Goodness-of-fit tests for multivariate copula-based time series models

Betina Berghaus, Axel Bücher

Goodness-of-fit tests for multivariate copula-based time series models

Betina Berghaus[*] and Axel Bücher[†]

May 9, 2014

Abstract

In recent years, stationary time series models based on copula functions became increasingly popular in econometrics to model nonlinear temporal and cross-sectional dependencies. Within these models, we consider the problem of testing the goodness-of-fit of the parametric form of the underlying copula. Our approach is based on a dependent multiplier bootstrap and it can be applied to any stationary, strongly mixing time series. The method extends recent i.i.d. results by Kojadinovic, Yan and Holmes [I. Kojadinovic, Y. Yan and M. Holmes, Fast large sample goodness-of-fit tests for copulas, Statistica Sinica 21 (2011), 841–871] and shares the same computational benefits compared to methods based on a parametric bootstrap. The finite-sample performance of our approach is investigated by Monte Carlo experiments for the case of copula-based Markovian time series models.

Keywords and Phrases: Empirical process, multiplier central limit theorem, Markov chain, multivariate observations, pseudo-observations, ranks, semiparametric copula model.

1 Introduction

It is an issue of fundamental importance in econometrics to understand and model both temporal and cross-sectional dependence between time series.

---

[*]Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany. E-mail: betina.berghaus@rub.de

[†]Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany. E-mail: axel.buecher@rub.de.

Nowadays, due to their capability of describing non-linear dependencies, copula functions provide a widely accepted tool to approach this problem, see, e.g., Frees and Valdez (1998); McNeil et al. (2005) for applications in actuarial sciences or finance. As a common practice, the copula C used to describe the dependence of interest is assumed to belong to a parametric copula family, parametrized by a finite-dimensional parameter $\theta \in \mathcal{P} \subset \mathbb{R}^p$, i.e.,

$$C \in \mathscr{C} = \{C_\theta : \theta \in \mathcal{P}\}.$$

In order to achieve maximal robustness with respect to potential marginal misspecification, the marginals are often allowed to follow arbitrary absolutely continuous cumulative distribution functions (c.d.f.s), in which case we speak of a semiparametric copula model.

In most applications hitherto, copulas are used to model the contemporary dependence of a stationary d-dimensional time series. Serial dependence is often either ignored or filtered out in a preliminary step. Relying on i.i.d. assumptions, several estimation techniques for the unknown parameter θ have been proposed, including the method of moments based on inversion of Spearman's rho or Kendall's tau (Berg and Quessy, 2009), minimum distance type estimators (Tsukahara, 2005) or the pseudo-maximum-likelihood estimator (Genest et al., 1995). Chen and Fan (2006a) and Rémillard (2010) developed the asymptotics for the pseudo-ML estimator and moment-based estimators, respectively, in case they are based on residuals of marginally fitted time series models. The asymptotics turn out to be the same as in the i.i.d. case which justifies the common practice of filtering out serial dependence features.

In recent years, it also became increasingly popular to model the temporal dependence of a stationary time series by copulas. Building upon models from Darsow et al. (1992), Chen and Fan (2006b) provide asymptotics for the pseudo-ML estimator for univariate stationary Markovian copula models. In these time series models, copulas are used to model the serial dependence at lag one. We refer to Beare (2010), who derived the corresponding mixing properties, for a number of references to applications of these models. Extensions to the multivariate case which also cover a joint modeling of the contemporary and the serial dependence are given in Rémillard et al. (2012).

All the afore-mentioned estimation methods depend on the underlying assumption that the unknown copula indeed stems from the model $\mathscr{C}$. Not surprisingly, much research has been devoted to test the assumption $H_0 : C \in \mathscr{C}$, see Genest et al. (2009) for a recent review in the i.i.d. case. Most of the proposed methods, including those that yield to the best finite-sample performance in large-scale simulation studies, are based on the parametric boot-

strap. This makes their application quite heavy with respect to computational feasibility. An attractive alternative has been proposed by Kojadinovic and Yan (2011) and Kojadinovic et al. (2011), based on a multiplier bootstrap technique.

With respect to temporal dependence and Markovian copula models, much less has been done. To the best of our knowledge, the only reference on goodness-of-fit tests in these models is given by Rémillard et al. (2012). These authors propose a parametric bootstrap procedure to test the goodness-of-fit and mention that, apart from their procedure, *there exist almost no formal goodness-of-fit tests for copulas in a serially dependent context.*

The present paper tries to fill this gap. Our paper can be regarded as an extension of the afore-mentioned, computationally attractive goodness-of-fit tests in Kojadinovic and Yan (2011) and Kojadinovic et al. (2011) to a general serial dependent setting. Our main assumptions are that the observed time series is strongly mixing and that the parametric estimator allows for a linear expansion in terms of a function of the marginal ranks. Among other settings, this covers the case of multivariate Markovian copula models along with the corresponding pseudo-ML estimator (Chen and Fan, 2006b; Rémillard et al., 2012). The method can also be used to test the contemporary dependence of the raw data rather than that of residuals from marginal time series models.

The test statistic is based on a suitable difference between the empirical copula C_n and a parametric estimator C_{θ_n}, details are given in Section 2 below. We prove weak convergence of the corresponding process $\sqrt{n}(C_n - C_{\theta_n})$. In order to get access to the critical values of the test statistcs, we adapt the dependent multiplier bootstrap of Bücher and Kojadinovic (2013) to the present setting, following ideas of Kojadinovic and Yan (2011) and Kojadinovic et al. (2011). The finite-sample performance of our methods are investigated by Monte Carlo experiments.

The remaining parts of this paper are organized as follows: in Section 2, we present the basic setting (Section 2.1), introduce the test statistics (Section 2.2) and show how critical values can be obtained via the dependent multiplier bootstrap (Section 2.3). The main result of the paper is given in Theorem 2.2. In Section 3, we discuss some implementation issues, while Section 4 presents simulation results. Finally, Section 5 provides an illustrative data application on gasoline price markups, while all technical details and proofs are deferred to a sequence of appendices.

2 Goodness-of-Fit Tests

2.1 The general framework

Even though the common copula-based Markovian time series models provide the main motivation of our work, our methods can be easily worked out in a slightly more general setting at no additional cost. The special Markovian case is treated in Example 2.1 below.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)'$ be a d-dimensional random vector with joint cumulative distribution function F and continuous marginal c.d.f.s $F_1, \ldots, F_d$. The copula C of F, or, equivalently, the copula of $\boldsymbol{X}$, is defined as the c.d.f. of the random vector $\boldsymbol{U} = (U_1, \ldots, U_d)'$ that arises from marginal application of the probability integral transform, i.e., $U_\mu = F_\mu(X_\mu)$ for $\mu = 1, \ldots, d$. By construction, the marginal c.d.f.s of C are standard uniform on $[0, 1]$. By Sklar's Theorem, C is the unique function for which we have

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$$

for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

Let $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ with $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})'$ be a stationary sequence of d-variate random vectors with joint c.d.f. F, continuous marginal c.d.f.s $F_1, \ldots, F_d$ and copula C as described above. For $-\infty \le a \le b \le \infty$, let $\mathcal{F}_a^b$ denote the sigma-field generated by those $\boldsymbol{X}_i$ for which $i \in \{a, a+1, \ldots, b\}$ and define, for $k \ge 1$,

$$\alpha_{\boldsymbol{X}}(k) = \sup\left\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_{-\infty}^i, B \in \mathcal{F}_{i+k}^\infty, i \in \mathbb{Z}\right\}$$

as the alpha-mixing coefficient of the time series $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$. Throughout this paper we will assume that $\alpha_{\boldsymbol{X}}(k) \to 0$ for $k \to \infty$ at a certain rate to be specified later on, i.e., we assume that $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ is strongly mixing.

Within this general framework, we observe a "sample" $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and want to do statistical inference on the unknown copula C. In particular, it is the main purpose of this paper to develop a test for the hypothesis

$$H_0 : C \in \mathscr{C} \quad \text{vs.} \quad H_1 : C \notin \mathscr{C}.$$

This shall be done in a semiparametric way treating the marginals as nuisance parameters, i.e., they remain unspecified.

Example 2.1. The general setting described above includes the important special case of a (potentially multivariate) copula-based Markovian time series model (Rémillard et al., 2012), in short *Markovian copula model*. This model is based on the assumption that the d-dimensional stationary time

series $(\boldsymbol{Y}_i)_{i\in\mathbb{Z}}$ of interest is Markovian of order 1 with continuous marginal c.d.f.s $F_1, \ldots, F_d$. Obviously, the distribution of the times series is determined through the marginal c.d.f.s and the $2d$-dimensional copula C of the $2d$-dim. time series $\boldsymbol{X}_i = (\boldsymbol{Y}_{i-1}, \boldsymbol{Y}_i), i \in \mathbb{Z}$. It follows from the results in Beare (2010) that $(\boldsymbol{Y}_i)_{i\in\mathbb{Z}}$ is alpha-mixing for many commonly used copula families, and it is easy to see that this property transfers to $(\boldsymbol{X}_i)_{i\in\mathbb{Z}}$.

2.2 The test statistic

The procedure described in the following is a suitable adaptation of two recent publications of Kojadinovic et al. (2011) and Kojadinovic and Yan (2011) which are restricted to the i.i.d. case. As a test statistic, i.e., as a measure of the goodness-of-fit, it is reasonable to calculate a suitable distance between a nonparametric estimator C_n and a parametric estimator C_{θ_n}. It will be a crucial intermediate step to determine the asymptotic distribution of

$$\sqrt{n}(C_n - C_{\theta_n}) = \sqrt{n}(C_n - C) + \sqrt{n}(C - C_{\theta_n}). \tag{2.1}$$

In the following, we will split the discussion into three steps: two for each summand of the latter decomposition and one for assembling both parts.

Step 1: The nonparametric part. For the first summand on the right-hand side of (2.1), the standardized nonparametric part, we can rely on the empirical copula dating back to Deheuvels (1979) and Rüschendorf (1976). For its definition, we first introduce pseudo-observations from the copula C, defined as $\hat{\boldsymbol{U}}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{id})$ with $\hat{U}_{i\mu} = (n+1)^{-1}R_{i\mu}$ and with $R_{i\mu}$ denoting the rank of $X_{i\mu}$ among $X_{1\mu}, \ldots, X_{n\mu}$, $\mu = 1, \ldots, d$. The empirical copula C_n is defined as the empirical distribution function of $\hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_n$, i.e.,

$$C_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\hat{\boldsymbol{U}}_i \le \boldsymbol{u}), \qquad \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d.$$

The associated standardized version, called the empirical copula process,

$$\mathbb{C}_n : [0,1]^d \to \mathbb{R}, \quad \boldsymbol{u} \mapsto \mathbb{C}_n(\boldsymbol{u}) = \sqrt{n}(C_n - C)(\boldsymbol{u}),$$

is well-understood and has been investigated under slightly different conditions in Rüschendorf (1976); Fermanian et al. (2004); Segers (2012); Bücher and Volgushev (2013), among others. Weak convergence with respect to the supremum-distance holds provided that there are no ties, marginally, that $\alpha_{\boldsymbol{X}}(k) = O(k^{-a})$ for some $a > 1$ and provided a smoothness condition on

C originating from Segers (2012) (see Condition I in Appendix A below) is satisfied, see Bücher and Volgushev (2013), Corollary 2.5, relying on Bücher (2013), Theorem 1.

More precisely, we have $\mathbb{C}_n = \sqrt{n}(C_n - C) \rightsquigarrow \mathbb{C}_C$ in $(\ell^\infty[0,1]^d, \|\cdot\|_\infty)$, where $\mathbb{C}_C$ can be represented as

$$\mathbb{C}_C(\boldsymbol{u}) = \mathbb{B}_C(\boldsymbol{u}) - \sum_{\mu=1}^{d} C^{[\mu]}(\boldsymbol{u})\mathbb{B}_C(\boldsymbol{u}^\mu), \quad \boldsymbol{u} \in [0,1]^d,$$

with $\boldsymbol{u}^\mu$ denoting the vector with u_μ in the μ-th component and the other components equal to 1 and with $C^{[\mu]}$ denoting the μth partial derivative of C. Moreover, $\mathbb{B}_C$ denotes a centered Gaussian process on $[0,1]^d$ with covariance function

$$\mathrm{Cov}\{\mathbb{B}_{\mathbb{C}}(\boldsymbol{u}), \mathbb{B}_C(\boldsymbol{v})\} = \sum_{i \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{1}(\boldsymbol{U}_0 \leq \boldsymbol{u}), \mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{v})\}, \quad \boldsymbol{u}, \boldsymbol{v} \in [0,1]^d,$$

where $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{id})'$ with $U_{i\mu} = F_\mu(X_{i\mu})$ for $\mu = 1, \ldots, d$.

Step 2: The parametric part. For the second summand on the right-hand side of (2.1), the parametric part, let us suppose that the hypothesis H_0 is met and let θ_0 denote the true, unknown parameter in Θ. Several semi-parametric estimators for θ_0 have been proposed, which we denote by θ_n in the following. Among the most popular ones are the maximum pseudo-likelihood estimator (Genest et al., 1995) or the moment-based estimator derived from inversion of the sample version of Spearman's rho (see, for instance, Berg and Quessy, 2009). Under suitable smoothness and mixing conditions, the behavior of $\sqrt{n}(C_{\theta_n} - C)$ can be derived from the asymptotics of $\Theta_n = \sqrt{n}(\theta_n - \theta_0)$. For the latter, we will assume that we can write

$$\Theta_n = \sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} J_{\theta_0}(\hat{\boldsymbol{U}}_i) + o_p(1),$$

where, for any $\theta \in \mathcal{P}$, $J_\theta : [0,1]^d \to \mathbb{R}^p$ is a (known) score-function satisfying certain smoothness conditions that are specified in Definition V in the appendix. Note that both of the afore-mentioned semi-parametric estimators allow for such a decomposition. We will show later that, under suitable conditions on the model and on the mixing rate, Θ_n converges weakly to some limit which we denote by Θ. Moreover, we will obtain that

$$\sqrt{n}(C_{\theta_n} - C) = \nabla C_{\theta_0}\Theta_n + o_P(1) \rightsquigarrow \nabla C_{\theta_0}\Theta \tag{2.2}$$

in $(\ell^\infty[0,1]^d, \|\cdot\|_\infty)$, where $\nabla C_\theta = \nabla C_\theta(\boldsymbol{u}) \in \mathbb{R}^{1 \times p}$ denotes the gradient of $\theta \mapsto C_\theta(\boldsymbol{u})$.

Step 3: Joining the nonparametric and the parametric part. It will be proved in Theorem 2.2 below that the two summands on the right-hand side of (2.1) converge jointly. Hence, we can conclude that

$$\sqrt{n}(C_n - C_{\theta_n}) \rightsquigarrow \mathbb{C}_C - \nabla C_{\theta_0}\Theta$$

in $(\ell^\infty[0,1]^d, \|\cdot\|_\infty)$. Based on this result, the continuous mapping theorem allows to derive the asymptotic distribution of suitable test statistics for H_0. Of particular interest are Kolomogorov-Smirnov-type tests

$$\mathcal{T}_n = \sqrt{n} \sup_{\boldsymbol{u}\in[0,1]^d} |C_n(\boldsymbol{u}) - C_{\theta_n}(\boldsymbol{u})| \qquad (2.3)$$

and Cramér-von-Mises tests

$$\mathcal{S}_n = n\int_{[0,1]^d} \{C_n(\boldsymbol{u}) - C_{\theta_n}(\boldsymbol{u})\}^2 \, dC_n(\boldsymbol{u}) = \sum_{i=1}^{n}\{C_n(\hat{\boldsymbol{U}}_i) - C_{\theta n}(\hat{\boldsymbol{U}}_i)\}^2. \quad (2.4)$$

We obtain that $\mathcal{T}_n \rightsquigarrow \mathcal{T}_C$ and $\mathcal{S}_n \rightsquigarrow \mathcal{S}_C$ where

$$\mathcal{T}_C = \sup_{\boldsymbol{u}\in[0,1]^d} |\mathbb{C}_C(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\Theta|,$$

$$\mathcal{S}_C = \int_{[0,1]^d} \{\mathbb{C}_C(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\Theta\}^2 \, dC(\boldsymbol{u}).$$

We reject the null hypothesis for large values of $\mathcal{T}_n$ or of $\mathcal{S}_n$. To get access to the critical values of the limiting distributions, we propose a bootstrap procedure for the process $\sqrt{n}(C_n - C_{\theta_n})$ in the following subsection.

2.3 Critical values via dependent multiplier bootstrap

In this subsection, we construct a bootstrap approximation for both processes of the decomposition in (2.1).

Step 1: The nonparametric part. The multiplier bootstrap procedure for the empirical copula process is a well-established tool for statistical inference on copulas. It has been proposed in Scaillet (2005) and Rémillard and Scaillet (2009) for i.i.d. data sets, and has been applied to different testing issues in Genest et al. (2011); Kojadinovic et al. (2011); Bücher et al. (2011, 2012), among others. The procedure has been generalized to the alpha-mixing case in Bücher and Ruppert (2013) with several further improvements in Bücher and Kojadinovic (2013).

Let $B \in \mathbb{N}$ be some large integer. For $b \in \{1, \ldots, B\}$ and $\boldsymbol{u} \in [0,1]^d$, set

$$\mathbb{C}_n^{(b)}(\boldsymbol{u}) = \mathbb{B}_n^{(b)}(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_n^{[\mu]}(\boldsymbol{u})\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu)$$

with

$$\mathbb{B}_n^{(b)}(\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)}\{\mathbb{1}(\hat{\boldsymbol{U}}_i \leq \boldsymbol{u}) - C_n(\boldsymbol{u})\}.$$

Here, the multipliers $(Z_{i,n}^{(b)})_{i \in \mathbb{Z}}$ are independent sequences of identically distributed random variables with mean 0 and variance 1, which are independent of $(\boldsymbol{X}_i)_{i \in \mathbb{N}}$ and can be chosen to be either i.i.d. (provided the original time series is serially independent) or to satisfy some weak dependence condition (see Definition A.1 in Appendix A). In the latter case, we speak of a dependent multiplier bootstrap. A possible way of simulating these multipliers is specified in Section 3.1. Moreover, $C_n^{[\mu]}$ denotes an appropriate consistent estimator of $C^{[\mu]}$, for instance an estimator based on finite-differencing as proposed in Rémillard and Scaillet (2009) or more precisely in Segers (2012), i.e.,

$$C_n^{[\mu]}(\boldsymbol{u}) = \begin{cases} \frac{C_n(\boldsymbol{u}+h_n\boldsymbol{e}_\mu) - C_n(\boldsymbol{u}-h_n\boldsymbol{e}_\mu)}{2h_n} & \text{if } u_\mu \in [h_n, 1-h_n] \\ C_n^{[\mu]}(u_1, \ldots, u_{\mu-1}, h_n, u_{\mu+1}, \ldots, u_d) & \text{if } u_\mu \in [0, h_n) \\ C_n^{[\mu]}(u_1, \ldots, u_{\mu-1}, 1-h_n, u_{\mu+1}, \ldots, u_d) & \text{if } u_\mu \in (1-h_n, 1] \end{cases}$$

with a sequence $h_n \to 0$ such that $\inf_n h_n\sqrt{n} > 0$ and with $\boldsymbol{e}_\mu$ denoting the μ-th unit vector in $\mathbb{R}^d$.

Now, under conditions on the mixing rate and on the dependent multipliers, we have

$$\left(\sqrt{n}(C_n - C), \mathbb{C}_n^{(1)}, \ldots, \mathbb{C}_n^{(B)}\right) \rightsquigarrow \left(\mathbb{C}_C, \mathbb{C}_C^{(1)}, \ldots, \mathbb{C}_C^{(B)}\right) \tag{2.5}$$

in $(\ell^\infty[0,1]^d, \|\cdot\|_\infty)^{B+1}$, where $\mathbb{C}_C^{(1)}, \ldots, \mathbb{C}_C^{(B)}$ denote independent copies of the process $\mathbb{C}_C$. This result can be interpreted in the way that, for n large, $\mathbb{C}_n^{(1)}, \ldots, \mathbb{C}_n^{(B)}$ are 'almost' independent copies of $\mathbb{C}_n$ and can hence be used to approximate its distribution. For a proof, see Bücher and Kojadinovic (2013) for the alpha-mixing case or Segers (2012) for the i.i.d. case.

