Econometric Theory, **33**, 2017, 1352–1386.
doi:10.1017/S026646661600044X

TESTING FOR CHANGES IN KENDALL'S TAU

HEROLD DEHLING
University of Bochum

DANIEL VOGEL
University of Aberdeen

MARTIN WENDLER
University of Greifswald

DOMINIK WIED
University of Cologne

For a bivariate time series $((X_i, Y_i))_{i=1,\dots,n}$, we want to detect whether the correlation between X_i and Y_i stays constant for all $i = 1, \dots, n$. We propose a nonparametric change-point test statistic based on Kendall's tau. The asymptotic distribution under the null hypothesis of no change follows from a new U-statistic invariance principle for dependent processes. Assuming a single change-point, we show that the location of the change-point is consistently estimated. Kendall's tau possesses a high efficiency at the normal distribution, as compared to the normal maximum likelihood estimator, Pearson's moment correlation. Contrary to Pearson's correlation coefficient, it shows no loss in efficiency at heavy-tailed distributions, and is therefore particularly suited for financial data, where heavy tails are common. We assume the data $((X_i, Y_i))_{i=1,\dots,n}$ to be stationary and P-near epoch dependent on an absolutely regular process. The P-near epoch dependence condition constitutes a generalization of the usually considered L_p-near epoch dependence allowing for arbitrarily heavy-tailed data. We investigate the test numerically, compare it to previous proposals, and illustrate its application with two real-life data examples.

1. INTRODUCTION

For risk management and portfolio optimization, the dependence between financial asset prices is of enormous importance. It is often assumed to be constant over the observed time period, which is a simplifying assumption that is evidently violated for longer observation periods. For good statistical modeling and successful decision making, it is essential to detect changes in the association of financial

The authors wish to thank their colleague Roland Fried for several very stimulating discussions that motivated this paper. Moreover, we are grateful for helpful comments from the editors and referees, which substantially improved a previous version of the paper. We are also indebted to Alexander Dürre, who did a thorough proofreading of the manuscript. The research was supported in part by the Collaborative Research Grant 823 *Statistical modelling of nonlinear dynamic processes* of the German Research Foundation. Address correspondence to Daniel Vogel, Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, UK; e-mail: daniel.vogel@abdn.ac.uk.

 © Cambridge University Press 2016

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

price processes and, within reasonable time frames, re-estimate the correlation parameters. Particularly, in times of global financial crises, the price processes of most financial assets tend to be highly dependent, united in their common downward trend, causing the hedging powers of investment diversification to cease—an effect for which the term *diversification meltdown* has been coined.

The problem of detecting changes in the distribution of sequential observations has a long history in statistics, see e.g., Csörgő and Horváth (1997). However, particularly detecting changes in the dependence structure of multivariate time series has attracted the focus of statistical research only recently. Examples for such detection procedures are provided by Loretan and Phillips (1994), who test for covariance stationarity of a possibly heavy-tailed time series, Giacomini, Härdle, and Spokoiny (2009), who consider tests for homogeneity of time-varying copulae, Aue, Hörmann, Horváth, and Reimherr (2009), who propose a test for a constant covariance matrix, and Wied, Krämer, and Dehling (2012), who suggest a change-point test for correlations between two random variables based on Pearson's correlation coefficient.

With this paper, we want to contribute to the literature by proposing a new test for constant Kendall's tau that can be applied to dependent series. We recommend to use the rank correlation measure Kendall's tau instead of Pearson's correlation coefficient because it is almost as efficient as the moment correlation at normality, but is significantly more efficient at heavy-tailed distributions. For details see Section 5. This issue is very important in finance and economics, where many key variables, including financial returns and foreign exchange rates, are commonly known to be heavy-tailed.

Gombay and Horváth (1999), Quessy, Saïd, and Favre (2013) study tests for changes in the dependence of multivariate time series based on Kendall's tau, but only consider independent observations. Moreover, these authors also use a bootstrap approximation for deriving critical values of a test statistic, while we provide a consistent long-run variance estimator, and do not need to rely on the bootstrap. A recent reference is Bücher and Kojadinovic (2016), who propose change-point tests for Kendall's tau under mixing conditions, but not under the concept of P-near epoch dependence (see below). Our change-point test does not require information on the position of potential break points. This is a structural similarity with many other tests in the econometrics and statistics literature, (e.g., Inoue, 2001). See also Section 5.1 in Stock (1994) and the references therein.

Kendall's tau is a U-statistic. The main tool in analyzing the asymptotic behavior of the test statistic is a new functional limit theorem for sequential U-statistics processes for dependent data. This theorem is of interest in its own right. Allowing unbounded kernels and placing no moment requirement on the data process, it is formulated in by far greater generality than necessary for Kendall's tau, where the U-kernel is bounded and, further, the whole analysis can be restricted to bounded data sequences by an invariance argument. This functional limit theorem provides the basis for constructing change-point tests in the same spirit for any quantity that may be expressed as a U-statistic.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Several authors have used U-statistics for change-point problems before (e.g., Gombay and Horváth, 1995; Gombay and Horváth, 1999). The main contribution of the present paper is the thorough treatment of dependent series. We consider approximating functionals of mixing processes, where the approximation is in probability, not in an L_p sense as in the usual near epoch dependence condition. We call this approximation concept P-near epoch dependence. It generalizes L_p-near epoch dependence, not requiring the existence of any moments, and is hence a fitting framework for nonparametric and robust data analysis.

Another appealing property of our approach in terms of broad applicability is the lack of assumptions on the copula between the two random variables X_i and Y_i for a fixed i. Although Kendall's tau is a dependence measure that only depends on the copula, there are no conditions on the existence of partial derivatives. In particular, distributions with nonzero tail dependence are included in our assumptions, which is important in empirical finance (e.g., Patton, 2006). See Segers (2012) for a discussion on this issue.

The paper is organized as follows: Section 2 contains the main theoretical results about U-statistics, that is, the functional limit theorem for U-statistics and the consistency result for the estimator of the long-run variance. In Section 3, the asymptotic properties of the test statistic under the null hypothesis are given. Section 4 deals with estimating the location of a potential change-point. In Section 5, the test is compared to previous proposals. Section 6 contains a numerical study, where we observe that the efficiency properties of Kendall's tau and Pearson's correlation coefficient translate into similar size and power properties of the corresponding change-point tests. Section 7 demonstrates applications to financial data examples. Appendix A contains further background on the concept of P-near epoch dependence, Appendix B investigates the behavior of the U-statistic process under a sequence of local alternatives, and Appendix C contains the proofs for the theorems of the main text. The proofs of the lemmas in the appendix and further technical results can be found in the online supplement. Readers may refer to the supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

We use bold type face to denote vector-valued objects. Throughout, $|\cdot|_p$ denotes the p-norm in $\mathbb{R}^d$, $p \in [1, \infty)$, $d \in \mathbb{N}$. To denote the L_p norm $(E|X|^p)^{1/p}$ of a real-valued random variable X, we occasionally write $\|X\|_p$, $p \in [1, \infty)$. All random variables are defined on a common probability space $(\Omega, \mathcal{F}, P)$.

2. INVARIANCE PRINCIPLE FOR U-STATISTICS OF DEPENDENT SERIES

We treat Kendall's tau in the framework of asymptotic U-statistic theory. We first give a functional central limit theorem for general U-statistics for multivariate and short-range dependent time series. We further devise an estimator for the long run variance term and show its consistency. Throughout this section, let $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of d-dimensional random variables with

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

(d-dimensional) distribution function F. Let further $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a measurable, symmetric function. We call

$$U_n = U_n(g) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} g\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right)$$

the U-statistic with kernel g. The essential tool to treat U-statistics asymptotically is the Hoeffding decomposition into a linear and degenerate part, i.e.,

$$U_n(g) = U + \frac{2}{n} \sum_{i=1}^{n} g_1(\boldsymbol{X}_i) + \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} g_2\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right),$$

where

$$\begin{aligned} &U = Eg(\boldsymbol{X}, \boldsymbol{Y}), \quad g_1(\boldsymbol{x}) = Eg(\boldsymbol{x}, \boldsymbol{Y}) - U, \\ &g_2(\boldsymbol{x}, \boldsymbol{y}) = g(\boldsymbol{x}, \boldsymbol{y}) - g_1(\boldsymbol{x}) - g_1(\boldsymbol{y}) - U, \end{aligned} \tag{1}$$

and $\boldsymbol{X}, \boldsymbol{Y}$ are independent copies of $\boldsymbol{X}_0$.

Concerning the serial dependence structure of the process $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$, we assume it to be near epoch dependent in probability (P-NED) on an absolutely regular process. For the formal statement of this short-range dependence assumption, which follows below, it is convenient to let the process $(\boldsymbol{X}_i)$ be indexed by $\mathbb{Z}$. The observed data is then the positive branch of the doubly infinite sequence. For two sub-σ-fields $\mathcal{A}, \mathcal{B}$ of $\mathcal{F}$, we define the absolute regularity coefficient

$$\beta(\mathcal{A}, \mathcal{B}) = E\left[\operatorname{ess\,sup}\{|P(A|\mathcal{B}) - P(A)| : A \in \mathcal{A}\}\right].$$

The absolute regularity coefficient is a measure of dependence between the σ-fields $\mathcal{A}$ and $\mathcal{B}$, it lies between 0 and 1, and equals 0 if $\mathcal{A}$ and $\mathcal{B}$ are independent.

DEFINITION 2.1. *Let $(\boldsymbol{X}_n)_{n \in \mathbb{Z}}$ and $(\boldsymbol{Z}_n)_{n \in \mathbb{Z}}$ be d- and r-variate stochastic processes on $(\Omega, \mathcal{F}, P)$, respectively, $d, r \ge 1$, such that the $(d+r)$-variate process $((\boldsymbol{X}_n, \boldsymbol{Z}_n))_{n \in \mathbb{Z}}$ is stationary. For $k \le n$, let $\mathcal{F}_k^n = \sigma(\boldsymbol{Z}_k, \ldots, \boldsymbol{Z}_n)$, where also $k = -\infty$ and $n = \infty$ are permitted.*

(*i*) *The process $(\boldsymbol{Z}_n)_{n \in \mathbb{Z}}$ is called* absolutely regular *if the absolute regularity coefficients*

$$\beta_k = \beta(\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty), \quad k \ge 1,$$

converge to zero as $k \to \infty$.

(*ii*) *The process $(\boldsymbol{X}_n)_{n \in \mathbb{Z}}$ is called* near epoch dependent in probability *or short* P-near epoch dependent (P-NED) *on the process $(\boldsymbol{Z}_n)_{n \in \mathbb{Z}}$ if there is a sequence of approximating constants $(a_k)_{k \in \mathbb{N}}$ with $a_k \to 0$ as $k \to \infty$, a sequence of functions $f_k : \mathbb{R}^{r \times (2k+1)} \to \mathbb{R}^d$, $k \in \mathbb{N}$, and a nonincreasing function $\Phi : (0, \infty) \to (0, \infty)$ such that*

$$P\left(|\boldsymbol{X}_0 - f_k(\boldsymbol{Z}_{-k}, \ldots, \boldsymbol{Z}_k)|_1 > \varepsilon\right) \le a_k \Phi(\varepsilon) \tag{2}$$

for all $k \in \mathbb{N}$ and $\varepsilon > 0$.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

By Lemma A.1(iii) of Appendix A, L_p-near epoch dependence ($p \geq 1$) implies near epoch dependence in probability. So P-NED can be viewed as a generalization of the more frequently considered L_2-NED. All the limit theorems in this article may be formulated for L_p-NED sequences as well. We prefer to use P-NED instead of L_p-NED, since we particularly want to include very heavy tailed data and do not want to assume the existence of even first moments. Further details and references on the different weak dependence concepts are given in Appendix A.

For the functional U-statistic limit theorem we require Assumptions 2.2, 2.3, and 2.4 to hold.

