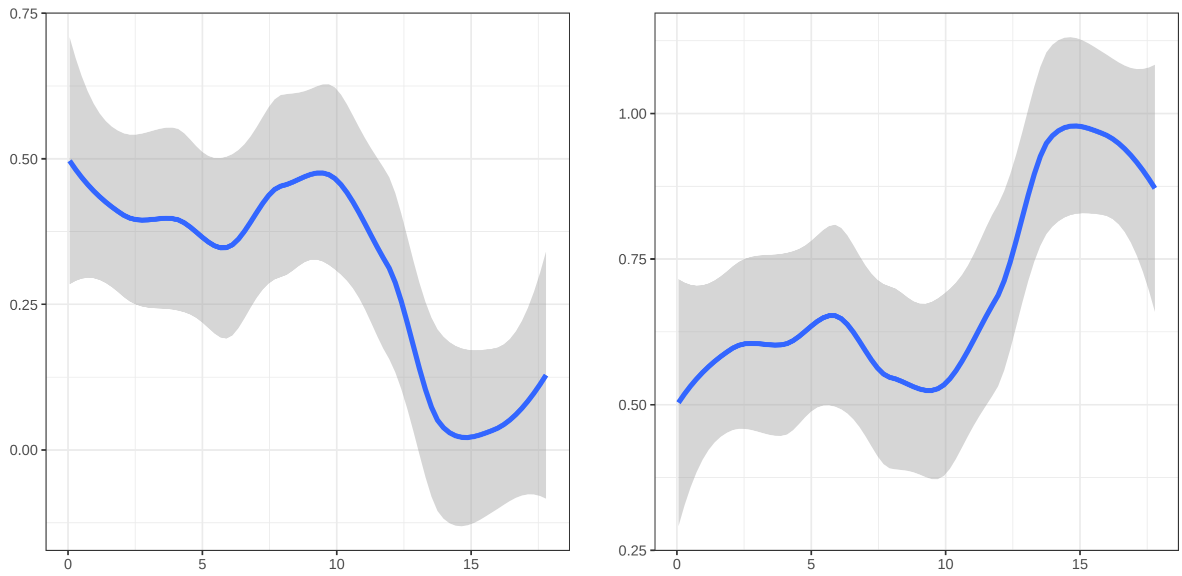
Notes. This figure displays the estimated paths for Kendall's τs and weights of the selected Gumbel-Frank mixture copula for CA-MA. The shaded area represents the 95% confidence interval. The data are at quarterly frequency and span from 1975:Q1 to 2022:Q4.

Figure 8: The estimated paths for Kendall's τs and weights of Gumbel and Frank for CO-MA



(a) CO-MA: Gumbel's τ

(b) CO-MA: Frank's τ

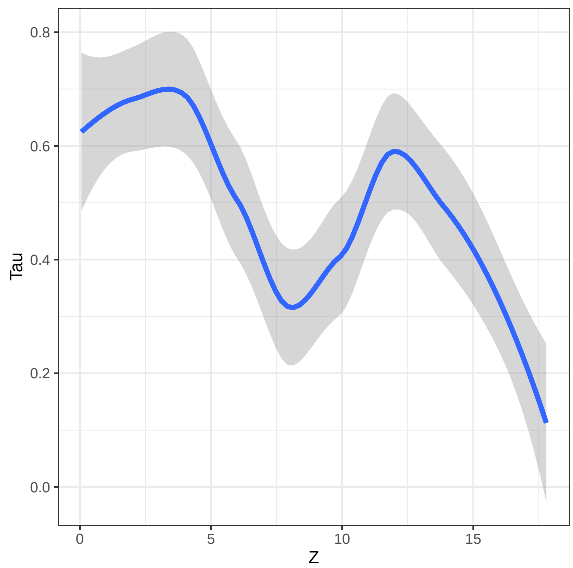


(c) CO-MA: Gumbel's weight

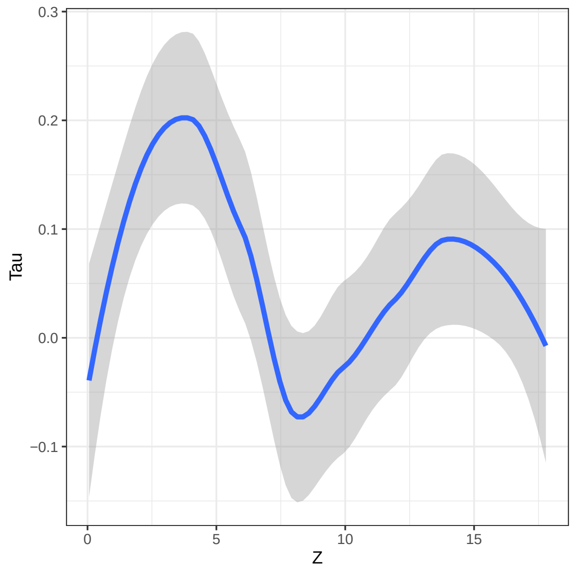
(d) CO-MA: Frank's weight

Notes. This figure displays the estimated paths for Kendall's τs and weights of the selected Gumbel-Frank mixture copula for CO-MA. The shaded area represents the 95% confidence interval. The data are at quarterly frequency and span from 1975:Q1 to 2022:Q4.

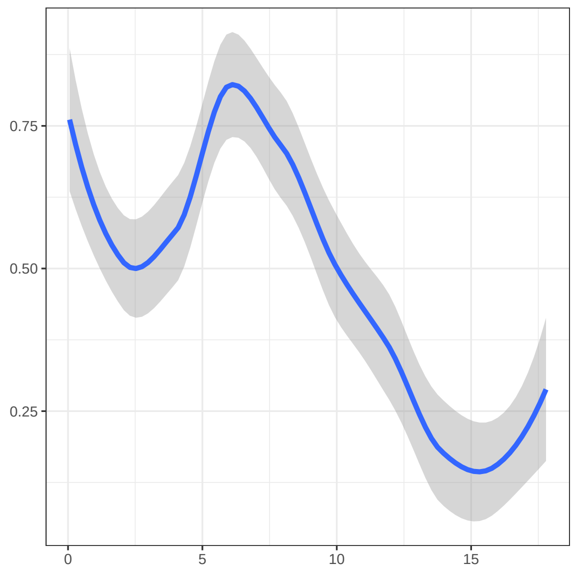
Figure 9: The estimated paths for Kendall's τs and weights of Gumbel and Frank for CA, CO, MA, and WA



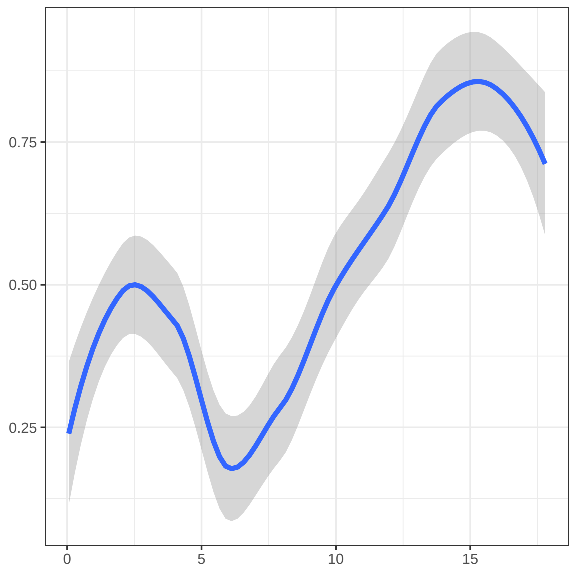
(a) 4-dimensional: Gumbel's τ



(b) 4-dimensional: Frank's τ



(c) 4-dimensional: Gumbel's weight



(d) 4-dimensional: Frank's weight

Notes. This figure displays the estimated paths for Kendall's τs and weights of the selected Gumbel-Frank mixture 4-dimensional copula for all four states. The shaded area represents the 95% confidence interval. The data are at quarterly frequency and span from 1975:Q1 to 2022:Q4.

Appendix: Mathematical Proofs

A1 Proof of Theorem 1

To prove the consistency of the local polynomial penalized maximum pseudo likelihood estimator $\hat{\boldsymbol{\eta}}(z)$, we can show that, with large probability, there is a local maximum in the sphere with center at $\boldsymbol{\eta}(z)$ for any sufficiently small radius. Specifically, the sphere is the union of $\varepsilon_{i,j}$-sphere of radius $\varepsilon_{i,j}$ centered at $\nu_{i,j}$ for $i = 1, 2$ and $j = 1, \cdots, d$, and denote the maximum radius by ε. That is, we want to show that the regularized kernel weighted pseudo log-likelihood function satisfies $\lim_{T\to\infty} P\Big(Q(\widetilde{\boldsymbol{\eta}}(z)) < Q(\boldsymbol{\eta}(z))\Big) \geq 1-\varepsilon$ for any point $\widetilde{\boldsymbol{\eta}}(z)$ on the surface of the ε-sphere. And $\{Q(\widetilde{\boldsymbol{\eta}}(z)) - Q(\boldsymbol{\eta}(z))\}$ can be written as

$$Q(\widetilde{\boldsymbol{\eta}}(z)) - Q(\boldsymbol{\eta}(z)) = L(\widetilde{\boldsymbol{\eta}}(z)) - L(\boldsymbol{\eta}(z)) + \tilde{L}(\widetilde{\boldsymbol{\omega}}(z)) - \tilde{L}(\boldsymbol{\omega}(z)), \qquad \text{(AE 1)}$$

where $\widetilde{\boldsymbol{\omega}}(z)$ consists of first dth elements of $\widetilde{\boldsymbol{\eta}}(z)$.