Step 2: The parametric part. For a bootstrap approximation of the parametric component of decomposition (2.1) we will show in the appendix that we can write

$$\Theta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_{\theta_0}(\boldsymbol{U}_i) + \sum_{\mu=1}^{d} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)(\hat{U}_{i,\mu} - U_{i,\mu}) \right\} + o_P(1). \tag{2.6}$$

8

Observing that $\hat{U}_{i,\mu} = n^{-1}\sum_{j=1}^n \mathbb{1}(U_{j,\mu} \leq U_{i,\mu})$ we can write the second (double) sum as

$$\sum_{\mu=1}^d \left\{ \frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)\{\mathbb{1}(U_{j,\mu} \leq U_{i,\mu}) - U_{i,\mu}\} \right\}.$$

Introducing the random multipliers $(Z_{i,n}^{(b)})_{i=1,\dots,n}$ $(b = 1,\dots,B)$ and replacing each term on the right-hand side of (2.6) by observable estimates we arrive at the following bootstrap approximation for Θ_n,

$$\Theta_n^{(b)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \times \{ J_{\theta_n}(\hat{\boldsymbol{U}}_i) + \hat{K}_{i,n,\theta_n} \}$$

where, for any $\theta \in \mathcal{P}$ and $i \in \{1,\dots,n\}$,

$$\hat{K}_{i,n,\theta} = \sum_{\mu=1}^d \left\{ \frac{1}{n} \sum_{j=1}^n J_\theta^{[\mu]}(\hat{\boldsymbol{U}}_j) \times \{\mathbb{1}(\hat{U}_{i,\mu} \leq \hat{U}_{j,\mu}) - \hat{U}_{j,\mu}\} \right\}. \qquad (2.7)$$

Note that the function J_θ is known for common copula families and common parametric estimators, see also Kojadinovic and Yan (2011); Kojadinovic et al. (2011). Under H_0 and regularity conditions, it will follow from Theorem 2.2 below that

$$\left(\Theta_n, \Theta_n^{(1)}, \dots, \Theta_n^{(B)}\right) \rightsquigarrow \left(\Theta, \Theta^{(1)}, \dots, \Theta^{(B)}\right) \qquad (2.8)$$

where $\Theta, \Theta^{(1)}, \dots, \Theta^{(B)}$ denote independent random vectors each of which has the same distribution as Θ. This result may be of independent interest and can be used to construct confidence bands for the estimation of θ_0 in the alpha-mixing case. In fact, we make use of these confidence bands in the empirical application in Section 5.

Step 3: Joining the nonparametric and the parametric part. As in Section 2.2, the 'nonparametric' convergence in (2.5) and the 'parametric' one in (2.8) actually holds jointly. More precisely, the following theorem summarizes the main results of this paper, including those already mentioned in the previous paragraphs.

Theorem 2.2. *For $b = 1, \dots, B$, let $(Z_{i,n}^{(b)})_{i \in \mathbb{Z}}$ be independent sequences of multipliers satisfying the conditions in Definition A.1.*
(i) Suppose that Conditions I, II and III in Appendix A are met. Then

$$\left(\sqrt{n}(C_n - C), \mathbb{C}_n^{(1)}, \dots, \mathbb{C}_n^{(B)}\right) \rightsquigarrow \left(\mathbb{C}_C, \mathbb{C}_C^{(1)}, \dots, \mathbb{C}_C^{(B)}\right) \qquad (2.9)$$

9

in $(\ell^\infty[0,1]^d, \|\cdot\|_\infty)^{B+1}$, where $\mathbb{C}_C^{(1)}, \dots, \mathbb{C}_C^{(B)}$ denote independent copies of the process $\mathbb{C}_C$.

(ii) If Conditions V and VI hold, if κ from Definition A.1 satisfies $\kappa > 1/\{2(1+\nu)\}$ with ν from Condition V and if $C \in \mathscr{C}$, then

$$\left(\Theta_n, \Theta_n^{(1)}, \dots, \Theta_n^{(B)}\right) \rightsquigarrow \left(\Theta, \Theta^{(1)}, \dots, \Theta^{(B)}\right) \qquad (2.10)$$

where $\Theta, \Theta^{(1)}, \dots, \Theta^{(B)}$ denote i.i.d. d-dimensional random vectors that are normally distributed with mean $\mathbf{0}$ and covariance matrix

$$\Sigma = \sum_{i=-\infty}^{\infty} \mathbb{E}[\{J_{\theta_0}(\boldsymbol{U}_0) + K_{0,\theta_0}\}\{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}^T],$$

where $K_{i,\theta_0} = \sum_{\mu=1}^d \int_{[0,1]^d} J_{\theta_0}^{[\mu]}(\boldsymbol{u})\{\mathbb{1}(U_{i,\mu} \le u_\mu) - u_\mu\}\mathrm{d}C_{\theta_0}(\boldsymbol{u})$ and where J_θ is defined in Condition V.

(iii) If Condition IV and all of the conditions in (i) and (ii) are satisfied, then the convergence in (2.9) and (2.10) is jointly in $\{\ell^\infty[0,1]^d\}^{B+1} \times \mathbb{R}^{(B+1)p}$. Moreover,

$$\left(\sqrt{n}(C_n - C_{\theta_n}), \mathbb{C}_n^{(1)} - \nabla C_{\theta_n}\Theta_n^{(1)}, \dots, \mathbb{C}_n^{(B)} - \nabla C_{\theta_n}\Theta_n^{(B)}\right)$$
$$\rightsquigarrow \left(\mathbb{C}_C - \nabla C_{\theta_0}\Theta, \mathbb{C}_C^{(1)} - \nabla C_{\theta_0}\Theta^{(1)}, \dots, \mathbb{C}_C^{(B)} - \nabla C_{\theta_0}\Theta^{(B)}\right) \quad (2.11)$$

in $\{\ell^\infty([0,1]^d)\}^{B+1}$.

This result suggests to use the following bootstrap approximations for the Kolomogorov-Smirnov-type statistic $\mathcal{T}_n$ and the Cramér-von-Mises-type statistic $\mathcal{S}_n$ defined in (2.3) and (2.4), respectively, namely

$$\mathcal{T}_n^{(b)} = \sqrt{n} \sup_{\boldsymbol{u} \in [0,1]^d} |\mathbb{C}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)}|$$

and

$$\mathcal{S}_n^{(b)} = \int_{[0,1]^d} \{\mathbb{C}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)}\}^2 \, dC_n(\boldsymbol{u})$$
$$= \frac{1}{n} \sum_{i=1}^n \{\mathbb{C}_n^{(b)}(\hat{\boldsymbol{U}}_i) - \nabla C_{\theta_n}(\hat{\boldsymbol{U}}_i)\Theta_n^{(b)}\}^2,$$

for $b = 1, \dots, B$. For a given level $\alpha \in (0,1)$, let $t_{n,B,1-\alpha}$ and $s_{n,B,1-\alpha}$ denote the empirical $(1-\alpha)$-quantile of the sample $\mathcal{T}_n^{(1)}, \dots, \mathcal{T}_n^{(B)}$ and $\mathcal{S}_n^{(1)}, \dots, \mathcal{S}_n^{(B)}$, respectively. The following theorem, whose proof can be found in Section F of the the supplement to Bücher and Kojadinovic (2013), shows that the tests which reject H_0 for $\mathcal{T}_n > t_{n,B,1-\alpha}$ or for $\mathcal{S}_n > s_{n,B,1-\alpha}$ asymptotically hold their level, for $B \to \infty$ and $n \to \infty$.

Theorem 2.3. *Under H_0, if the regularity conditions of Theorem 2.2 are met,*

$$\lim_{B \to \infty} \lim_{n \to \infty} \mathbb{P}(\mathcal{T}_n > t_{n,B,1-\alpha}) = \lim_{B \to \infty} \lim_{n \to \infty} \mathbb{P}(\mathcal{S}_n > s_{n,B,1-\alpha}) = \alpha.$$

An approximate p-value for the test based on $\mathcal{T}_n$ can hence be defined as $\hat{p} = B^{-1} \sum_{b=1}^{B} \mathbb{1}(\mathcal{T}_n^{(b)} > \mathcal{T}_n)$, and similarly for the test based on $\mathcal{S}_n$.

3 Implementation Issues

3.1 Simulating dependent multipliers

The bootstrap version of the test statistics in the previous section depends crucially on a sequence of dependent multipliers satisfying the conditions in Definition A.1 in the appendix. A detailed discussion of how to simulate such a sequence can be found in Bücher and Kojadinovic (2013). For the sake of completeness, we briefly summarize one possible approach originating from Bühlmann (1993), which the former authors found to lead to a reasonable finite-sample performance.

First of all, choose a bandwidth parameter ℓ_n (see Section 3.2 below) and set $b_n = \text{round}\{(\ell_n + 1)/2\}$. Let κ be a positive, bounded function which is symmetric around zero such that $\kappa(x) > 0$ for all $x \in (-1, 1)$. In our simulation study, we use the Parzen-kernel defined as

$$\kappa_P(x) = (1 - 6x^2 + 6|x|^3)\mathbb{1}(|x| \le 1/2) + 2(1 - |x|)^3 \mathbb{1}(1/2 < |x| \le 1).$$

Now, for $j = 1, \ldots, \ell_n$, define weights $\tilde{\omega}_{jn} = \omega_{jn}(\sum_{i=1}^{\ell_n} \omega_{in}^2)^{-1/2}$ where $\omega_{jn} = \kappa\{(j - b_n)/b_n\}$. Finally, let $\xi_1, \ldots, \xi_{n+2b_n-2}$ be i.i.d. random variables which are independent from the sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with $\mathbb{E}[\xi_1] = 0$, $\mathbb{E}[\xi_1^2] = 1$ and $\mathbb{E}[|\xi_1|^m] < \infty$, for any $m \ge 1$. Then the sequence of random variables $Z_{1,n}, \ldots, Z_{n,n}$ defined as

$$Z_{i,n} = \sum_{j=1}^{\ell_n} \tilde{\omega}_{jn} \xi_{j+i-1}$$

asymptotically satisfies the conditions in Definition A.1, see Bücher and Kojadinovic (2013). For a detailed discussion on the properties of these multipliers and an alternative simulation procedure for the multpliers see Bücher and Kojadinovic (2013).

3.2 Choosing the bandwidth parameter ℓ_n

The simulation method described in the previous section depends on the choice of the 'bandwidth' parameter ℓ_n. Within the context of empirical copulas, Bücher and Kojadinovic (2013) derived a closed-form formula for a theoretically optimal choice, where optimality is to be understood as optimality with respect to a certain MSE-minimizing criterion. Moreover, they proposed a data-adaptive estimation procedure for this (theoretically) optimal bandwidth. In the following, we adapt their approach to the processes underlying the goodness-of-fit tests in Section 2.

For $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$, let $\sigma(\boldsymbol{u}, \boldsymbol{v})$ denote the characterizing covariance kernel of the process $\boldsymbol{u} \mapsto \mathbb{C}_C(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\Theta$. Its bootstrap approximation, for a fixed $b \in \{1, \ldots, B\}$, is given by $\mathbb{C}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)}$. The main idea of the subsequent developments is as follows: if $\hat{\sigma}_n$ denotes the covariance kernel of the bootstrap process, conditional on the data, then a theoretically optimal choice of ℓ_n is given by the minimizer of the integrated mean squared error of $\hat{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})$, seen as an estimator for $\sigma(\boldsymbol{u}, \boldsymbol{v})$, with respect to ℓ_n. Unfortunately, necessary closed form expressions for the mean or the variance of $\hat{\sigma}_n$ are out of reach, whence we follow the proposal of Bücher and Kojadinovic (2013) and consider an asymptotically equivalent form of $\mathbb{C}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)}$ instead, for which calculations are feasible.

More precisely, define $\tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}) = n^{-1/2} \sum_{i=1}^n Z_{i,n}^{(b)}\{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\}$ and $\tilde{\Theta}_n^{(b)} = n^{-1/2} \sum_{i=1}^n Z_{i,n}^{(b)}\{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}$, where K_{i,θ_0} is defined in Theorem 2.2, and let $\tilde{\mathbb{C}}_n^{(b)}(\boldsymbol{u}) = \tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}) - \sum_{\mu=1}^d C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}^\mu)$. It follows from the arguments in the proof of Theorem 2.2 that

$$\sup_{\boldsymbol{u} \in [0,1]^d} \left| \mathbb{C}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)} - \{\tilde{\mathbb{C}}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\tilde{\Theta}_n^{(b)}\} \right| = o_P(1).$$

In contrast to $\hat{\sigma}_n$, the (conditional) covariance kernel of the (unobservable) process $\tilde{\mathbb{C}}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\tilde{\Theta}_n^{(b)}$ can be calculated explicitly, at least up to the first-order terms. We have

$$\begin{aligned}
\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v}) &= \text{Cov}_Z\{\tilde{\mathbb{C}}_n^{(b)}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\tilde{\Theta}^{(b)}, \tilde{\mathbb{C}}_n^{(b)}(\boldsymbol{v}) - \nabla C_{\theta_0}(\boldsymbol{v})\tilde{\Theta}^{(b)}\} \\
&= \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}_Z[Z_{i,n}^{(b)} Z_{j,n}^{(b)}] f(\boldsymbol{U}_i, \boldsymbol{u}) f(\boldsymbol{U}_j, \boldsymbol{v}) \\
&= \frac{1}{n} \sum_{i,j=1}^n \varphi\{(i-j)/\ell_n\} f(\boldsymbol{U}_i, \boldsymbol{u}) f(\boldsymbol{U}_j, \boldsymbol{v})
\end{aligned}$$

where Cov_Z and $\mathbb{E}_Z$ denote covariance and expectation conditional on the

data, respectively, and where, for $i = 1, \ldots, n$ and $\boldsymbol{u} \in [0,1]^d$,

$$f(\boldsymbol{U}_i, \boldsymbol{u}) = \mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\{\mathbb{1}(U_{i\mu} \leq u_\mu) - u_\mu\}$$
$$- \nabla C_{\theta_0}(\boldsymbol{u})\{J_{\theta_0}(\boldsymbol{U}_i) + K_{\theta_0,i}\}.$$

Mimicking the proofs of Proposition 5.1 and Proposition 5.2 in Bücher and Kojadinovic (2013), we obtain the following results regarding bias and variance of $\tilde{\sigma}_n$, seen as an estimator for σ.

Lemma 3.1. *Additionally to the conditions assumed in (iii) of Theorem 2.2 suppose that a defined in Condition VI satisfies $a > 3(2 + \nu)/\nu$ and that φ defined in Definition A.1 is twice continuously differentiable on $[-1,1]$ with $\varphi''(0) \neq 0$. Then, for any $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$,*

$$\mathbb{E}[\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})] - \sigma(\boldsymbol{u}, \boldsymbol{v}) = \frac{\Gamma(\boldsymbol{u}, \boldsymbol{v})}{\ell_n^2} + r_{n,1}(\boldsymbol{u}, \boldsymbol{v}),$$

where $\sup_{\boldsymbol{u},\boldsymbol{v} \in [0,1]^d} |r_{1,n}(\boldsymbol{u}, \boldsymbol{v})| = o(\ell_n^{-2})$ and $\Gamma(\boldsymbol{u}, \boldsymbol{v}) = \frac{\varphi''(0)}{2} \sum_{k \in \mathbb{Z}} k^2 \gamma(k, \boldsymbol{u}, \boldsymbol{v})$ with $\gamma(k, \boldsymbol{u}, \boldsymbol{v}) = \text{Cov}\{f(\boldsymbol{U}_0, \boldsymbol{u}), f(\boldsymbol{U}_k, \boldsymbol{v})\}$.

Additionally, provided the function φ is Lipschitz-continuous and provided $\int_{[0,1]^d} \prod_{\mu=1}^{d} r_\mu(u_\mu)^{4+2\nu} dC_{\theta_0}(\boldsymbol{u}) < \infty$, where $r_1, \ldots, r_d$ are defined in Condition V, then

$$\text{Var}\{\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})\} = \frac{\ell_n}{n}\Delta(\boldsymbol{u}, \boldsymbol{v}) + r_{n,2}(\boldsymbol{u}, \boldsymbol{v}),$$

where $\Delta(\boldsymbol{u}, \boldsymbol{v}) = \int_{-1}^{1} \varphi(x)^2 dx\{\sigma(\boldsymbol{u}, \boldsymbol{u})\sigma(\boldsymbol{v}, \boldsymbol{v}) + \sigma(\boldsymbol{u}, \boldsymbol{v})^2\}$ and where the remainder term satisfies $\sup_{\boldsymbol{u},\boldsymbol{v} \in [0,1]^d} |r_{n,2}(\boldsymbol{u}, \boldsymbol{v})| = o(\ell_n/n)$.

As a consequence of Lemma 3.1, the (pointwise) mean integrated squared error of $\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})$ can be written as

$$\text{MSE}\{\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})\} = \frac{\{\Gamma(\boldsymbol{u}, \boldsymbol{v})\}^2}{\ell_n^4} + \Delta(\boldsymbol{u}, \boldsymbol{v})\frac{\ell_n}{n} + r_{n,1}^2(\boldsymbol{u}, \boldsymbol{v}) + r_{n,2}(\boldsymbol{u}, \boldsymbol{v}).$$

Furthermore, the integrated mean squared error is given by

$$\text{IMSE}(\tilde{\sigma}_n) = \int_{[0,1]^{2d}} \text{MSE}\{\tilde{\sigma}_n(\boldsymbol{u}, \boldsymbol{v})\} d(\boldsymbol{u}, \boldsymbol{v}) = \frac{\bar{\Gamma}^2}{\ell_n^4} + \bar{\Delta}\frac{\ell_n}{n} + o(\ell_n^{-4}) + o(\ell_n/n),$$

where $\bar{\Gamma} = \int_{[0,1]^{2d}} \Gamma(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}(\boldsymbol{u}, \boldsymbol{v})$ and $\bar{\Delta} = \int_{[0,1]^{2d}} \Delta(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}(\boldsymbol{u}, \boldsymbol{v})$. Obviously, the function $\ell_n \mapsto \bar{\Gamma}/\ell_n^4 + \bar{\Delta}\ell_n/n$ is minimized for

$$\ell_{n,opt} = \left(\frac{4\bar{\Gamma}^2}{\bar{\Delta}}\right)^{1/5} n^{1/5},$$

which can be considered as a theoretically optimal choice for the bandwidth parameter ℓ_n.

In practice, the unknown quantities in $\ell_{n,opt}$ need to be estimated, namely $\sigma(\boldsymbol{u}, \boldsymbol{v}) = \sum_{k \in \mathbb{Z}} \gamma(k, \boldsymbol{u}, \boldsymbol{v})$ and $M(\boldsymbol{u}, \boldsymbol{v}) = \sum_{k \in \mathbb{Z}} k^2 \gamma(k, \boldsymbol{u}, \boldsymbol{v})$. For that purpose, we can closely follow Bücher and Kojadinovic (2013) again. Let $L \in \{1, \dots, n\}$ be the smallest number such that the marginal autocorrelations at lag L appear to be negligible, see Bücher and Kojadinovic (2013) and Politis and White (2004) for details. Let K denote the trapezoidal kernel, defined as $K(x) = [\{2(1 - |x|)\} \vee 0] \wedge 1$. Then, set

$$\hat{\sigma}_n(\boldsymbol{u}, \boldsymbol{v}) = \sum_{k=-L}^{L} K(k/L)\hat{\gamma}_n(k, \boldsymbol{u}, \boldsymbol{v}), \quad \hat{M}_n(\boldsymbol{u}, \boldsymbol{v}) = \sum_{k=-L}^{L} K(k/L)k^2\hat{\gamma}_n(k, \boldsymbol{u}, \boldsymbol{v}),$$

where, for $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$ and $k \in \{-L, \dots, L\}$,

$$\hat{\gamma}_n(k, \boldsymbol{u}, \boldsymbol{v}) = \begin{cases} n^{-1} \sum_{i=1}^{n-k} \hat{f}(\hat{\boldsymbol{U}}_i, \boldsymbol{u})\hat{f}(\hat{\boldsymbol{U}}_{i+k}, \boldsymbol{v}), & k \geq 0 \\ n^{-1} \sum_{i=1-k}^{n} \hat{f}(\hat{\boldsymbol{U}}_i, \boldsymbol{u})\hat{f}(\hat{\boldsymbol{U}}_{i+k}, \boldsymbol{v}), & k < 0 \end{cases},$$

and where

$$\hat{f}(\hat{\boldsymbol{U}}_i, \boldsymbol{u}) = \mathbb{1}(\hat{\boldsymbol{U}}_i \leq \boldsymbol{u}) - C_n(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_n^{[\mu]}(\boldsymbol{u})\{\mathbb{1}(\hat{U}_{i\mu} \leq u_\mu) - u_\mu\}$$

$$- \nabla C_{\theta_n}(\boldsymbol{u})\{J_{\theta_0}(\hat{\boldsymbol{U}}_i) + \hat{K}_{i,n,\theta_n}\}.$$

The plug-in principle finally yields an estimator $\hat{\ell}_{n,opt}$ for $\ell_{n,opt}$.