Assumption 2.2. The process $(\boldsymbol{X}_i)_{i\in\mathbb{Z}}$ is P-NED on an absolutely regular sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{Z}}$, and there is a $\delta > 0$ such that

$$a_k \Phi\left(k^{-6}\right) = O\left(k^{-6(2+\delta)/\delta}\right) \text{ and } \sum_{k=1}^{\infty} k \beta_k^{\delta/(2+\delta)} < \infty.$$

Furthermore, a moment condition on $g(\boldsymbol{X}_i, \boldsymbol{X}_j)$ is required. Note that we do not impose any moment conditions on the data sequence $(\boldsymbol{X}_i)_{i\in\mathbb{Z}}$ itself.

Assumption 2.3. There is a constant $M > 0$ such that for all $k, n \in \mathbb{N}$

$$E\left|g\left(f_k(\boldsymbol{Z}_{-k}, \ldots, \boldsymbol{Z}_k), f_k(\boldsymbol{Z}_{n-k}, \ldots, \boldsymbol{Z}_{n+k})\right)\right|^{2+\delta} \leq M \text{ and } E\left|g(\boldsymbol{X}_0, \boldsymbol{X}_n)\right|^{2+\delta} \leq M.$$

Note that Assumptions 2.2 and 2.3 are linked via δ. Weaker moment conditions have to be paid for by a faster decay of the short-range dependence coefficients and vice versa. The next assumption is also known as the variation condition and was introduced by Denker and Keller (1986). It can be understood as a form of Lipschitz continuity of the kernel g with respect to F.

Assumption 2.4. There are constants $L, \epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$

$$E\left(\sup_{|\boldsymbol{x}-\boldsymbol{X}|\leq\epsilon, |\boldsymbol{y}-\boldsymbol{Y}|\leq\epsilon} \left|g(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{X}, \boldsymbol{Y})\right|\right)^2 \leq L\epsilon,$$

where $\boldsymbol{X}, \boldsymbol{Y}$ are independent with the same distribution as $\boldsymbol{X}_0$.

We are now ready to state the following weak invariance principle for the sequential U-process. The proof is given in Appendix C.

THEOREM 2.5. *Under Assumptions 2.2, 2.3, and 2.4, we have*

$$\left(\frac{[ns]}{\sqrt{n}}\left(U_{[ns]}(g) - U\right)\right)_{s\in[0,1]} \xrightarrow{d} 2\sigma W$$

in $D[0, 1]$, *where* W *denotes a standard Brownian motion, and the long run variance is given by*

$$\sigma^2 = \sum_{r=-\infty}^{\infty} \operatorname{Cov}\left(g_1(\boldsymbol{X}_0), g_1(\boldsymbol{X}_r)\right).$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Without specific assumptions on the distribution of the whole process $(\boldsymbol{X}_i)_{i\in\mathbb{Z}}$, the long run variance term σ^2 is unknown, and, even if one is willing to make such assumptions, it may yet be cumbersome to evaluate it. Thus, for statistical applications, an estimator of σ^2 is desired. For the sample mean, the problem of estimating the long run variance has already been studied by many authors. Our proposal for an estimate of σ^2 is based upon combining the HAC (heteroscedasticity and autocorrelation consistent) kernel estimator by de Jong and Davidson (2000) with an empirical version of the Hoeffding decomposition.

For the kernel g, we define the empirical version $\hat{g}_1$ of g_1 as

$$\hat{g}_1(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}, \boldsymbol{X}_i) - \frac{1}{n^2}\sum_{i,j=1}^{n} g(\boldsymbol{X}_i, \boldsymbol{X}_j),$$

and the empirical covariance for lag r as $\hat{\rho}(r) = \frac{1}{n}\sum_{i=1}^{n-r} \hat{g}_1(\boldsymbol{X}_i)\hat{g}_1(\boldsymbol{X}_{i+r})$. We then estimate σ^2 by

$$\hat{\sigma}_n^2 = \hat{\rho}(0) + 2\sum_{r=1}^{n-1} \kappa\left(\frac{r}{b_n}\right)\hat{\rho}(r), \tag{3}$$

where κ is a weight function (or HAC kernel function) and b_n a bandwidth depending on n. In order achieve consistency, κ and b_n have to fulfill some regularity conditions.

Assumption 2.6. The function $\kappa : [0,\infty) \to [-1,1]$ is continuous at 0 and at all but a finite number of points and $\kappa(0) = 1$. Furthermore, $|\kappa|$ is dominated by a nonincreasing, integrable function and

$$\int_0^\infty \left| \int_0^\infty \kappa(t)\cos(xt)dt \right| dx < \infty.$$

The bandwidth b_n satisfies $b_n \to \infty$ as $n \to \infty$ and $b_n/\sqrt{n} \to 0$.

Assumption 2.6 mainly coincides with Assumption 1 of de Jong and Davidson (2000). It is satisfied by a large class of kernels, in particular the popular Bartlett kernel $\kappa(t) = (1-|t|)\mathbf{1}_{\{|t|\le 1\}}$. The proof of the following consistency result is also given in Appendix C.

THEOREM 2.7. *Under Assumptions 2.2, 2.3, 2.4, and 2.6, we have* $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$ *as* $n \to \infty$, *where* σ^2 *is as in Theorem 2.5.*

Theorems 2.5 and 2.7 give rise to a general test statistic

$$\hat{T}_n = \frac{1}{2\hat{\sigma}_n} \max_{2\le k\le n-1} \frac{k}{\sqrt{n}} |U_k - U_n|$$

for CUSUM-type change-point tests based on U-statistics. By combining the continuous mapping theorem (applied to the functional which maps $x \in D[0,1]$ to the real number $\sup_{0\le t\le 1}|x(t) - tx(1)|$) and Slutsky's lemma we arrive at the following result.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

COROLLARY 2.8. *Under Assumptions 2.2, 2.3, 2.4, and 2.6, and if* $\sigma^2 > 0$, *we have* $\hat{T}_n \xrightarrow{d} \sup_{0\le\lambda\le 1}|B(\lambda)|$, *where* B *is a standard Brownian bridge on* $[0,1]$.

The assumptions of Theorems 2.5 and 2.7 are very broad and easy to verify. The kernel g as well as the marginal distribution F are, except for the variation condition, completely arbitrary. Furthermore, all time series models relevant in financial applications fulfill our short-range dependence condition with exponential decay of a_k and β_k. For further details on how P-NED is related to the usual L_2 near epoch dependence, see Appendix A.

Invariance principles for U-statistics similar to Theorem 2.5 were established by Yoshihara (1976) for absolutely regular processes, which do not cover many time series models. Central limit theorems for U-statistics have been investigated under more general conditions: Denker and Keller (1986) considered Lipschitz continuous functionals of absolutely regular sequences and Borovkova, Burton, and Dehling (2001) L_1-NED processes. As far as we know, functional central limit theorems (invariance principles) for U-statistics have not been studied under more general dependence conditions.

The potential applications of Corollary 2.8 are manifold. Several U-statistics have gained popularity as estimators that combine high efficiency under normality with appealing invariance and robustness properties (in the classical sense of robust statistics). The leading example is certainly Kendall's tau, which we will study in depth in Sections 3 through 7. Another prominent example is Gini's mean difference

$$g_n = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} |x_i - x_j|$$

for univariate data $x_1, \dots, x_n$. Similarly to Kendall's tau, Gini's mean difference is, as a measure of scale, almost as efficient as the maximum likelihood estimator at normality (the standard deviation), but is more efficient than the latter at heavy-tailed distributions and less sensitive to single outlying observations (Gerstenberger and Vogel, 2015). Thus Corollary 2.8 directly yields the asymptotic null distribution of a Gini's mean difference based change-point test for scale, which requires only $2+\delta$ moments as compared to $4+\delta$ moments for a sample-variance-based test. Assumption 2.4 is automatically fulfilled in this example as the corresponding kernel $g(x,y) = |x-y|$ is Lipschitz continuous.

Furthermore, the results can be straightforwardly extended to multivariate U-statistics. The process convergence of a p-dimensional U-statistic

$$\boldsymbol{U}_n = \left(U_n^{(1)}, \dots, U_n^{(p)}\right),$$

where p is generally different from the data dimension d, is obtained by means of the Cramér–Wold device by considering the univariate U-statistic $\tilde{U}_n = \sum_i \lambda_i U_n^{(i)}$ for arbitrary $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$. Similarly, a multivariate version of Theorem 2.7

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

follows from an entry-wise consideration as convergence in probability of a random matrix is implied by the convergence of its marginals. Besides vector-valued versions of Kendall's tau or Gini's mean difference to test for changes in the rank correlation or scale, respectively, of several time series jointly, we can also consider the following estimator,

$$\boldsymbol{K}_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{(\boldsymbol{X}_i - \boldsymbol{X}_j)(\boldsymbol{X}_i - \boldsymbol{X}_j)^\top}{|\boldsymbol{X}_i - \boldsymbol{X}_j|_2^2},$$

which is sometimes referred to as the *spatial Kendall's tau matrix*. It is also known to possess a rather high efficiency at normality, which has led several authors to consider this estimator in various contexts (e.g., Fan, Liu, and Wang, 2015). This estimator allows to consistently estimate the eigenvectors and the ordering of the eigenvalues of the covariance matrix. Thus Theorems 2.5 and 2.7 also provide the theoretical foundation for a robust change-point test for, say, detecting changes in the leading eigenvector of the marginal covariance matrix of a multivariate time series.

Further, these U-statistic results encompass all linear statistics, i.e., U-statistics of order one, such as the classical CUSUM test based on the sample mean. The classical change-point test for detecting changes in scale, as studied by Inclán and Tiao (1994), is essentially the CUSUM test applied to the squares of the centered data. It requires some additional technical effort to thoroughly deal with the centering for the data, which is ignored by some authors and dealt with in different ways by other authors. Employing the U-statistic representation of the sample variance with kernel $g(x, y) = (x - y)^2/2$, our results provide another elegant method of proof.

A referee raised the question whether the tail dependence coefficient could also be treated in the current U-statistic framework. In this situation, one would consider a U-statistic of order one with kernel

$$g(F_X(x_i), F_Y(y_i)) = \sqrt{\frac{n}{k}} \mathbf{1}_{\{F_X(x_i) \le k/n,\, F_Y(y_i) \le k/n\}} \tag{4}$$

for a bivariate time series $((X_i, Y_i))_{i \in \mathbb{Z}}$, where F_X and F_Y denote the marginal distributions of X_0 and Y_0, respectively, and $k = k(n)$ converges to ∞ and k/n converges to 0. Using this kernel in Theorem 2.5, one obtains the process $\mathbb{B}_n$ from Bücher, Jäschke, and Wied (2015), which provides the basis for a CUSUM-type test statistic for detecting changes in tail dependence. Thus one would need to extend the current setting to a situation where the kernel g may depend on n. More crucially, one would need a theorem in which boundedness of the second moments of the kernel is sufficient. While g^2 converges to the tail dependence coefficient, $g^{2+\delta}$ diverges unless the tail dependence coefficient is zero. This seems difficult to achieve in the current serial dependence setting. A solution could be to restrict the dependency conditions, e.g., to uniformly mixing processes. An alternative way of treating tail dependence is by employing the direct link between the tail

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

dependence coefficient and Kendall's tau in specific parametric models. By testing for constant Kendall's tau and assuming a specific model is true, one automatically tests for constant tail dependence as well. However, such an approach may be considered not optimal since, in contrast to an approach based on (4), it strongly relies on the model assumption, see the discussion in Bücher et al. (2015).