Applying Taylor's expansion to $L(\widetilde{\boldsymbol{\eta}}(z))$ around the point $\boldsymbol{\eta}(z)$ firstly, we have

$$L(\widetilde{\boldsymbol{\omega}}(z)) = L(\boldsymbol{\eta}(z)) + TS_{1T} + TS_{2T}(1 + o(1)),$$

and

$$S_{1T} = \frac{1}{T}\sum_{r=1}^{4d} A_r^T(\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}_r(z)), \qquad S_{2T} = \frac{1}{2T}\sum_{r=1}^{4d}\sum_{s=1}^{4d} B_{r,s}^n(\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}_r(z))(\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}_s(z)),$$

where

$$A_r^T = \sum_{t=1}^{T}(Z_t - z)^{D_r}\ell_{1(r)}\left(\boldsymbol{\eta}(z); \hat{\boldsymbol{U}}(z)\right)K_h(Z_t - z),$$
$$B_{r,s}^T = \sum_{t=1}^{T}(Z_t - z)^{D_r+D_s}\ell_{2(r,s)}\left(\breve{\boldsymbol{\eta}}(z); \hat{\boldsymbol{U}}(z)\right)K_h(Z_t - z),$$

with indicator functions $D_r = I(r \geq 2d), D_s = I(s \geq 2d)$, and $\breve{\boldsymbol{\eta}}(z)$ is a convex combination of $\widetilde{\boldsymbol{\eta}}(z)$ and $\boldsymbol{\eta}(z)$. Note that we use $\ell_{1(r)}(\cdot)$ and $\ell_{2(r,s)}(\cdot)$ to denote the rth element of vector $\ell_1(\cdot)$ and the (r, s)th element of matrix $\ell_2(\cdot)$, respectively. In addition, S_{1T} can be rewritten as

$$S_{1T} = \frac{1}{T}\sum_{r=1}^{4d}(A_r^T - A_r)(\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z)) + \frac{1}{T}\sum_{r=1}^{4d} A_r(\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z)),$$

where

$$A_r = \sum_{t=1}^{T}(Z_t - z)^{D_r}\ell_{1(r)}\left(\boldsymbol{\omega}(Z_t), \boldsymbol{\theta}(Z_t); \boldsymbol{U}(z)\right) K_h(Z_t - z).$$

Further,

$$S_{1T} \le \frac{4d}{T}\varepsilon \max_{r\in\{1,\cdots,4d\}} \left|A_r^T - A_r\right| + \frac{4d}{T}\varepsilon \max_{r\in\{1,\cdots,4d\}} |A_r|\,. \qquad \text{(AE 2)}$$

For the order of S_{1T}, let's consider the first term of (AE 2) firstly, then we have

$$\begin{aligned}
\frac{1}{T}\left|A_r^T - A_r\right| &\le \frac{1}{T}\sum_{t=1}^{T}|Z_t - z|^{D_r} K_h(Z_t - z)\left|\ell_{1(r)}\left(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2; \hat{\boldsymbol{U}}(z)\right) - \ell_{1,(r)}\left(\boldsymbol{\omega}(Z_t), \boldsymbol{\theta}(Z_t); \boldsymbol{U}(z)\right)\right| \\
&\le \frac{1}{T}\sum_{t=1}^{T}|Z_t - z|^{D_r} K_h(Z_t - z)\left[c\Big\{\sup_{z,k}\left|(Z_t - z)^2\omega_k^{(2)}(z)\right| + \sup_{z,k}\left|(Z_t - z)^2\theta_k^{(2)}(z)\right|\right. \\
&\quad \left. + \sum_{j=1}^{p}\sup_{x_j\in\mathbb{R}}\left|\widehat{F}_{X_j|z}(x_j) - F_{X_j|z}(x_j)\right|\Big\}\right],
\end{aligned}$$

where the second inequality holds by the definition of Lipschitz continuity. According to the facts $\widehat{F}_{X_j|z}(x_j) - F_{X_j|z}(x_j) = \frac{1}{2}h_j^2\mu_2 F_{X_j|z}^{(2)}(x_j) + O_p\left(\frac{1}{\sqrt{nh_j}}\right) + o_p(h_j^2)$ for $j \in \{1, \cdots, p\}$ from Cai (2002), the order is given as

$$\begin{aligned}
\frac{1}{T}\max_r \mid A_r^T - A_r \mid &= \max_r O_p\left\{\left(\frac{h^{2+D_r}}{T}\right) + O_p\left(\frac{h^{D_r}}{T}\sum_{j=1}^{p}\left(\frac{1}{\sqrt{Th_j}} + h_j^2\mu_2 F_{X_j|z}^{(2)}(x_j)\right)\right)\right\} \\
&= O_p\left(\frac{h^2}{T}\right) + O_p\left(\frac{1}{T}\sum_{j=1}^{p}\left(\frac{1}{\sqrt{Th_j}} + h_j^2\mu_2 F_{X_j|z}^{(2)}(x_j)\right)\right) \\
&= o_p(1) \qquad \text{(AE 3)}
\end{aligned}$$

by the details of the proof of Theorem 6.1 in Fan and Yao (2003).

For the second term of (AE 2), we have

$$E\left(\frac{1}{T}A_r\right) = \frac{1}{T}\sum_{t=1}^{T}\int\Big[\int \ell_{1(r)}\left(\boldsymbol{\omega}(z_t),\boldsymbol{\theta}(z_t);\widehat{F}_{X_1|z}(x_{1t}),\cdots,\widehat{F}_{X_p|z}(x_{pt})\right)dF_{\mathbf{X}|z}(\boldsymbol{x}_t)\Big]$$
$$\times(z_t-z)^{D_r}K_h(z_t-z)dF_Z(z_t),$$

and it is always equal to 0 for each r, since the first order term of log-likelihood function with true parameters $\boldsymbol{\omega}(z_t)$ and $\boldsymbol{\theta}(z_t)$ would equal zero. In addition, by condition A3, we obtain

$$\begin{aligned}
&Var\left(\frac{1}{T}A_r\right)\\
=\ &\frac{1}{T^2}Var\left[\sum_{t=1}^{T}(Z_t-z)^{D_r}\ell_{1(r)}\left(\boldsymbol{\omega}(Z_t),\boldsymbol{\theta}(Z_t);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)K_h(Z_t-z)\right]\\
=\ &\frac{1}{T^2}\sum_{t=1}^{T}E\left((Z_t-z)^{2D_r}K_h^2(Z_t-z)\ell_{1(r)}^2\left(\boldsymbol{\omega}(Z_t),\boldsymbol{\theta}(Z_t);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)\right)\\
&+\frac{1}{T^2}\sum_{t=2}^{T}\sum_{j=1}^{t-1}E\big[(Z_t-z)^{D_r}(Z_j-z)^{D_r}K_h(Z_t-z)K_h(Z_j-z)\times\\
&\ell_{1,(r)}\left(\boldsymbol{\omega}(Z_t),\boldsymbol{\theta}(Z_t);\hat{\mathbf{u}}(z)\right)\ell_{1(r)}\left(\boldsymbol{\omega}(Z_j),\boldsymbol{\theta}(Z_j);\boldsymbol{U}(z)\right)\big]\\
&+\frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{j=t+1}^{T}E\big[(Z_t-z)^{D_r}(Z_j-z)^{D_r}K_h(Z_t-z)K_h(Z_j-z)\times\\
&\ell_{1,(r)}\left(\boldsymbol{\omega}(Z_t),\boldsymbol{\theta}(Z_t);\boldsymbol{U}(z)\right)\ell_{1(r)}\left(\boldsymbol{\omega}(Z_j),\boldsymbol{\theta}(Z_j);\boldsymbol{U}(z)\right)\big]\\
\le\ &O\left(\frac{h^{2D_r}}{Th}\right),
\end{aligned}$$

where the inequality follows from the fact

$$\begin{aligned}
&E\left[(Z-z)^{2D_r}K_h^2(Z-z)\ell_{1(r)}^2\left(\boldsymbol{\omega}(Z),\boldsymbol{\theta}(Z);F_{X_1|z}(X_1),\cdots,F_{X_p|z}(X_p)\right)\right]\\
=\ &\int\frac{1}{h^2}(hv)^{2D_r}K^2(v)\ell_{1(r)}^2\left(\boldsymbol{\omega}(Z),\boldsymbol{\theta}(Z);F_{X_1|z}(x_1),\cdots,F_{X_p|z}(x_p)\right)dF_{\mathbf{X},Z}(\boldsymbol{x},hv+z)\\
\le\ &C\frac{h^{2D_r}}{h}\int\frac{1}{h}v^{2D_r}K^2(v)dF_{\mathbf{X},Z}(\boldsymbol{x},hv+z)\\
=\ &O\left(\frac{h^{2D_r}}{Th}\right)
\end{aligned}$$

by conditions A3 and A6. Therefore, $S_{1T}\xrightarrow{p}0$ as $T\to\infty$ and the order of S_{1T} is smaller than $O_p\left(\frac{1}{\sqrt{Th}}+\frac{h^2}{T}+\sum_{j=1}^{p}\frac{1}{T\sqrt{Th_j}}\right)$.

The term S_{2T} can be rewritten as follows:

$$S_{2T} = \frac{1}{2T}\sum_{r,s=1}^{4d}(B_{rs}^T - B_{rs})(\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z))(\widetilde{\boldsymbol{\eta}}_s(z) - \boldsymbol{\eta}_s(z)) + \frac{1}{2n}\sum_{r,s=1}^{4d} B_{rs}(\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z))(\widetilde{\boldsymbol{\eta}}_s(z) - \boldsymbol{\eta}_s(z)),$$

with

$$B_{rs} = \sum_{t=1}^{T}(Z_t - z)^{D_r+D_s}\ell_{2(r,s)}\left(\boldsymbol{\omega}(Z_t), \boldsymbol{\theta}(Z_t); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right) K_h(Z_t - z).$$