3.3 Score functions for common estimators

3.3.1 Inversion of Spearman's rho for alpha-mixing observations

One example of an estimator which fulfills the smoothness assumptions in Condition V in the bivariate case provides the moment estimator defined by $\theta_n = \rho^{-1}(\rho_n)$, where ρ_n is the sample version of Spearman's rho and where

$$\rho : \mathcal{P} \to [-1, 1], \quad \rho(\theta) = 12 \int_{[0,1]^2} C_\theta(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - 3.$$

If the function ρ is bijective and continuously differentiable, this leads to a consistent estimator of the parameter θ_0, with score function given by

$$J_\theta(\boldsymbol{u}) = \frac{1}{\rho'(\theta)}\{12u_1u_2 - 3 - \rho(\theta)\}.$$

To see the latter, note that, by the mean value theorem,

$$\sqrt{n}(\theta_n - \theta_0) = \sqrt{n}\{\rho^{-1}(\rho_n) - \rho^{-1}(\rho(\theta_0))\} = A_n + R_n,$$

where $A_n = \sqrt{n}\frac{1}{\rho'(\theta_0)}\{\rho_n - \rho(\theta_0)\}$ and $R_n = \sqrt{n}(\frac{1}{\rho'(\theta_n')} - \frac{1}{\rho'(\theta_0)})\{\rho_n - \rho(\theta_0)\}$ with θ_n' being an intermediate point between θ_0 and θ_n.

We have $R_n = o_P(1)$ by continuity of ρ', consistency of θ_n' and by the fact that $\sqrt{n}(\rho_n - \rho_0) = O_P(1)$ (which follows from the continuous mapping theorem and weak convergence of the empirical copula process for mixing rates $\alpha_{\boldsymbol{X}}(n) = O(n^{-a})$ for some $a > 1$, see Section 2.2).

Moreover, by the definition of ρ_n,

$$A_n = \frac{\sqrt{n}}{\rho'(\theta_0)}\left\{1 - \frac{6\sum_{i=1}^n(\hat{U}_{i1} - \hat{U}_{i2})^2}{n(n-1)^2} - \rho(\theta_0)\right\}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{1}{\rho'(\theta_0)}\left\{12\hat{U}_{i1}\hat{U}_{i2} - 3 - \rho(\theta_0)\right\}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n J_{\theta_0}(\hat{U}_i).$$

Since J_θ is bounded in $\boldsymbol{u} \in [0, 1]^2$, all assumptions in Condition V can be easily shown to hold for most of the most common copula families. For example, for elliptical copulas, we have $\rho(\theta) = \frac{6}{\pi}\arcsin(\frac{\theta}{2})$ whence the corresponding score function is given by

$$J_\theta(\boldsymbol{u}) = \frac{\pi}{3}\sqrt{1 - \frac{\theta^2}{4}}\left\{12u_1u_2 - 3 - \frac{6}{\pi}\arcsin(\theta/2)\right\}.$$

3.3.2 The pseudo-ML estimator in Markovian copula models

Another estimator satisfying the smoothness assumptions in Condition V is provided by the pseudo-ML estimator in Markovian copula models as described in Example 2.1 above. If M denotes the d-dimensional copula of $\boldsymbol{Y}_i$, then, by stationarity, M is also the copula of $\boldsymbol{Y}_{i-1}$ and for all $\boldsymbol{u} \in [0, 1]^d$

$$C(\boldsymbol{u}, \boldsymbol{1}) = M(\boldsymbol{u}) = C(\boldsymbol{1}, \boldsymbol{u}).$$

To test for the hypothesis $C \in \{C_\theta : \theta \in \mathcal{P}\}$, where C_θ has density c_θ and where m_θ denotes the density corresponding to the marginal copula

$M_\theta = C_\theta(\mathbf{1}, \cdot) = C_\theta(\cdot, \mathbf{1})$ for any $\theta \in \mathcal{P}$, one may use the maximum-pseudo-likelihood estimator for the parameter θ defined as

$$\hat{\theta}_n = \arg \max_{\theta \in \mathcal{P}} \sum_{i=2}^{n} \log \left\{ \frac{c_\theta(\hat{\boldsymbol{U}}_i, \hat{\boldsymbol{U}}_{i-1})}{m_\theta(\hat{\boldsymbol{U}}_{i-1})} \right\}.$$

From the results in Rémillard et al. (2012) it follows that the corresponding score function $J_\theta : [0,1]^{2d} \to \mathbb{R}^p$ is given by

$$J_\theta(\boldsymbol{u}) = \left\{ \mathbb{E}_{C_\theta} \left[\nabla^2 \frac{c_\theta(\boldsymbol{U}_1, \boldsymbol{U}_0)}{m_\theta(\boldsymbol{U}_0)} \right] \right\}^{-1} \nabla \left[\log \left\{ \frac{c_\theta(\boldsymbol{u})}{m_\theta(u_1, \dots, u_d)} \right\} \right].$$

It can be shown that the assumptions in Condition V hold for most of the commonly used copulas. For a discussion of similar assumptions in the case of a one-dimensional Markovian time series, we refer the reader to Chen and Fan (2006b).

4 Numerical results

In this section, we study the finite-sample performance of the goodness-of-fit test by means of Monte Carlo simulations. We restrict ourselves to the investigation of the one-dimensional Markovian copula model described in Example 2.1.

The experimental design is as follows: we consider four common bivariate copula models (the Gaussian, the t_4, the Clayton and the Gumbel–Hougaard copula), six different levels of dependence as measured through Kendall's rank correlation coefficient ($\tau = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$) and five different sample sizes ($n = 50, 100, 200, 400, 800$). Due to the high level of serial dependence introduced through a strongly dependent copula, analogue simulations for higher values of τ did not lead to satisfying results. For a visual exploration of the strength of the serial dependence, see Figure 1.

For the parametric estimator, we choose the pseudo-ML estimator from Chen and Fan (2006b) as described in Section 3.3.2. The multipliers are generated as described in Section 3.1 (with ξ_i standard normally distributed) and the bandwidth parameter is chosen according to the description in Section 3.2.

The corresponding results can be found in Tables 1 and 2 where we state the empirical rejection probabilities for a nominal level of 5% based on $N = 1,000$ simulation runs, each of which being based on $B = 200$ bootstrap replications. The main findings can be summarized as follows:
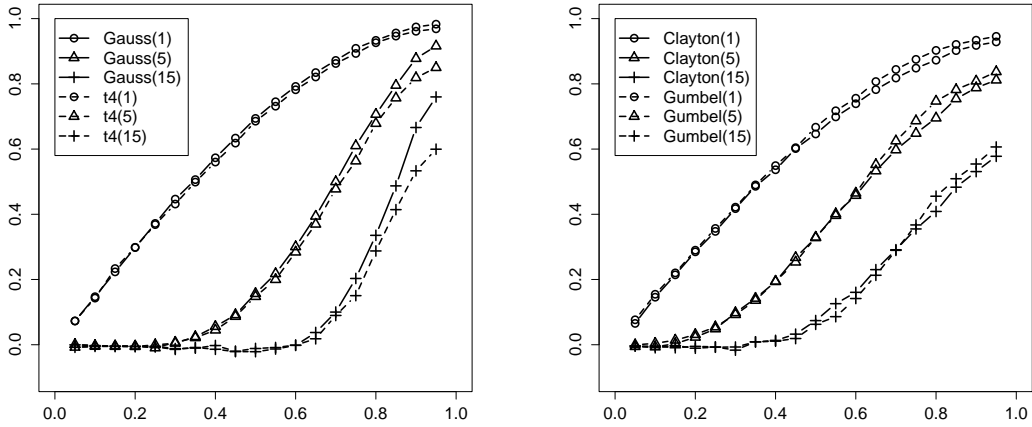
Figure 1: Estimated autocorrelations at lag $1\,(\circ)$, $5\,(\triangle)$ and $15\,(+)$ for the Markovian copula model based on the Gaussian copula (left panel, solid lines), the t_4-copula (left panel, dashed lines), the Clayton copula (right panel, solid lines) and the Gumbel copula (right panel, dashed lines).

- The approximation of the nominal level is accurate and gets better with increasing sample size. For the t_4 and the Gaussian copula, the test seems to be globally conservative, while it is slightly too liberal for the Clayton copula.

- We observe decent power properties for Gaussian or t_4 vs. Clayton or Gumbel and vice versa. Also, Gumbel and Clatyon can be well-distinguished. On the other hand, a distinction between the t_4 and the Gaussian copula seems to be more difficult. The latter behavior has also been observed by Kojadinovic et al. (2011) in the i.i.d. context.

- The power of the test is increasing in the strength of the dependence as measured through Kendall's tau. This can be explained by the fact the independence copula is included in all four models under consideration, whence the difference between the models tends to be very small for small values of τ.

5 Data application

In this section, we illustrate the use of our goodness-of-fit test by applying it to a time series of 267 weakly gasoline price markups in Windsor, Ontario, from August 20, 1989, to September 25, 1994. As described in Beare

| H_0 | n | True copula: Gaussian | | | | | |
		$\tau = .05$	$\tau = .1$	$\tau = .15$	$\tau = .2$	$\tau = .25$	$\tau = .3$
Gaussian	50	14.1	10.7	6.5	5.9	5.6	3.1
	100	6.8	5.2	3.2	3.1	2.4	1.3
	200	5.8	4.2	3.6	3.3	2.3	2.2
	400	4.5	4.6	4.1	3.8	3.2	2.4
	800	4.8	4.8	3.9	2.6	3.2	4.4
t_4	50	6.8	6.7	4.9	4.5	4.1	3.9
	100	3.0	2.9	2.9	4.2	1.7	1.6
	200	5.3	2.9	3.4	3.4	3.3	3.1
	400	5.8	8.7	8.3	9.1	9.4	8.5
	800	15.7	15.7	23.0	29.5	31.0	28.7
Clayton	50	24.2	28.3	26.9	33.2	34.3	38.5
	100	15.2	23.3	26.5	35.3	45.3	50.7
	200	14.3	25.9	39.3	56.2	66.9	79.1
	400	14.4	35.1	58.6	80.0	91.8	97.2
	800	24.0	56.0	84.3	95.9	99.7	100.0
Gumbel	50	4.8	6.3	7.6	5.9	5.4	6.8
	100	5.4	6.6	9.7	8.7	8.6	7.5
	200	8.1	12.3	14.6	16.9	17.9	21.8
	400	10.4	18.5	27.0	38.3	45.6	49.3
	800	16.5	34.4	54.0	71.5	82.5	87.2

| H_0 | n | True copula: t_4 | | | | | |
		$\tau = .05$	$\tau = .1$	$\tau = .15$	$\tau = .2$	$\tau = .25$	$\tau = .3$
Gaussian	50	20.6	15.6	14.6	12.3	10.5	7.3
	100	12.0	10.6	9.4	8.8	6.4	6.0
	200	12.0	11.5	12.0	8.4	7.5	7.9
	400	15.9	14.3	12.9	12.6	12.7	8.8
	800	32.4	28.4	29.0	27.2	23.3	23.9
t_4	50	8.5	5.6	5.7	3.9	2.8	2.9
	100	2.9	2.1	2.6	1.8	1.6	0.6
	200	2.6	2.4	2.0	1.7	1.3	1.0
	400	1.9	1.6	2.0	2.0	0.7	1.4
	800	1.4	2.2	1.1	1.3	1.3	1.9
Clayton	50	21.2	21.0	24.2	30.0	30.9	34.2
	100	14.9	18.7	23.7	25.3	32.7	40.9
	200	11.8	20.8	28.8	42.0	55.0	68.3
	400	11.4	27.4	45.3	64.6	81.6	89.1
	800	25.2	46.3	73.2	92.0	98.5	99.9
Gumbel	50	6.7	7.6	7.9	8.1	8.1	8.0
	100	7.2	7.2	6.5	8.2	9.0	9.4
	200	6.9	7.9	11.1	12.0	15.1	19.2
	400	10.3	14.3	17.4	23.9	28.2	34.8
	800	29.2	27.2	41.5	49.5	60.4	69.7

Table 1: Simulated rejection probabilities for the goodness-of-fit test in the Markovian copula model based on the Gaussian and the t_4-copula.

| H_0 | n | True copula: Clayton | | | | | |
		$\tau = .05$	$\tau = .1$	$\tau = .15$	$\tau = .2$	$\tau = .25$	$\tau = .3$
Gaussian	50	12.8	11.5	9.2	9.3	7.6	6.7
	100	6.5	5.9	7.6	9.1	10.4	12.6
	200	5.6	8.8	16.5	23.4	30.6	34.2
	400	6.8	16.0	31.4	50.7	68.1	72.8
	800	11.9	34.2	66.6	85.3	95.3	98.3
t_4	50	5.9	5.7	3.5	4.0	3.2	3.2
	100	2.2	4.4	4.1	4.3	3.6	3.4
	200	2.8	5.3	7.1	11.1	14.1	15.0
	400	6.6	14.2	21.7	30.0	38.1	41.0
	800	17.5	35.5	51.0	64.0	69.0	73.0
Clayton	50	16.6	17.4	12.6	13.0	14.2	13.8
	100	10.2	7.6	9.2	8.2	6.6	9.3
	200	8.3	5.0	5.3	6.6	7.3	9.3
	400	4.4	5.7	5.5	5.7	6.4	7.8
	800	3.9	5.0	5.9	6.5	5.9	8.1
Gumbel	50	7.9	10.2	14.6	17.6	19.9	17.5
	100	7.4	13.2	25.1	31.6	38.5	40.0
	200	13.7	30.4	44.3	64.2	79.3	83.9
	400	22.5	56.5	81.0	94.0	98.4	99.1
	800	39.1	84.3	99.2	100.0	100.0	100.0

| H_0 | n | True copula: Gumbel–Hougaard | | | | | |
		$\tau = .05$	$\tau = .1$	$\tau = .15$	$\tau = .2$	$\tau = .25$	$\tau = .3$
Gaussian	50	14.0	13.4	10.8	10.7	7.3	4.6
	100	8.3	7.4	6.9	6.1	4.9	3.3
	200	5.7	10.3	12.7	11.6	11.4	9.2
	400	7.5	13.7	21.1	19.3	23.3	23.6
	800	8.5	22.0	31.9	46.1	57.1	59.9
t_4	50	7.7	6.5	6.2	5.6	5.6	5.0
	100	4.5	3.9	3.8	4.9	4.3	3.7
	200	5.3	4.7	6.5	6.5	5.7	5.8
	400	6.6	10.2	12.4	13.4	12.6	13.0
	800	17.0	19.1	25.5	32.7	33.4	33.8
Clayton	50	27.7	30.0	34.7	36.0	40.9	35.5
	100	19.0	25.0	35.7	45.0	45.9	52.7
	200	22.0	38.4	54.9	65.2	79.9	84.7
	400	24.9	55.8	77.1	90.9	96.1	97.9
	800	38.7	81.7	96.7	99.4	100.0	100.0
Gumbel	50	6.2	5.1	8.3	6.1	5.7	4.2
	100	5.5	5.6	5.2	4.6	3.6	4.8
	200	5.4	6.2	5.8	5.5	4.5	4.6
	400	4.9	4.8	6.2	5.5	5.0	3.8
	800	6.6	6.2	6.9	5.5	5.4	5.7

Table 2: Simulated rejection probabilities for the goodness-of-fit test in the Markovian copula model based on the Clayton and the Gumbel–Hougaard copula.

and Seo (2013), these markups were obtained by dividing the average retail price across several gasoline stations in Windsor by the wholesale price of unbranded gasoline at the terminal in Toronto, Ontario. A plot of the time series can be found in Figure 2 below. The data set was investigated by several authors, see, e.g., Eckert (2002) and McCausland (2007). Recently, Beare and Seo (2013) used Markovian copula models to investigate time reversibility of the time series of price markups. Economic and financial time series frequently exhibit time irreversibility and these authors found indeed overwhelming evidence for time irreversibility of the markup data.

Loosely speaking, time reversibility means that the probabilistic behavior of a time series $(X_i)_{i \in \mathbb{Z}}$ does not alter by reversing the direction of time. Formally, it is defined by the requirement that

$$(X_{i_1}, \ldots, X_{i_n}) \overset{d}{=} (X_{i_n}, \ldots, X_{i_1}) \text{ for all } i_1 < \cdots < i_n.$$

For a Markovian copula model, it is equivalent to exchangeability of the copula of (X_{i-1}, X_i), i.e., to the fact that $C(u,v) = C(v,u)$ for all $u,v \in [0,1]$, see, e.g., Beare and Seo (2013). Therefore, the results on the markup data in the last-named paper suggest that our goodness-of-fit tests should reject any exchangeable copula model for C.

We run our procedure based on $B = 1,000$ bootstrap replications for the four copulas considered in the simulation study in Section 4. The corresponding p-values are $p = 0.002$ for the Gaussian copula, $p = 0.471$ for the t_4-copula, $p = 0.016$ for the Clayton copula, and $p = 0.011$ for the Gumbel–Hougaard copula. Hence, there is indeed rather strong evidence against three out of four of these models

Non-exchangeable copula models can for instance be derived by Khoudraji's device, see Khoudraji (1995). Based on two arbitrary copulas C_1 and C_2, one can construct a new copula through

$$C(u,v) = C_1(u^\alpha, v^\beta) C_2(u^{1-\alpha}, v^{1-\beta}), \quad (u,v) \in [0,1]^2,$$

where $\alpha, \beta \in [0,1]$, see also Liebscher (2008). We focus on the special case where C_1 is either from the Gumbel–Hougaard or the Clayton family and where C_2 is equal to the independence copula (see also Beare and Seo, 2013, for the Gumbel–Hougaard choice). For both of the resulting 3-parameter copula families, we did not find any evidence against H_0, with p-values of $p = 0.561$ and $p = 0.941$ for the Gumbel and Clayton extensions, respectively. For the extension of the Clayton copula, we obtained parameter estimates $\alpha = 0.868, \beta = 0.726$ and $\theta = 7.362$ (the parameter of the Clayton copula), with corresponding 95%-multiplier bootstrap confidence bands $(0.77, 0.97)$, $(0.63, 0.83)$ and $(3.49, 11.25)$, respectively.
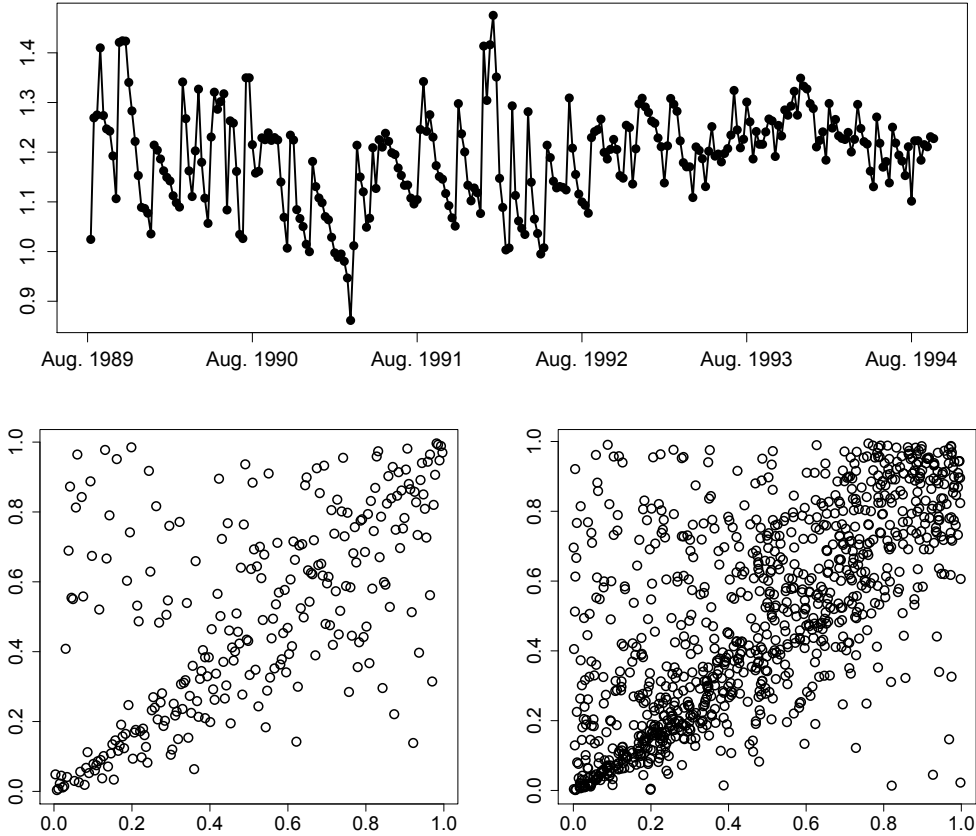
Figure 2: Upper panel: the time series of markups. Lower left panel: pseudo observation $(\hat{U}_{i-1}, \hat{U}_i)$ for the markups. Lower right panel: a simulated sample of size $1,000$ from the fitted non-exchangeable extension of the Clayton copula with parameters $\alpha = 0.868, \beta = 0.726$ and $\theta = 7.362$.

In Figure 2, we show a plot of the pseudo-observations based upon the observations (X_{i-1}, X_i), along with a simulated sample from the extension of the Clayton copula based on the fitted parameters. Indeed, we observe a very good approximation of the data by the model. Overall, our results are consistent with the findings in Beare and Seo (2013).

A Technical assumptions

In this section, we will introduce the technical conditions needed for the proof of Theorem 2.2. Moreover, we will briefly comment on each of the conditions.