3. CHANGE-POINT DETECTION FOR KENDALL'S TAU

Let $((X_i, Y_i))_{i\in\mathbb{Z}}$ be a strictly stationary series of bivariate random vectors with marginal distribution function $F(x, y) = P(X_0 \le x, Y_0 \le y)$. Throughout the remainder of the article, we assume F to be Lipschitz continuous. This is fulfilled if F possesses a bounded density, but, e.g., $X_0 = Y_0$ almost surely is also allowed. Kendall's rank correlation coefficient, also known as Kendall's tau, is defined as

$$\tau = P((X' - X)(Y' - Y) > 0) - P((X' - X)(Y' - Y) < 0),$$

where (X, Y) and (X', Y') are two independent random variables with distribution function F. Kendall's tau is a measure of correlation between the random variables X and Y, where we understand correlation generally as *monotone dependence*, which, loosely speaking, carries information on to what extent one variable on average increases or decreases as the value of the other increases. Kendall's tau, as well as the related dependence measure Spearman's rho, is a function of the copula only. In particular, it does not depend on the marginal distributions and is hence invariant to monotone marginal transformations, see Nelsen (e.g., 2006, Chap. 5). Consequently, its sample version only depends on the ranks of the data, which is the reason for Kendall's tau being also referred to as *rank correlation measure*. The sample version of Kendall's tau is defined as

$$\hat{\tau}_n = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \left[\mathbf{1}_{(0,\infty)}\left\{(X_j - X_i)(Y_j - Y_i)\right\} - \mathbf{1}_{(-\infty,0)}\left\{(X_j - X_i)(Y_j - Y_i)\right\}\right],$$

which is a U-statistic with kernel function $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by

$$g\left((x, y), (x', y')\right) = \mathbf{1}_{(0,\infty)}\{(x' - x)(y' - y)\} - \mathbf{1}_{(-\infty,0)}\{(x' - x)(y' - y)\}. \qquad \textbf{(5)}$$

For dependent data, $\hat{\tau}_n$ is not necessarily unbiased, but under the weak dependence conditions we consider, it is (strongly) consistent for τ. We will study a test for change in the dependence structure of the marginals by the test statistic

$$\hat{T}_{\tau,n} = \max_{k=1,\dots,n} \frac{k}{\sqrt{n}} |\hat{\tau}_k - \hat{\tau}_n|,$$

which rejects the null hypothesis of constant rank correlation if $\hat{T}_{\tau,n}$ is too large. By means of the functional limit theorem for sequential U-statistics (Theorem 2.5), we know that

$$\left(\frac{[tn]}{\sqrt{n}}(\hat{\tau}_{[tn]} - \tau)\right)_{t\in[0,1]}$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

converges under the assumption of no change and for weakly dependent series weakly to a Brownian motion $2\sigma_\tau W$, where

$$\sigma_\tau^2 = \sum_{j=-\infty}^{\infty} E\left[\psi(X_0, Y_0)\psi(X_j, Y_j)\right] \tag{6}$$

with $\psi(x, y) = 4F(x, y) - 2F_X(x) - 2F_Y(y) + 1 - \tau$. The function ψ can be seen to be the linear part g_1 of the Hoeffding decomposition (1) for the specific kernel g in (5) for any continuous distribution function F. By the continuous mapping theorem, we have the weak convergence

$$\hat{T}_{\tau,n} = \max_{k=1,\ldots,n} \frac{k}{\sqrt{n}} |\hat{\tau}_k - \hat{\tau}_n| \xrightarrow{d} 2\sigma_\tau \sup_{0\le\lambda\le 1} |B(\lambda)|,$$

where B is a standard Brownian Bridge. The distribution of $\sup_{0\le\lambda\le 1}|B(\lambda)|$ is known and sometimes referred to as Kolmogorov distribution. We use the estimator for the long run variance proposed in Section 2. Let F_n, $F_{X,n}$, and $F_{Y,n}$ denote the empirical distribution functions of $((X_i, Y_i))_{i=1,\ldots,n}$, $(X_i)_{i=1,\ldots,n}$, and $(Y_i)_{i=1,\ldots,n}$, respectively. Then $\hat{\psi}_{n,i} = 4F_n(X_i, Y_i) - 2F_{X,n}(X_i) - 2F_{Y,n}(Y_i) + 1 - \hat{\tau}_n$ can be seen to equal $\hat{g}_1((X_i, Y_i))$, and the variance estimator (3) can be written as

$$\hat{\sigma}_{\tau,n}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_{n,i}^2 + \frac{2}{n}\sum_{j=1}^{n-1} \kappa\left(\frac{j}{b_n}\right) \sum_{i=1}^{n-j} \hat{\psi}_{n,i}\hat{\psi}_{n,i+j}, \tag{7}$$

where κ and b_n are the same as in Section 2. Corollary 3.1 below gives the asymptotic distribution of the test statistic $\hat{T}_n/(2\hat{\sigma}_{\tau,n})$ under the null hypothesis of no change.

COROLLARY 3.1. *Let $((X_i, Y_i))_{i\in\mathbb{Z}}$ be a two-dimensional, stationary process with a Lipschitz continuous marginal distribution function. Assume that $((X_i, Y_i))_{i\in\mathbb{Z}}$ is P-NED with approximating constants $(a_k)_{k\ge 1}$ on an absolutely regular process with absolute regularity coefficients $(\beta_k)_{k\ge 1}$ satisfying Assumption 2.2 for some $\delta > 0$. Let further Assumption 2.6 hold. Then, if $\sigma_\tau^2 > 0$,*

$$\frac{\hat{T}_{\tau,n}}{2\hat{\sigma}_{\tau,n}} \xrightarrow{d} \sup_{0\le\lambda\le 1} |B(\lambda)|, \tag{8}$$

where $(B(\lambda))_{0\le\lambda\le 1}$ is, as before, a standard Brownian bridge.

4. CHANGE-POINT ESTIMATION AND LOCAL POWER

If the test rejects the null hypothesis of constant correlation, and if it is furthermore reasonable to assume that there is one sudden change-point, it is of interest to locate this change-point. An intuitive estimator, which is common when

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

dealing with CUSUM-type change-point tests, is the position at which the weighted differences take their maximum, that is

$$\hat{k}_n = \operatorname*{arg\,max}_{1 \le k \le n} \frac{k}{\sqrt{n}} |\hat{\tau}_k - \hat{\tau}_n|.$$

We will show in the following that this is indeed a reasonable estimator. We will assume that the following model holds.

Model 4.1 (Change-point model)**.** Let $0 < \lambda^* < 1$. For $n \in \mathbb{N}$ let $((X_i^{(n)}, Y_i^{(n)}))_{1 \le i \le [\lambda^* n]}$ and $((X_i^{(n)}, Y_i^{(n)}))_{[\lambda^* n]+1 \le i \le n}$ be two bivariate, stationary stochastic processes with marginal distribution functions F and G, respectively. Let furthermore $((X_i^{(n)}, Y_i^{(n)}))_{1 \le i \le n}$ be P-near epoch dependent[1] on an absolutely regular process with coefficients satisfying Assumption 2.2 uniformly for all n.

The goal is to estimate λ^*. Let τ_F and τ_G denote Kendall's tau of F and G, respectively. Moreover, let $\tau_{F,G} = Eg((X_1, Y_1), (X_2, Y_2))$, cf. (5), where $(X_1, Y_1) \sim F$ and $(X_2, Y_2) \sim G$ are independent.

THEOREM 4.2. *If* $((X_i^{(n)}, Y_i^{(n)}))_{1 \le i \le n, n \in \mathbb{N}}$ *follows Model 4.1, furthermore* $\tau_F \neq \tau_G$ *and*

$$\frac{(1-\lambda^*)^2}{2\left((1-\lambda^*)^2 + \lambda^*\right)} \le \frac{\tau_{F,G} - \tau_F}{\tau_G - \tau_F} < 1, \tag{9}$$

then $\hat{k}_n / n \xrightarrow{p} \lambda^*$ *as* $n \to \infty$.

Condition (9) prohibits $\tau_{F,G}$ to be too close to τ_F compared to τ_G. It is an open research question which values of $\tau_{F,G}$ are possible for given τ_F and τ_G, and in particular if $\tau_{F,G}$ may at all lie outside the interval $[\tau_F, \tau_G]$.

We further analyze the power of the proposed change-point test under local alternatives. We give a formula for the asymptotic distribution of the test statistic within a specific class of local alternatives.

Model 4.3. Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a strictly stationary, bivariate process that satisfies Assumption 2.2, and let $0 < \lambda^* < 1$ and $\Delta > 0$. We define for any integer $n \ge 1$

$$\left(X_i^{(n)}, Y_i^{(n)}\right) = \begin{cases} (X_i, Y_i) & 1 \le i \le [n\lambda^*], \\ \left(X_i + \frac{\Delta}{\sqrt{n}} Y_i, Y_i\right) & [n\lambda^*] \le i \le n. \end{cases}$$

We assume that (X_i, Y_i) has an absolutely continuous distribution F with bounded density f satisfying $\lim_{x \to \infty} f(x, y) = 0$ and $\lim_{y \to \infty} f(x, y) = 0$ for all $x, y \in \mathbb{R}$ and

$$\int \int \sup_{0 \le |\alpha| \le \varepsilon} |y \frac{\partial}{\partial x} f(x + \alpha y, y)| dx dy < \infty$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

for some $\varepsilon > 0$. By F_n we denote the distribution of $(X_i^{(n)}, Y_i^{(n)})$. The density of F_n is given by

$$f_n(x, y) = f\left(x - \frac{\Delta}{\sqrt{n}} y, y\right).$$

THEOREM 4.4. *If $(X_i, Y_i)_{i \geq 1}$ follows Model 4.3, we have*

$$\hat{T}_{\tau,n} \xrightarrow{d} \sup_{0 \leq \lambda \leq 1} |2\sigma_\tau B(\lambda) + \Delta(\phi_{\lambda^*}(\lambda) - \lambda \phi_{\lambda^*}(1))|,$$

where σ_τ is defined in (6) and the function $\phi_{\lambda^*} : [0, 1] \to \mathbb{R}$ is defined as

$$\phi_{\lambda^*}(s) = \begin{cases} 4\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y(2F(y|x) - 1)F(dy|x)\right) f_X^2(x)dx\right)(s - \lambda^*) & s \geq \lambda^*, \\ 0 & s \leq \lambda^*. \end{cases}$$

From Theorem 4.4 we can conclude the consistency of the test against local alternatives of the type studied in Model 4.3, for which it suffices to observe that the integral occurring in the limit in Theorem 4.4 is nonzero: For any absolutely continuous distribution function F with density f, and with finite expectation, we have

$$\begin{aligned} \int_{-\infty}^{\infty} y(2F(y) - 1)f(y)dy &= \int_{-\infty}^{\infty} 2yF(y)f(y)dy - \int_{-\infty}^{\infty} yf(y)dy \\ &= \int ydF^2(y) - \int ydF(y) \;=\; E(\max(Y_1, Y_2)) - E(Y_1), \end{aligned}$$

where Y_1, Y_2 are independent random variables with distribution F. Now, $E(\max(Y_1, Y_2)) - E(Y_1) > 0$, unless Y_1 is a constant, in which case $E(\max(Y_1, Y_2)) - E(Y_1) = 0$. Thus, the inner integral $\int_{-\infty}^{\infty} y(2F(y|x) - 1)F(dy|x)$ is positive, unless the conditional distribution of Y_1 given $X_1 = x$ is degenerate, i.e., Y_1 takes only one value. Hence the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y(2F(y|x) - 1)F(dy|x)f_X^2(x)dx$$

is positive unless Y_1 is a deterministic function of X_1.

5. PREVIOUS PROPOSALS

Wied et al. (2012) considered the test statistic

$$\hat{T}_{\varrho,n} = \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} |\hat{\varrho}_k - \hat{\varrho}_n|$$

based on Pearson's linear correlation coefficient $\hat{\varrho}_n$. This test will serve as the main benchmark method for our test. Our motivation for using Kendall's tau is the wish to efficiently detect structural changes in arbitrarily heavy-tailed and

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

potentially contaminated data. Both tests are constructed in a similar way. The differences between the two tests are largely due to the different properties of the estimators $\hat{\varrho}_k$ and $\hat{\tau}_k$. It is therefore worthwhile to have a brief look at these two correlation measures.