By the Lipschitz continuity of ℓ_2, we obtain

$$\begin{aligned}
& \frac{1}{T}\left|B_{rs}^T - B_{rs}\right| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T}\left|Z_t - z\right|^{D_r+D_s} K_h(Z_t - z)\left|\ell_{2(r,s)}\left(\breve{\boldsymbol{\nu}}_1, \breve{\boldsymbol{\nu}}_2; \widehat{F}_{X_1|z}(X_{1t}), \cdots, \widehat{F}_{X_p|z}(X_{pt})\right)\right. \\
& \left. -\ell_{2(r,s)}\left(\boldsymbol{\omega}(Z_t), \boldsymbol{\theta}(Z_t); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right)\right| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T}\left|Z_t - z\right|^{D_r+D_s} K_h(Z_t - z)\left[c\Big\{\sup_{z,k}\left|(Z_t - z)^2\omega_k^{(2)}(z)\right| + \sup_{z,k}\left|(Z_t - z)^2\theta_k^{(2)}(z)\right|\right. \\
& \left. +\sum_{j=1}^{p}\sup_{x_j\in\mathbb{R}}\left|\widehat{F}_{X_j|z}(x_j) - F_{X_j|z}(x_j)\right|\Big\}\right] \\
=\ & O_p\left(\frac{h^{2+D_r+D_s}}{T}\right) + O_p\left(\sum_{j=1}^{p}\frac{h_j^2 h^{D_r+D_s}}{T} + \frac{h^{D_r+D_s}}{T\sqrt{Th_j}}\right),
\end{aligned}$$

where the definition of $\breve{\boldsymbol{\nu}}_j$ is similar to $\breve{\boldsymbol{\eta}}$ for $j = 1, 2$. Moreover, we have

$$\begin{aligned}
E\left(\frac{1}{T}B_{rs}\right) &= E\left(\frac{1}{T}\sum_{t=1}^{T}(Z_t - z)^{D_r+D_s}\ell_{2(r,s)}\left(\boldsymbol{\omega}(Z_t), \boldsymbol{\theta}(Z_t); \boldsymbol{u}(z)\right) K_h(Z_t - z)\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\int \ell_{2(r,s)}\left(\boldsymbol{\omega}(z_t), \boldsymbol{\theta}(z_t); \boldsymbol{u}(z)\right)(z_t - z)^{D_r+D_s}K_h(z_t - z)dF_{\boldsymbol{X},Z}(\boldsymbol{x}_t, z_t) \\
&= h^{D_r+D_s}\int \ell_{2(r,s)}\left(\boldsymbol{\omega}(z), \boldsymbol{\theta}(z); \boldsymbol{u}(z)\right)(v)^{D_r+D_s}K(v)dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z)(1 + o(1)) \\
&\equiv h^{D_r+D_s}\mathcal{X}_{r,s}(1 + o(1)),
\end{aligned}$$

where the third equality holds by the Taylor expansion. Define $r_0 \leq d$ as the smallest index such that $\nu_{1,r_0} - \omega_{r_0}(z) \neq 0$ or $\nu_{2,r_0} - \theta_{r_0}(z) \neq 0$, then

$$
\begin{aligned}
&\frac{1}{2T}\sum_{r=1}^{4d}\sum_{s=1}^{4d} B_{rs}(\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z))(\widetilde{\boldsymbol{\eta}}_s(z) - \boldsymbol{\eta}_s(z)) \\
= &\frac{1}{2}\left[E\left(\frac{1}{T}B_{r_0r_0}\right) + O_p\left(\sqrt{Var\left(\frac{1}{T}B_{r_0r_0}\right)}\right)\right](\widetilde{\boldsymbol{\eta}}_{r_0}(z) - \boldsymbol{\eta}_{r_0}(z))^2 \\
&+\frac{1}{2}\sum_{r\neq r_0, s\neq r_0}\left[E\left(\frac{1}{T}B_{rs}\right) + O_p\left(\sqrt{Var\left(\frac{1}{T}B_{rs}\right)}\right)\right](\widetilde{\boldsymbol{\eta}}_r(z) - \boldsymbol{\eta}_r(z))(\widetilde{\boldsymbol{\eta}}_s(z) - \boldsymbol{\eta}_s(z)) \\
= &ch^{2D_{r_0}}\mathcal{X}_{r_0,r_0} + o_p(h^{2D_{r_0}}),
\end{aligned}
$$

where constant c is positive. Therefore, we have $S_{2T} + ch^{2D_{r_0}}\mathcal{X}_{r_0,r_0} \xrightarrow{p} 0$ as $T \to +\infty$. Combining the result that S_{1T} converges to 0 in probability, then we obtain the probability that$L(\widetilde{\boldsymbol{\eta}}(z)) - L(\boldsymbol{\eta}(z)) < 0$ tends to 1 for any points $\widetilde{\boldsymbol{\eta}}(z))$ on the surface of sphere.

Since $P_{\gamma_T}(0) = 0$ and $\sum_{k=1}^{d}(\widetilde{\eta}_k(z) - \eta_k(z)) = 0$ where $\widetilde{\eta}_k(z)$ and $\eta_k(z)$ are the kth elements of $\widetilde{\boldsymbol{\eta}}(z)$ and $\boldsymbol{\eta}(z)$ respectively, then we have

$$
\begin{aligned}
\tilde{L}(\widetilde{\boldsymbol{\omega}}(z)) - \tilde{L}(\boldsymbol{\omega}(z)) &= T\sum_{k=1}^{d}P_{\gamma_T}(\widetilde{\omega}_k(z)) - \rho\Big(1 - \sum_{k=1}^{d}\widetilde{\omega}_k(z)\Big) - T\sum_{k=1}^{d}P_{\gamma_T}(\omega_k(z)) \\
&\quad +\rho\Big(1 - \sum_{k=1}^{d}\omega_k(z)\Big) \\
&= T\sum_{k=1}^{d}P_{\gamma_T}(\widetilde{\omega}_k(z)) - T\sum_{k=1}^{d}P_{\gamma_T}(\omega_k(z)).
\end{aligned}
$$

In order to obtain the upper bound, we apply the Taylor's expansion at $\boldsymbol{\omega}(z)$, and obtain that

$$
\begin{aligned}
\frac{1}{T}\left|\tilde{L}(\widetilde{\boldsymbol{\omega}}(z)) - \tilde{L}(\boldsymbol{\omega}(z))\right| &= \left|\sum_{k=1}^{d}[P_{\gamma_T}(\widetilde{\omega}_k(z)) - P_{\gamma_T}(\omega_k(z))]\right| \\
&\leq d\|\widetilde{\boldsymbol{\omega}}(z) - \boldsymbol{\omega}(z)\|_2 \max_{1\leq k\leq d}\left\{\left|P'_{\gamma_T}(\omega_k(z))\right|\right\} \\
&\quad +\frac{d}{2}\|\widetilde{\boldsymbol{\omega}}(z) - \boldsymbol{\omega}(z)\|_2^2 \max_{1\leq k\leq d}\{|\ P''_{\gamma_T}(\omega_k(z))\ |\}\{1 + o(1)\}.
\end{aligned}
$$

According to the fact in Fan and Li (2001) that $\max_{1\leq k\leq d}\{|\partial P_{\gamma_T}(\omega_k(z))/\partial\omega_k(z)|\} = 0$ for the SCAD penalty function when $\gamma_T \to 0$, and by assumption that the maximum value of second derivative of the penalty function goes to zero, then $ch^{2D_{r_0}}\mathcal{X}_{r_0,r_0}(1 + o_p(1))$ is the sole dominant term in the right side of (AE 1) by comparing the order of terms.

Therefore, there is a local maximum in the ε-sphere for any given $\varepsilon > 0$. This completes the proof. $\square$

A2 Proof of Theorem 2 (a)

We proceed to the proof of part (a) in Theorem 2 first. Consider the local maximizer $\hat{\boldsymbol{\eta}}(z) = \left(\boldsymbol{\omega}^{*\top}(z), \hat{\bar{\boldsymbol{\eta}}}^{\top}(z)\right)^{\top}$ of the penalized kernel weighted pseudo log-likelihood function $Q(\boldsymbol{\eta}(z))$, where $\boldsymbol{\omega}^{*}(z) = \mathbf{0}$ and $\hat{\bar{\boldsymbol{\eta}}}(z)$ is the remaining part of vector $\hat{\boldsymbol{\eta}}(z)$ after removing $\boldsymbol{\omega}^{*}(z)$. To prove the sparsity we need to show that, as $T \to \infty$,

$$P\left(Q\left(\widetilde{\boldsymbol{\omega}}(z), \widetilde{\bar{\boldsymbol{\eta}}}(z)\right) < Q\left(\boldsymbol{\omega}^{*}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right)\right) \to 1 \qquad \text{(AE 4)}$$

for arbitrary $\widetilde{\boldsymbol{\eta}}(z) = \left(\widetilde{\boldsymbol{\omega}}^{\top}(z), \widetilde{\bar{\boldsymbol{\eta}}}^{\top}(z)\right)^{\top}$.We start by rewriting the leading term in (AE 4) as

$$\begin{aligned}
& Q\left(\widetilde{\boldsymbol{\omega}}(z), \widetilde{\bar{\boldsymbol{\eta}}}(z)\right) - Q\left(\boldsymbol{\omega}^{*}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right) \\
= \; & Q\left(\widetilde{\boldsymbol{\omega}}(z), \widetilde{\bar{\boldsymbol{\eta}}}(z)\right) - Q\left(\widetilde{\boldsymbol{\omega}}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right) + Q\left(\widetilde{\boldsymbol{\omega}}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right) - Q\left(\boldsymbol{\omega}^{*}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right) \\
\leq \; & Q\left(\widetilde{\boldsymbol{\omega}}(z), \widetilde{\bar{\boldsymbol{\eta}}}(z)\right) - Q\left(\widetilde{\boldsymbol{\omega}}(z), \hat{\bar{\boldsymbol{\eta}}}(z)\right).
\end{aligned}$$

It suffices to show that, if $\lim \inf_{T\to\infty} \lim \inf_{\pi_k \to \omega_k^{*}(z)+} \frac{P'_{\gamma_T}(\pi_k)}{\gamma_T}$ is positive and $\sqrt{Th}\gamma_T \to +\infty$, then $\partial Q(\widetilde{\boldsymbol{\eta}}(z))/\partial \widetilde{\omega}_k(z) < 0$ for any given $\widetilde{\boldsymbol{\eta}}(z)$ satisfying $\|\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}(z)\| = O_p\left(\frac{1}{\sqrt{Th}}\right)$ and $0 < \widetilde{\omega}_k(z) - \omega_k^{*}(z) < \frac{c}{\sqrt{Th}}$ with constant $c > 0$, where $\widetilde{\omega}_k(z)$ and $\omega_k^{*}(z)$ are the kth element of $\widetilde{\boldsymbol{\omega}}(z)$ and $\boldsymbol{\omega}^{*}(z)$ respectively. By Taylor expansion, we have