First of all, weak convergence of the empirical copula process in Section 2.2 requires a smoothness condition on C, see Segers (2012).

Condition I. For $\mu = 1, \ldots, d$, the partial derivative $C^{[\mu]}(\boldsymbol{u}) = \frac{\partial}{\partial u_\mu} C(\boldsymbol{u})$ exists and is continuous for any $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0, 1]^d$ with $u_\mu \in (0, 1)$.

For completeness, we define $C^{[\mu]}(\boldsymbol{u})$ as zero for any $\boldsymbol{u}$ such that $u_\mu \in \{0, 1\}$, see also Segers (2012); Bücher and Volgushev (2013).

Throughout the proofs, we need to be able to easily go back and forth between normalized ranks and empirical quantile functions. For that purpose, ties must not occur. In the case of serial independence, it is sufficient to assume that the marginal distributions are continuous. In the case of serial dependence, continuity of the marginal distributions is not sufficient to guarantee the absence of ties which brings us to the following condition, see also Bücher and Kojadinovic (2013).

Condition II. For any $\mu \in \{1, \ldots, d\}$, there are no ties in the marginal time series $X_{1\mu}, \ldots, X_{n\mu}$ with probability one.

Since we allow for alpha-mixing observations, the multipliers $Z_{i,n}^{(b)}$ have to be chosen serially dependent in an appropriate way. We follow Bücher and Kojadinovic (2013), relying on previous results from Bühlmann (1993); Bücher and Ruppert (2013).

Definition A.1. A sequence $(Z_{i,n})_{i \in \mathbb{Z}}$ is called a *dependent multiplier sequence* if the following three conditions are met:

(M1) $(Z_{i,n})_{i \in \mathbb{Z}}$ is independent of the process $(\boldsymbol{X}_i)_{i \in \mathbb{N}}$.

(M2) The sequence $(Z_{i,n})_{i \in \mathbb{Z}}$ in ℓ_n-dependent, i.e., $Z_{i,n}^{(b)}$ is independent of $Z_{i+h,n}^{(b)}$ for any $h > \ell_n$ and any $i \in \mathbb{Z}$, where $\ell_n \in \mathbb{N}$ tends to infinity and $\ell_n = O(n^{1/2-\kappa})$ for some $\kappa \in (0, 1/2)$, for $n \to \infty$.

(M3) The sequence $(Z_{i,n})$ is strictly stationary with $\mathbb{E}[Z_{i,n} Z_{i+h,n}] = \varphi(h/\ell_n)$, where φ is a symmetric function which is strictly positive on $(-1, 1)$, zero on $\mathbb{R} \setminus (-1, 1)$ and continuous in 0. Furthermore, $\sup_{n \in \mathbb{N}} \mathbb{E}[|Z_{i,n}|^p] < \infty$ for all $p \geq 1$.

For the proof of (2.5), Bücher and Kojadinovic (2013) imposed the following condition on the mixing rate of $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$, where $\boldsymbol{U}_i = (\boldsymbol{U}_{i1}, \ldots, \boldsymbol{U}_{id})'$ with $U_{i\mu} = F_\mu(X_{i\mu})$ for $\mu \in \{1, \ldots, d\}$.

Condition III. $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ is a stationary, strongly mixing process satisfying $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 3 + 3d/2$.

The next condition, ensuring uniform continuity of the partial derivative of the copula with respect to θ, is needed for the proof of (2.2).

Condition IV. As $\varepsilon \downarrow 0$,

$$\sup_{\|\theta' - \theta\| < \varepsilon} \sup_{\boldsymbol{u} \in [0,1]^d} \|\nabla C_{\theta'}(\boldsymbol{u}) - \nabla C_{\theta}(\boldsymbol{u})\| \to 0.$$

Regarding the parametric estimator θ_n, we opt for general conditions that include the pseudo-maximum likelihood estimator. Following Tsukahara (2005), we begin by defining q-functions and u-shaped functions.

Definition A.2. *(q-functions)* Let $\mathcal{Q}$ be the class of all continuous functions $q : [0, 1] \to [0, 1]$ which are positive on $(0, 1)$, symmetric around $1/2$, increasing on $(0, 1/2]$ and for which $\int_{[0,1]} q(u)^{-2} \mathrm{d}u < \infty$. The functions in $\mathcal{Q}$ are called q-functions .

For instance, for any $0 < \delta < 1/2$, the function $q(u) = \{u(1-u)\}^{1/2-\delta}$. is an element of $\mathcal{Q}$.

Definition A.3. *(u-shaped functions)* A function $r : (0, 1) \to (0, \infty)$, which is symmetric around $1/2$, is called a u-shaped function if it is decreasing on $(0, 1/2]$.

A typical example of a u-shaped function is $r(u) = \{u(1-u)\}^{-\tau}$ for $\tau \geq 0$.

Condition V. The estimator θ_n allows for the decomposition

$$\Theta_n := \sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_{\theta_0}(\hat{\boldsymbol{U}}_i) + o_p(1),$$

where, for $\theta \in \mathcal{P}$, $J_\theta : [0, 1]^d \to \mathbb{R}^p$ denotes a score-function with the following properties:

(i) J_θ is standardized, i.e.,

$$\int_{[0,1]^d} J_\theta(\boldsymbol{u}) \mathrm{d}C_\theta(\boldsymbol{u}) = 0.$$

23

(ii) For any $\theta \in \mathcal{P}$ and for any $\mu = 1, \ldots, d$ the derivatives $J_\theta^{[\mu]} = \frac{\partial}{\partial u_\mu} J_\theta \in$
$\mathbb{R}^p$, $\nabla J_\theta \in \mathbb{R}^{p \times p}$ and $\nabla(J_\theta^{[\mu]}) \in \mathbb{R}^{p \times p}$ exist, are continuous on $(0,1)^d$ and
there is a constant k_θ such that for all $\theta' \in \mathcal{P}$ with $\|\theta - \theta'\| \leq k_\theta$

$$\|J_{\theta'}(\boldsymbol{u})\| \leq \prod_{\mu=1}^d r_\mu(u_\mu) \quad \text{and} \quad \|\nabla J_{\theta'}(\boldsymbol{u})\|_{op} \leq \prod_{\mu=1}^d r_\mu(u_\mu),$$

where $r_1, \ldots, r_d$ are u-shaped functions with

$$\int_{[0,1]^d} \prod_{\mu=1}^d r_\mu(u_\mu)^{2+\nu} \mathrm{d}C_{\theta_0}(\boldsymbol{u}) < \infty, \tag{A.1}$$

$$\int_{[0,1]^d} \prod_{\mu=1}^d \sup_{\frac{|v-u_\mu|}{q_\mu(u_\mu)} \leq \delta} r_\mu(v)^2 \mathrm{d}C_{\theta_0}(\boldsymbol{u}) < \infty \tag{A.2}$$

for some ν and $\delta > 0$. Additionally, for any $\mu = 1, \ldots, d$, there exists a
u-shaped function $\tilde{r}_\mu$ with

$$\|J_{\theta_0}^{[\mu]}(\boldsymbol{u})\| \leq \tilde{r}_\mu(u_\mu) \prod_{\mu' \neq \mu} r_{\mu'}(u_{\mu'}) \quad \text{and} \quad \|\nabla(J_{\theta'}^{[\mu]})(\boldsymbol{u})\|_{op} \leq \tilde{r}_\mu(u_\mu) \prod_{\mu' \neq \mu} r_{\mu'}(u_{\mu'})$$

such that for some q-function q_μ with $\int_0^1 q_\mu(u)^{-(2+\nu)} \mathrm{d}u < \infty$

$$\int_{[0,1]^d} q_\mu(u_\mu) \sup_{\frac{|v-u_\mu|}{q_\mu(u_\mu)} \leq \delta} r_\mu(v) \prod_{\mu' \neq \mu} \sup_{\frac{|v-u_{\mu'}|}{q_{\mu'}(u_{\mu'})} \leq \delta} r_{\mu'}(v) \mathrm{d}C_{\theta_0}(\boldsymbol{u}) < \infty. \tag{A.3}$$

A discussion of estimators satisfying these assumptions is given in Section 3.3. In general, the components of the score function J_θ are not bounded. Therefore, in order to derive the asymptotic behavior of Θ_n by a suitable application of the central limit theorem for alpha-mixing sequences, we need a stronger condition on the mixing coefficients of $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ which depends on the moments of the score function.

Condition VI. $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ is a stationary, strongly mixing process satisfying $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > \frac{2(2+\nu)(1+\nu)}{\nu^2} = 2 + \frac{6}{\nu} + \frac{4}{\nu^2}$, where ν is defined in Condition V.

B Proofs of the main results

Let us first recall and introduce some notation. For $\boldsymbol{u} \in [0,1]^d$, let

$$\tilde{\mathbb{B}}_n(\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\}, \tag{B.1}$$

$$\tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)}\{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\}, \tag{B.2}$$

and, recalling the definition of K_{i,θ_0} in Theorem 2.2, set

$$\tilde{\Theta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}, \qquad \tilde{\Theta}_n^{(b)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)}\{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}. \tag{B.3}$$

Proof of Theorem 2.2. As the other two items can be shown by similar, but simpler proofs, we only prove (iii). Based on the central result in Lemma C.6, we begin by showing that

$$\mathbb{H}_n = \left(\mathbb{C}_n, \Theta_n, \mathbb{C}_n^{(1)}, \Theta_n^{(1)}, \ldots, \mathbb{C}_n^{(B)}, \Theta_n^{(B)}\right)$$
$$\rightsquigarrow \mathbb{H}_C = \left(\mathbb{C}_C, \Theta, \mathbb{C}_C^{(1)}, \Theta^{(1)}, \ldots, \mathbb{C}_C^{(B)}, \Theta^{(B)}\right) \tag{B.4}$$

in $\left\{\ell^\infty([0,1]^d) \times \mathbb{R}^p\right\}^{B+1}$. For the parametric components of $\mathbb{H}_n$ in (B.4), note that $\Theta_n = \tilde{\Theta}_n + o_p(1)$ for $n \to \infty$ as a consequence of Lemma C.1 and C.2. Similarly, $\Theta_n^{(b)} = \tilde{\Theta}_n^{(b)} + o_p(1)$ for any $b = 1, \ldots, B$ from Lemma C.3, C.4 and C.5.

For the nonparametric components of $\mathbb{H}_n$, we have

$$\sup_{\boldsymbol{u} \in [0,1]^d} \left|\mathbb{C}_n(\boldsymbol{u}) - \tilde{\mathbb{B}}_n(\boldsymbol{u}) + \sum_{\mu=1}^d C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n(\boldsymbol{u}^\mu)\right| = o_p(1)$$

and

$$\sup_{\boldsymbol{u} \in [0,1]^d} \left|\mathbb{C}_n^{(b)}(\boldsymbol{u}) - \tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}) + \sum_{\mu=1}^d C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{u}^\mu)\right| = o_p(1),$$

by simple adaptations of the proof of Proposition 3.1 in Segers (2012) and the proof of Proposition 4.2 in Bücher and Kojadinovic (2013), respectively.

As a consequence,

$$
\mathbb{H}_n(\boldsymbol{u}) = \Big(\tilde{\mathbb{B}}_n(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n(\boldsymbol{u}^\mu), \tilde{\Theta}_n,
$$

$$
\tilde{\mathbb{B}}_n^{(1)}(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n^{(1)}(\boldsymbol{u}^\mu), \tilde{\Theta}_n^{(1)}, \ldots,
$$

$$
\tilde{\mathbb{B}}_n^{(B)}(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\tilde{\mathbb{B}}_n^{(B)}(\boldsymbol{u}^\mu), \tilde{\Theta}_n^{(B)} \Big) + o_P(1),
$$

where the remainder is uniform in $\boldsymbol{u} \in [0,1]^d$. An application of the continuous mapping Theorem and Lemma C.6 proves (B.4).

It remains to prove assertion (2.11) in Theorem 2.2. Applying Lemma C.7 and the continuous mapping Theorem to (B.4) yields that the process

$$
\Big(\sqrt{n}(C_n - C_{\theta_n}), \mathbb{B}_n^{(1)}(\boldsymbol{u}) - \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\mathbb{B}_n^{(1)}(\boldsymbol{u}^\mu) - \nabla C_{\theta_0}(\boldsymbol{u})\Theta_n^{(1)}, \ldots
$$

$$
\ldots, \mathbb{B}_n^{(B)}(\boldsymbol{u}) \sum_{\mu=1}^{d} C_{\theta_0}^{[\mu]}(\boldsymbol{u})\mathbb{B}_n^{(B)}(\boldsymbol{u}^\mu) - \nabla C_{\theta_0}(\boldsymbol{u})\Theta_n^{(B)} \Big)
$$

converges in $\{\ell^\infty([0,1]^d)\}^{B+1}$ to the limiting process in (2.11). Hence, it remains to be shown that replacing $C_{\theta_0}^{[\mu]}$ and ∇C_{θ_0} by $C_n^{[\mu]}$ and ∇C_{θ_n}, respectively, in the last display does not change the limiting distribution. To see this, fix μ and bound

$$
\sup_{\boldsymbol{u}\in[0,1]^d} \Big| C_{\theta_0}^{[\mu]}(\boldsymbol{u})\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu) + \nabla C_{\theta_0}(\boldsymbol{u})\Theta_n^{(b)} - C_n^{[\mu]}(\boldsymbol{u})\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu) - \nabla C_{\theta_n}(\boldsymbol{u})\Theta_n^{(b)} \Big|
$$

$$
\leq \max\Big\{ \sup_{u_\mu \in [\delta,1-\delta]} |\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu)||C_n^{[\mu]}(\boldsymbol{u}) - C_{\theta_0}^{[\mu]}(\boldsymbol{u})|,
$$

$$
\sup_{u_\mu \notin [\delta,1-\delta]} |\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu)||C_n^{[\mu]}(\boldsymbol{u}) - C_{\theta_0}^{[\mu]}(\boldsymbol{u})| \Big\}
$$

$$
+ \sup_{\boldsymbol{u}\in[0,1]^d} |\nabla C_{\theta_n}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})||\Theta_n^{(b)}|.
$$

Due to Condition III and consistency of θ_n as a consequence of Lemma C.1, C.2 and C.6, ∇C_{θ_n} provides an uniformly consistent estimator of ∇C_{θ_0}. Therefore, the last summand on the right-hand side of the last display converges to zero in probability as $n \to \infty$. The first argument in the maximum

in that display converges to zero in probability since the estimator $C_n^{[\mu]}$ is uniformly consistent on the set $\{\boldsymbol{u} \in [0,1]^d : u_\mu \in [\delta, 1-\delta]\}$ for any $0 < \delta < 1/2$ and since $\mathbb{B}_n^{(b)} = O_P(1)$. For the estimation of the second argument in the maximum note that, since $\|C_n^{[\mu]}\|_\infty \leq 2$ and $\|C_\theta^{[\mu]}\|_\infty \leq 1$, we have

$$\mathbb{P}\Big(\sup_{u_\mu \notin [\delta, 1-\delta]} |\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu)||C_n^{[\mu]}(\boldsymbol{u}) - C_{\theta_0}^{[\mu]}(\boldsymbol{u})| > \eta \Big) \leq \mathbb{P}\Big(\sup_{u_\mu \notin [\delta, 1-\delta]} |\mathbb{B}_n^{(b)}(\boldsymbol{u}^\mu)| > \eta/3 \Big)$$

which can be made arbitrarily small by decreasing δ. This proves the Theorem. $\qquad\square$

C Auxiliary results

Let us first introduce some additional notation: for $\boldsymbol{u} \in [0,1]^d$, let

$$G_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}), \qquad G_n^\star(\boldsymbol{u}) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u})$$

and $G_{n\mu}(u_\mu) = G_n(1,\ldots,1,u_\mu,1,\ldots,1)$ with u_μ at the μth coordinate, for $\mu = 1,\ldots,d$, and similar for $G_{n\mu}^\star$. By Theorem 2.2 in Shao and Yu (1996) and a simple adaptation of Lemma 2.3.1 in Ruymgaart (1973) to the alpha-mixing case we can find, for any $\gamma > 0$, some constant $D > 0$ such that $\mathbb{P}(\Omega_{n,D}) \geq 1 - \gamma$ for any $n \in \mathbb{N}$, where

$$\Omega_{n,D} = \left\{ \omega : \sup_{u \in [U_{1:n,\mu}, U_{n:n,\mu}]} \left| \frac{\sqrt{n}\{G_{n\mu}^\star(u) - u\}}{q_\mu(u)} \right| \vee \sup_{u \in (0,1)} \left| \frac{\sqrt{n}\{G_{n\mu}(u) - u\}}{q_\mu(u)} \right| \leq D \right.$$
$$\left. \text{for all } \mu = 1,\ldots,d \right\} \quad \text{(C.1)}$$

and where $U_{i:n,\mu}$ denotes the ith order statistic of $U_{1,\mu},\ldots,U_{n,\mu}$. Note that, on the event $\Omega_{n,D}$, we have $\max_{i=1}^n |\hat{U}_{i,\mu} - U_{i,\mu}|/q_\mu(U_{i,\mu}) \leq D/\sqrt{n}$ for all $\mu = 1,\ldots,d$.

Lemma C.1. *Suppose that $(\boldsymbol{U}_i)_{i\in\mathbb{Z}}$ satisfies $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 1 + \sqrt{2}$ (which is a consequence of Condition III). Moreover, suppose that Condition V holds. Then*

$$L_{n1} = \Theta_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,n,\theta_0}\} = o_P(1),$$

27

where, for $\theta \in \mathcal{P}$ and $i \in \{1, \dots, n\}$,

$$K_{i,n,\theta} = \frac{1}{n} \sum_{j=1}^{n} \sum_{\mu=1}^{d} J_\theta^{[\mu]}(\boldsymbol{U}_j) \times \{\mathbb{1}(U_{i,\mu} \leq U_{j,\mu}) - U_{j,\mu}\}.$$

Proof. Recall the definition of $\Omega_{n,D}$ in (C.1). Since we can make $\mathbb{P}(\Omega \setminus \Omega_{n,D})$ arbitrary small by increasing D, it is sufficient to show that $L_{n1}\mathbb{1}_{\Omega_{n,D}} = o_P(1)$, for any $D > 0$. For the ease of notation, we will suppress this indicator in the following and we will assume without loss of generality, that the inequality in (C.1) holds for any $\omega \in \Omega$.

Now, the mean value theorem allows to write

$$J_{\theta_0}(\hat{\boldsymbol{U}}_i) = J_{\theta_0}(\boldsymbol{U}_i) + \sum_{\mu=1}^{d} J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)(\hat{U}_{i,\mu} - U_{i,\mu}) + R_{ni},$$

where the remainder term is given by

$$R_{ni} = \sum_{\mu=1}^{d} \{J_{\theta_0}^{[\mu]}(\bar{\boldsymbol{U}}_{i,n}) - J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)\}(\hat{U}_{i,\mu} - U_{i,\mu})$$

and where $\bar{\boldsymbol{U}}_{i,n}$ denotes an intermediate point lying between $\boldsymbol{U}_i$ and $\hat{\boldsymbol{U}}_i$. Let us first show that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{ni} = o_P(1)$. We have

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{ni} \right\| \leq D \sum_{\mu=1}^{d} \left\{ \frac{1}{n} \sum_{i=1}^{n} S_{ni\mu} \right\}$$

where $S_{ni\mu} = \|J_{\theta_0}^{[\mu]}(\bar{\boldsymbol{U}}_{i,n}) - J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)\| q_\mu(U_{i,\mu})$. It remains to be shown that each of the d sums inside the curly brackets converges to 0 in probability. For $\eta > 0$, set $M_\eta = [\eta, 1 - \eta]^d$, and write

$$\frac{1}{n} \sum_{i=1}^{n} S_{ni\mu} = \frac{1}{n} \sum_{i=1}^{n} S_{ni\mu} \mathbb{1}(\boldsymbol{U}_i \in M_\eta) + \frac{1}{n} \sum_{i=1}^{n} S_{ni\mu} \mathbb{1}(\boldsymbol{U}_i \in M_\eta^C) \qquad \text{(C.2)}$$

Let us first consider the second sum on the right. By Condition V, for sufficiently large n such that $D/\sqrt{n} < \delta$ with δ as in the condition, we can bound its expected value by

$$2 \int_{M_\eta^C} q_\mu(u_\mu) \sup_{\frac{|v-u_\mu|}{q_\mu(u_\mu)} \leq \delta} \tilde{r}_\mu(v) \prod_{\mu' \neq \mu} \sup_{\frac{|v-u_{\mu'}|}{q_{\mu'}(u_{\mu'})} \leq \delta} r_{\mu'}(v) \, dC_{\theta_0}(\boldsymbol{u}) < \infty.$$

28

The latter can be made arbitrary small by decreasing η, whence it remains to consider the first sum on the right-hand side of (C.2) for fixed $\eta > 0$. Obviously, the sum can be bounded by

$$\max\{\|J_{\theta_0}^{[\mu]}(\bar{\boldsymbol{U}}_{i,n}) - J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)\| : i = 1, \ldots, n, \boldsymbol{U}_i \in M_\eta\}$$

This expression converges to 0 by uniform continuity of $J_{\theta_0}^{[\mu]}$ on $M_{\eta/2}$ and by the fact that $\max_{i=1}^n \|\bar{\boldsymbol{U}}_{i,n} - \boldsymbol{U}_i\| \leq D/\sqrt{n}$ as a consequence of assuming $\Omega = \Omega_{n,D}$. To conclude, $\|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ni}\| = o_P(1)$.