Kendall's tau is invariant with respect to strictly monotonic, componentwise transformations of F. For continuous F, it can be written as $\tau = 4EF(X,Y) - 1$, which can be seen to be a function of the copula of F alone. The same applies to its asymptotic variance at i.i.d. data, $ASV(\hat{\tau}_n) = 16\mathrm{Var}(\psi(X,Y))$ with ψ being defined in Section 3. For further details see, e.g., Nelsen (2006, Chap. 5). Consequently, no matter how heavy the tails of the distribution F are, as long as the marginals are joined by a Gauss copula, the variance of $\hat{\tau}_n$ is the same as in the normal model.

In the normal model there is a one-to-one correspondence between τ and Pearson's moment correlation ϱ, which is given by

$$\tau = (2/\pi)\arcsin(\varrho), \quad -1 \le \varrho \le 1. \tag{10}$$

Thus by letting $\hat{\varrho}_{\tau,n} = \sin(\pi(\hat{\tau}_n - 1/2))$, the estimators $\hat{\varrho}_n$ and $\hat{\varrho}_{\tau,n}$ are, under normality, both Fisher-consistent for the same quantity ϱ. Comparing their asymptotic variances

$$ASV(\hat{\varrho}_n) = (1-\varrho^2)^2, \quad ASV(\hat{\varrho}_{\tau,n}) = (1-\varrho^2)(\pi^2/9 - 4\arcsin^2(\varrho/2)) \tag{11}$$

(e.g., Croux and Dehon, 2010), allows a prognosis concerning the efficiency relation of the corresponding change-point tests. For example, for two independent random variables we have $ASV(\hat{\varrho}_n) = 1$ and $ASV(\hat{\varrho}_{\tau,n}) = \pi^2/9 = 1.097$.

The scope of the identity (10) extends to all elliptical distributions with finite second moments. (e.g., McNeil, Frey, and Embrechts, 2005, p. 97). For example, at the two-dimensional elliptical t_ν-distribution with ν degrees of freedom and uncorrelated margins, we have

$$ASV(\hat{\varrho}_n) = (\nu-2)/(\nu-4), \quad \nu > 4,$$

$$ASV(\hat{\varrho}_{t(\nu),n}) = (\nu+4)/(\nu+2), \quad \nu > 0,$$

where $\hat{\varrho}_{t(\nu),n}$ is the maximum likelihood estimator of ϱ at the two-dimensional t_ν family. (e.g., Bilodeau and Brenner, 1999, p. 221). Dengler (2010) gives values for the asymptotic variance of $\hat{\varrho}_{\tau,n}$ at uncorrelated t_ν-distributions. It is a decreasing function of ν, it equals 1.922 and 1.296 for $\nu = 1$ and $\nu = 5$, respectively, and is smaller than $ASV(\hat{\varrho}_n)$ for $\nu \le 16$. We further take note of the remarkable fact that for all uncorrelated t- and normal distributions, the asymptotic relative efficiency of Kendall's tau with respect to the respective MLE is above 90% for $\nu \ge 2$. It is more than 99% at an uncorrelated t_5 distribution.

The other popular nonparametric correlation measure, Spearman's rho, which is often considered alongside Kendall's tau, is defined as Pearson's linear correlation of the ranks of the data. It can be written as

$$\hat{r}_n = \frac{12}{(n-1)n(n+1)} \sum_{i=1}^{n} R_n(X_i) R_n(Y_i) \; - \; 3\frac{n+1}{n-1},$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

where $R_n(X_i)$ denotes the rank of the ith observation X_i among $X_1, \ldots, X_n$, likewise $R_n(Y_i)$. The population version of Spearman's rho, $s = 12E(F_X(X)F_Y(Y)) - 3$, is also a function of the copula. Generally, Kendall's tau and Spearman's rho have similar statistical properties. See, e.g., Nelsen (2006, Chap. 5) for details on their relationship. Croux and Dehon (2010) compare both with respect to robustness and efficiency and arrive at the conclusion, that in both respects their performance is comparable, but Kendall's tau is slightly favorable. Wied, Dehling, van Kampen, and Vogel (2014) propose a nonparametric, robust change-point test for constant correlation for strongly mixing sequences that is closely related to Spearman's rho. They consider the test statistic

$$\hat{T}_{s,n} = \max_{1\le k\le n} \frac{k}{\sqrt{n}} |\hat{s}_k - \hat{s}_n|, \quad \text{where} \quad \hat{s}_k = 12n^{-3} \sum_{i=1}^{k} R_n(X_i) R_n(Y_i) - 3 - 12/n, \quad 1 \le k \le n, \tag{12}$$

along with a suitable long-run variance estimator in the same vein as $\hat{\sigma}^2_{\tau,n}$ in (7) above. The proof of its convergence is based on an invariance principle for the multivariate sequential empirical process. Despite the mentioned practical parity of Kendall's tau and Spearman's rho, this test has a low efficiency compared to our Kendall's tau based test (see Section 6). The reason lies in the usage of $R_n(\cdot)$ instead of $R_k(\cdot)$ in (12). Recently, Kojadinovic, Quessy, and Rohmer (2016) proposed a new constancy test for Spearman's rho which fixes this drawback and has therefore considerably higher power. Spearman's rho is asymptotically equivalent to a U-statistic of order 3, and an asymptotic analysis of the related test statistic

$$\hat{T}_{r,n} = \max_{1\le k\le n} \frac{k}{\sqrt{n}} |\hat{r}_k - \hat{r}_n|$$

by means of U-statistics theory is mathematically much more involved. Since Spearman's rho, on the other hand, exhibits no pronounced advantage over Kendall's tau, we do not pursue this further here. Finally, we note that both estimators require a comparable computing effort. Both can be computed in $O(n \log n)$ time. Simple algorithms to compute the test statistics require $O(n^2)$.

6. SIMULATION RESULTS

In this section we give some numerical results, comparing the performance of the test to the previous proposals by Wied et al. (2012) and Wied et al. (2014). These are based on the test statistics $\hat{T}_{\varrho,n}$ and $\hat{T}_{s,n}$, respectively, (cf. Section 5) and referred to as *Pearson test* and *Spearman test* in the following. We call the new proposal *Kendall test*. Moreover, we consider the *improved Spearman test* by Kojadinovic et al. (2016).

Throughout, we estimate the long-run variance σ^2_τ of the Kendall test by the estimator $\hat{\sigma}^2_{\tau,n}$ given by (7), where we choose the bandwidth $b_n = \lfloor 2n^{1/3} \rfloor$ and κ to be the quartic kernel

$$\kappa(x) = (1 - x^2)^2 \mathbf{1}_{[-1,1]}(x).$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Altogether we found neither the choice of the kernel nor the bandwidth to be very critical. Generally, choosing the bandwidth too small in strong dependence scenario tends to have a bigger impact than choosing it too large in the case of little or no dependence. We recommend the quartic kernel since it is smooth and has more of a flat-top-like behavior than the popular Bartlett kernel, giving more weight to small-lag autocorrelations. The variance estimation for $\hat{T}_{\varrho,n}$ and $\hat{T}_{s,n}$ is done according to the authors' proposals, which both also implement kernel estimators following de Jong and Davidson (2000). For the improved Spearman test by Kojadinovic et al. (2016), we use implementation in the R-package npcp (Kojadinovic, 2015) with the Bartlett kernel for the long-run variance estimation.

For all tests and all dependence scenarios, we take the bandwidth $b_n = \lfloor 2n^{1/3} \rfloor$. This is a relatively large bandwidth, which is suitable for strong dependence as in our AR(1) example below. A smaller bandwidth would be more appropriate for no or little serial dependence. The problem of selecting an optimal, data-adaptive bandwidth is a ubiquitous one. It affects the long-run variance estimation of all change-point tests in a similar way and is analogous to the optimal blocklength selection problem for subsampling or bootstrapping, which may also be employed to obtain critical values. We do not discuss the problem of automatic bandwidth selection in the present article. Our concern is how the use of different estimators affects the behavior of the change-point tests. The methods under consideration use the same technique for long-run variance estimation and, hence, choosing the same bandwidth in all cases allows a fair comparison of the methods. However, any bandwidth selection procedure previously proposed, see, e.g., Kojadinovic et al. (2016, Section 3.3) and the references therein, can also be put to use for the Kendall test and is likely to further improve the power of the test as compared to the results presented below.

We consider three data models that implement three different types of serial dependence: independent observations, a multivariate AR(1) process and a constant conditional correlation (CCC) GARCH(1,1) process, as considered, e.g., in Aue et al. (2009). All three models are based on a sequence of independent and identically distributed innovations $(\delta_i, \varepsilon_i)$, $i \in \mathbb{Z}$, having a bivariate, centered elliptical distribution $\mathcal{E}_2(\mathbf{0}, S)$ with

$$S = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix},$$

where the parameter $|\varrho| \leq 1$ is equal to the usual moment correlation if the second moments of $(\delta_1, \varepsilon_1)$ are finite.

Model 6.1. $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix}, \quad i \in \mathbb{Z}.$

Model 6.2. The series $((X_i, Y_i))_{i \in \mathbb{Z}}$ follows the AR(1) process

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = 0.8 \begin{pmatrix} X_{i-1} \\ Y_{i-1} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix}, \quad i \in \mathbb{Z},$$

with AR parameter $\varphi = 0.8$.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Model 6.3. The series $((X_i, Y_i))_{i\in\mathbb{Z}}$ follows the CCC-GARCH(1,1) process

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} \sigma_i \\ \tau_i \end{pmatrix} \circ \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \text{ with } \begin{pmatrix} \sigma_i^2 \\ \tau_i^2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} \circ \begin{pmatrix} X_{i-1}^2 \\ Y_{i-1}^2 \end{pmatrix} + \begin{pmatrix} 0.84 \\ 0.84 \end{pmatrix} \circ \begin{pmatrix} \sigma_{i-1}^2 \\ \tau_{i-1}^2 \end{pmatrix}, i \in \mathbb{Z},$$

where $\circ$ denotes the Hadamard product, i.e., component-wise vector multiplication.

Most simulation results below are for $n = 500$. At this sample size, the distribution of the test statistic $\hat{T}_{\tau,n}/(2\hat{\sigma}_{\tau,n})$ is well approximated by its limit distribution under the null hypothesis in all dependence scenarios considered. For the first half of the data, we sample independent realizations $(\delta_i, \varepsilon_i)$, $i = 1, \ldots, 250$, with correlation parameter $\varrho_1 = 0.4$. For the second half of the data, we use the correlation parameter ϱ_2, for which we allow the values 0.4 (null hypothesis), 0.6, 0.8, 0.2, 0, -0.2, -0.4. Thus in Model 6.1, we have a constant correlation of 0.4 at the beginning and then a sudden jump, whereas in Models 6.2 and 6.3, there is a gradual but quick change in the correlation of the observed process (X_i, Y_i). Also note that, under the null, the data process $((X_i, Y_i))_{i\in\mathbb{Z}}$ has the same marginal correlation as the innovation process $((\delta_i, \varepsilon_i))_{i\in\mathbb{Z}}$ in Models 6.1 and 6.2, but generally not in Model 6.3.