$$\begin{aligned}
\frac{\partial Q(\widetilde{\boldsymbol{\eta}}(z))}{\partial \widetilde{\omega}_k(z)} &= \frac{\partial L(\widetilde{\boldsymbol{\eta}}(z)))}{\partial \widetilde{\omega}_k(z)} + \frac{\partial \tilde{L}(\widetilde{\boldsymbol{\omega}}(z)))}{\partial \widetilde{\omega}_k(z)} \\
&= \frac{\partial L(\boldsymbol{\eta}(z))}{\partial \omega_k^{*}(z)} + \frac{\partial^2 L(\boldsymbol{\eta}(z))}{\partial \omega_k^{*}(z) \partial \boldsymbol{\eta}^{\top}(z)}(\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}(z)) \\
&\quad + (\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}(z))^{\top} \frac{\partial^3 L(\breve{\boldsymbol{\eta}}(z))}{2\partial \omega_k^{*}(z) \partial \boldsymbol{\eta}(z) \partial \boldsymbol{\eta}^{\top}(z)}(\widetilde{\boldsymbol{\eta}}(z) - \boldsymbol{\eta}(z)) + \frac{\partial \tilde{L}(\widetilde{\boldsymbol{\omega}}(z)))}{\partial \widetilde{\omega}_k(z)} \\
&= \frac{\partial L(\boldsymbol{\eta}(z))}{\partial \omega_k^{*}(z)} + \sum_{i=1}^{4d} \frac{\partial^2 L(\boldsymbol{\eta}(z))}{\partial \omega_k^{*}(z) \partial \eta_i(z)}(\widetilde{\eta}_i(z) - \eta_i(z)) \\
&\quad + \frac{1}{2} \sum_{i=1}^{4d} \sum_{j=1}^{4d} \frac{\partial^3 L(\breve{\boldsymbol{\eta}}(z))}{2\partial \omega_k^{*}(z) \partial \eta_i(z) \partial \eta_j(z)}(\widetilde{\eta}_i(z) - \eta_i(z))(\widetilde{\eta}_j(z) - \eta_j(z)) - TP'_{\gamma_T}(\widetilde{\omega}_k(z)) - \rho,
\end{aligned}$$

where $\breve{\boldsymbol{\eta}}(z)$ is the convex combination of $\boldsymbol{\eta}(z)$ and $\widetilde{\boldsymbol{\eta}}(z)$. Note that,

$$\frac{1}{T}\frac{\partial L(\boldsymbol{\eta}(z))}{\partial \omega_k^*(z)} = O_p\left(\frac{1}{\sqrt{Th}}\right) \quad \text{and} \quad \frac{1}{T}\sum_{i=1}^{4d}\frac{\partial^2 L(\boldsymbol{\eta}(z))}{\partial \omega_k^*(z)\partial \eta_i(z)} = E\left(\frac{\partial^2 L(\boldsymbol{\eta}(z))}{\partial \omega_k^*(z)\partial \eta_i(z)}\right) + o_p(1),$$

then we obtain

$$\begin{aligned}
\frac{\partial Q(\widetilde{\boldsymbol{\eta}}(z))}{\partial \widetilde{\omega}_k(z)} &= O_p\left(\sqrt{\frac{T}{h}}\right) - TP'_{\gamma_T}(\widetilde{\omega}_k(z)) \\
&= T\gamma_T\left(O_p\left(\frac{1}{\gamma_T\sqrt{Th}}\right) - \frac{P'_{\gamma_T}(\widetilde{\omega}_k(z))}{\gamma_T}\right) \\
&< 0.
\end{aligned}$$

Further, the inequality (AE 4) holds as T increases to infinity. This completes the proof of part (a).

A3 Proof of Theorem 2 (b)

We now embark on proving part (b) of Theorem 2. Similar to $\boldsymbol{\omega}^\dagger(z)$ and $\boldsymbol{\omega}^*(z)$, we use $\boldsymbol{\nu}_j^\dagger$ to denote the corresponding vector with nonzero weight in $\boldsymbol{\nu}_j$ for $j = 1, 2$. The objective function $Q(\breve{\boldsymbol{\eta}}(z))$ could be written as $Q(\breve{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) + \boldsymbol{\eta}^\dagger(z))$ by the sparsity. Equivalently, we can find the maximizer $\hat{\tilde{\boldsymbol{\eta}}}^\dagger(z)$ to maximize $Q(\breve{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) + \boldsymbol{\eta}^\dagger(z)) - Q(\boldsymbol{\eta}^\dagger(z))$, and the maximizer is the maximum likelihood estimate $\hat{\boldsymbol{\eta}}^\dagger(z)$, and that satisfies the likelihood equation, for $k = 1, \cdots, 4d^\dagger$,

$$\begin{aligned}
& \frac{\partial Q(\breve{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) + \boldsymbol{\eta}^\dagger(z)) - Q(\boldsymbol{\eta}^\dagger(z))}{\partial \breve{\eta}_k^\dagger(z)}\Big|_{\breve{\boldsymbol{\eta}}^\dagger(z)=\hat{\boldsymbol{\eta}}^\dagger(z)} \\
= & \frac{\partial L(\breve{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) + \boldsymbol{\eta}^\dagger(z)) - L(\boldsymbol{\eta}^\dagger(z))}{\partial \breve{\eta}_k^\dagger(z)}\Big|_{\breve{\boldsymbol{\eta}}^\dagger(z)=\hat{\boldsymbol{\eta}}^\dagger(z)} \\
& + TI(k \le d^\dagger)[P'_{\gamma_T}(\hat{\omega}_k^\dagger(z) - \omega_k^\dagger(z) + \omega_k^\dagger(z)) - P'_{\gamma_T}(\omega_k^\dagger(z))].
\end{aligned}$$

Let $\tilde{\boldsymbol{\eta}}^\dagger(z) = \arg\max_{\breve{\boldsymbol{\eta}}^\dagger(z)\in\Gamma} L\left(\breve{\boldsymbol{\eta}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z) + \boldsymbol{\eta}^\dagger(z)\right) - L\left(\boldsymbol{\eta}^\dagger(z)\right) \equiv \arg\max_{\breve{\boldsymbol{\eta}}^\dagger(z)\in\Gamma} \widetilde{Q}(\breve{\boldsymbol{\eta}}^\dagger(z))$ and $\hat{\tilde{\boldsymbol{\eta}}}^\dagger(z) = \arg\max_{\breve{\boldsymbol{\eta}}^\dagger(z)\in\Gamma_c} \widetilde{Q}(\breve{\boldsymbol{\eta}}^\dagger(z))$, where Γ is a neighborhood of $\boldsymbol{\eta}^\dagger(z)$ and Γ_c is a cone. Consider the fact that

$$\begin{aligned}\left\|\hat{\boldsymbol{\eta}}^{\dagger}(z)-\boldsymbol{\eta}^{\dagger}(z)\right\|_2 &\leq \left\|\hat{\boldsymbol{\eta}}^{\dagger}(z)-\tilde{\boldsymbol{\eta}}^{\dagger}(z)\right\|_2+\left\|\tilde{\boldsymbol{\eta}}^{\dagger}(z)-\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)\right\|_2+\left\|\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)-\boldsymbol{\eta}^{\dagger}(z)\right\|_2 \\ &= \left\|\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)-\boldsymbol{\eta}^{\dagger}(z)\right\|_2+o_p(1)\end{aligned}$$

for penalized mixture copula estimator in Cai and Wang (2014), then we need to investigate the asymptotic distribution of $\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)$. Define the normalized estimator $\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z) \equiv D_T\left(\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)-\boldsymbol{\eta}^{\dagger}(z)\right)$, two vectors $\hat{\hat{\boldsymbol{\omega}}}^{\dagger}(z)$ and $\hat{\hat{\boldsymbol{\theta}}}^{\dagger}(z)$ are corresponding to $\boldsymbol{\omega}^{\dagger}(z)$ and $\boldsymbol{\theta}^{\dagger}(z)$ respectively in $\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)$, then we have the likelihood equation

$$\begin{aligned}\frac{\partial\widetilde{Q}(\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z))}{\partial\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)} &= \frac{\partial}{\partial\hat{\hat{\boldsymbol{\eta}}}^{\dagger}(z)}\sum_{t=1}^{T}[\ell_1^{\top}(\boldsymbol{\omega}^{\dagger}(z),\boldsymbol{\theta}^{\dagger}(z);\widehat{F}_{X_1|z}(X_{1t}),\cdots,\widehat{F}_{X_p|z}(X_{pt}))\Delta_t\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z)+ \\ &\quad (\Delta_t\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z))^{\top}\ell_2(\dot{\boldsymbol{\omega}}^{\dagger}(z),\dot{\boldsymbol{\theta}}^{\dagger}(z);\widehat{F}_{X_1|z}(X_{1t}),\cdots,\widehat{F}_{X_p|z}(X_{pt}))\Delta_t\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z)]K_h(Z_t-z) \\ &= \sum_{t=1}^{T}\Delta_t\ell_1\left(\boldsymbol{\omega}^{\dagger}(z),\boldsymbol{\theta}^{\dagger}(z);\widehat{F}_{X_1|z}(X_{1t}),\cdots,\widehat{F}_{X_p|z}(X_{pt})\right)K_h(Z_t-z)+ \\ &\quad 2\sum_{t=1}^{T}\Delta_t\ell_2\left(\dot{\boldsymbol{\omega}}^{\dagger}(z),\dot{\boldsymbol{\theta}}^{\dagger}(z);\widehat{F}_{X_1|z}(X_{1t}),\cdots,\widehat{F}_{X_p|z}(X_{pt})\right)\Delta_tK_h(Z_t-z)\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z) \\ &\equiv \overline{W}_1+\overline{W}_2\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z) \\ &= 0\end{aligned}$$

by using Taylor expansion of $L(\hat{\hat{\boldsymbol{\eta}}}^{\dagger}_T(z))$ around $\boldsymbol{\eta}^{\dagger}(z)$, where $\Delta_t = (diag(1,Z_t-z)\otimes \mathrm{I}_{2d^{\dagger}})D_T^{-1}$, and $\dot{\boldsymbol{\omega}}^{\dagger}(z)$ is a convex combination of $\hat{\hat{\boldsymbol{\omega}}}^{\dagger}(z)$ and $\boldsymbol{\omega}^{\dagger}(z)$, and $\dot{\boldsymbol{\theta}}^{\dagger}(z)$ is a convex combination of $\hat{\hat{\boldsymbol{\theta}}}^{\dagger}(z)$ and $\boldsymbol{\theta}^{\dagger}(z)$. For the purpose of obtaining the asymptotic property, we first define $W_1 = h\overline{W}_1, W_2 = h\overline{W}_2$, and show that $W_2 = W_2^*+o_p(1)$, where

$$W_2^* = 2h\sum_{t=1}^{T}\Delta_t\ell_2\left(\dot{\boldsymbol{\omega}}^{\dagger}(z),\dot{\boldsymbol{\theta}}^{\dagger}(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)\Delta_tK_h(Z_t-z).$$