To finalize the proof of Lemma C.1, it remains to be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{\mu=1}^d J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)(\hat{U}_{i,\mu} - U_{i,\mu}) - \frac{1}{\sqrt{n}}\sum_{i=1}^n K_{i,n,\theta_0}$$

converges to 0 in probability. By exchanging the indices in the double sum resulting from the sum over the K_{i,n,θ_0}, the last display can be written as

$$\frac{1}{n}\sum_{i=1}^n \sum_{\mu=1}^d J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i) \times \sqrt{n}[\{G_{n\mu}^\star(U_{i,\mu}) - U_{i,\mu}\} - \{G_{n\mu}(U_{i,\mu}) - U_{i,\mu}\}].$$

We can bound the norm of the last display by $\frac{2D}{n}\sum_{i=1}^n \sum_{\mu=1}^d J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)q(U_{i,\mu})$, which is integrable by Condition V. Then, since $\sup_{u \in [0,1]} |\sqrt{n}\{G_{n\mu}^\star(u) - G_{n\mu}(u)\}| \leq n^{-1/2}$, the assertion follows from dominated convergence. $\square$

Lemma C.2. *Suppose that $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ satisfies $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 1 + \sqrt{2}$ (which is a consequence of Condition III). Moreover, suppose that Condition V holds. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (K_{i,n,\theta_0} - K_{i,\theta_0}) \xrightarrow{P} 0,$$

where K_{i,n,θ_0} and K_{i,θ_0} are defined in Lemma C.1 and Theorem 2.2, respectively.

Proof. As in the proof of Lemma C.1, we may assume without loss of generality that $\Omega = \Omega_{n,D}$ for some $D > 0$. Note that we can write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (K_{i,n,\theta_0} - K_{i,\theta_0}) = \sum_{\mu=1}^d I_{n,\mu}([0,1]^d),$$

where, for some Borel-measurable subset $A \subset [0,1]^d$,

$$I_{n,\mu}(A) = \int_A J_{\theta_0}^{[\mu]}(\boldsymbol{u}) \sqrt{n} \{G_{n\mu}(u_\mu) - u_\mu\} \, \mathrm{d}\{G_n(\boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\}.$$

It suffices to show that $I_{n,\mu}([0,1]^d) = o_P(1)$ for fixed $\mu \in \{1,\dots,d\}$. For $\eta \in (0,1/2)$, set $M_\eta = [\eta, 1-\eta]^d$, decompose $I_{n,\mu}([0,1]^d) = I_{n,\mu}(M_\eta) + I_{n,\mu}(M_\eta^C)$ and treat both summands separately.

For $I_{n,\mu}(M_\eta^C)$, we proceed similar as in the proof of Lemma C.1: since we can work on the event $\Omega_{n,D}$, we can bound $\|I_{n,\mu}(M_\eta^C)\|$ by

$$\frac{D}{n} \sum_{i=1}^n \|J_{\theta_0}^{[\mu]}(\boldsymbol{U}_i)\| q_\mu(U_{i,\mu}) \mathbb{1}(\boldsymbol{U}_i \in M_\eta^C) + D \int_{M_\eta^C} \|J_{\theta_0}^{[\mu]}(\boldsymbol{u})\| q_\mu(u_\mu) \, \mathrm{d}C_{\theta_0}(\boldsymbol{u}),$$

and the expectation of this expression can be made arbitrary small by decreasing η.

Thus, it remains to be shown that $I_{n,\mu}(M_\eta) = o_P(1)$ for each fixed $\eta > 0$. This is a consequence of Lemma C.8, the continuous mapping theorem, weak convergence of $\boldsymbol{u} \mapsto J_{\theta_0}^{[\mu]}(\boldsymbol{u}) \sqrt{n} \{G_{n\mu}(u_\mu) - u_\mu\}$ as an element of $\ell^\infty(M_\eta)$ and the fact that $\sup_{\boldsymbol{u} \in [0,1]^d} |G_n(\boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})| = o_P(1)$ as a consequence of, for instance, Theorem 7.3 in Rio (2000). $\qquad\square$

Lemma C.3. *Suppose that $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ satisfies $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 6$ (which is a consequence of Condition III). Moreover, suppose that Condition V holds. Then, for any $b = 1, \dots, B$,*

$$L_{n1}^{(b)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \left[\{J_{\theta_n}(\hat{\boldsymbol{U}}_i) + \hat{K}_{i,n,\theta_n}\} - \{J_{\theta_0}(\hat{\boldsymbol{U}}_i) + \hat{K}_{i,n,\theta_0}\} \right] = o_P(1),$$

where $\hat{K}_{i,n,\theta}$ is defined in (2.7).

Proof. As in the proof of Lemma C.1, we may assume without loss of generality that $\Omega = \Omega_{n,D}$ for some $D > 0$. By definition, we have

$$\hat{K}_{i,n,\theta_n} - \hat{K}_{i,n,\theta_0} = \sum_{\mu=1}^d \frac{1}{n} \sum_{j=1}^n \{J_{\theta_n}^{[\mu]}(\hat{\boldsymbol{U}}_j) - J_{\theta_0}^{[\mu]}(\hat{\boldsymbol{U}}_j)\} \times \{\mathbb{1}(\hat{U}_{i,\mu} \le \hat{U}_{j,\mu}) - \hat{U}_{j,\mu}\}.$$

Therefore, by the mean value theorem, there exist some $\theta_{i,n}'$ and $\theta_{j,n}^\mu$, $i,j = 1,\dots,n$, $\mu = 1,\dots,d$, all lying between θ_n and θ_0, such that

$$J_{\theta_n}(\hat{\boldsymbol{U}}_i) - J_{\theta_0}(\hat{\boldsymbol{U}}_i) = \nabla J_{\theta_{i,n}'}(\hat{\boldsymbol{U}}_i)(\theta_n - \theta)$$

$$\hat{K}_{i,n,\theta_n} - \hat{K}_{i,n,\theta_0} = \sum_{\mu=1}^d \frac{1}{n} \sum_{j=1}^n \nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)(\theta_n - \theta_0) \times \{\mathbb{1}(U_{i,\mu} \le U_{j,\mu}) - \hat{U}_{j,\mu}\},$$

30

where we used the fact that $\mathbb{1}(U_{i,\mu} \leq U_{j,\mu}) = \mathbb{1}(\hat{U}_{i,\mu} \leq \hat{U}_{j,\mu})$. This allows to write $L_{n1}^{(b)} = A_n + \sum_{\mu=1}^{d} B_{n,\mu}$, where

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)(\theta_n - \theta_0),$$

$$B_{n,\mu} = \frac{1}{n} \sum_{j=1}^{n} \nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)(\theta_n - \theta_0) \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \{\mathbb{1}(U_{i,\mu} \leq U_{j,\mu}) - \hat{U}_{j,\mu}\}.$$

Let us first treat A_n, which we decompose as $A_n = A_{n,1} \times \Theta_n + A_{n,2}$, where

$$A_{n,1} = \frac{1}{n} \sum_{i=1}^{n} Z_{i,n}^{(b)} \nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i) \mathbb{1}\{\|\nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)\|_{op} \leq \prod_{\mu=1}^{d} r_\mu(\hat{U}_{i,\mu})\}$$

$$A_{n,2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)(\theta_n - \theta_0) \mathbb{1}\{\|\nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)\|_{op} > \prod_{\mu=1}^{d} r_\mu(\hat{U}_{i,\mu})\},$$

with $r_1, \ldots, r_d$ being defined in Condition V. We begin with the estimation of the first term in this decomposition. Let $A_{n,1}^{(s,s')}$ denote the (s, s')-entry of the $p \times p$-matrix $A_{n,1}$. Obviously, its mean is equal to 0 and its variance can be estimated by

$$\mathbb{E}\big[\{A_{n,1}^{(s,s')}\}^2\big]$$
$$\leq \frac{1}{n} \sum_{i,j=1}^{n} |\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]| \Big\{\mathbb{E}\big[\prod_{\mu=1}^{d} r_\mu(\hat{U}_{i,\mu})^2\big]\Big\}^{1/2} \Big\{\mathbb{E}\big[\prod_{\mu=1}^{d} r_\mu(\hat{U}_{j,\mu})^2\big]\Big\}^{1/2}$$
$$\leq \frac{1}{n^2} \sum_{i,j=1}^{n} |\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]| \int_{[0,1]^d} \prod_{\mu=1}^{d} \sup_{\frac{|v-u_\mu|}{q_\mu(u_\mu)} \leq Bn^{-1/2}} r_\mu(v)^2 \, dC_{\theta_0}(\boldsymbol{u})$$
$$= O(n^{-2}) O(n\ell_n) = o(1),$$

where we used the Cauchy-Schwarz inequality, (A.2) and the fact that

$$\frac{1}{n} \sum_{i,j=1}^{n} |\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]| = \frac{1}{n} \sum_{i,j=1}^{n} \varphi\{(i-j)/\ell_n\}$$
$$= \sum_{k=-\ell_n}^{\ell_n} (1 - |k|/n)\varphi(k/\ell_n) \leq \{2\ell_n + 1\}\varphi(0). \quad \text{(C.3)}$$

Thus, $A_{n,1} \times \sqrt{n}(\theta_n - \theta_0) = o_P(1)$. For the estimation of $A_{n,2}$, let $\varepsilon > 0$ be

31

arbitrary. By Condition V, we can find $n_0 \in \mathbb{N}$ such that, for any $n \geq n_0$,

$$1 - \varepsilon < \mathbb{P}\Big(\|\theta_n - \theta_0\| < k_\theta \Big) \leq \mathbb{P}\Big(\cap_{i=1}^n \{|\theta'_{i,n} - \theta_0| < k_\theta\} \Big)$$

$$\leq \mathbb{P}\big(\cap_{i=1}^n \{\|\nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)\|_{op} \leq \prod_{\mu=1}^d r_\mu(\hat{U}_{i,\mu})\}\big).$$

This implies that, for any $\eta > 0$ and for all $n \geq n_0$,

$$\mathbb{P}(|A_{n,2}| > \eta) \leq \mathbb{P}\big(\cup_{i=1}^n \{\|\nabla J_{\theta'_{i,n}}(\hat{\boldsymbol{U}}_i)\|_{op} > \prod_{\mu=1}^d r_\mu(\hat{U}_{i,\mu})\}\big) < \varepsilon,$$

that is $A_{n,2} = o_p(1)$ and hence $A_n = o_P(1)$.

It remains to be shown, that $B_{n,\mu} = o_P(1)$ for all $\mu = 1,\ldots,d$. For this purpose, we bound $B_{n,\mu}$ as follows

$$\|B_{n,\mu}\|$$
$$= \left\| \frac{1}{n} \sum_{j=1}^n \nabla J_{\theta_{j,\mu}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)(\theta_n - \theta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)}\{I(U_{i,\mu} \leq U_{j,\mu}) - U_{j,\mu} + U_{j,\mu} - \hat{U}_{j,\mu}\} \right\|$$

$$\leq \|\theta_n - \theta_0\| \times \left\{ \frac{1}{n} \sum_{j=1}^n \|\nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)\|_{op} \, q_\mu(U_{j,\mu}) \right\} \times$$

$$\left\{ \sup_{u \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \frac{\mathbb{1}(U_{i,\mu} \leq u) - u}{q_\mu(u)} \right| \right.$$

$$\left. + \sup_{u \in [U_{1:n,\mu}, U_{n:n,\mu}]} \sqrt{n} \left| \frac{G_{n\mu}^\star(u) - u}{q_\mu(u)} \right| \times \left| \frac{1}{n} \sum_{i=1}^n Z_{i,n}^{(b)} \right| \right\}$$

Since $\|\theta_n - \theta_0\|$ converges to zero in probability, it suffices to show that both factors in curly brackets are stochastically bounded. For the second factor, this follows from Lemma D.1, Theorem 2.2 in Shao and Yu (1996), a simple adaptation of Lemma 2.3.1 in Ruymgaart (1973) to the alpha-mixing case and from (C.3). Regarding the first factor, we decompose

$$\frac{1}{n} \sum_{j=1}^n \|\nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)\|_{op} \, q_\mu(U_{j,\mu})$$

$$= \frac{1}{n} \sum_{j=1}^n \|\nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)\|_{op} \, q_\mu(U_{j,\mu}) \times \mathbb{1}(\|\theta_n - \theta_0\| \leq k_{\theta_0})$$

$$+ \frac{1}{n} \sum_{j=1}^n \|\nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{\boldsymbol{U}}_j)\|_{op} \, q_\mu(U_{j,\mu}) \times \mathbb{1}(\|\theta_n - \theta_0\| > k_{\theta_0}).$$

32

Clearly, the second sum on the right is $o_P(1)$ by consistency of θ_n. By Condition V (in particular (A.3)), Markov's inequality and $|\hat{U}_{j,\mu} - U_{j,\mu}|/q_\mu(U_{j,\mu}) \leq D/\sqrt{n}$ for all μ as a consequence of the fact that we may assume $\Omega = \Omega_{n,D}$, we have

$$\frac{1}{n} \sum_{j=1}^n \|\nabla J_{\theta_{j,n}^\mu}^{[\mu]}(\hat{U}_j)\|_{op} q_\mu(U_{j,\mu}) \times \mathbb{1}(\|\theta_n - \theta_0\| \leq k_{\theta_0})$$

$$\leq \frac{1}{n} \sum_{j=1}^n \tilde{r}_\mu(\hat{U}_{j,\mu}) q_\mu(U_{j,\mu}) \prod_{\mu' \neq \mu} r_{\mu'}(\hat{U}_{j,\mu'}) = O_P(1).$$

To conclude, $B_{n,\mu} = o_P(1)$ which proves the lemma. $\qquad\square$

Lemma C.4. *Suppose that $(\boldsymbol{U}_i)_{i\in\mathbb{Z}}$ satisfies $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 6$ (which is a consequence of Condition III). Moreover, suppose that Condition V holds. Then, for any $b = 1, \ldots, B$,*

$$L_{n2}^{(b)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \left[\{J_{\theta_0}(\hat{\boldsymbol{U}}_i) + \hat{K}_{i,n,\theta_0}\} - \{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,n,\theta_0})\} \right] = o_P(1).$$

Proof. As in the proof of Lemma C.1, we may assume without loss of generality that $\Omega = \Omega_{n,D}$ for some $D > 0$. Let us write $L_{n2}^{(b)} = A_n + B_n$, where

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \{J_{\theta_0}(\hat{\boldsymbol{U}}_i) - J_{\theta_0}(\boldsymbol{U}_i)\}, \quad B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \{\hat{K}_{i,n,\theta_0} - K_{i,n,\theta_0}\},$$

and consider each term separately. For $\eta \in (0, 1/2)$, set $M_\eta = [\eta, 1 - \eta]^d$. By the mean value theorem, there exist intermediate values $\tilde{\boldsymbol{U}}_i$ between $\hat{\boldsymbol{U}}_i$ and $\boldsymbol{U}_i$ such that we can write $A_n = \sum_{\mu=1}^d \{A_{n,\mu}(M_\eta) + A_{n,\mu}(M_\eta^C)\}$, where, for any $M \subset [0,1]^d$,

$$A_{n,\mu}(M) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} J_{\theta_0}^{[\mu]}(\tilde{\boldsymbol{U}}_i)(\hat{U}_{i,\mu} - U_{i,\mu}) \mathbb{1}(\boldsymbol{U}_i \in M).$$

We begin with the treatment of $A_{n,\mu}(M_\eta^C)$, for a fixed μ. Since we may assume $\Omega = \Omega_{n,D}$, we can bound $\|A_{n,\mu}(M_\eta^C)\|$ by

$$\frac{D}{n} \sum_{i=1}^n |Z_{i,n}^{(b)}| q_\mu(U_{i,\mu}) \sup_{\frac{|u-U_{i,\mu}|}{q_\mu(U_{i,\mu})} \leq \frac{B}{\sqrt{n}}} \tilde{r}_\mu(u) \prod_{\mu' \neq \mu} \sup_{\frac{|u-U_{i,\mu'}|}{q_{\mu'}(U_{i,\mu'})} \leq \frac{B}{\sqrt{n}}} r_{\mu'}(u) \mathbb{1}(\boldsymbol{U}_i \in M_\eta^C)$$

33

As in the proof of Lemma C.1, by Condition V, for sufficiently large n, the expectation of the latter expression converges to zero as $\eta \to 0$. Therefore, it remains to consider $A_{n,\mu}(M_\eta)$ for fixed $\eta \in (0, 1/2)$.

Since $\max_{i=1}^n \|\hat{U}_i - U_i\| \leq D/\sqrt{n}$, we get that $A_{n,\mu}(M_\eta) = \bar{A}_{n,\mu}(M_\eta) + o_P(1)$ for $n \to \infty$, where

$$\bar{A}_{n,\mu}(M_\eta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} J_{\theta_0}^{[\mu]}(\tilde{U}_i)(\hat{U}_{i,\mu} - U_{i,\mu}) \mathbb{1}(U_i \in M_\eta, \|\hat{U}_i - U_i\| \leq \eta/2).$$

By the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}\big[\|\bar{A}_{n,\mu}(M_\eta)\|^2\big] = \mathbb{E}\big[\{\bar{A}_{n,\mu}(M_\eta)\}'\{\bar{A}_{n,\mu}(M_\eta)\}\big]$$

$$\leq \sup_{\boldsymbol{u} \in M_{\eta/2}} \|J_{\theta_0}^{[\mu]}(\boldsymbol{u})\|^2 \times \mathbb{E}\big[\max_{i=1}^n |\hat{U}_{i,\mu} - U_{i,\mu}|^2\big] \times \frac{1}{n} \sum_{i,j=1}^n |\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]|.$$

The first factor on the right-hand side is bounded by Condition V. The second factor is bounded by D^2/n as we may assume that $\Omega = \Omega_{n,D}$. Regarding the third factor, note that $\frac{1}{n} \sum_{i,j=1}^n |\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]| = O(\ell_n)$ as shown in (C.3). We can conclude that $\bar{A}_{n,\mu}(M_\eta)$ is of order $O_p(n^{-1/4 - \kappa/2}) = o_P(1)$, and therefore also $A_n = o_P(1)$.

For the proof of the lemma, it remains to be shown that $B_n = o_P(1)$. We can decompose $B_n = \sum_{\mu=1}^d (B_{n,1,\mu} + B_{n,2,\mu} + B_{n,3,\mu})$, with

$$B_{n,1,\mu} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \frac{1}{n} \sum_{j=1}^n \{J_\theta^{[\mu]}(\hat{U}_j) - J_\theta^{[\mu]}(\boldsymbol{U}_j)\}\{\mathbb{1}(U_{i,\mu} \leq U_{j,\mu}) - U_{j,\mu}\}$$

$$B_{n,2,\mu} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,n}^{(b)} \frac{1}{n} \sum_{j=1}^n \{J_\theta^{[\mu]}(\hat{U}_j) - J_\theta^{[\mu]}(\boldsymbol{U}_j)\}(U_{j,\mu} - \hat{U}_{j,\mu})$$

$$B_{n,3,\mu} = \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n J_\theta^{[\mu]}(\boldsymbol{U}_j)(\hat{U}_{j,\mu} - U_{j,\mu}) \right\} \times \left\{ \frac{1}{n} \sum_{i=1}^n Z_{i,n}^{(b)} \right\}.$$

$B_{n,3,\mu}$ converges in probability to zero: the first factor is of order $O_P(1)$ by a similar argumentation as before based on the fact that we may assume $\Omega = \Omega_{n,D}$, and the second factor is of order $O_P(\ell_n^{1/2}/n^{1/2})$ by (C.3).

Regarding $B_{n,1,\mu}$, we can bound

$$\|B_{n,1,\mu}\| \leq \left\{ \frac{1}{\sqrt{n}} \sup_{u \in [0,1]} \Big| \sum_{i=1}^n Z_{i,n}^{(b)} \frac{\mathbb{1}(U_{i,\mu} \leq u) - u}{q_\mu(u)} \Big| \right\}$$

$$\times \left\{ \frac{1}{n} \sum_{j=1}^n \|J_\theta^{[\mu]}(\hat{U}_j) - J_\theta^{[\mu]}(\boldsymbol{U}_j)\| q_\mu(U_{j,\mu}) \right\}.$$

The first factor on the right-hand side is of order $O_P(1)$ by Lemma D.1. The second factor converges to 0 in probability by the same argumentation as for the treatment of (C.2) in the proof Lemma C.1.