In Models 6.1 and 6.2, we consider five different elliptical distributions for $(\delta_i, \varepsilon_i)$: the bivariate normal distribution and bivariate t_ν-distribution with $\nu = 20, 5, 3, 1$. The parameter $\nu > 0$ is called the *degrees of freedom* or the *tail index*. The t_{20} distribution has slightly heavier tails than the normal, whereas t_5, t_3, and t_1 serve as examples of very heavy-tailed distributions. The t_ν distribution possesses finite moments of order α for any $\alpha < \nu$. Thus, the Pearson test by Wied et al. (2012), which requires finite fourth moments, does not work for $\nu = 3$ and $\nu = 1$. From an economic point of view, however, the tail indices 3 and 1 are most interesting. There is evidence that financial returns on many stocks, stock indices and foreign exchange rates in developed economies typically have tail indices ν in the interval $\nu \in (2, 4)$ and thus have finite variances but infinite fourth moments. It has emerged that $\nu = 3$ is an appropriate choice for financial returns in developed markets and Ibragimov, Ibragimov, and Kattuman (2013) provide empirical evidence that ν may be even smaller than 2 for foreign exchange rates in emerging economies.

In Model 6.3, we used the normal distribution and t_ν distributions with $\nu = 20, 8, 5$. The CCC-GARCH model generates heavy tails also for normal innovations. For t_ν innovations with $\nu = 3, 1$, the process explodes. For $\nu = 5$, the CCC-GARCH process of Model 6.3 generates very pronounced volatility clusters.

For each combination of model, jump height and marginal distribution we generate 1,000 samples and compute the three test statistics from each sample. The observed rejection frequencies at the significance level 0.05 for sample size $n = 500$ are given in Tables 1, 2, and 3 for Models 6.1, 6.2, and 6.3, respectively. At Table 1 (independence scenario) we note the following.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

TABLE 1. Efficiency comparison of several correlation change-point tests under Model 6.1. Different marginal distributions, 500 observations, different jump sizes in the middle of the sample. Empirical rejection frequencies at the asymptotic 0.05 level based on 1,000 repetitions

	Change at $n/2$:	None	−0.2	+0.2	−0.4	+0.4	−0.6	−0.8
Distribution	Test							
Normal	Pearson	0.04	0.46	0.70	0.97	1.00	1.00	1.00
	Spearman	0.04	0.06	0.07	0.22	0.20	0.47	0.78
	Improved Spearman	0.03	0.43	0.58	0.95	1.00	1.00	1.00
	Kendall	0.05	0.44	0.65	0.96	1.00	1.00	1.00
t_{20}	Pearson	0.04	0.42	0.65	0.97	1.00	1.00	1.00
	Spearman	0.03	0.07	0.08	0.22	0.20	0.47	0.77
	Improved Spearman	0.02	0.42	0.56	0.94	1.00	1.00	1.00
	Kendall	0.04	0.46	0.63	0.97	1.00	1.00	1.00
t_5	Pearson	0.04	0.24	0.41	0.73	0.95	0.95	0.98
	Spearman	0.04	0.08	0.08	0.22	0.20	0.46	0.76
	Improved Spearman	0.03	0.38	0.50	0.91	1.00	1.00	1.00
	Kendall	0.04	0.41	0.55	0.95	1.00	1.00	1.00
t_3	Pearson	0.06	0.14	0.25	0.39	0.69	0.64	0.79
	Spearman	0.03	0.08	0.08	0.21	0.18	0.43	0.72
	Improved Spearman	0.04	0.32	0.46	0.88	1.00	1.00	1.00
	Kendall	0.03	0.39	0.52	0.91	1.00	1.00	1.00
t_1	Pearson	0.47	0.48	0.50	0.49	0.56	0.52	0.51
	Spearman	0.03	0.06	0.07	0.17	0.17	0.38	0.63
	Improved Spearman	0.03	0.25	0.33	0.74	0.96	0.98	1.00
	Kendall	0.04	0.29	0.38	0.83	0.98	1.00	1.00

(1) The Pearson test is slightly better than the Kendall test for the normal distribution. Both tests lose power with increasing tails, but the loss is much smaller for the Kendall test. For the t_{20} distribution, the results are comparable. The Kendall test is clearly better for heavier tails. These observations are fully in line with our expectations considering the efficiency comparison of the respective correlation measures in Section 5.
(2) Throughout, the naive Spearman test has a very low power, which is not true for the improved Spearman test. The improved Spearman test and Kendall test show comparable results under normality as well as heavy tails, with advantages for the Kendall test. This is in line with the efficiency comparison of both estimators by Croux and Dehon (2010).
(3) For the t_3 distribution, the Pearson test yields still approximate results, whereas for the t_1 distribution it is completely useless.

Analyzing Table 2 (AR scenario) we find that

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

TABLE 2. Efficiency comparison of several correlation change-point tests. $((X_i, Y_i))_{i=1,\ldots,n}$ AR(1) process with AR-parameter $\varphi = 0.8$. Different innovation distributions, 500 observations, several alternatives. Empirical rejection frequencies at the asymptotic 0.05 level based on 1,000 repetitions

	Change at $n/2$:	None	−0.2	+0.2	−0.4	+0.4	−0.6	−0.8
Distribution	Test							
Normal	Pearson	0.07	0.13	0.27	0.45	0.77	0.78	0.94
	Spearman	0.05	0.06	0.04	0.07	0.07	0.11	0.18
	Improved Spearman	0.03	0.11	0.15	0.39	0.73	0.73	0.95
	Kendall	0.05	0.14	0.19	0.46	0.73	0.79	0.96
t_{20}	Pearson	0.06	0.11	0.31	0.41	0.77	0.77	0.94
	Spearman	0.05	0.06	0.06	0.08	0.08	0.12	0.20
	Improved Spearman	0.03	0.12	0.15	0.36	0.70	0.70	0.95
	Kendall	0.04	0.13	0.23	0.42	0.74	0.79	0.95
t_5	Pearson	0.08	0.12	0.26	0.34	0.67	0.67	0.89
	Spearman	0.06	0.05	0.06	0.08	0.09	0.12	0.18
	Improved Spearman	0.03	0.13	0.15	0.35	0.67	0.68	0.90
	Kendall	0.05	0.15	0.19	0.40	0.67	0.73	0.95
t_3	Pearson	0.10	0.11	0.26	0.25	0.56	0.50	0.67
	Spearman	0.06	0.08	0.06	0.08	0.09	0.09	0.17
	Improved Spearman	0.05	0.09	0.12	0.28	0.58	0.58	0.84
	Kendall	0.05	0.12	0.18	0.34	0.62	0.67	0.90
t_1	Pearson	0.46	0.49	0.52	0.50	0.56	0.50	0.54
	Spearman	0.08	0.07	0.08	0.09	0.11	0.12	0.14
	Improved Spearman	0.03	0.05	0.06	0.13	0.20	0.24	0.41
	Kendall	0.06	0.12	0.11	0.18	0.34	0.34	0.53

(4) the power of all tests is lower for the AR(1) process than in the independent case, and
(5) the observations made at Table 1 concerning the comparison of the tests generally also apply here. The Kendall test is slightly better than the improved Spearman test.
(6) For normal, t_{20}, t_5, and t_3 innovations, the performance of the Kendall, the improved Spearman and the Pearson test are rather similar. The effect of the heavy tails is less pronounced than in the independent case. This is not entirely surprising. In Model 6.2, the marginal distribution of the process can be expressed as a sum of independent random variables. Although it is generally not normal, it is, purely heuristically speaking, closer to a normal distribution than the innovations (if these possess finite second moments).

At Table 3 (CCC-GARCH scenario) we observe that

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

TABLE 3. Efficiency comparison of several correlation change-point tests. $((X_i, Y_i))_{i=1,\ldots,n}$ multivariate GARCH process. Different innovation distributions, 500 observations, several alternatives. Empirical rejection frequencies at the asymptotic 0.05 level based on 1,000 repetitions

	Change at $n/2$:	None	−0.2	+0.2	−0.4	+0.4	−0.6	−0.8
Distribution	Test							
Normal	Pearson	0.06	0.47	0.63	0.94	0.99	1.00	1.00
	Spearman	0.04	0.07	0.09	0.22	0.17	0.47	0.76
	Improved Spearman	0.03	0.43	0.56	0.96	1.00	1.00	1.00
	Kendall	0.04	0.45	0.59	0.96	1.00	1.00	1.00
t_{20}	Pearson	0.07	0.46	0.55	0.90	0.95	0.98	1.00
	Spearman	0.04	0.07	0.07	0.22	0.19	0.46	0.78
	Improved Spearman	0.02	0.41	0.51	0.94	1.00	1.00	1.00
	Kendall	0.05	0.46	0.57	0.96	1.00	1.00	1.00
t_8	Pearson	0.13	0.42	0.42	0.76	0.78	0.90	0.94
	Spearman	0.05	0.07	0.08	0.21	0.19	0.42	0.72
	Improved Spearman	0.03	0.38	0.50	0.90	1.00	1.00	1.00
	Kendall	0.04	0.45	0.52	0.94	0.99	1.00	1.00
t_5	Pearson	0.30	0.44	0.36	0.66	0.62	0.75	0.81
	Spearman	0.06	0.09	0.10	0.20	0.23	0.41	0.69
	Improved Spearman	0.03	0.34	0.39	0.88	0.99	1.00	1.00
	Kendall	0.07	0.40	0.41	0.89	0.94	0.99	0.99

(7) the efficiencies of the test are comparable to the independence case. GARCH processes are white noise with zero autocorrelations. Serial dependence in the second-order characteristics appear to influence the tests less.
(8) The Pearson test has difficulties keeping the size. A size-adjusted power comparison shows a strict superiority of the Kendall test over the Pearson test.

To give an impression of the power of the test in other data situations, we also include some limited simulation results for sample sizes $n = 250, 500, 1,000$ and change locations ranging from $n/8$ to $n/2$. The data generating process is CCC-GARCH (Model 6.3) with normal innovations and $\rho_1 = 0.4$ and $\rho_2 = 0$. The parameters for the long-run variance estimation are as before. The results are summarized in Table 4. We find the general picture mediated by Table 3 concerning the comparison of the tests confirmed. The power of all change-point tests is lower for changes occurring nearer to either end of the data sequence.

Altogether the simulation results are favorable for the Kendall test. It is clearly superior to the Pearson test at heavy-tailed data and qualitatively noninferior under normality. It is noninferior to the improved Spearman test in all scenarios considered. Kojadinovic et al. (2016) also propose a bootstrapping procedure to obtain critical values for the improved Spearman test, which their simulation

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

TABLE 4. Efficiency comparison of several correlation change-point tests at Model 6.3 (multivariate GARCH). Innovation correlation changes from $\varrho_1 = 0.4$ to $\varrho_2 = 0$, normal innovations. Different sample sizes, different jump locations. Empirical rejection frequencies at the asymptotic 0.05 level based on 1,000 repetitions

	Change at :	None	$n/8$	$n/4$	$3n/8$	$n/2$
Sample size	Test					
$n = 250$	Pearson	0.06	0.12	0.35	0.59	0.64
	Spearman	0.03	0.04	0.07	0.10	0.10
	Improved Spearman	0.02	0.04	0.20	0.47	0.58
	Kendall	0.04	0.11	0.38	0.60	0.69
$n = 500$	Pearson	0.06	0.22	0.76	0.92	0.94
	Spearman	0.04	0.06	0.11	0.15	0.22
	Improved Spearman	0.03	0.12	0.68	0.91	0.96
	Kendall	0.04	0.22	0.79	0.94	0.96
$n = 1,000$	Pearson	0.06	0.55	0.99	1.00	1.00
	Spearman	0.05	0.10	0.26	0.40	0.43
	Improved Spearman	0.03	0.44	0.99	1.00	1.00
	Kendall	0.04	0.56	1.00	1.00	1.00

results suggest improves the power slightly. A similar procedure applicable to the Kendall test has been proposed recently by Bücher and Kojadinovic (2016).