Similar to the (AE 3), we obtain

$$\begin{aligned}|(W_2-W_2^*)_{r,s}| &= \Big|\sum_{t=1}^{T}2hK_h(Z_t-z)[\Delta_t(\ell_2(\dot{\boldsymbol{\omega}}^{\dagger}(z),\dot{\boldsymbol{\theta}}^{\dagger}(z);\widehat{F}_{X_1|z}(X_{1t}),\cdots,\widehat{F}_{X_p|z}(X_{pt})) \\ &\quad -\ell_2(\dot{\boldsymbol{\omega}}^{\dagger}(z),\dot{\boldsymbol{\theta}}^{\dagger}(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})))\Delta_t]_{r,s}\Big|\end{aligned}$$

for any $1 \leq r \leq 4d^\dagger$ and $1 \leq s \leq 4d^\dagger$. We now study each of the elements for the four parts of the matrix respectively, that is, when $r \leq 2d^\dagger$ and $s \leq 2d^\dagger$,

$$
\begin{aligned}
|(W_2 - W_2^*)_{r,s}| &= \Big| \sum_{t=1}^{T} 2hK_h(Z_t - z)\frac{1}{Th}[\ell_{2(r,s)}(\dot{\boldsymbol{\omega}}^\dagger(z), \dot{\boldsymbol{\theta}}^\dagger(z); \widehat{F}_{X_1|z}(X_{1t}), \cdots, \widehat{F}_{X_p|z}(X_{pt})) \\
&\quad -\ell_{2(r,s)}(\dot{\boldsymbol{\omega}}^\dagger(z), \dot{\boldsymbol{\theta}}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt}))]\Big| \\
&\leq \left| \sum_{t=1}^{T} \frac{2}{Th} K\left(\frac{Z_t - z}{h}\right)\right| \times \sum_{j=1}^{p} \sup_{x_j} \left|\widehat{F}_{X_j|z}(x_j) - F_{X_j|z}(x_j)\right| \\
&= O_p\left(h^2 + \frac{1}{\sqrt{Th}}\right) \cdot O_p\left(\sum_{j=1}^{p} h_j^2 \mu_2 F^{(2)}_{X_j|z}(x_j) + \frac{1}{\sqrt{Th_j}}\right) \\
&= o_p(1).
\end{aligned}
$$

Similarly, when $r > 2d^\dagger$ and $s > 2d^\dagger$, we have

$$
\begin{aligned}
|(W_2 - W_2^*)_{r,s}| &= \Big| \sum_{t=1}^{T} 2hK_h(Z_t - z)\frac{(Z_t - z)^2}{Th^3}[\ell_{2(r,s)}(\dot{\boldsymbol{\omega}}^\dagger(z), \dot{\boldsymbol{\theta}}^\dagger(z); \widehat{F}_{X_1|z}(X_{1t}), \cdots, \widehat{F}_{X_p|z}(X_{pt})) \\
&\quad -\ell_{2(r,s)}(\dot{\boldsymbol{\omega}}_{\mathbf{1}}^\dagger(z), \dot{\boldsymbol{\theta}}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt}))]\Big| \\
&= O_p\left(\sqrt{\frac{h}{T}}\right) \cdot O_p\left(\sum_{j=1}^{p} h_j^2 \mu_2 F^{(2)}_{X_j|z}(x_j) + \frac{1}{\sqrt{Th_j}}\right) \\
&= o_p(1).
\end{aligned}
$$

As for other two parts of matrix $W_2 - W_2^*$, we can obtain the similar results. Therefore, W_2 converges to matrix W_2^* in probability.

As for its expectation, we have

$$
\begin{aligned}
E[(W_2^*)_{r,s}] &= E\left[2K_h(Z - z)\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z); F_{X_1|z}(X_1), \cdots, F_{X_p|z}(X_p)\right)\right]\right] \\
&= \int 2\frac{1}{h}K(\frac{z_0 - z}{h})\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z); F_{X_1|z}(x_1), \cdots, F_{X_p|z}(x_p)\right)\right] dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z_0) \\
&= \int 2K(v)\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z); F_{X_1|z}(x_1), \cdots, F_{X_p|z}(x_p)\right)\right] dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z + hv),
\end{aligned}
$$

where z_0 is a realized velue of random variable Z when $r \leq 2d^\dagger$ and $s \leq 2d^\dagger$. In the same way, we have

$$\begin{aligned} E[(W_2^*)_{r,s}] &= E\left[2hK_h(Z-z)\frac{(Z-z)^2}{h^3}\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z);F_{X_1|z}(X_1),\cdots,F_{X_p|z}(X_p)\right)\right]\right] \\ &= \int \frac{2}{h}K(\frac{z_0-z}{h})\frac{(z_0-z)^2}{h^2}\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z);F_{X_1|z}(x_1),\cdots,F_{X_p|z}(x_p)\right)\right]dF_{\boldsymbol{X},Z}(\boldsymbol{x},z_0) \\ &= \int 2K(v)v^2\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z);F_{X_1|z}(x_1),\cdots,F_{X_p|z}(x_p)\right)\right]dF_{\boldsymbol{X},Z}(\boldsymbol{x},z+hv) \end{aligned}$$

when $r>2d^\dagger$ and $s>2d^\dagger$. For the other two parts, $r>2d^\dagger, s\le 2d^\dagger$ and $r\le 2d^\dagger, s>2d^\dagger$, of matrix W_2^*, it is clear that each element is equal to 0. Thus $E(W_2^*) = 2E\left\{[diag(K(V),V^2K(V))\otimes \mathrm{I}_{2d^\dagger}]\ell_2\left(\boldsymbol{\eta}^\dagger(Z_t);F_{X_1|Z_t}(X_1),\cdots,F_{X_p|Z_t}(X_p)\right)|Z_t=z\right\}$.

Given that W_2 converges to the term W_2^*, we need compute the variance of W_2^* to show that it tends to its expectation which is a constant matrix. For $r\le 2d^\dagger$ and $s\le 2d^\dagger$, we have

$$\begin{aligned} & Var\left\{\sum_{t=1}^{T}\frac{2}{T}K_h(Z_t-z)\left[\ell_{2(r,s)}\left(\dot{\boldsymbol{\omega}}^\dagger(z),\dot{\boldsymbol{\theta}}^\dagger(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)\right]\right\} \\ = & \frac{4}{Th^2}Var\left\{\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)\right]K\left(\frac{Z-z}{h}\right)\right\} \\ & +\frac{8}{Th^2}\sum_{j=1}^{n-1}\left(1-\frac{j}{T}\right)Cov(\ell^z_{2(r,s)}(t),\ell^z_{2(r,s)}(t+j)) \\ \equiv & \frac{4}{Th^2}\Sigma(0)+\frac{8}{T^2h^2}\sum_{j=1}^{T-1}(T-j)\Sigma(j) \\ \le & \frac{4}{Th^2}\Sigma(0)+\frac{c}{Th^2}\sum_{j=1}^{T-1}\alpha(j)^{\frac{\varsigma}{4+\varsigma}}\left[E\left|\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z,Z);\boldsymbol{U}(z)\right)\right]K\left(\frac{Z-z}{h}\right)\right|^{\frac{4+\varsigma}{2}}\right]^2 \end{aligned}$$

where $\ell^z_{2(r,s)}(t) = h\left[\ell_{2(r,s)}(\boldsymbol{\eta}^\dagger(z,Z_t);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt}))\right]K_h(Z_t-z)$, then the variance of any elements in this part equals $o_p(1)$ since $\sum_{j=1}^{\infty}\alpha(j)^{\frac{\varsigma}{4+\varsigma}}<\infty$ and condition A6. Similarly, when $r>2d^\dagger$ and $s>2d^\dagger$, we know the variance is smaller than

$$\frac{4}{Th^2}\Sigma(0)+\frac{c}{Th^2}\sum_{j=1}^{T-1}\alpha(j)^{\frac{\varsigma}{4+\varsigma}}\left[E\left|\left(\frac{Z-z}{h}\right)^2\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z,Z);\boldsymbol{U}(z)\right)\right]K\left(\frac{Z-z}{h}\right)\right|^{\frac{4+\varsigma}{2}}\right]^2$$

by the condition A6. For the other two parts, the similar inequality still holds,

$$\begin{aligned}
& Var\left\{\sum_{t=1}^{T} 2\frac{Z_t - z}{Th} K_h(Z_t - z)\left[\ell_{2(r,s)}\left(\dot{\boldsymbol{\omega}}^\dagger(z), \dot{\boldsymbol{\theta}}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right)\right]\right\} \\
\leq & \frac{4}{Th^2}\Sigma(0) + \frac{c}{Th^2}\sum_{j=1}^{T-1}\alpha(j)^{\frac{\varsigma}{4+\varsigma}}\left[E\left|\left[\ell_{2(r,s)}\left(\boldsymbol{\eta}^\dagger(z,Z);\boldsymbol{U}(z)\right)\right] K\left(\frac{Z-z}{h}\right)\right|^{\frac{4+\varsigma}{2}}\right]^2 \\
= & o_p(1).
\end{aligned}$$

Hence, W_2 converges to the constant matrix $E(W_2^*)$.