A similar argumentation also works for $B_{n,2,\mu}$: on the set $\Omega_{n,D}$, we have

$$\|B_{n,2,\mu}\| \leq \left\{ \frac{1}{n} \sum_{i=1}^{n} |Z_{i,n}^{(b)}| \right\} \times \left\{ \frac{D}{n} \sum_{j=1}^{n} \|J_\theta^{[\mu]}(\hat{\boldsymbol{U}}_j) - J_\theta^{[\mu]}(\boldsymbol{U}_j)\| q_\mu(U_{j,\mu}) \right\},$$

and this expression is $o_P(1)$ since $\mathbb{E}[|Z_{i,n}^{(b)}|] < \infty$ and by the same reasons as in the proof of Lemma C.1. $\qquad\square$

Lemma C.5. *Suppose that $(\boldsymbol{U}_i)_{i\in\mathbb{Z}}$ satisfies $\alpha_{\boldsymbol{U}}(n) = O(n^{-a})$ for some $a > 6$ (which is a consequence of Condition III). Moreover, suppose that Condition V holds. Then, for any $b = 1, \ldots, B$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \left\{ K_{i,n,\theta_0} - K_{i,\theta_0} \right\} \xrightarrow{P} 0.$$

Proof. For $\mu \in \{1, \ldots, d\}$ and $A \subset [0,1]^d$, set

$$\tilde{I}_{n,\mu}(A) = \int_A J_{\theta_0}^{[\mu]}(\boldsymbol{u}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \{\mathbb{1}(U_{i,\mu} \leq u_\mu) - u_\mu\} \mathrm{d}\{G_n(\boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\}.$$

As in the proof of Lemma C.2, it suffices to show that $\tilde{I}_{n,\mu}([0,1]^d) = o_P(1)$ for any fixed $\mu = 1, \ldots, d$. This follows along the same lines as in the proof of Lemma C.2 with the weak convergence of $\sqrt{n}\{G_{n\mu}(u_\mu) - u_\mu\}$ and Theorem 2.2 in Shao and Yu (1996) replaced by the weak convergence of the process $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \{I(U_{i,\mu} \leq u_\mu) - u_\mu\}$ and Lemma D.1, respectively. $\qquad\square$

Lemma C.6. *If Conditions III, V and VI hold and if $\kappa > \frac{1}{2(1+\nu)}$ with κ and ν as defined in Definition A.1 and Condition V, respectively, then*

$$\left(\tilde{\mathbb{B}}_n, \tilde{\Theta}_n, \tilde{\mathbb{B}}_n^{(1)}, \tilde{\Theta}_n^{(1)}, \ldots, \tilde{\mathbb{B}}_n^{(B)}, \tilde{\Theta}_n^{(B)} \right) \rightsquigarrow \left(\mathbb{B}_C, \Theta, \mathbb{B}_C^{(1)}, \Theta^{(1)}, \ldots, \mathbb{B}_C^{(B)}, \Theta^{(B)} \right).$$

in $\left\{ \ell^\infty([0,1]^d) \times \mathbb{R}^p \right\}^{B+1}$, where the expressions on the left are defined in (B.1), (B.2) and (B.3).

Proof. Using Condition III, tightness of the vector of processes follows from marginal tightness of $\tilde{\mathbb{B}}_n$ and $\tilde{\mathbb{B}}_n^{(b)}$, see Theorem 3.1 in Bücher and Kojadinovic (2013).

Regarding weak convergence of the finite dimensional distributions, we only consider $(\tilde{\mathbb{B}}_n, \tilde{\Theta}_n, \tilde{\mathbb{B}}_n^{(b)}, \tilde{\Theta}_n^{(b)})$ for the ease of reading. By the Cramér-Wold device, we have to show that, for any $q, q' \in \mathbb{N}$, any $c_1, \ldots, c_q, \bar{c}_1, \ldots, \bar{c}_{q'} \in \mathbb{R}$, $c, \bar{c} \in \mathbb{R}^p$ and any $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_q, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{q'} \in [0,1]^d$,

$$V_n := \sum_{s=1}^{q} c_s \tilde{\mathbb{B}}_n(\boldsymbol{u}_s) + \sum_{s=1}^{q'} \bar{c}_s \tilde{\mathbb{B}}_n^{(b)}(\boldsymbol{v}_s) + c' \tilde{\Theta}_n + \bar{c}' \tilde{\Theta}_n^{(b)}$$

$$\rightsquigarrow V := \sum_{s=1}^{q} c_s \mathbb{B}_C(\boldsymbol{u}_s) + \sum_{s=1}^{q'} \bar{c}_s \mathbb{B}_C^{(b)}(\boldsymbol{v}_s) + c' \Theta + \bar{c}' \Theta^{(b)}.$$

First of all, we decompose

$$V_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i + W_i^{(b)} + T_i + T_i^{(b)}$$

with $W_i = \sum_{s=1}^{q} c_s \{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}_s) - C_{\theta_0}(\boldsymbol{u}_s)\}, W_i^{(b)} = \sum_{s=1}^{q'} Z_{i,n}^{(b)} \bar{c}_s \{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{v}_s) - C_{\theta_0}(\boldsymbol{v}_s)\}, T_i = c'\{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}$ and $T_i^{(b)} = Z_{i,n}^{(b)} \bar{c}' \{J_{\theta_0}(\boldsymbol{U}_i) + K_{i,\theta_0}\}$. The subsequent proof is based on the 'big block-small block'-technique. The assumption on a in Condition VI is equivalent to $\frac{1}{2(1+\nu)} < \frac{1}{2} - \frac{2+\nu}{a\nu}$ whence, noting that also $\kappa > \frac{1}{2(1+\nu)}$ by assumption, we may choose $0 < \eta_1 < \eta_2 < \kappa$ such that $\frac{1}{2(1+\nu)} < \eta_1 < \eta_2 < \frac{1}{2} - \frac{2+\nu}{a\nu}$. Now, set $b_n = \lfloor n^{1/2-\eta_1} \rfloor$ (the length of the big blocks), $s_n = \lfloor n^{1/2-\eta_2} \rfloor$ (the length of the small blocks) and $k_n = \lfloor n/(b_n + s_n) \rfloor$ (the number of big or small blocks). Notice, that $k_n = O(n^{1/2+\eta_1})$. For $j = 1, \ldots, k_n$, set

$$B_{jn} = \sum_{i=(j-1)(b_n+s_n)+1}^{(j-1)(b_n+s_n)+b_n} W_i + W_i^{(b)} + T_i + T_i^{(b)},$$

$$S_{jn} = \sum_{i=(j-1)(b_n+s_n)+b_n+1}^{j(b_n+s_n)} W_i + W_i^{(b)} + T_i + T_i^{(b)},$$

such that we can write

$$V_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B_{jn} + \frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} S_{jn} + \frac{1}{\sqrt{n}} R_n,$$

where $R_n = \sum_{i=k_n(b_n+s_n)+1}^{n} W_i + W_i^{(b)} + T_i + T_i^{(b)}$ is the sum over the remaining indices that are not part of a big or a small block.

First of all, let us show that the variance of V_n is equal to $\text{Var}(\frac{1}{\sqrt{n}}\sum_{j=1}^{k_n} B_{jn})+o(1)$ as $n \to \infty$, i.e., that

$$\frac{1}{n}\text{Var}\left(\sum_{j=1}^{k_n} S_{jn}\right) + \frac{1}{n}\text{Var}(R_n) + \frac{2}{n}\sum_{j,j'=1}^{k_n}\text{Cov}(B_{jn}, S_{j'n})$$

$$+ \frac{2}{n}\sum_{j=1}^{k_n}\text{Cov}(B_{jn}, R_n) + \frac{2}{n}\sum_{j=1}^{k_n}\text{Cov}(S_{jn}, R_n) \quad \text{(C.4)}$$

vanishes as $n \to \infty$. To this end, we will frequently exploit the following bounds which are consequences of Lemma 3.9 and Lemma 3.11 in Dehling and Philipp (2002)

$$|\mathbb{E}[W_i W_{i'}]| = \left|\sum_{s,s'=1}^q c_s c_{s'}\mathbb{E}[\{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}_s) - C_{\theta_0}(\boldsymbol{u}_s)\}\{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}_{s'}) - C_{\theta_0}(\boldsymbol{u}_{s'})\}]\right|$$

$$\leq 4\sum_{s,s'=1}^q |c_s c_{s'}|\alpha(|i-i'|) \leq \text{const} \times \alpha(|i-i'|)$$

$$|\mathbb{E}[W_i T_{i'}]| = \left|\sum_{s=1}^q c_s c'\mathbb{E}[\{\mathbb{1}(\boldsymbol{U}_i \leq \boldsymbol{u}_s) - C_\theta(\boldsymbol{u}_s)\}J_{\theta_0}(\boldsymbol{U}_{i'})]\right|$$

$$\leq 20\|c\|\sum_{s=1}^q |c_s|\alpha(|i-i'|)^{\frac{1+\nu}{2+\nu}}\mathbb{E}[\|J_{\theta_0}\|^{2+\nu}]^{\frac{1}{2+\nu}} \leq \text{const} \times \alpha(|i-i'|)^{\frac{1+\nu}{2+\nu}}$$

$$|\mathbb{E}[T_i T_{i'}]| = |c'\mathbb{E}[J_{\theta_0}(\boldsymbol{U}_i)J_{\theta_0}(\boldsymbol{U}_{i'})']c| \leq 40\|c\|^2\alpha(|i-i'|)^{\frac{\nu}{2+\nu}}\mathbb{E}[\|J_{\theta_0}\|^{2+\nu}]^{\frac{2}{2+\nu}}$$

$$\leq \text{const} \times \alpha(|i-i'|)^{\frac{\nu}{2+\nu}}.$$

Analogously, $|\mathbb{E}[W_i^{(b)}W_{i'}^{(b)}]| \leq \text{const} \times \alpha(|i-i'|))$, $|\mathbb{E}[W_i^{(b)}T_{i'}^{(b)}]| \leq \text{const} \times \alpha(|i-i'|)^{\frac{1+\nu}{2+\nu}}$ and $|\mathbb{E}[T_i^{(b)}T_{i'}^{(b)}]| \leq \text{const} \times \alpha(|i-i'|)^{\frac{\nu}{2+\nu}}$. Notice that all the other pairs of random variables are uncorrelated and that largest bound is a constant multiple of $\alpha(|i-i'|)^{\frac{\nu}{2+\nu}}$. Now, we can begin with the discussion of the first summand in (C.4). We have

$$\frac{1}{n}\text{Var}\left(\sum_{j=1}^{k_n} S_{jn}\right) = \frac{1}{n}\sum_{j=1}^{k_n}\text{Var}(S_{jn}) + \frac{2}{n}\sum_{j\neq j'}\text{Cov}(S_{jn}, S_{j'n}). \quad \text{(C.5)}$$

Since the distance between any two summands in S_{jn} and $S_{j'n}$ for $j \neq j'$ is at least b_n, their covariance is of order $\alpha(b_n)^{\frac{\nu}{2+\nu}} = O(b_n^{-a\nu/(2+\nu)})$. Observing that S_{jn} consists of s_n summands, we obtain that the second term in the last display is of order $O(k_n^2 s_n^2 b_n^{-a\nu/(2+\nu)} n^{-1}) = O(n^{1-a\nu/(4+2\nu)+\eta_1(2+a\nu/(2+\nu))-2\eta_2}) =$

$O(n^{1-a\nu/(4+2\nu)+\eta_2 a\nu/(2+\nu)}) = o(1)$ since, by construction, $\eta_1 < \eta_2 < \frac{1}{2} - \frac{2+\nu}{a\nu}$. For the first sum on the right-hand side of (C.5), we have, by dominated convergence,

$$\frac{1}{n}\sum_{j=1}^{k_n}\mathrm{Var}(S_{nj}) = \frac{1}{n}\sum_{j=1}^{k_n}\mathbb{E}[S_{nj}^2] \le \mathrm{const}\times\frac{1}{n}\sum_{j=1}^{k_n}\sum_{i=-s_n}^{s_n}(s_n-|i|)\alpha(|i|)^{\frac{\nu}{2+\nu}}$$

$$= O(k_n s_n n^{-1}) = O(n^{\eta_1-\eta_2}) = o(1).$$

For the second term of (C.4), we have

$$\frac{1}{n}\mathrm{Var}(R_n) \le \mathrm{const}\times\frac{1}{n}\sum_{i,i'=k_n(b_n+s_n)+1}^{n}\alpha(|i-i'|)^{\frac{\nu}{2+\nu}}$$

$$\le \mathrm{const}\times\frac{1}{n}\sum_{i=-\{n-k_n(b_n+s_n)\}}^{n-k_n(b_n+s_n)}(n-k_n(s_n+b_n)-|i|)\alpha(|i|)^{\frac{\nu}{2+\nu}}$$

$$= O(\{n-k_n(b_n+s_n)\}/n) = O((b_n+s_n)/n) = o(1),$$

where we used that $k_n \ge n/(b_n+s_n)-1$.

Now, let us bound the third term in (C.4). First we notice that, if $j = j'$ or $j' = j-1$, we have $|\mathbb{E}[B_{jn}S_{j'n}]| \le \mathrm{const}\sum_{i=1}^{b_n}\sum_{i'=b_n+1}^{s_n+b_n}\alpha(|i-i'|)^{\frac{\nu}{2+\nu}} \le \mathrm{const}\sum_{i=1}^{s_n+b_n}i\alpha(i)^{\frac{\nu}{2+\nu}} \le \mathrm{const} < \infty$, since $a\nu/(2+\nu) > 2+1/\nu$ by Condition VI. In the other cases the distance between the blocks B_{jn} and $S_{j'n}$ is at least b_n, such that $|\mathbb{E}[B_{jn}S_{j'n}]| = O(b_n s_n\alpha(b_n)^{\frac{\nu}{2+\nu}})$. Together, this yields

$$\frac{2}{n}\sum_{j,j'=1}^{k_n}\mathbb{E}[B_{jn}S_{j'n}] = O(n^{-1}k_n) + O(k_n^2 b_n s_n b_n^{-a\frac{\nu}{2+\nu}}n^{-1})$$

$$= O(n^{-1/2+\eta_1}) + O(n^{1-\frac{a}{2}\frac{\nu}{2+\nu}+\eta_1(1+a\frac{\nu}{2+\nu})-\eta_2}) = o(1),$$

where the last equality follows exactly as above for the treatment of the first summand in (C.4). In the same manner, we get $\frac{2}{n}\sum_{j=1}^{k_n}\mathbb{E}[B_{jn}R_n] = O(n^{\frac{1}{2}-\frac{a}{2}\frac{\nu}{2+\nu}-\eta_1+a\frac{\nu}{2+\nu}\eta_2}) = o(1)$ and $\frac{2}{n}\sum_{j=1}^{k_n}\mathbb{E}[S_{nj}R_n] = O(n^{\frac{1}{2}-\frac{a}{2}\frac{\nu}{2+\nu}-\eta_2+a\frac{\nu}{2+\nu}\eta_1})+O(\{b_n+s_n\}/n) = o(1)$.

For the next step of the proof, let $B'_{jn}, j = 1,\ldots,k_n$ denote independent random variables such that each B'_{jn} has the same distribution as B_{jn}. We will show that the characteristic function of $n^{-1/2}\sum_{j=1}^{k_n}B_{jn}$ is asymptotically equivalent to the characteristic function of $n^{-1/2}\sum_{j=1}^{k_n}B'_{jn}$. For $t \in \mathbb{R}$, define $\Psi_{jn}(t) = \exp(itn^{-1/2}B_{jn})$ and notice that $\mathbb{E}[\prod_{j=1}^{k_n}\Psi_{jn}(t)]$ and $\prod_{j=1}^{k_n}\mathbb{E}[\Psi_{jn}(t)]$

are the characteristic functions of $n^{-1/2} \sum_{j=1}^{k_n} B_{jn}$ and $n^{-1/2} \sum_{j=1}^{k_n} B'_{jn}$, respectively. The difference of the two characteristic functions can be decomposed as follows

$$
\begin{aligned}
\left| \mathbb{E}\Big[\prod_{j=1}^{k_n} \Psi_{jn}(t) \Big] - \prod_{j=1}^{k_n} \mathbb{E}[\Psi_{jn}(t)] \right| &\leq \left| \mathbb{E}\Big[\prod_{j=1}^{k_n} \Psi_{jn}(t) \Big] - \mathbb{E}[\Psi_{1n}(t)]\mathbb{E}[\prod_{j=2}^{k_n} \Psi_{jn}(t)] \right| \\
&+ |\mathbb{E}[\Psi_{1n}(t)]| \times \left| \mathbb{E}\Big[\prod_{j=2}^{k_n} \Psi_{jn}(t) \Big] - \mathbb{E}[\Psi_{2n}(t)]\mathbb{E}[\prod_{j=3}^{k_n} \Psi_{jn}(t)] \right| \\
&+ \cdots + \left| \prod_{j=1}^{k_n-2} \mathbb{E}[\Psi_{jn}(t)] \right| \times \left| \mathbb{E}\Big[\prod_{k_n-1}^{k_n} \Psi_{jn}(t) \Big] - \prod_{j=k_n-1}^{k_n} \mathbb{E}[\Psi_{jn}(t)] \right|
\end{aligned}
$$

Applying Lemma 3.9 in Dehling and Philipp (2002) $(k_n - 1)$ times, we get

$$
\begin{aligned}
\left| \mathbb{E}\Big[\prod_{j=1}^{k_n} \Psi_{jn}(t) \Big] - \prod_{j=1}^{k_n} \mathbb{E}[\Psi_{jn}(t)] \right| \\
\leq 2\pi \times (k_n - 1) \max_{i=1,\ldots,k_n-1} \alpha(\sigma(\Psi_{in}), \sigma\{ \prod_{j=i+1}^{k_n} \Psi_{jn}(t) \}),
\end{aligned}
$$

which is of order $O(k_n \alpha(s_n)) = O(n^{1/2 - a/2 + a\eta_2 + \eta_1})$. Using that $\eta_1 < \eta_2$, this expression converges to 0 by the choice of η_2 and the fact that $(2 + v)/(a\nu) > 1/a > 1/(a+1)$. As a consequence, provided $n^{-1/2} \sum_{j=1}^{k_n} B'_{jn}$ converges weakly, then so does $n^{-1/2} \sum_{j=1}^{k_n} B_{jn}$ with the same limiting distribution.

Therefore, in order to finalize the proof, it remains to be shown that $n^{-1/2} \sum_{j=1}^{k_n} B'_{jn}$ converges weakly to V. This will be accomplished by proving that the variance of $n^{-1/2} \sum_{j=1}^{k_n} B'_{jn}$ converges to $\mathrm{Var}(V)$ as $n \to \infty$ and that the Lindeberg-condition from the Lindeberg-Feller central limit theorem for independent triangular arrays is met. We begin with the convergence of the variance and note that

$$
\begin{aligned}
\mathrm{Var}(V) = {} & \sum_{s,s'=1}^{q} c_s c_{s'} \sum_{i \in \mathbb{Z}} \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'}) + \sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \sum_{i \in \mathbb{Z}} \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'}) \\
& + \sum_{s=1}^{q} c_s c' \sum_{i \in \mathbb{Z}} \bar{\gamma}(i, \boldsymbol{u}_s) + \sum_{s=1}^{q'} \bar{c}_s \bar{c}' \sum_{i \in \mathbb{Z}} \bar{\gamma}(i, \boldsymbol{v}_s) \\
& + c' \sum_{i \in \mathbb{Z}} \tilde{\gamma}(i) c + \bar{c}' \sum_{i \in \mathbb{Z}} \tilde{\gamma}(i) \bar{c},
\end{aligned}
$$

where $\gamma(i, \boldsymbol{u}, \boldsymbol{v}) = \text{Cov}\{\mathbb{1}(\boldsymbol{U}_1 \le \boldsymbol{u}), \mathbb{1}(\boldsymbol{U}_{1+i} \le \boldsymbol{v})\}$, $\bar{\gamma}(i, \boldsymbol{u}) = \text{Cov}\{J_{\theta_0}(\boldsymbol{U}_1) + K_{1,\theta_0}, \mathbb{1}(\boldsymbol{U}_{1+i} \le \boldsymbol{u})\}$ and $\tilde{\gamma}(i) = \text{Cov}\{J_{\theta_0}(\boldsymbol{U}_1) + K_{1,\theta_0}, J_{\theta_0}(\boldsymbol{U}_{1+i}) + K_{1+i,\theta_0}\}$. Now, let us show that $\text{Var}(\frac{1}{\sqrt{n}} \sum_{j=1}^{k_n} B'_{jn}) = \text{Var}(V_n) + o(1)$. To this end, note that

$$\text{Var}\left(n^{-1/2} \sum_{j=1}^{k_n} B'_{jn} \right) = \frac{1}{n} \sum_{j=1}^{k_n} \text{Var}(B'_{jn}) = \frac{1}{n} \sum_{j=1}^{k_n} \text{Var}(B_{jn})$$

$$= \text{Var}\left(n^{-1/2} \sum_{j=1}^{k_n} B_{jn} \right) - \frac{1}{n} \sum_{j \ne j'} \text{Cov}(B_{jn}, B_{j'n}).$$

Since we have already shown in the beginning of the proof that the variance on the right-hand side equals $\text{Var}(V_n) + o(1)$, it remains to be shown that $\frac{1}{n} \sum_{j \ne j'} \text{Cov}(B_{jn}, B_{j'n}) = o(1)$. Since the distance between the random variables within the two blocks is at least s_n, we have $\mathbb{E}[B_{jn}B_{j'n}] = O(b_n^2 \alpha(s_n)^{\frac{\nu}{2+\nu}})$ for $j \ne j'$. Therefore, $n^{-1} \sum_{j \ne j'} \text{Cov}(B_{jn}, B_{j'n}) = O(n^{-1} k_n^2 b_n^2 s_n^{-a\frac{\nu}{2+\nu}}) = O(n^{1-\frac{a}{2}\frac{\nu}{2+\nu} + a\frac{\nu}{2+\nu}\eta_2})$, which is $o(1)$ as shown above.