The above data generating processes are all based on elliptical innovations. Any change in the generalized correlation coefficient ϱ constitutes a shift of the same magnitude for all dependence measures considered, and the corresponding tests generally behave similarly. However, in data situations where a change occurs that affects the various dependence measures to a different degree, the tests may behave qualitatively differently. A class of examples where a change occurs in the Pearson correlation but neither in Kendall's tau nor Spearman's rho is generated by leaving the copula constant but changing the marginal distributions. Similarly, one may change the copula in such a way that Kendall's tau or Spearman's rho remain fixed. (A change-point test for the whole copula has recently been proposed by Bücher, Kojadinovic, Rohmer, and Segers (2014)). To illustrate this point, consider the following example: Suppose in the first half of the sequence that the observations are i.i.d., following the distribution of

$$\begin{pmatrix} X \\ Y \end{pmatrix} = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}\right).$$

In the second half, the observations are i.i.d. with distribution of $(X, Y^3|Y|/\sqrt{105})$. Thus, before and after the change, the data have mean zero, marginal variances one, and a Kendall's tau coefficient of $\tau = 2\pi^{-1}\arcsin(0.6) \approx 0.41$. However, the Pearson correlation changes from 0.6 to $\sqrt{2/(105\pi)}24/5 \approx 0.37$.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

TABLE 5. Empirical rejection frequencies at the asymptotic 0.05 level based on 1,000 repetitions; Sample size $n = 500$

	Kendall test	Pearson test
Constant copula example	0.23	0.94
Constant correlation example	1.00	0.07

Consequently, the Kendall test has no power against this change whereas the Pearson test does. If the distribution of the observations changes from that of (X, Y) in the first half to that of (U, V) in the second half, where $U = 2\sqrt{3}(\tilde{U} - 1/2)$ and $V = 42\sqrt{3}/13(\tilde{U}^{13} - 1/14)$ for $\tilde{U} \sim U[0, 1]$, then, as before, mean and marginal variances remain the same, also the Pearson correlation remains at 0.6, but the Kendall's tau coefficient changes from 0.41 to 1. In this case, the Pearson test has no power, whereas the Kendall test has. In Table 5, we give simulated rejection frequencies (at the 0.05 level) for the two examples for $n = 500$ with the settings for the long-run variance estimation as before. Generally, the differences in the dependence measures tend to be of comparable size in realistic data models and not as pronounced as in the artificial examples above.

7. DATA EXAMPLES

Wied et al. (2012) analyze the dependence between the German stock index (DAX) and the Standard and Poor's 500 (S&P 500). We apply the Kendall test and the Pearson test (with the same parameter choices for the variance estimation as in the simulation section) to the daily log returns of the two financial indices in the years 2006 through 2009 (1,043 observations). The second half of this period covers what has been termed the Global Financial Crisis. The processes $\left(\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|\right)_{k=1,\ldots,n}$ and $\left(\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|\right)_{k=1,\ldots,n}$ are depicted in Figure 1.

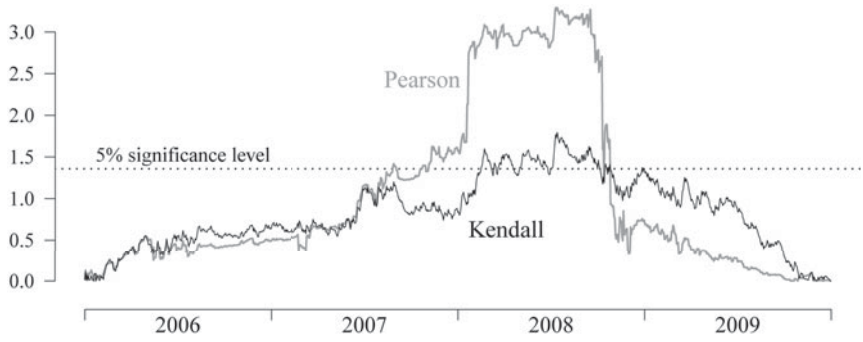


FIGURE 1. Processes $\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|$ (grey) and $\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|$ (black), $k = 1, \ldots, n$, computed from log returns of DAX and S&P 500 between January 1, 2006, and December 31, 2009.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Their maxima are the values of the test statistics of the Pearson and the Kendall test, respectively. Both tests give a p-value below 0.005, and both attain their maximum on July 14, 2008, at the height of the financial crisis. (Lehman Brothers filed for bankruptcy on September 14, 2008.) The tests behave similarly and their outcome supports the assumption that the dependence between both indices considered can not be assumed to be identical before and during the financial crisis of 2008.

With a second set of data, the differences between both tests become apparent. We consider the Dow Jones Industrial Average and the Nasdaq Composite in the years 1987 and 1988 (Figure 2). The most notable feature of both time series is the heavy loss on October 19, 1987, commonly known as Black Monday. Here we may ask in particular the question if the market conditions substantially changed after this date. Does Black Monday constitute a break in the correlation between the two time series? The Pearson test reports a p-value indistinguishable from zero by machine accuracy. The underlying processes of the Pearson and the Kendall test are shown in Figure 3. The outcome of the Pearson test is determined by the peak on October 19, 1987, which is explained as follows. On October 19, both indices suffered heavy losses, suggesting a strong positive correlation of their log returns. The following day the Dow Jones recovered to some small degree, whereas the Nasdaq experienced an even larger drop, suggesting strong negative correlation. Thus the process of successive sample correlations jumps up and immediately down again.

The Kendall test gives a p-value of 0.24, indicating one can assume the correlation between the two time series to be constant over the observed time period. The empirical Kendall's tau is 0.52 prior to Black Monday, and 0.56 afterwards. Indeed, the market conditions turned out to be not much different from before, the DJIA even closed positive for 1987.

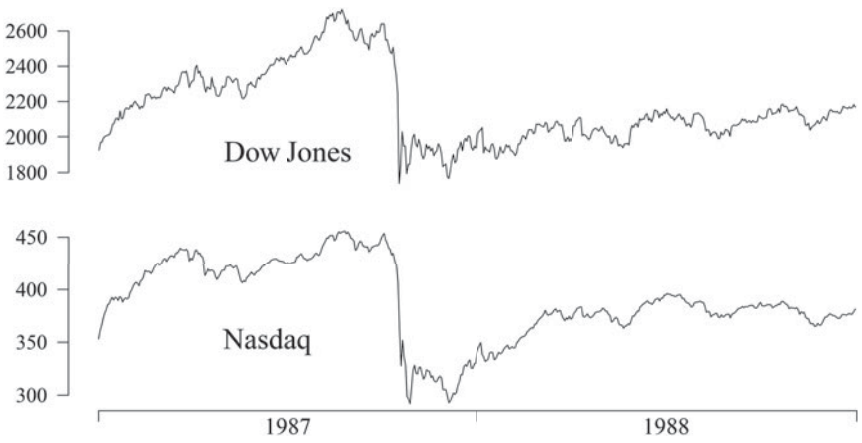


FIGURE 2. Daily closings of Dow Jones Industrial Average and Nasdaq Composite from January 1, 1987, to December 31, 1988.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press



FIGURE 3. Processes $\frac{k}{\sqrt{n}}|\hat{r}_k - \hat{r}_n|$(grey) and $\frac{k}{\sqrt{n}}|\hat{\tau}_k - \hat{\tau}_n|$ (black), $k = 1, \ldots, n$, computed from log returns of DJIA and Nasdaq Composite between January 1, 1987, and December 31, 1988.

The strong impact of a few or even a single extreme observation on the Pearson test is in line with Stock and Watson (2010, Chap. 2), who illustrate the inappropriateness of normal distribution assumptions and the necessity of considering heavy-tailed distributions in modeling financial time series such as the DJIA during the Black Monday crisis.

8. CONCLUSION

We have presented a fluctuation test for detecting changes in the dependence between two time series based on Kendall's rank correlation coefficient. We have demonstrated the noninferiority of the test in terms of efficiency and the clear superiority in terms of robustness and applicability to a similar, previously proposed test, which is based on Pearson's moment correlation. To allow arbitrarily heavy-tailed data and very weak assumptions concerning the serial dependence, we have introduced the concept of near epoch dependence in probability. We have studied the asymptotic behavior of the test statistic under stationarity by means of limit theorems for U-statistics for weakly dependent, stationary processes.

We showed simulation results for various dependence scenarios with elliptical innovations. The elliptical model is of wide-spread use and its intrinsic symmetry allows a meaningful comparison of the two change-point tests. Outside ellipticity, where *monotone* dependence (which is measured, e.g., by Kendall's τ) and *linear* dependence (which is measured by the Pearson correlation) are not necessarily equivalent, it is a matter of debate which of the two types of dependence is more relevant. A detailed discussion goes beyond the scope of this paper, but it appears that one is often interested in monotone dependence rather than linear dependence, and the prevalent use of the sample correlation to measure monotone dependence is presumably due to its simplicity and historical dominance.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Furthermore, simulations also show that the proposed test possesses a variety of advantageous features that have not been discussed in this paper. It has power against gradual or fluctuating changes in the correlation, not only sudden jumps, as presented in Section 6. It also exhibits a much better robustness against heteroscedasticity than the Pearson test. However, a thorough theoretical assessment of these properties as well as constructing tests that explicitly allow heteroscedasticity require the study of U-statistics at nonstationary sequences, which is mathematically rather involved, as our study of the one-change-point model 4.3 illustrates.

Further future research directions include, e.g., the extension to more than two dimensions or guidelines for an on-line application of the test with results about the detection time of a change. Another interesting research question, which is related to the one studied here, is to devise a robust test for detecting changes in the *coherence* of two time series. For example, a series of i.i.d. variables, shifted by one observation, is highly coherent to the original series, but our test, which only compares observations at the same time point, does not detect that type of dependence.

NOTE

1. For nonstationary processes the short-range dependence conditions have to be formulated slightly more generally than in Definition 2.1. The absolute regularity coefficients $(\beta_k)_{k\in\mathbb{N}}$ are defined as $\beta_k = \sup_{t\in\mathbb{Z}} \beta(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+k}^{\infty})$, and the P-NED approximation coefficients $(a_k)_{k\in\mathbb{N}}$ must satisfy

$$\sup_{t\in\mathbb{Z}} P\left(\left|\boldsymbol{X}_t - f_{k,t}(\boldsymbol{Z}_{t-k},\dots,\boldsymbol{Z}_{t+k})\right|_1 > \varepsilon\right) \le a_k \Phi_t(\varepsilon),$$

where the functions $f_{k,t}$ and Φ_t may also depend on t. The underlying process $(\boldsymbol{Z}_t)_{t\in\mathbb{Z}}$ is not required to be stationary.

REFERENCES

Aue, A., S. Hörmann, L. Horváth, & M. Reimherr (2009) Break detection in the covariance structure of multivariate time series models. *Annals of Statistics* 37, 4046–4087.

Berkes, I., S. Hörmann, & J. Schauer (2009) Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications* 119, 1298–1324.

Bilodeau, M. & D. Brenner (1999) *Theory of Multivariate Statistics.* Springer.

Borovkova, S., R.M. Burton, & H. Dehling (2001) Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society* 353, 4261–4318.

Bücher, A., S. Jäschke, & D. Wied (2015) Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics* 187, 154–168.

Bücher, A. & I. Kojadinovic (2016) Dependent multiplier bootstraps for non-degenerate U-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference* 170, 83–105.

Bücher, A., I. Kojadinovic, T. Rohmer, & J. Segers (2014) Detecting changes in cross-sectional dependence in multivariate time series. *Journal of Multivariate Analysis* 132, 111–128.

Croux, C. & C. Dehon (2010) Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications* 19, 497–515.

Csörgő, M. & L. Horváth (1997) *Limit Theorems in Change-Point Analysis.* Wiley.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

de Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–424.

Dehling, H. & R. Fried (2012) Asymptotic distribution of two-sample empirical U-quantiles with applications to robust tests for shifts in location. *Journal of Multivariate Analysis* 105, 124–140.

Dengler, B. (2010) On the Asymptotic Behaviour of the Estimator of Kendall's Tau. Ph.D. thesis, Technische Universität Wien, Austria.

Denker, M. & G. Keller (1986) Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics* 44, 67–93.

Fan, J., H. Liu, & W. Wang (2015) Large covariance estimation through elliptical factor models, submitted.