Next, we discuss the vector W_1, and this term can be rewritten as follows:

$$\begin{aligned}
W_1 = & \sum_{t=1}^{T} h\Delta_t \ell_1\left(\boldsymbol{\omega}^\dagger(z), \boldsymbol{\theta}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right) K_h(Z_t - z) \\
\approx & \sum_{t=1}^{T} h\Delta_t \ell_1\left(\boldsymbol{\omega}^\dagger(z), \boldsymbol{\theta}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right) K_h(Z_t - z) \\
& + \sum_{j=1}^{p}\sum_{t=1}^{T} h\Delta_t \frac{\partial \ell_1}{\partial U_1}\left(\boldsymbol{\omega}^\dagger(z), \boldsymbol{\theta}^\dagger(z); \boldsymbol{U}(z)\right)\left(\widehat{F}_{X_j|z}(X_{jt}) - F_{X_j|z}(X_{jt})\right) K_h(Z_t - z) \\
\equiv & \sum_{t=1}^{T} W_{1,1}^t + \sum_{j=1}^{p}\sum_{t=1}^{T} W_{1,j+1}^t \\
\equiv & W_{1,1} + \sum_{j=1}^{p} W_{1,j+1}. \qquad \text{(AE 5)}
\end{aligned}$$

In the same vein, we need to calculate its expectation by partition each sub-term of W_1. Similar to the derivation of W_2^*, for any element which belongs to the first half and its index is r, we have

$$E[(W_{1,1})_r] = \sqrt{Th}\int\left[\ell_{1(r)}(\boldsymbol{\eta}^\dagger(z); \boldsymbol{u}(z))\right] h^2 K(v) dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z + hv),$$

$$E[(W_{1,j+1})_r] = \sqrt{Th}\int\left[\ell_{1,j}(\boldsymbol{\eta}^\dagger(z); \boldsymbol{u}(z))\right]_r K(v)\tfrac{\mu_2}{2} h_j^2 F_{X_j|z}^{(2)}(x_j) dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z + hv)(1 + o(1))$$

under condition A7 for $j \in \{1, \cdots, p\}$, and $[\boldsymbol{a}]_r$ denotes the rth element of the arbitrary vector $\boldsymbol{a}$. For the second half, the index $r > 2d$, then

$$E[(W_{1,1})_r] = \sqrt{Th^3}\int\left[\ell_{1(r)}(\boldsymbol{\eta}^\dagger(z); \boldsymbol{u}(z))\right] h^2 v K(v) dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z + hv),$$

$$E[(W_{1,j+1})_r] = \sqrt{Th^3}\int\left[\ell_{1,j}(\boldsymbol{\eta}^\dagger(z); \boldsymbol{u}(z))\right]_r vK(v)\tfrac{\mu_2}{2} h_j^2 F_{X_j|z}^{(2)}(x_j) dF_{\boldsymbol{X},Z}(\boldsymbol{x}, z + hv)(1 + o(1)).$$

Hence we can obtain the asymptotic expectation vector $D_T\Xi_z^*$ of W_1. Note first of all that $Var(W_1)=\sum_{j=1}^{p+1}Var(W_{1,j})+\sum_{\substack{i,j=1\\ i\neq j}}^{p+1}Cov(W_{1,i},W_{1,j})$ and hence we need to calculate the second moment in order to approximate the variance of term W_1. When $r\le 2d^\dagger$, we can get

$$\begin{aligned}
Var[(W_{1,1})_r] &= Var\left\{\sum_{t=1}^{T} h[\Delta_t\ell_1\left(\boldsymbol{\omega}^\dagger(z),\boldsymbol{\theta}^\dagger(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)]_r K_h(Z_t-z)\right\}\\
&= hVar\left\{[\ell_{1(r)}\left(\boldsymbol{\omega}^\dagger(z),\boldsymbol{\theta}^\dagger(z);F_{X_1|z}(X_1),\cdots,F_{X_p|z}(X_p)\right)]K_h(Z-z)\right\}\\
&\quad +2h\sum_{j=1}^{T-1}(1-\frac{j}{T})Cov(\ell_{1(r)}(t)K_h(Z_t-z),\ell_{1(r)}(t+j)K_h(Z_{t+j}-z))\\
&= h\Sigma(0)+\frac{2h}{T}\sum_{j=1}^{T-1}(T-j)\Sigma(j)\\
&\le h\Sigma(0)+ch\sum_{j=1}^{T-1}\alpha(j)^{\frac{\gamma}{4+\gamma}}\left[E\left|[\boldsymbol{\omega}^\dagger(z),\boldsymbol{\theta}^\dagger(z);F_{X_1|z}(X_1),\cdots,F_{X_p|z}(X_p)]_r K_h(Z-z)\right|\right]^2,
\end{aligned}$$

where $\ell_{1(r)}(t)=\ell_{1(r)}\left(\boldsymbol{\omega}^\dagger(z),\boldsymbol{\theta}^\dagger(z);F_{X_1|z}(X_{1t}),\cdots,F_{X_p|z}(X_{pt})\right)$.

As for the terms $W_{1,j+1}$ for $j\in\{1,\cdots,p\}$, we have

$$Var[(W_{1,j+1})_r]=O_p\left(hh_j^2+\frac{h}{\sqrt{Th_j}}\right).$$

It can be seen that the contribution of W_1 to the variance is $W_{1,1}$, now consider the covariance,

$$|Cov[(W_{1,1})_r,(W_{1,2})_r]|\le\sqrt{Var(W_{1,1})\cdot Var(W_{1,2})}=O_p\left(h\sqrt{(h_1^2+\frac{1}{\sqrt{Th_1}})(h_2^2+\frac{1}{\sqrt{Th_2}})}\right),$$

and we could obtain the similar results for other terms, and the covariances of $W_{1,j+1}$ and $W_{1,i+1}$ have the same order of the variance of $W_{1,j+1}$ if $h_i=h_j$ for $i\neq j$. For the another case $r>2d^\dagger$, we can obtain the similar results. Therefore, the variance of $W_{1,1}$ is the leading term of the variance of W_1. Further, following the same technology on pages 251-252 of Fan and Gijbels (1996), we have

$$\begin{aligned}
E(W_{1,1}W_{1,1}^\top) &= h^2E\left[\sum_{t=1}^{T}\Delta_t\ell_1(t)\ell_1^\top(t)\Delta_t K_h^2(Z_t-z)\right]\\
&= E\left\{\left[diag(K^2(V),V^2K^2(V))\otimes I_{2d^\dagger}\right]\ell_1(t)\ell_1^\top(t)\right\}+o(1),
\end{aligned}$$

where $\ell_1(t) = \ell_1\left(\boldsymbol{\omega}^\dagger(z), \boldsymbol{\theta}^\dagger(z); F_{X_1|z}(X_{1t}), \cdots, F_{X_p|z}(X_{pt})\right)$. Hence, it follows that

$$Var(\hat{\tilde{\boldsymbol{\eta}}}_T^\dagger(z)) = Var(D_T\left(\hat{\tilde{\boldsymbol{\eta}}}^\dagger(z) - \boldsymbol{\eta}^\dagger(z)\right)) = W_2^{*-1}E(W_{1,1}W_{1,1}^\top)(W_2^{*-1})^\top.$$

Now it suffices to prove W_1 is asymptotically normal. To establish the asymptotic normality of W_1, it is equivalent to show that for any nonzero vector $\mathbf{a} \in \mathbb{R}^{4d^\dagger}$, the linear combination $S_T = \mathbf{a}^\top W_1$ is asymptotically normal. Specifically, we employ the small- and large-block arguments as part 2.7.7 of Fan and Yao (2003), and partition S_T into $2k_T + 1$ blocks with large blocks of size l_T, small blocks of size s_T and the last remaining block of size $T - k_T(l_T + s_T)$. Put $l_T = \left[\sqrt{T}/\ln(T)\right]$, $s_T = \left[(\sqrt{T}\ln(T))^\chi\right]$ and $k_T = [T - l_T - s_T]$ with $\varsigma/(4+\varsigma) \le \chi < 1$, where ς is a positive constant specified in condition A6, then $\alpha(T) = o\left(T^{(4+\varsigma)/\varsigma}\right)$ and $k_T\alpha(s_T) = o(1)$ by using harmonic series and $\sum_{j=1}^{\infty}\alpha(j)^{(4+\varsigma)/\varsigma} < \infty$. It follows that

$$\begin{aligned} S_T &= \mathbf{a}^\top W_{1,1} + \mathbf{a}^\top W_{1,2} + \mathbf{a}^\top W_{1,3} \\ &= \mathbf{a}^\top \sum_{r=1}^{k_T}\left(\xi_{1,r}^{(1)} + \sum_{j=1}^{p}\xi_{j+1,r}^{(j+1)}\right) + \mathbf{a}^\top \sum_{r=1}^{k_T}\left(\zeta_{1,r}^{(1)} + \sum_{j=1}^{p}\zeta_{j+1,r}^{(j+1)}\right) + \mathbf{a}^\top\left(\psi_{1,r}^{(1)} + \sum_{j=1}^{p}\psi_{j+1,r}^{(j+1)}\right), \end{aligned}$$

where

$$\xi_{s,r}^{(s)} = \sum_{t=(r-1)(l_T+s_T)+1}^{rl_T+(r-1)s_T} W_{1,s}^t, \quad \zeta_{s,r}^{(s)} = \sum_{t=rl_T+(r-1)s_T+1}^{r(l_T+s_T)} W_{1,s}^t, \quad \psi_{s,r}^{(s)} = \sum_{t=k_T(l_T+s_T)+1}^{T} W_{1,s}^t$$

for $s \in \{1, \cdots, p+1\}$. By applying the proposition 2.7 of Fan and Yao (2003), we have

$$\sum_{r=1}^{k_n}\zeta_{s,r}^{(s)} = o_p(1) \qquad \text{and} \qquad \psi_{s,r}^{(s)} = o_p(1)$$

directly for $s \in \{1, \cdots, p+1\}$. Hence,

$$S_T = \mathbf{a}^\top\sum_{r=1}^{k_T}\xi_{1,r}^{(1)} + \mathbf{a}^\top\sum_{j=1}^{p}\sum_{r=1}^{k_T}\xi_{j+1,r}^{(j+1)} + o_p(1) \equiv T_T + o_p(1).$$

Similarly, we can partition T_T into 2 parts via truncation. Define

$$\xi_{s,r}^{(s)L} = \sum_{t=(r-1)(l_T+s_T)+1}^{rl_T+(r-1)s_T} W_{1,s}^t I(\|W_{1,s}^t\|_2 \le L), \quad \xi_{s,r}^{(s)R} = \sum_{t=(r-1)(l_T+s_T)+1}^{rl_T+(r-1)s_T} W_{1,s}^t I(\|W_{1,s}^t\|_2 > L)$$

for a fixed positive constant L and $s \in \{1, \cdots, p+1\}$. Then

$$\begin{aligned} T_T &= \mathbf{a}^\top \left(\sum_{r=1}^{k_T} \xi_{1,r}^{(1)L} + \sum_{j=1}^{p} \sum_{r=1}^{k_T} \xi_{j+1,r}^{(j+1)L} \right) + \mathbf{a}^\top \left(\sum_{r=1}^{k_T} \xi_{1,r}^{(1)R} + \sum_{j=1}^{p} \sum_{r=1}^{k_T} \xi_{j+1,r}^{(j+1)R} \right) \\ &\equiv T_T^L + T_T^R. \end{aligned}$$