Now, let us show that $\text{Var}(V_n) \to \text{Var}(V)$ as $n \to \infty$. We can write $\text{Var}(V_n)$ as

$$\frac{1}{n} \sum_{i,i'=1}^{n} \left[\left\{ \sum_{s,s'=1}^{q} c_s c_{s'} \gamma(i - i', \boldsymbol{u}_s, \boldsymbol{u}_{s'}) + \sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \varphi\{(i - i')/\ell_n\} \gamma(i - i', \boldsymbol{v}_s, \boldsymbol{v}_{s'}) \right\} \right.$$

$$+ \left\{ \sum_{s=1}^{q} c' c_s \bar{\gamma}(i - i', \boldsymbol{u}_s) + \sum_{s=1}^{q'} \bar{c}' \bar{c}_s \varphi\{(i - i')/\ell_n\} \bar{\gamma}(i - i', \boldsymbol{v}_s) \right\}$$

$$\left. + \left\{ c' \tilde{\gamma}(i - i')c + \varphi\{(i - i')/\ell_n\} \bar{c}' \tilde{\gamma}(i - i') \bar{c} \right\} \right]$$

$$= \sum_{i=-n}^{n} \frac{n - |i|}{n} \left[\left\{ \sum_{s,s'=1}^{q} c_s c_{s'} \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'}) + \sum_{s,s'=1}^{q'} c_s c_{s'} \varphi(i/\ell_n) \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'}) \right\} \right.$$

$$+ \left\{ \sum_{s=1}^{q} c' c_s \bar{\gamma}(i, \boldsymbol{u}_s) + \sum_{s=1}^{q'} \bar{c}' \bar{c}_s \varphi(i/\ell_n) \bar{\gamma}(i, \boldsymbol{v}_s) \right\}$$

$$\left. + \left\{ c' \tilde{\gamma}(i)c + \varphi(i/\ell_n) \bar{c}' \tilde{\gamma}(i) \bar{c} \right\} \right]. \tag{C.6}$$

For the sake of brevity, we will only show convergence of the terms in the first curly brackets on the right-hand side of the last display to the respective

terms in $\mathrm{Var}(V)$. We have

$$\sum_{i=-n}^{n} \frac{n-|i|}{n} \sum_{s,s'=1}^{q} c_s c_{s'} \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'})$$

$$= \sum_{s,s'=1}^{q} c_s c_{s'} \sum_{i=-n}^{n} \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'}) - \sum_{s,s'=1}^{q} c_s c_{s'} \frac{1}{n} \sum_{i=-n}^{n} |i| \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'}).$$

Since $|\gamma(i, \boldsymbol{u}, \boldsymbol{v})| \leq \mathrm{const} \times \alpha(|i|)$ and $\sum_{i=1}^{\infty} |i| \alpha(i) < \infty$, the second term on the right-hand side vanishes as $n \to \infty$, whereas the first term converges to $\sum_{s,s'=1}^{q} c_s c_{s'} \sum_{i \in \mathbb{Z}} \gamma(i, \boldsymbol{u}_s, \boldsymbol{u}_{s'})$.

Moreover, as $\varphi(h) = 0$ for $|h| > 1$ and $\ell_n = o(n)$, we have

$$\sum_{i=-n}^{n} \frac{n-|i|}{n} \sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \varphi(i/\ell_n) \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'})$$

$$= \sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \sum_{i \in \mathbb{Z}} \frac{n-|i|}{n} \varphi(i/\ell_n) \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'}),$$

By continuity of φ in 0, we have $\frac{n-|i|}{n} \varphi\{i/\ell_n\} \to 1$ as $n \to \infty$ for any fixed $i \in \mathbb{Z}$. Moreover, $|\frac{n-|i|}{n} \varphi(i/\ell_n) \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'})| \leq \mathrm{const} \times \alpha(i)$ for all $i \in \mathbb{Z}$ and all $s, s' = 1, \dots, q'$. Therefore, by dominated convergence, as $n \to \infty$,

$$\sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \sum_{i=-n}^{n} \frac{n-|i|}{n} \varphi(i/\ell_n) \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'}) \to \sum_{s,s'=1}^{q'} \bar{c}_s \bar{c}_{s'} \sum_{i \in \mathbb{Z}} \gamma(i, \boldsymbol{v}_s, \boldsymbol{v}_{s'}).$$

The convergence of the remaining summands in (C.6) follows along similar lines, exploiting that $\|\bar{\gamma}(i, \boldsymbol{u})\| \leq \mathrm{const} \times \alpha(i)^{\frac{1+\nu}{2+\nu}}$, $\|\tilde{\gamma}(i)\|_{op} \leq \mathrm{const} \times \alpha(i)^{\frac{\nu}{2+\nu}}$ and that $\sum_{i=1}^{\infty} |i| \alpha(i)^{\frac{\nu}{2+\nu}} < \infty$.

Finally, let us prove the Lindeberg condition, i.e., that, for any $\varepsilon > 0$,

$$\frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{jn}'^2 \mathbb{1}(|B_{jn}'| > \sqrt{n}\varepsilon)] = \frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{jn}^2 \mathbb{1}(|B_{jn}| > \sqrt{n}\varepsilon)] \to 0$$

as $n \to \infty$. To bound the former expression we use Hölder's inequality with

41

$p = (2 + \nu)/2$ and $q = (2 + \nu)/\nu$ and Markov's inequality to obtain

$$\frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{jn}^2 \mathbb{1}(|B_{jn}| \leq \sqrt{n}\varepsilon)] \leq \frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[|B_{jn}|^{2+\nu}]^{\frac{2}{2+\nu}} \mathbb{E}[\mathbb{1}(|B_{jn}| \leq \sqrt{n}\varepsilon)^{\frac{2+\nu}{\nu}}]^{\frac{\nu}{2+\nu}}$$

$$= \frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[|B_{jn}|^{2+\nu}]^{\frac{2}{2+\nu}} \mathbb{P}(|B_{jn}| \leq \sqrt{n}\varepsilon)^{\frac{\nu}{2+\nu}}$$

$$\leq \frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[|B_{jn}|^{2+\nu}]^{\frac{2}{2+\nu}} \mathbb{E}[|B_{jn}|^{2+\nu}]^{\frac{\nu}{2+\nu}} (\sqrt{n}\varepsilon)^{-\nu}$$

$$= \frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[|B_{jn}|^{2+\nu}](\sqrt{n}\varepsilon)^{-\nu}$$

By Minkowski's inequality, we can bound $\mathbb{E}[|B_{jn}|^{2+\nu}]^{\frac{1}{2+\nu}}$ by a sum over b_n summands of the form $\mathbb{E}[|W_i|^{2+\nu}]^{\frac{1}{2+\nu}} + \mathbb{E}[|W_i^{(b)}|^{2+\nu}]^{\frac{1}{2+\nu}} + \mathbb{E}[|T_i|^{2+\nu}]^{\frac{1}{2+\nu}} + \mathbb{E}[|T_i^{(b)}|^{2+\nu}]^{\frac{1}{2+\nu}}$, whence $\mathbb{E}[|B_{jn}|^{2+\nu}] = O(b_n^{2+\nu})$. This finally implies

$$\frac{1}{n} \sum_{j=1}^{k_n} \mathbb{E}[B_{jn}^2 \mathbb{1}(|B_{jn}| \leq \sqrt{n}\varepsilon)] = O(b_n^{2+\nu} k_n n^{-\nu/2-1}) = O(n^{1/2 - \eta_1(1+\nu)}) = o(1),$$

by the definition of η_1 and the Lemma is proved. $\qquad\square$

Lemma C.7. *Under Conditions IV, V and VI, we have*

$$\sqrt{n}\{C_{\theta_n}(\boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\} = \nabla C_{\theta_0}(\boldsymbol{u})\Theta_n + R_n(\boldsymbol{u}),$$

where $\Theta_n = \sqrt{n}(\theta_n - \theta_0)$ and $\sup_{\boldsymbol{u} \in [0,1]^d} \|R_n(\boldsymbol{u})\| = o_p(1)$.

Proof. It follows from the mean value theorem that

$$\sqrt{n}\{C_{\theta_n}(\boldsymbol{u}) - C_{\theta_0}(\boldsymbol{u})\} = \nabla C_{\theta_0}(\boldsymbol{u})\Theta_n + R_n(\boldsymbol{u}),$$

where $R_n(\boldsymbol{u}) = \sqrt{n} \sum_{s=1}^{p} \left\{ \frac{\partial}{\partial \theta_s} C_{\tilde{\theta}}(\boldsymbol{u}) - \frac{\partial}{\partial \theta_s} C_{\theta_0}(\boldsymbol{u}) \right\}(\theta_{ns} - \theta_{0s})$ and where $\tilde{\theta}$ denotes an intermediate point lying between θ_0 and θ_n. The Cauchy-Schwarz inequality allows to estimate

$$\sup_{\boldsymbol{u} \in [0,1]^d} \|R_n(\boldsymbol{u})\| \leq \sup_{\|\theta' - \theta_0\| < \|\theta_n - \theta_0\|} \sup_{\boldsymbol{u} \in [0,1]^d} \|\nabla C_{\theta'}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\| \times \|\Theta_n\|.$$

Since $\Theta_n = \tilde{\Theta}_n + o_p(1)$ for $n \to \infty$ as a consequence of Lemma C.1 and C.2, we obtain that $\|\Theta_n\| = O_P(1)$ from Lemma C.6 (note that Condition III was

used in Lemma C.6 only for the tightness part, whence we do not need to assume it here). Fix $\eta > 0$. By Condition IV, we may choose $\delta > 0$ such that $\sup_{\|\theta' - \theta_0\| < \delta} \sup_{\boldsymbol{u} \in [0,1]^d} \|\nabla C_{\theta'}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\| \leq \eta$. Therefore,

$$\mathbb{P}(\sup_{\boldsymbol{u} \in [0,1]^d} \sup_{\|\theta' - \theta_0\| < \|\theta_n - \theta\|} \|\nabla C_{\theta'}(\boldsymbol{u}) - \nabla C_{\theta_0}(\boldsymbol{u})\| > \eta) \leq \mathbb{P}(\|\theta_n - \theta_0\| > \delta),$$

which converges to 0 as $n \to \infty$, since θ_n is a consistent estimator of θ_0. $\quad\square$

Lemma C.8. *Let A be a continuous function on $[\boldsymbol{a}, \boldsymbol{b}]$ and B_n be a sequence of functions on $[\boldsymbol{a}, \boldsymbol{b}]$ with $\int_{[\boldsymbol{a},\boldsymbol{b}]} |dB_n| \leq V$, uniformly in n, for some $V > 0$. If $\sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |B_n(\boldsymbol{x})| \to 0$ and if there is a sequence of functions A_n such that $\sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |A_n(\boldsymbol{x}) - A(\boldsymbol{x})| \to 0$ as $n \to \infty$, then we have*

$$\Phi(A_n, B_n) \to 0$$

as $n \to \infty$, where $\Phi(A, B) = \int_{[\boldsymbol{a},\boldsymbol{b}]} A(\boldsymbol{x}) \mathrm{d}B(\boldsymbol{x})$.

Proof. For $m \in \mathbb{N}$, $j = 1, \ldots, m$ and $\mu = 1, \ldots d$, let $y_{j,\mu,m} = a_\mu + (b_\mu - a_\mu)j/m$. Define a piecewise constant approximation of A through

$$\tilde{A}_m(\boldsymbol{x}) = \sum_{k_1,\ldots,k_d=1}^{m} \mathbb{1}(\boldsymbol{x} \in (\boldsymbol{y}_{k_1-1,\ldots,k_d-1}, \boldsymbol{y}_{k_1,\ldots,k_d}]) A(\boldsymbol{y}_{k_1,\ldots k_d}),$$

where $\boldsymbol{y}_{k_1,\ldots,k_d} = (y_{k_1,1,m}, \ldots, y_{k_d,d,m})$. Note, that $r(m) = \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |A(\boldsymbol{x}) - \tilde{A}_m(\boldsymbol{x})| \to 0$ as $m \to \infty$ by uniform continuity of A on $[\boldsymbol{a}, \boldsymbol{b}]$.

Now, write $\Phi(A_n, B_n) = I_{n1} + I_{n2} + I_{n3}$, where

$$I_{n1} = \int_{[\boldsymbol{a},\boldsymbol{b}]} \{A_n(\boldsymbol{x}) - A(\boldsymbol{x})\} \mathrm{d}B_n(\boldsymbol{x}),$$

$$I_{n2} = \int_{[\boldsymbol{a},\boldsymbol{b}]} \{A(\boldsymbol{x}) - \tilde{A}_m(\boldsymbol{x})\} \mathrm{d}B_n(\boldsymbol{x}), \quad I_{n3} = \int_{[\boldsymbol{a},\boldsymbol{b}]} \tilde{A}_m(\boldsymbol{x}) \mathrm{d}B_n(\boldsymbol{x}).$$

We begin by bounding I_{n1}. We have

$$|I_{n1}| \leq \int_{[\boldsymbol{a},\boldsymbol{b}]} |A_n(\boldsymbol{x}) - A(\boldsymbol{x})| |\mathrm{d}B_n(\boldsymbol{x})| \leq \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |A_n(\boldsymbol{x}) - A(\boldsymbol{x})| \times V$$

which converges to zero as $n \to \infty$ by uniform convergence of A_n to A. Similarly, $|I_{n2}|$ can be bounded by $r(m) \times V$, which can be made smaller than any given $\varepsilon > 0$ by increasing m. Finally, for any fixed m, $|I_{n3}|$ can be bounded by $m^d \times \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |A(\boldsymbol{x})| \times \sup_{\boldsymbol{x} \in [\boldsymbol{a},\boldsymbol{b}]} |\mathbb{B}_n(\boldsymbol{x})|$. The latter expression converges to 0 for $n \to \infty$ by uniform convergence of B_n. Assembling everything, for any $\varepsilon > 0$, $\limsup_{n \to \infty} |\Phi(A_n, B_n)| \leq \varepsilon$, which proves the Lemma. $\quad\square$

D Weighted Convergence of a dependent multiplier bootstrap empirical process

Lemma D.1. *Let $(U_i)_{i \in \mathbb{Z}}$ be a stationary, strongly mixing sequence of uniformly distributed random variables with $\alpha_U(k) = O(k^{-a})$, where $a > 6$, and suppose that $Z_{i,n}^{(b)}$ satisfies the conditions in Definition A.1. For $u \in [0,1]$, set*

$$\bar{\mathbb{A}}_n^{(b)}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,n}^{(b)} \{\mathbb{1}(U_i \leq u) - u\}$$

and let q be a q-function with $q(u) \geq K\{u(1-u)\}^\gamma$ for some constant $K > 0$ and some $0 \leq \gamma < (a-3)/(2a)$. Then, in $\ell^\infty([0,1])$,

$$\bar{\mathbb{A}}_n^{(b)}(u)/q(u) \rightsquigarrow \mathbb{A}(u)/q(u)$$

where $\mathbb{A}$ denotes the weak limit of the process $u \mapsto n^{-1/2} \sum_{i=1}^{n} \{\mathbb{1}(U_i \leq u) - u\}$.

Proof. The proof is similar to the proof of Theorem 2.1 in Shao and Yu (1996). First, note that $\bar{\mathbb{A}}_n^{(b)} \rightsquigarrow \mathbb{A}$ by Theorem 3.1 in Bücher and Kojadinovic (2013). Hence, by Theorem 3.2 in Billingsley (1999) it suffices to show that, for any $\varepsilon > 0$,

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{0 < u \leq \eta} \left| \bar{\mathbb{A}}_n^{(b)}(u)/q(u) \right| \geq \varepsilon \right) = 0 \quad \text{and}$$

$$\lim_{\eta \to 0} \mathbb{P}\left(\sup_{0 < u \leq \eta} |\mathbb{A}(u)/q(u)| \geq \varepsilon \right) = 0, \quad \text{(C.7)}$$

and that

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{1-\eta \leq u < 1} \left| \bar{\mathbb{A}}_n^{(b)}(u)/q(u) \right| \geq \varepsilon \right) = 0 \quad \text{and}$$

$$\lim_{\eta \to 0} \mathbb{P}\left(\sup_{1-\eta \leq u < 1} |\mathbb{A}(u)/q(u)| \geq \varepsilon \right) = 0. \quad \text{(C.8)}$$

Since the proofs for (C.7) and (C.8) are very similar, we only treat (C.7).

We begin with a proof of the left-hand side of (C.7). Let $K \geq 1$ be a constant and let us first assume that, for any $n \geq 1$ and $i \in \{1, \dots, n\}$, we have $Z_{i,n}^{(b)} \geq -K$. The general case will be treated afterwards. We begin by showing the following three inequalities, which we will need in the course of this proof. For any $p > 2a/(a-1)$ and for any $u, v \in [0,1]$

$$\mathbb{E}[|\bar{\mathbb{A}}_n^{(b)}(u) - \bar{\mathbb{A}}_n^{(b)}(v)|^2] \leq \text{const} \times |u - v|^{\frac{2}{p}}. \quad \text{(C.9)}$$

44

For any $p' \in (\frac{2a}{a-3}, 4)$ and for any $u, v \in [0, 1]$

$$\mathbb{E}[|\bar{\mathbb{A}}_n^{(b)}(u) - \bar{\mathbb{A}}_n^{(b)}(v)|^4] \leq \text{const} \times (|u - v|^{4/p'} + n^{-1}|u - v|^{2/p'}). \qquad \text{(C.10)}$$

For any $0 \leq u \leq v \leq u + h \leq 1$,

$$\begin{aligned}
|\bar{\mathbb{A}}_n^{(b)}(u) - \bar{\mathbb{A}}_n^{(b)}(v)| &\leq |\bar{\mathbb{A}}_n^{(b)}(u + h) - \bar{\mathbb{A}}_n^{(b)}(u)| \\
&\quad + K|\mathbb{A}_n(u + h) - \mathbb{A}_n(u)| + \sqrt{n}h\{2K + n^{-1}\sum_{i=1}^n Z_{i,n}^{(b)}\}, \quad \text{(C.11)}
\end{aligned}$$

where $\mathbb{A}_n(u) = n^{-1/2}\sum_{i=1}^n\{\mathbb{1}(U_i \leq u) - u\}$.

The assertion in (C.10) follows directly from Lemma B.2 in Bücher and Kojadinovic (2013), whereas (C.11) can be shown by the same arguments as in the proof of (31) in Bücher and Kojadinovic (2013).

Hence, it remains to show (C.9). We have, for any $u \leq v$,

$$\mathbb{E}[|\bar{\mathbb{A}}_n^{(b)}(u) - \bar{\mathbb{A}}_n^{(b)}(v)|^2] = \mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{i,n}^{(b)}\{\mathbb{1}(u < U_i \leq v) - (v - u)\}\right|^2\right]$$

$$= \left|\frac{1}{n}\sum_{i,j=1}^n \mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}] \times \mathbb{E}[\{\mathbb{1}(u < U_i \leq v) - (v - u)\}\{\mathbb{1}(u < U_j \leq v) - (v - u)\}]\right|.$$

The fact that $|\mathbb{E}[Z_{i,n}^{(b)} Z_{j,n}^{(b)}]|$ is bounded by 1 together with an application of Lemma 3.11 in Dehling and Philipp (2002) with $r = s = p$ and $t = p/(p-2)$ allows to bound the right-hand side of the last display by

$$\frac{10}{n}\sum_{i,j=1}^n \alpha(|i - j|)^{\frac{p-2}{p}} \mathbb{E}[\mathbb{1}(u < U_i \leq v)]^{\frac{2}{p}} = |v - u|^{\frac{2}{p}}\frac{10}{n}\sum_{i,j=1}^n \alpha(|i - j|)^{\frac{p-2}{p}}$$

$$= |v - u|^{\frac{2}{p}}\frac{10}{n}\sum_{i=0}^n (n - |i|)\alpha(i)^{\frac{p-2}{p}} \leq |v - u|^{\frac{2}{p}}K_1,$$

where $K_1 < \infty$ since $p > 2a/(a-1)$.