Gerstenberger, C. & D. Vogel (2015) On the efficiency of gini's mean difference. *Statistical Methods and Applications* 24, 569–596.

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

Gombay, E. & L. Horváth (1995) An application of U-statistics to change-point analysis. *Acta Scientiarum Mathematicarum* 60, 345–357.

Gombay, E. & L. Horváth (1999) Change-points and bootstrap. *Environmetrics* 10, 725–736.

Ibragimov, M., R. Ibragimov, & P. Kattuman (2013) Emerging markets and heavy tails. *Journal of Banking and Finance* 37, 2546–2559.

Inclán, C. & G.C. Tiao (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* 89, 913–923.

Inoue, A. (2001) Testing for distributional change in time series. *Econometric Theory* 17, 156–187.

Kojadinovic, I. (2015) *npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multivariate Observations*. R package version 0.1-6. Available at: http://CRAN.R-project.org/package=npcp.

Kojadinovic, I., J.-F. Quessy, & T. Rohmer (2016) Testing the constancy of Spearman's rho in multivariate time series. *Annals of the Institute of Statistical Mathematics* 68, 929–954.

Loretan, M. & P.C.B. Phillips (1994) Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1, 211–248.

McNeil, A., R. Frey, & P. Embrechts (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press.

Nelsen, R.B. (2006) *An Introduction to Copulas*, 2nd ed. Springer.

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

Pötscher, B.M. & I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models.* Springer-Verlag.

Quessy, J.-F., M. Saïd, & A.-C. Favre (2013) Multivariate Kendall's tau for change-point detection in copulas. *Canadian Journal of Statistics* 41, 65–82.

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

Stock, J. (1994) Unit roots, structural breaks and trends. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2739–2841. North-Holland.

Stock, J. & M. Watson (2010) *Introduction to Econometrics*, 3rd ed. Addison-Wesley.

Stout, W.F. (1974) *Almost Sure Convergence.* Academic Press.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer-Verlag.

Wied, D., H. Dehling, M. van Kampen, & D. Vogel (2014) A fluctuation test for constant Spearmans rho with nuisance-free limit distribution. *Computational Statistics and Data Analysis* 76, 723–736.

Wied, D., W. Krämer, & H. Dehling (2012) Testing for a change in correlation at an unknown point in time using an extended delta method. *Econometric Theory* 28, 570–589.

Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

Yoshihara, K.-I. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete* 35, 237–252.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

APPENDIX

The Appendix is split into three parts. Appendix A provides background on the concept of P-near epoch dependence. In Appendix B, we study the asymptotic behavior of the U-statistic process under local alternatives, which provides the basis for the proof of Theorem 4.4. Appendix C contains all proofs of the results of the main text. Results labeled D.1, D.2, … refer to the online supplement (Appendix D), available at Cambridge Journals Online (journals.cambridge.org/ect).

A. P-NED: Near Epoch Dependence in Probability

In this section we add some background on the newly introduced P-near epoch dependence and connect it to the usual L_p-near epoch dependence (Lemma A.1). Similar conditions that embody the idea of approximating functionals in a probability sense are S-mixing considered by Berkes, Hörmann, and Schauer (2009) and the L_0-approximability of Pötscher and Prucha (1997, Chap. 6).

First note that the P-NED condition (Definition 2.1(ii)) is equivalent to convergence in probability of $f_k(\boldsymbol{Z}_{-k},\dots,\boldsymbol{Z}_k)$ to $\boldsymbol{X}_0$ for $k\to\infty$. If the latter is true, i.e., if

$$\varepsilon_k = \inf\left\{\varepsilon \;\middle|\; P\left(\left|\boldsymbol{X}_0 - f_k(\boldsymbol{Z}_{-k},\dots,\boldsymbol{Z}_k)\right|_1 > \varepsilon\right) \le \varepsilon\right\} \;\to\; 0 \qquad (k\to\infty),$$

then setting $a_k = \varepsilon_k$ and

$$\Phi(\varepsilon) = \left(\sup_{l\le k} \frac{P(|\boldsymbol{X}_0 - f_l(\boldsymbol{Z}_{-l},\dots,\boldsymbol{Z}_l)|_1 > \varepsilon_k)}{\varepsilon_l}\right)\vee 1 \qquad \text{for } \varepsilon\in[\varepsilon_k,\varepsilon_{k-1})$$

fulfills (2). The requirement of the bound on $P\left(\left|\boldsymbol{X}_0 - f_k(\boldsymbol{Z}_{-k},\dots,\boldsymbol{Z}_k)\right|_1 > \varepsilon\right)$ in (2) to factorize into an ε-part and a k-part is not a restriction, but facilitates rate computations.

Recall that a process $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is called L_p *near epoch dependent* (L_p-NED), $p\ge 1$, on the process $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ if the approximating constants $a_{p,k} = \left(E\left|\boldsymbol{X}_0 - E(\boldsymbol{X}_0|\mathcal{F}_{-k}^{k})\right|_p^p\right)^{1/p}$, $k\ge 1$, converge to zero as $k\to\infty$.

Compared to L_2-NED, the P-NED condition substantially enlarges the class of processes for which the condition is easily checked by many heavy-tailed distributions. For example, the autoregressive process

$$\boldsymbol{X}_n = \sum_{k=0}^{\infty} a^k \boldsymbol{Z}_{n-k},$$

where $a\in(-1,1)$, and $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ is an i.i.d. sequence of $\mathbb{R}^d$-valued random variables, can easily be seen to be P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with $a_k = |a|^{\alpha k}$ and $\Phi(\varepsilon) = K\varepsilon^{-\alpha}$ for some constant $K>0$, as long as $P(|\boldsymbol{Z}_n|_1\ge t) = O(t^{-\alpha})$ is satisfied. This is a very weak condition on the innovation distribution, that should be compared to analogous conditions for an AR(1) process to be mixing. In particular, the existence of a density is not required, and all standard examples of discrete or heavy-tailed (e.g., Pareto, Cauchy, geometric) distributions are permitted. Furthermore, P-NED substantially enlarges the class of Lipschitz functionals considered by Denker and Keller (1986, p. 74). The above example would satisfy their condition only if $\boldsymbol{Z}_n$ was bounded. To further detail this example, let, e.g., $a = 1/2$ and $\boldsymbol{Z}_i = (Z_i^{(1)}, Z_i^{(2)})$ have the following discrete distribution

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

$$P\left((Z_i^{(1)}, Z_i^{(2)}) = (0, \tfrac{1}{2})\right) = P\left((Z_i^{(1)}, Z_i^{(2)}) = (\tfrac{1}{2}, 0)\right) = (1-\varrho)/4,$$

$$P\left((Z_i^{(1)}, Z_i^{(2)}) = (0, 0)\right) = P\left((Z_i^{(1)}, Z_i^{(2)}) = (\tfrac{1}{2}, \tfrac{1}{2})\right) = (1+\varrho)/4.$$

The parameter $\varrho \in [-1, 1]$ is the moment correlation of this distribution. The Kendall's tau coefficient of $(Z_i^{(1)}, Z_i^{(2)})$ is $\tau = \varrho/2$. The process $(\boldsymbol{X}_i)_{i\in\mathbb{Z}} = ((X_i^{(1)}, X_i^{(2)}))_{i\in\mathbb{Z}}$ is not strongly mixing (since $\boldsymbol{X}_{i-1}$ is a deterministic function of $\boldsymbol{X}_i$) and hence not absolutely regular, but it is P-NED on $(\boldsymbol{Z}_i)_{i\in\mathbb{Z}}$ with exponentially decreasing approximation coefficients. The distribution of $(X_i^{(1)}, X_i^{(2)})$ has the same moment correlation ϱ as $(Z_i^{(1)}, Z_i^{(2)})$, its Kendall rank correlation is $\tau = 2\varrho/(3-\varrho^2)$, and the margins $X_i^{(1)}$ and $X_i^{(2)}$ are uniformly distributed on $(0, 1)$. Then consider the process $((\tilde{X}_i, \tilde{Y}_i))_{i\in\mathbb{Z}}$ with $\tilde{X}_i^{(1)} = H(X_i^{(1)})$, $\tilde{X}_i^{(2)} = H(X_i^{(2)})$, where H denotes the quantile function of a Pareto (type I) distribution with shape parameter $1/2$ and location parameter 1, i.e., $H(x) = (1-x)^{-2}$, $x \in [0, 1)$. This strictly increasing transformation leaves the P-NED coefficients as well as Kendall's tau unchanged. The margins $\tilde{X}_i^{(1)}$, $\tilde{X}_i^{(2)}$ are Pareto distributed and have finite moments only up to order less than 1/2. The stationary process $((\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}))_{i\in\mathbb{Z}}$ is neither mixing nor L_p-NED for any $p \geq 1$, but fulfills the assumptions of Corollary 3.1. The next lemma connects P-NED and L_p-NED.

LEMMA A.1. *Let $((\boldsymbol{X}_n, \boldsymbol{Z}_n))_{n\in\mathbb{Z}}$ be as in Definition 2.1.*

(*i*) *If $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with functions Φ and f_k, $k \in \mathbb{N}$, and approximating constants $(a_k)_{k\in\mathbb{N}}$, and $g : \mathbb{R}^r \to \mathbb{R}^d$ is a Lipschitz continuous function with Lipschitz constant L, then the process $(g(\boldsymbol{X}_n))_{n\in\mathbb{Z}}$ is P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with functions $\tilde{\Phi}(\varepsilon) = \Phi(\varepsilon/L)$ and $g \circ f_k$, $k \in \mathbb{N}$, and the same approximating constants $(a_k)_{k\in\mathbb{N}}$.*

(*ii*) *Let $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ be bounded and P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with functions Φ and f_k, $k \in \mathbb{N}$, and approximating constants $(a_k)_{k\in\mathbb{N}}$. Then $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is L_p-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ for any $p \geq 1$. If there is further a sequence $(s_k)_{k\in\mathbb{N}}$ of non-negative numbers such that*

$$a_k \Phi(s_k) = O(s_k) \qquad (k \to \infty), \tag{A.1}$$

then the L_p-NED approximating constants $(a_{p,k})_{k\in\mathbb{N}}$ of $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ satisfy $a_{p,k}^p = O(s_k)$ for $k \to \infty$.

(*iii*) *Let $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ be L_p-NED, $p \geq 1$, on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with approximating constants $(a_{p,k})_{k\in\mathbb{N}}$. Then $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$. If there is further a non-increasing function $\Phi : (0, \infty) \to (0, \infty)$ and a sequence $(s_k)_{k\in\mathbb{N}}$ of non-negative numbers converging to zero that satisfy*

$$\Phi(\varepsilon) s_k \geq q^{p-1} a_{p,k}^p \varepsilon^{-p},$$

then $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ is P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with approximation constants $(s_k)_{k\in\mathbb{N}}$ and function Φ. The functions f_k can be chosen as $f_k(\boldsymbol{Z}_{-k}, \ldots, \boldsymbol{Z}_k) = E(\boldsymbol{X}_0 | \mathcal{F}_{-k}^k)$, $k \in \mathbb{N}$.

Since Φ is nonincreasing, condition (A.1) puts an upper bound on the speed of decay of $(s_k)_{k\in\mathbb{N}}$ in the sense that, if (A.1) is fulfilled by some sequence $(s_k)_{k\in\mathbb{N}}$, then it is also fulfilled by any sequence $(\tilde{s}_k)_{k\in\mathbb{N}}$ for which $\tilde{s}_k \leq s_k$ for all k larger than some $n \in \mathbb{N}$.

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

The next lemma shows that the near epoch dependence is preserved under transformations that satisfy the variation condition.