Similar to (AE 5), we have $W_{1,s}^L$ and $W_{1,s}^R$ for $s \in \{1, \cdots, p+1\}$. Let

$$\mathbf{\Gamma}_z^{*L} = D_T^{-1} \int H(K^2(v), v^2 K^2(v)) \ell_1^L \ell_1^{L\top} dF_{\mathbf{X},Z}(\boldsymbol{x}, z + hv)$$

where the vector $\ell_1^L = I(|\ell_1| \leq L)$. Then $\mathbf{\Gamma}_z^{*L} \to \mathbf{\Gamma}_z^*$ as $L \to \infty$ and $Var\left(T_T^L / \sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^{*L} \mathbf{a}}\right) \to 1$ as $T \to \infty$. In the same vein, we can show that $\mathbf{\Gamma}_z^{*R} \to 0$ as $L \to \infty$ and $Var\left(T_T^R / \sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^{*R} \mathbf{a}}\right) \to 1$ as $T \to \infty$. Define

$$\Upsilon_T = \left| E \exp\left(\frac{it(T_T - \mathbf{a}^\top D_T \mathbf{\Xi}_z^*)}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right) - \exp\left(-\frac{t^2}{2} \right) \right|,$$

where $i = \sqrt{-1}$. Note that

$$\begin{aligned} \Upsilon_T \leq\; & E \left| \exp\left(\frac{it(T_T^L - \mathbf{a}^\top D_T \mathbf{\Xi}_z^{*L})}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right) \left[\exp\left(\frac{it(T_T^R - \mathbf{a}^\top D_T \mathbf{\Xi}_z^{*R})}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right) - 1 \right] \right| \\ & + \left| E \exp\left(\frac{it(T_T^L - \mathbf{a}^\top D_T \mathbf{\Xi}_z^{*L})}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right) - \prod_{r=1}^{k_T} E \exp\left[\frac{it\mathbf{a}^\top \left(\xi_{1,r}^{(1)L} + \sum_{j=1}^{p} \xi_{j+1,r}^{(j+1)L} - \frac{1}{k_T} D_T \mathbf{\Xi}_z^{*L} \right)}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right] \right| \\ & + \left| \prod_{r=1}^{k_T} E \exp\left[\frac{it\mathbf{a}^\top \left(\xi_{1,r}^{(1)L} + \sum_{j=1}^{p} \xi_{j+1,r}^{(j+1)L} - \frac{1}{k_T} D_T \mathbf{\Xi}_z^{*L} \right)}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right] - \exp\left(-\frac{t^2}{2} \frac{\mathbf{a}^\top \mathbf{\Gamma}_z^{*L} \mathbf{a}}{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}} \right) \right| \\ & + \left| \exp\left(-\frac{t^2}{2} \frac{\mathbf{a}^\top \mathbf{\Gamma}_z^{*L} \mathbf{a}}{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}} \right) - \exp\left(-\frac{t^2}{2} \right) \right|. \end{aligned}$$

For the first term on the right-hand side of the expression above, we can obtain its bound

$$E \left(\left| \exp\left(\frac{it\left(T_T^R - \mathbf{a}^\top D_T \mathbf{\Xi}_z^{*R}\right)}{\sqrt{\mathbf{a}^\top \mathbf{\Gamma}_z^* \mathbf{a}}} \right) - 1 \right| \right)$$

by the property of characteristic function, and the upper bound may tend to zero by choosing a large enough L. The last term may also have the same result by choosing a large L as well. The upper bound of the second term is given as $16(k_T - 1)\alpha(s_T)$ by proposition 2.6 of Fan and Yao (2003). Then, the asymptotic normality follows from the statement that the third term converges to zero, and we can express this term as

$$\begin{aligned}
&\prod_{r=1}^{k_T} E\exp\left[\frac{it\mathbf{a}^\top\left(\xi_{1,r}^{(1)L}+\sum_{j=1}^{p}\xi_{j+1,r}^{(j+1)L}-\frac{1}{k_T}D_T\boldsymbol{\Xi}_z^{*L}\right)}{\sqrt{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}}\right]-\exp\left(-\frac{t^2}{2}\frac{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}}{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}\right)\\
=\;& E\exp\left[\frac{it\mathbf{a}^\top\sum_{j=1}^{k_n}\sum_{r=1}^{k_T}\left(\xi_{1,r}^{(1)L}+\sum_{j=1}^{p}\xi_{j+1,r}^{(j+1)L}-\frac{1}{k_T}D_T\boldsymbol{\Xi}_z^{*L}\right)}{\sqrt{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}}\right]-\exp\left(-\frac{t^2}{2}\frac{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}}{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}\right)\\
=\;& E\exp\left[\frac{it\left(T_T^L-\mathbf{a}^\top D_T\boldsymbol{\Xi}_z^{*L}\right)}{\sqrt{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}}\right]-\exp\left(-\frac{t^2}{2}\frac{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}}{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*}\mathbf{a}}\right).
\end{aligned}$$

Now it suffices to show that $T_T^L - \mathbf{a}^\top D_T\boldsymbol{\Xi}_z^{*L} \xrightarrow{d} N\left(0, \mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}\right)$ while treating $\left\{\xi_{s,r}^{(s)L}\right\}$ as a sequence of independent random variables. It follows from Theorem 2.20 of Fan and Yao (2003) that, for any given $\epsilon > 0$, the Lindberg condition

$$\lim_{k_T\to\infty}\frac{1}{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}}\sum_{j=1}^{k_T}E\left(\varrho_j^\top\varrho_j I\left(\left|\varrho_j^\top\varrho_j\right|>\epsilon\sqrt{\mathbf{a}^\top\boldsymbol{\Gamma}_z^{*L}\mathbf{a}}\right)\right)\to 0,$$

where $\varrho_j = \mathbf{a}^\top\left(\xi_{1,r}^{(1)L}+\sum_{j=1}^{p}\xi_{j+1,r}^{(j+1)L}-\frac{1}{k_T}D_T\boldsymbol{\Xi}_z^{*L}\right)$. Therefore, combining the results of W_1 and W_2, we finish the proof of the theorem. □

A4 Proof of Theorem 3

Note that the function $L_t(\hat{\boldsymbol{\eta}}) = L_t\left(\hat{\boldsymbol{\eta}}; \widehat{F}_{\boldsymbol{X}|z}(\boldsymbol{x}_t)\right)$, then the GLRT statistic can be written as

$$\begin{aligned}
M_T &= \sum_{t=1}^{T}L_t\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t)\right)-\sum_{t=1}^{T}L_t\left(\widetilde{\boldsymbol{\eta}}^\dagger(Z_t)\right)\\
&= \sum_{t=1}^{T}\left[L_t\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t)\right)-L_t\left(\boldsymbol{\eta}^\dagger(Z_t)\right)\right]-\sum_{t=1}^{T}\left[L_t\left(\widetilde{\boldsymbol{\eta}}^\dagger(Z_t)\right)-L_t\left(\boldsymbol{\eta}^\dagger(Z_t)\right)\right]\\
&\equiv M_{1T}-M_{2T}.
\end{aligned}$$

The second term M_{2T} is the canonical likelihood ratio statistic and it has an asymptotic chi-quare distribution, then we need to investigate the behavior of the first term M_{1T}.

Applying the Taylor's expansion to $L_t\left(\hat{\boldsymbol{\eta}}(Z_t)\right)$ around the point $\boldsymbol{\eta}(Z_t)$, then we have

$$\begin{aligned}
M_{1T} &= \sum_{t=1}^{T}\left[L_t\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t)\right) - L_t\left(\boldsymbol{\eta}^{\dagger}(Z_t)\right)\right] \\
&\approx \sum_{t=1}^{T}\ell_1^{\top}\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) H(1, Z_t - Z_s)\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t) - \boldsymbol{\eta}^{\dagger}(Z_t)\right) K_h(Z_t - Z_s).
\end{aligned}$$

By Theorem 2, it is equivalent to establish the asymptotic distribution of the term

$$\begin{aligned}
\widetilde{M}_{1T} &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\ell_1^{\top}\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^{\dagger}(Z), \boldsymbol{U}(Z)))\right] \times \\
&\quad H\left(K(\frac{Z_t - Z_s}{h}), \frac{Z_t - Z_s}{h}K(\frac{Z_t - Z_s}{h})\right) H(1, Z_t - Z_s)\ell_1\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) \times \\
&\quad K_h(Z_t - Z_s)\left(h^2 + \frac{1}{\sqrt{Th}}\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\ell_1^{\top}\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^{\dagger}(Z), \boldsymbol{U}(Z)))\right] \times \\
&\quad H\left(h, (Z_t - Z_s)^2\right)\ell_1\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) K^2\left(\frac{Z_t - Z_s}{h}\right)\left(1 + \frac{1}{\sqrt{Th^5}}\right) \\
&= \widetilde{M}_{1T}^1 + \widetilde{M}_{1T}^2,
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{M}_{1T}^1 &= \frac{1}{T}\sum_{t=1}^{T}\ell_1^{\top}\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^{\dagger}(Z), \boldsymbol{U}(Z)))\right] \times \\
&\quad H(h, 0)\,\ell_1\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) K^2(0)\left(1 + \frac{1}{\sqrt{Th^5}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{M}_{1T}^2 &= \frac{2}{T(T-1)}\sum_{s<t}\ell_1^{\top}\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) E^{-1}\left[H(K(V), V^2K(V))(-2\ell_2(\boldsymbol{\eta}^{\dagger}(Z), \boldsymbol{U}(Z)))\right] \times \\
&\quad H\left(h, (Z_t - Z_s)^2\right)\ell_1\left(\hat{\boldsymbol{\eta}}^{\dagger}(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) K^2\left(\frac{Z_t - Z_s}{h}\right)\left(1 + \frac{1}{\sqrt{Th^5}}\right) \\
&\equiv \frac{2}{T(T-1)}\sum_{s<t}\widetilde{\mathcal{M}}(Z_t, Z_s).
\end{aligned}$$

For the term $\widetilde{M}_{1T}^{1}$, we further have it converges to

$$\begin{aligned}\mu_T &= \mathbf{1}_{4d^\dagger}^\top E\left[\ell_1\left(\boldsymbol{\eta}^\dagger(Z),\boldsymbol{U}(Z)\right)\ell_1^\top\left(\boldsymbol{\eta}^\dagger(Z),\boldsymbol{U}(Z)\right)\right]H\left(h,0\right)\times\\ &\quad E^{-1}\left[H(K(V),V^2K(V))(-2\ell_2(\boldsymbol{\eta}^\dagger(Z),\boldsymbol{U}(Z)))\right]\mathbf{1}_{4d^\dagger}K^2\left(0\right)\left(1+\frac{1}{\sqrt{Th^5}}\right)\end{aligned}$$

by Slutsky's theorem.