In the following, we choose $p = p' \in (\frac{2a}{a-3}, 4 \wedge \gamma^{-1})$.

Let $A_n \subset \Omega$ denote the event that $\{n^{-1}\sum_{i=1}^n Z_{i,n}^{(b)} \leq K\}$. By Markov's inequality, we have $\mathbb{P}(\Omega \setminus A_n) = o(1)$.

Now, consider the probability on the left-hand side of (C.7). We have

$$\mathbb{P}\left(\sup_{0 < u \leq \eta} |\bar{\mathbb{A}}_n^{(b)}(u)/q(u)| \geq \varepsilon\right) = \mathbb{P}\left(\sup_{0 < u \leq \eta} |\bar{\mathbb{A}}_n^{(b)}(u)/q(u)| \geq \varepsilon, A_n\right) + o(1),$$

for $n \to \infty$. The probability on the right-hand side can be bounded by

$$\sum_{j=1}^{\infty} \mathbb{P}\Big(\sup_{\eta 2^{-j} < u \leq \eta 2^{-j+1}} \big| \bar{\mathbb{A}}_n^{(b)}(u)/q(u) \big| \geq \varepsilon, A_n \Big)$$

$$\leq \sum_{j=1}^{\infty} \mathbb{P}\Big(\sup_{\eta 2^{-j} < u \leq \eta 2^{-j+1}} \big| \bar{\mathbb{A}}_n^{(b)}(u) \big| \geq \varepsilon q(\eta 2^{-j}), A_n \Big)$$

$$\leq \sum_{j=1}^{\infty} \mathbb{P}\Big(\sup_{0 < u \leq \eta 2^{-j+1}} \big| \bar{\mathbb{A}}_n^{(b)}(u) \big| \geq \varepsilon q(\eta 2^{-j}), A_n \Big).$$

We will now split the sum according to $j \in G_n$ or $j \in H_n$, where

$$G_n = \{ j \in \mathbb{N} : n^{1/2} 3K \eta 2^{-j+1} \leq \varepsilon_j/2 \},$$
$$H_n = \{ j \in \mathbb{N} : n^{1/2} 3K \eta 2^{-j+1} > \varepsilon_j/2 \},$$

and where $\varepsilon_j = \varepsilon q(\eta 2^{-j})$. First, consider those summands with $j \in G_n$. In this case, (C.11) yields

$$\mathbb{P}\Big(\sup_{0 < u \leq \eta 2^{-j+1}} |\bar{\mathbb{A}}_n^{(b)}(u)| \geq \varepsilon_j, A_n \Big)$$

$$\leq \mathbb{P}\Big(|\bar{\mathbb{A}}_n^{(b)}(\eta 2^{-j+1})| + |\mathbb{A}_n(\eta 2^{-j+1})| + \sqrt{n} \eta 2^{-j+1} \big\{ 2K + \frac{1}{n} \sum_{i=1}^{n} Z_{i,n}^{(b)} \big\} \geq \varepsilon_j, A_n \Big)$$

$$\leq \mathbb{P}\Big(|\bar{\mathbb{A}}_n^{(b)}(\eta 2^{-j+1})| + |\mathbb{A}_n(\eta 2^{-j+1})| \geq \varepsilon_j/2 \Big)$$

$$\leq \mathbb{P}\Big(|\bar{\mathbb{A}}_n^{(b)}(\eta 2^{-j+1})| \geq \varepsilon_j/4 \Big) + \mathbb{P}\Big(|\mathbb{A}_n(\eta 2^{-j+1})| \geq \varepsilon_j/4 \Big).$$

The first probability on the right can be bounded with the help of (C.9) by

$$(\varepsilon_j/4)^{-2} \mathbb{E}[|\bar{\mathbb{A}}_n^{(b)}(\eta 2^{-j+1})|^2] \leq \text{const} \times (\varepsilon_j/4)^{-2} \times (\eta 2^{-j+1})^{\frac{2}{p}}$$

$$\leq \text{const} \times \{\varepsilon q(\eta 2^{-j})\}^{-2} \times (\eta 2^{-j})^{\frac{2}{p}}$$

$$\leq \text{const} \times (\eta 2^{-j})^{2(\frac{1}{p} - \gamma)}$$

$$= \text{const} \times \eta^{2\beta} (2^{2\beta})^{-j},$$

where $\beta = 1/p - \gamma > 0$, by the choice of p. Hence, for any $n \in \mathbb{N}$,

$$\sum_{j \in G_n} \mathbb{P}\Big(|\bar{\mathbb{A}}_n^{(b)}(\eta 2^{-j+1})| \geq \varepsilon_j/4 \Big) \leq \text{const} \times \eta^{2\beta} \sum_{j \in G_n} (2^{2\beta})^{-j} \leq \text{const} \times \eta^{2\beta}$$

which converges to 0 for $\eta \to 0$. For the estimation of $\mathbb{P}(|\mathbb{A}_n(\eta 2^{-j+1})| \geq \varepsilon_j/4)$ notice that the conditions of Theorem 2.2 in Shao and Yu (1996) are met

46

(since $\gamma < (a-3)/(2a)$ and since $\alpha(n) = O(n^{-a})$ for some $a > 6$). Therefore, also the conditions in Theorem 2.1 in the latter reference are met and we can estimate the probability exactly as in the proof of this Theorem.

Now, let $j \in H_n$ and define

$$\delta = \delta_{n,j} = \frac{1}{4} \frac{\varepsilon_j}{3Kn^{1/2}} = \frac{\varepsilon}{4} \frac{q(\eta 2^{-j})}{3Kn^{1/2}}$$

It follows form (C.11) that

$$\mathbb{P}\Big(\sup_{0<u\leq\eta 2^{-j+1}} |\bar{\mathbb{A}}_n^{(b)}(u)| \geq \varepsilon_j, A_n \Big) \leq \mathbb{P}\Big(\max_{1\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+1}\rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta)| \geq \varepsilon_j/2 \Big)$$

$$+ \mathbb{P}\Big(\max_{0\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+1}\rfloor} \sup_{i\delta<u\leq(i+1)\delta} |\bar{\mathbb{A}}_n^{(b)}(u) - \bar{\mathbb{A}}_n^{(b)}(i\delta)| \geq \varepsilon_j/2, A_n \Big)$$

The second probability on the right can be further estimated by

$$\mathbb{P}\Big(\max_{0\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+1}\rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta) - \bar{\mathbb{A}}_n^{(b)}((i+1)\delta)| + |\mathbb{A}_n(i\delta) - \mathbb{A}_n((i+1)\delta)|$$

$$+ \sqrt{n}\delta\{2K + \frac{1}{n}\sum_{i=1}^{n} Z_{i,n}^{(b)}\} \geq \varepsilon_j/2, A_n \Big)$$

$$\leq \mathbb{P}\Big(\max_{0\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+1}\rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta) - \bar{\mathbb{A}}_n^{(b)}((i+1)\delta)| + |\mathbb{A}_n(i\delta) - \mathbb{A}_n((i+1)\delta)| \geq \varepsilon_j/4 \Big)$$

$$\leq 2\, \mathbb{P}\Big(\max_{1\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+2}\rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta)| \geq \varepsilon_j/16 \Big) + 2\, \mathbb{P}\Big(\max_{1\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+2}\rfloor} |\mathbb{A}_n(i\delta)| \geq \varepsilon_j/16 \Big).$$

By (C.10) (note that $p' = p$) we have, for any $0 \leq i < k \leq \lfloor\delta^{-1}\eta 2^{-j+2}\rfloor$, that

$$\mathbb{E}[|\bar{\mathbb{A}}_n^{(b)}(i\delta) - \bar{\mathbb{A}}_n^{(b)}(k\delta)|^4] \leq \text{const} \times \{(|k-i|\delta)^{4/p} + n^{-1}(|k-i|\delta)^{2/p}\}$$

$$\leq \text{const} \times \{(|k-i|\delta)^{4/p} + n^{-1}|k-i|\delta^{2/p}\}.$$

Then, the main result in Móricz (1982) (with $\gamma = 4$, $f(k,m) = \text{const} \times m\delta$ and $\varphi(t,m) = \{(m\delta)^{4/p-1} + n^{-1}\delta^{2/p-1}\}^{1/4}$), which can be applied since $\bar{\mathbb{A}}_n^{(b)}(i\delta) = \sum_{k=1}^{i} \bar{\mathbb{A}}_n^{(b)}(k\delta) - \bar{\mathbb{A}}_n^{(b)}((k-1)\delta)$, implies

$$\mathbb{E}\Big[\max_{0\leq i\leq\lfloor\delta^{-1}\eta 2^{-j+2}\rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta)| \Big]$$

$$\leq \text{const} \times \eta 2^{-j+2} \times \Bigg\{ \sum_{k=0}^{\lfloor\log(\lfloor\delta^{-1}\eta 2^{-j+2}\rfloor)\rfloor-1} \{n^{-1}\delta^{2/p-1} + (\eta 2^{-j-k+1})^{4/p-1}\}^{1/4} \Bigg\}^4$$

$$\leq \text{const} \times \eta 2^{-j+2} \times \Bigg\{ \Bigg(\sum_{k=0}^{\lfloor\log(\delta^{-1}\eta 2^{-j+2})\rfloor} (n^{-1}\delta^{2/p-1})^{1/4} \Bigg)^4$$

47

$$+ \left(\sum_{k=0}^{\lfloor \log(\delta^{-1}\eta 2^{-j+2}) \rfloor} (\eta 2^{-j+1-k})^{1/p-1/4} \right)^4 \Bigg\}$$

$$\leq \text{const} \times \eta 2^{-j+2} \times \Bigg\{ n^{-1}\delta^{2/q-1}\log^4(\delta^{-1}\eta 2^{-j+2})$$

$$+ (\eta 2^{-j+1})^{4/p-1} \left(\sum_{k=0}^{\lfloor \log(\delta^{-1}\eta 2^{-j+2}) \rfloor} 2^{-k(1/p-1/4)} \right)^4 \Bigg\}$$

$$\leq \text{const} \times \Bigg\{ n^{-1}\delta^{2/p-1}(\eta 2^{-j})\log^4(\eta 2^{-j+2}\delta^{-1}) + (\eta 2^{-j})^{4/p} \Bigg\}$$

Now, we can estimate by Markov's inequality and the fact, that $\log(x)^4 \leq$ const $\times x^{1+2/p}$ for all $x \geq 1$ and that, for $j \in H_n$, $\eta 2^{-j+2}\delta^{-1} > 4$,

$$\mathbb{P}\Big(\max_{0 \leq i \leq \lfloor \delta^{-1}\eta 2^{-j+2} \rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta)| \geq \varepsilon_j/16 \Big)$$

$$\leq \text{const} \times \varepsilon_j^{-4}\{n^{-1}\delta^{2/p-1}\eta 2^{-j}\log^4(\eta 2^{-j+2}\delta^{-1}) + (\eta 2^{-j})^{4/p}\}$$

$$\leq \text{const} \times \varepsilon_j^{-4}\{n^{-1}\delta^{2/p-1}\eta 2^{-j}(\eta 2^{-j+2})^{1+2/p}\delta^{-1-2/p} + (\eta 2^{-j})^{4/p}\}$$

$$\leq \text{const} \times \varepsilon_j^{-4}\{\varepsilon_j^{-2}(\eta 2^{-j})^{2+2/p} + (\eta 2^{-j})^{4/p}\}$$

$$= \text{const} \times \varepsilon^{-4}\{\varepsilon^{-2}q(\eta 2^{-j})^{-6}(\eta 2^{-j})^{2+2/p} + q(\eta 2^{-j})^{-4}(\eta 2^{-j})^{4/p}\}$$

$$\leq \text{const} \times \{(\eta 2^{-j})^{-6\gamma+2+2/p} + (\eta 2^{-j})^{4/p-4\gamma}\}$$

Since $\gamma < 1/p$, the right-hand side is bounded by const $\times \eta^{\beta} \times (2^{\gamma})^{-j}$, where $\beta = 1/p - \gamma > 0$.

Thus, for any $n \in \mathbb{N}$,

$$\sum_{j \in H_n} \mathbb{P}\Big(\max_{0 \leq i \leq \lfloor \delta^{-1}\eta 2^{-j+2} \rfloor} |\bar{\mathbb{A}}_n^{(b)}(i\delta)| \geq \varepsilon_j/16 \Big) \leq \text{const} \times \eta^{\beta} \sum_{j \in G_n} (2^{\beta})^{-j} \leq \text{const} \times \eta^{\beta}$$

which, as to be shown, converges to 0 for $\eta \to 0$.

For the estimation of $\mathbb{P}(\max_{1 \leq i \leq \lfloor \delta^{-1}\eta 2^{-j+2} \rfloor} |\mathbb{A}_n(i\delta)| \geq \varepsilon_j/16)$ we can again proceed exatly as in the proof of Theorem 2.1 in Shao and Yu (1996).

Now, consider the general case where the multipliers are not bounded from below. Let $Z_{i,n}^{(b),+} = \max\{0, Z_{i,n}^{(b)}\}$, $Z_{i,n}^{(b),-} = \max\{0, -Z_{i,n}^{(b)}\}$, $K^+ = \mathbb{E}[Z_{i,n}^{(b),+}]$ and $K^- = \mathbb{E}[Z_{i,n}^{(b),-}]$ and note that $K^+ = K^-$. Furthermore, set $M_{i,n}^{(b),+} = Z_{i,n}^{(b),+} - K^+$ and $M_{i,n}^{(b),-} = Z_{i,n}^{(b),-} - K^-$, such that $Z_{i,n}^{(b)} = Z_{i,n}^{(b),+} - Z_{i,n}^{(b),-}$. Then we can write $\mathbb{A}_n^{(b)}(u) = \mathbb{A}_n^{(b),+}(u) - \mathbb{A}_n^{(b),-}(u)$, where $\bar{\mathbb{A}}_n^{(b),\pm}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{i,n}^{(b),\pm}\{\mathbb{1}(U_i \leq u) - u\}$, and it is sufficient to show that (C.7) and (C.8) hold with $\mathbb{A}_n^{(b)}$ replaced by $\mathbb{A}_n^{(b),+}$ and $\mathbb{A}_n^{(b),-}$. This, however, follows from what have shown so far, observing that both $(M_{i,n}^{(b),+})_{i \in \mathbb{Z}}$ and $(M_{i,n}^{(b),+})_{i \in \mathbb{Z}}$ satisfy the conditions in Definition A.1 and are bounded below by $-K^* = -K^-$.

48

Finally, let us treat $\mathbb{P}(\sup_{0<u\le\eta}|\mathbb{A}(u)/q(u)| \ge \varepsilon)$. Since $\mathbb{A}$ is sub-Gaussian with respect to the semimetric $d(u,v) = |u-v|^{1/p}$, we can apply Corollary 8.5 in Kosorok (2008) to get the bound

$$\mathbb{E}\Big[\sup_{0\le u\le\eta 2^{-j+1}}|\mathbb{A}(u)|\Big] \le \mathbb{E}|\mathbb{A}(0)| + \text{const}\int_0^{\eta 2^{-j+1}}\sqrt{\log(2x^{-p})}\mathrm{d}x,$$

for any $j \in \mathbb{N}_0$. Since $\mathbb{A}(0) = 0$, almost surely, and since $\log(x) \le \text{const}\times x^{(1-\gamma)/p}$ for all $x \ge 1$, we can bound the right-hand side by

$$\text{const}\times\int_0^{\eta 2^{-j+1}}x^{(\gamma-1)/2}\mathrm{d}x \le \text{const}\times(\eta 2^{-j})^{(\gamma+1)/2}.$$

Therefore, by Markov's inequality,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{0<u\le\eta}|\mathbb{A}(u)/q(u)| \ge \varepsilon\Big) &\le \sum_{j=0}^{\infty}\mathbb{P}\Big(\sup_{\eta 2^{-j}<u\le\eta 2^{-j+1}}|\mathbb{A}(u)/q(u)| \ge \varepsilon\Big)\\
&\le \sum_{j=0}^{\infty}\mathbb{P}\Big(\sup_{0\le u\le\eta 2^{-j+1}}|\mathbb{A}(u)| \ge \varepsilon q(\eta 2^{-j})\Big)\\
&\le \sum_{j=0}^{\infty}\mathbb{E}\Big[\sup_{0\le u\le\eta 2^{-j+1}}|\mathbb{A}(u)|\Big]q(\eta 2^{-j})^{-1}\varepsilon^{-1}\\
&\le \text{const}\times\sum_{j=0}^{\infty}(\eta 2^{-j})^{(\gamma+1)/2}q(\eta 2^{-j})^{-1}\\
&\le \text{const}\times\eta^{(1-\gamma)/2}\sum_{j=0}^{\infty}2^{-j(1-\gamma)/2},
\end{aligned}$$

which converges to 0 for $\eta \to 0$ since $(1-\gamma)/2 > 0$. $\qquad\square$

References

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica 78*(1), 395–410.

Beare, B. K. and J. Seo (2013). Time irreversible copula-based markov models. *Econometric Theory*, to appear.

Berg, D. and J.-F. Quessy (2009). Local power analyses of goodness-of-fit tests for copulas. *Scand. J. Stat. 36*(3), 389–412.

Billingsley, P. (1999). *Convergence of probability measures* (Second ed.). Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication.

Bücher, A. (2013). A note on weak convergence of the sequential multivariate empirical process under strong mixing. *Journal of Theoretical Probability*, in press.

Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist. 39*(4), 1963–2006.

Bücher, A., H. Dette, and S. Volgushev (2012). A test for Archimedeanity in bivariate copula models. *J. Multivariate Anal. 110*, 121–132.

Bücher, A. and I. Kojadinovic (2013). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *arXiv:1306.3930v1*.

Bücher, A. and M. Ruppert (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis 116*, 208–229.

Bücher, A. and S. Volgushev (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis 119*, 61–70.

Bühlmann, P. L. (1993). *The blockwise bootstrap in time series and empirical processes*. ProQuest LLC, Ann Arbor, MI. Thesis (Dr.Sc.Math)–Eidgenoessische Technische Hochschule Zürich (Switzerland).

Chen, X. and Y. Fan (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econometrics 135*(1-2), 125–154.

Chen, X. and Y. Fan (2006b). Estimation of copula-based semiparametric time series models. *J. Econometrics 130*(2), 307–335.

Darsow, W. F., B. Nguyen, and E. T. Olsen (1992). Copulas and Markov processes. *Illinois J. Math. 36*(4), 600–642.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5) 65*(6), 274–292.

Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pp. 3–113. Boston, MA: Birkhäuser Boston.

Eckert, A. (2002). Retail price cycles and response asymmetry. *Canadian Journal of Economics/Revue canadienne d'conomique 35*(1), 52–77.

Fermanian, J.-D., D. Radulović, and M. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli 10*(5), 847–860.

Frees, E. W. and E. A. Valdez (1998). Understanding relationships using copulas. *N. Am. Actuar. J. 2*(1), 1–25.

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

Genest, C., J. Nešlehová, and J.-F. Quessy (2011). Tests of symmetry for bivariate copulas. *The Annals of the Institute of Statistical Mathematics 64*, 811–834.

Genest, C., B. Rémillard, and D. Beaudoin (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insurance Math. Econom. 44*(2), 199–213.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées.* Phd thesis, Université Laval, Québec, Canada.

Kojadinovic, I., J. Segers, and J. Yan (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canad. J. Statist. 39*(4), 703–720.

Kojadinovic, I. and J. Yan (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Stat. Comput. 21*(1), 17–30.

Kojadinovic, I., J. Yan, and M. Holmes (2011). Fast large-sample goodness-of-fit tests for copulas. *Statist. Sinica 21*(2), 841–871.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer Series in Statistics. New York: Springer.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *J. Multivariate Anal. 99*(10), 2234–2250.

McCausland, W. J. (2007). Time reversibility of stationary regular finite-state markov chains. *Journal of Econometrics 136*(1), 303 – 318.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative risk management.* Princeton Series in Finance. Princeton, NJ: Princeton University Press. Concepts, techniques and tools.

Móricz, F. (1982). A general moment inequality for the maximum of partial sums of single series. *Acta Sci. Math. (Szeged) 44*(1-2), 67–75.

Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev. 23*(1), 53–70.

Rémillard, B. (2010). Goodness-of-fit tests for copulas of multivariate time series. Technical report, HEC Montréal.

Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semi-parametric models for multivariate time series. *J. Multivariate Anal. 110*, 30–42.

Rémillard, B. and O. Scaillet (2009). Testing for equality between two copulas. *J. Multivariate Anal. 100*(3), 377–386.

Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, Volume 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Berlin: Springer-Verlag.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Annals of Statistics 4*, 912–923.

Ruymgaart, F. H. (1973). *Asymptotic theory of rank tests for independence.* Amsterdam: Mathematisch Centrum. Mathematical Centre Tracts, 43.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canad. J. Statist. 33*(3), 415–427.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli 18*(3), 764–782.

Shao, Q.-M. and H. Yu (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab. 24*(4), 2098–2127.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist. 33*(3), 357–375.