To establish the asymptotic distribution of the U-statistic $\widetilde{M}_{1T}^{2}$, we apply a general Central Limit Theorem (CLT) from Fan and Li (1999) for dependent observations. It is clear that $E\left(\widetilde{M}_{1T}^{2}\right)=E\left(E\left(\widetilde{M}_{1T}^{2}\right)|Z_t\right)=0$ and the covariate is α-mixing process. In order to construct large and small blocks, we follow Fan and Li (1999) and define the parameters l and m as $l=[T^{\frac{1}{4}}]$ and $m=o(l)$, where l and m represent the number of elements in the large and small blocks, respectively. Without loss of generality, we take $m=o(h^{-\frac{1}{2}})$, then the value of m also diverges to infinity as T goes to infinity. Let $\left\{\widetilde{Z}_t\right\}_{t=1}^{T}$ be a sequence of independent and identically distributed variables which share the same marginal distribution as that of $\{Z_t\}_{t=1}^{T}$, and then we have

$$\begin{aligned}&E\left(\widetilde{\mathcal{M}}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)\\ =\ &E\Bigg\{\Bigg[\ell_1^\top\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_t),\widehat{\boldsymbol{U}}(\widetilde{Z}_t)\right)E^{-1}\left[H(K(V),V^2K(V))(-2\ell_2(\boldsymbol{\eta}^\dagger(\widetilde{Z}),\boldsymbol{U}(\widetilde{Z})))\right]\times\\ &H\left(h,(\widetilde{Z}_t-\widetilde{Z}_s)^2\right)\ell_1\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s),\widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)K^2\left(\frac{\widetilde{Z}_t-\widetilde{Z}_s}{h}\right)\left(1+\frac{1}{\sqrt{Th^5}}\right)\Bigg]^2\Bigg\}.\end{aligned}$$

Specifically, when $k\le 2d^\dagger$, it follows

$$\begin{aligned}&\int\ \left[\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_t),\widehat{\boldsymbol{U}}(\widetilde{Z}_t)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s),\widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)hK^2\left(\frac{\widetilde{Z}_t-\widetilde{Z}_s}{h}\right)\left(1+\frac{1}{\sqrt{Th^5}}\right)\right]^2\times\\ &\quad f_{\boldsymbol{X},\widetilde{Z}_t,\widetilde{Z}_s}(\boldsymbol{X},\widetilde{Z}_t,\widetilde{Z}_s)d\widetilde{Z}_td\widetilde{Z}_sd\boldsymbol{X}\\ =&\int\left[\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s+hY),\widehat{\boldsymbol{U}}(\widetilde{Z}_s+hY)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s),\widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)hK^2\left(Y\right)\left(1+\frac{1}{\sqrt{Th^5}}\right)\right]^2\times\\ &\quad f_{\boldsymbol{X},\widetilde{Z}_s+hY,\widetilde{Z}_s}(\boldsymbol{X},\widetilde{Z}_s+hY,\widetilde{Z}_s)hd\widetilde{Z}_sdYd\boldsymbol{X}\\ =&\ O\left(h^3\right),\end{aligned}$$

where $hY=\widetilde{Z}_t-\widetilde{Z}_s$ and $\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s),\widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)$ is the kth element of vector $\ell_1\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s),\widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)$.

And when $k > 2d^\dagger$, we have

$$\begin{aligned}
& \int \left[\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_t), \widehat{\boldsymbol{U}}(\widetilde{Z}_t)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s), \widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right)(\widetilde{Z}_t - \widetilde{Z}_s)^2 K^2\left(\frac{\widetilde{Z}_t - \widetilde{Z}_s}{h}\right)\left(1 + \frac{1}{\sqrt{Th^5}}\right)\right]^2 \times \\
& \quad f_{\boldsymbol{X}, \widetilde{Z}_t, \widetilde{Z}_s}(\boldsymbol{X}, \widetilde{Z}_t, \widetilde{Z}_s) d\widetilde{Z}_t d\widetilde{Z}_s d\boldsymbol{X} \\
= & \int \left[\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s + hY), \widehat{\boldsymbol{U}}(\widetilde{Z}_s + hY)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(\widetilde{Z}_s), \widehat{\boldsymbol{U}}(\widetilde{Z}_s)\right) h^2 Y^2 K^2(Y)\left(1 + \frac{1}{\sqrt{Th^5}}\right)\right]^2 \times \\
& \quad f_{\boldsymbol{X}, \widetilde{Z}_s + hY, \widetilde{Z}_s}(\boldsymbol{X}, \widetilde{Z}_s + hY, \widetilde{Z}_s) h d\widetilde{Z}_s dY d\boldsymbol{X} \\
= & \; O\left(h^5\right).
\end{aligned}$$

Next, we focus on the fourth order of $\widetilde{\mathcal{M}}(\widetilde{Z}_t, \widetilde{Z}_s)$, and we then obtain the orders of $E\left(\widetilde{\mathcal{M}}^2(\widetilde{Z}_t, \widetilde{Z}_s)\right)$ are equal to $O(h^5)$ when $k \le 2d^\dagger$ and $O(h^9)$ when $k > 2d^\dagger$, respectively.

Further, for the term $\int E^2\left[\widetilde{\mathcal{M}}(Z_1, Z_t)\widetilde{\mathcal{M}}(Z_1, Z_s) | Z_t, Z_s\right] f_{Z_t, Z_s}(Z_t, Z_s) dZ_t dZ_s$, when $k \le 2d^\dagger$, its order is equal to

$$\begin{aligned}
& \int\left\{\int \ell_{1(k)}^2\left(\hat{\boldsymbol{\eta}}^\dagger(Z_1), \widehat{\boldsymbol{U}}(Z_1)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right) K^2\left(\frac{Z_1 - Z_t}{h}\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) \times \right. \\
& \left. K^2\left(\frac{Z_1 - Z_s}{h}\right) h^2\left(1 + \frac{1}{\sqrt{Th^5}}\right)^2 f_{\boldsymbol{X}, Z_1 | Z_t, Z_s}(\boldsymbol{X}, Z_1) dZ_1 d\boldsymbol{X}\right\}^2 f_{Z_t, Z_s}(Z_t, Z_s) dZ_t dZ_s \\
= & \int\left\{\int \ell_{1(k)}^2\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t + h\tilde{Y}), \widehat{\boldsymbol{U}}(Z_t + h\tilde{Y})\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_t), \widehat{\boldsymbol{U}}(Z_t)\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) \times \right. \\
& \left. K^2\left(\tilde{Y}\right) K^2\left(\frac{Z_t - Z_s}{h} + \tilde{Y}\right) h^3\left(1 + \frac{1}{\sqrt{Th^5}}\right)^2 f_{\boldsymbol{X}, Z_t + h\tilde{Y} | Z_t, Z_s}(\boldsymbol{X}, Z_t + h\tilde{Y}) d\tilde{Y} d\boldsymbol{X}\right\}^2 \times \\
& f_{Z_t, Z_s}(Z_t, Z_s) dZ_t dZ_s \\
= & \int\left\{\int \ell_{1(k)}^2\left(\hat{\boldsymbol{\eta}}^\dagger(Z_s + h\bar{Y} + h\tilde{Y}), \widehat{\boldsymbol{U}}(Z_s + h\bar{Y} + h\tilde{Y})\right)\ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_s + h\bar{Y}), \widehat{\boldsymbol{U}}(Z_s + h\bar{Y})\right) \times \right. \\
& \ell_{1(k)}\left(\hat{\boldsymbol{\eta}}^\dagger(Z_s), \widehat{\boldsymbol{U}}(Z_s)\right) K^2\left(\tilde{Y}\right) K^2\left(\bar{Y} + \tilde{Y}\right) h^3\left(1 + \frac{1}{\sqrt{Th^5}}\right)^2 \times \\
& \left. f_{\boldsymbol{X}, Z_s + h\bar{Y} + h\tilde{Y} | \bar{Y}, Z_s}(\boldsymbol{X}, Z_s + h\bar{Y} + h\tilde{Y}) d\tilde{Y} d\boldsymbol{X}\right\}^2 h f_{Z_s + h\bar{Y}, Z_s}(Z_s + h\bar{Y}, Z_s) d\bar{Y} dZ_s \\
= & \; O(h^7),
\end{aligned}$$

where $h\tilde{Y} = Z_1 - Z_t$ and $h\bar{Y} = Z_t - Z_s$. Similarly, its order equals $O(h^{11})$ when $k > 2d^\dagger$.

Moreover,

$$\max_{t\neq s}\int E^2\left[\widetilde{\mathcal{M}}(Z_1,Z_t)\widetilde{\mathcal{M}}(Z_1,Z_s)|Z_t,Z_s\right]f_{Z_t,Z_s}(Z_t,Z_s)dZ_tdZ_s\times\frac{m^4}{E^2\left(\widetilde{\mathcal{M}}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)}=O(\sqrt{h}).$$

Therefore, we know

$$\widetilde{M}_{1T}^2\xrightarrow{d}N\left(0,\frac{2}{T^2}E\left(\widetilde{\mathcal{M}}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)\right).$$

Since $\widetilde{\boldsymbol{\eta}}^{\dagger}(Z_t)$ is the maximum likelihood estimator under the null hypothesis, the term M_{2T} vanishes compared to M_{1T} under the null hypothesis, and $\widetilde{\mathcal{M}}(\widetilde{Z}_t,\widetilde{Z}_s)$ converges to $\mathcal{M}(\widetilde{Z}_t,\widetilde{Z}_s)$, we then have

$$\frac{T^2\mu_T}{E\left(\mathcal{M}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)}M_T\xrightarrow{d}N\left(\frac{(T\mu_T)^2}{E\left(\mathcal{M}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)},\frac{2(T\mu_T)^2}{E\left(\mathcal{M}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)}\right)$$

by the Theorem 2.1 in Fan and Li (1999). Hence, we obtain

$$\frac{T^2\mu_T}{E\left(\mathcal{M}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)}M_T\xrightarrow{d}\chi^2\left(\frac{(T\mu_T)^2}{E\left(\mathcal{M}^2(\widetilde{Z}_t,\widetilde{Z}_s)\right)}\right).$$

The stated conclusion follows. □