LEMMA A.2. *Let $(\boldsymbol{X}_n)_{n\in\mathbb{Z}}$ be P-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{Z}}$ with $a_k\Phi(s_k) = O\big(s_k^{(2+\delta)/\delta}\big)$. Furthermore, let g be a U-statistic kernel with uniform $(2+\delta)$ moments for some $\delta > 0$, i.e., $E|g(\boldsymbol{X}_0, \boldsymbol{X}_n)|^{2+\delta} < M$ for all $n \in \mathbb{N}$, and let g satisfy the variation condition with respect to the distribution F of $\boldsymbol{X}_0$ (Assumption 2.4). Then the sequence $(g_1(\boldsymbol{X}_n))_{n\in\mathbb{Z}}$ is L_2-NED on $(\boldsymbol{Z}_n)_{n\in\mathbb{N}}$ with approximation constants $a_{k,2} = O\big(s_k^{\delta/(2+2\delta)}\big)$.*

B. U-statistics Process Under Local Alternatives

In this section, we provide the basis for the proof of Theorem 4.4 in Section C. We analyze the U-statistic process $\sum_{1\le i<j\le[ns]} g(X_i, X_j)$, $0 \le s \le 1$, in the case of a change point at time $[n\lambda^*]$, and calculate the asymptotic distribution under local alternatives. We consider the following model

$$X_i = X_i^{(n)} = \begin{cases} \xi_i & 1 \le i \le [n\lambda^*], \\ \xi_i^{(n)} & [n\lambda^*] < i \le n, \end{cases}$$

where $(\xi_i)_{i\ge1}$ and $\big(\xi_i^{(n)}\big)_{i\ge1}$, $n \ge 1$, are stationary processes such that for any $n \ge 1$ the bivariate processes $\big(\xi_i, \xi_i^{(n)}\big)_{i\ge1}$ are P-NED on an absolutely regular process, with mixing coefficients and NED-coefficients independent of n.

For the corresponding U-statistic process $\sum_{1\le i<j\le[ns]} g\big(X_i^{(n)}, X_j^{(n)}\big)$, we have to distinguish the cases $s \le \lambda^*$ and $s \ge \lambda^*$. When $s \le \lambda^*$, we have

$$\sum_{1\le i<j\le[ns]} g\big(X_i^{(n)}, X_j^{(n)}\big) = \sum_{1\le i<j\le[ns]} g(\xi_i, \xi_j),$$

while for $s \ge \lambda^*$ we obtain

$$\begin{aligned} &\sum_{1\le i<j\le[ns]} g\big(X_i^{(n)}, X_j^{(n)}\big) \\ &= \sum_{1\le i<j\le[n\lambda^*]} g(\xi_i, \xi_j) + \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=1}^{[n\lambda^*]} g\big(\xi_i, \xi_j^{(n)}\big) + \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=[n\lambda^*]+1}^{j-1} g\big(\xi_i^{(n)}, \xi_j^{(n)}\big). \end{aligned} \quad \textbf{(A.2)}$$

Note that of the three terms on the r.h.s., the first and the last term are one-sample U-statistics, while the second term is a two-sample U-statistic. We now define the constant terms and the first order terms of the Hoeffding decompositions,

$$\begin{aligned} \theta &= E(g(\xi, \eta)), \\ \theta_1^{(n)} &= E\big(g\big(\xi^{(n)}, \eta\big)\big), & \theta_2^{(n)} &= E\big(g\big(\xi^{(n)}, \eta^{(n)}\big)\big), \\ g_1(x) &= E(g(x, \xi_j)), & g_1^{(n)}(x) &= E\big(g\big(x, \xi_j^{(n)}\big)\big). \end{aligned}$$

Here, ξ, η, $\xi^{(n)}$, and $\eta^{(n)}$ are independent random variables such that ξ and η have the same distribution as ξ_1, while $\xi^{(n)}$ and $\eta^{(n)}$ have the same distribution as $\xi_1^{(n)}$. Note that the functions g_1 and $g_1^{(n)}$ are not centered. In fact, they require different centerings for each of the three terms on the r.h.s. of (A.2).

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

In this way, we obtain the following Hoeffding decompositions of the summands in (A.2):

$$
\begin{aligned}
g(\xi_i,\xi_j) &= \theta + (g_1(\xi_i)-\theta) + (g_1(\xi_j)-\theta) + h(\xi_i,\xi_j),\\
g(\xi_i,\xi_j^{(n)}) &= \theta_1^{(n)} + \bigl(g_1^{(n)}(\xi_i)-\theta_1^{(n)}\bigr) + \bigl(g_1(\xi_j^{(n)})-\theta_1^{(n)}\bigr) + h_1^{(n)}(\xi_i,\xi_j^{(n)}),\\
g(\xi_i^{(n)},\xi_j^{(n)}) &= \theta_2^{(n)} + \bigl(g_1^{(n)}(\xi_i^{(n)})-\theta_2^{(n)}\bigr) + \bigl(g_1^{(n)}(\xi_j^{(n)})-\theta_2^{(n)}\bigr) + h_2^{(n)}(\xi_i^{(n)},\xi_j^{(n)}).
\end{aligned}
$$

The functions h, $h_1^{(n)}$, and $h_2^{(n)}$ are defined in the obvious way, and a straightforward calculation shows that they are degenerate, i.e., that the integral with respect to one of the arguments vanishes.

Plugging this into (A.2), we obtain the Hoeffding decomposition of the U-statistic

$$
\begin{aligned}
\sum_{1\le i<j\le [ns]} g(X_i^{(n)},X_j^{(n)}) &= \binom{[n\lambda^*]}{2}\theta + [n\lambda^*]([ns]-[n\lambda^*])\theta_1^{(n)} + \binom{[ns]-[n\lambda^*]}{2}\theta_2^{(n)}\\
&\quad + ([n\lambda^*]-1)\sum_{i=1}^{[n\lambda^*]}(g_1(\xi_i)-\theta) + ([ns]-[n\lambda^*])\sum_{i=1}^{[n\lambda^*]}\bigl(g_1^{(n)}(\xi_i)-\theta_1^{(n)}\bigr)\\
&\quad + [n\lambda^*]\sum_{i=[n\lambda^*]+1}^{[ns]}\bigl(g_1(\xi_i^{(n)})-\theta_1^{(n)}\bigr) + ([ns]-[n\lambda^*]-1)\\
&\quad \times \sum_{i=[n\lambda^*]+1}^{[ns]}\bigl(g_1^{(n)}(\xi_i^{(n)})-\theta_2^{(n)}\bigr)\\
&\quad + \sum_{1\le i<j\le [n\lambda^*]} h(\xi_i,\xi_j) + \sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=1}^{[n\lambda^*]} h_1^{(n)}(\xi_i,\xi_j^{(n)})\\
&\quad + \sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=[n\lambda^*]+1}^{j-1} h_2^{(n)}(\xi_i^{(n)},\xi_j^{(n)}).
\end{aligned}
$$

We make a number of assumptions regarding the asymptotic behaviour of the terms in the Hoeffding expansion that have to be checked in particular examples.

Assumption B.1. There are real numbers c_1, c_2 such that $\sqrt{n}(\theta_1^{(n)}-\theta)\to c_1$ and $\sqrt{n}(\theta_2^{(n)}-\theta)\to c_2$.

Assumption B.2.

$$
\max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\Bigl((g_1(\xi_i)-\theta)-\bigl(g_1^{(n)}(\xi_i)-\theta_1^{(n)}\bigr)\Bigr)\right|\to 0,
$$

$$
\max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\Bigl((g_1(\xi_i)-\theta)-\bigl(g_1(\xi_i^{(n)})-\theta_1^{(n)}\bigr)\Bigr)\right|\to 0,
$$

$$
\max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\Bigl((g_1(\xi_i)-\theta)-\bigl(g_1^{(n)}(\xi_i^{(n)})-\theta_2^{(n)}\bigr)\Bigr)\right|\to 0.
$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

THEOREM B.3. *Under Assumptions B.1 and B.2,*

$$\frac{2}{\sqrt{n}([nt]-1)} \sum_{1\le i<j\le [ns]} \big(g(X_i^{(n)}, X_j^{(n)}) - \theta\big) \xrightarrow{d} 2\sigma W + \phi_{\lambda^*}(s),$$

where σ and W are defined as in Theorem 2.5, and where the function ϕ_λ^* is defined as follows:

$$\phi_{\lambda^*}(s) = \begin{cases} \frac{2\lambda^*(s-\lambda^*)}{s} c_1 + \frac{(s-\lambda^*)^2}{s} c_2 & s \ge \lambda^*, \\ 0 & s \le \lambda^* \end{cases}$$

with c_1 and c_2 being defined in Assumption B.1.

Proof of Theorem B.3. We first consider the Hoeffding decomposition of the U-statistic under the null hypothesis, i.e.,

$$\sum_{1\le i<j\le [ns]} \big(g(\xi_i, \xi_j) - \theta\big) = ([ns]-1) \sum_{i=1}^{[ns]} (g_1(\xi_i) - \theta) + \sum_{1\le i<j\le [ns]} h(\xi_i, \xi_j).$$

Comparing this with the Hoeffding decomposition under the local alternative, we obtain for $s \ge \lambda^*$,

$$\begin{aligned}
&\sum_{1\le i<j\le [ns]} \big(g(X_i^{(n)}, X_j^{(n)}) - \theta\big) - \sum_{1\le i<j\le [ns]} \big(g(\xi_i, \xi_j) - \theta\big) \\
&\quad = [n\lambda^*]([ns]-[n\lambda^*])(\theta_1^{(n)} - \theta) + \frac{1}{2}([ns]-[n\lambda^*])([ns]-[n\lambda^*]-1)\big(\theta_2^{(n)} - \theta\big) \\
&\qquad + ([ns]-[n\lambda^*]) \sum_{i=1}^{[n\lambda^*]} \Big(\big(g_1^{(n)}(\xi_i) - \theta_1^{(n)}\big) - (g_1(\xi_i) - \theta)\Big) \\
&\qquad + ([ns]-[n\lambda^*]-1) \sum_{i=[n\lambda^*]+1}^{[ns]} \Big(\big(g_1(\xi_i^{(n)} - \theta_1^{(n)}\big) - (g_1(\xi_i) - \theta)\Big) \\
&\qquad + ([ns]-[n\lambda^*]-1) \sum_{i=[n\lambda^*]+1}^{[ns]} \Big(\big(g_1^{(n)}\big(\xi_i^{(n)}\big) - \theta_2^{(n)}\big) - (g(\xi_i) - \theta)\Big) \\
&\qquad + \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=1}^{[n\lambda^*]} h_1^{(n)}\big(\xi_i, \xi_j^{(n)}\big) + \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=[n\lambda^*]+1}^{[ns]} h_2^{(n)}\big(\xi_i^{(n)}, \xi_j^{(n)}\big) \\
&\qquad - \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=1}^{[n\lambda^*]} h(\xi_i, \xi_j) - \sum_{j=[n\lambda^*]+1}^{[ns]} \sum_{i=[n\lambda^*]+1}^{[ns]} h(\xi_i, \xi_j),
\end{aligned}$$

while for $s \le \lambda^*$, the term on the l.h.s. equals zero. Thus we obtain, again for $s \ge \lambda^*$,

$$\begin{aligned}
&\frac{2}{\sqrt{n}([ns]-1)} \Big(\sum_{1\le i<j\le [ns]} \big(g(X_i^{(n)}, X_j^{(n)}) - \theta\big) - \sum_{1\le i<j\le [ns]} \big(g(\xi_i, \xi_j) - \theta\big) \Big) - \phi_\lambda^*(s) \\
&\quad = \frac{2[n\lambda^*]([ns]-[n\lambda^*])}{n([ns]-1)} \sqrt{n}\big(\theta_1^{(n)} - \theta\big) + \frac{([ns]-[n\lambda^*])([ns]-[n\lambda^*]-1)}{n([ns]-1)} \sqrt{n}\big(\theta_2^{(n)} - \theta\big) - \phi_\lambda^*(s)
\end{aligned}$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

$$
\begin{aligned}
&+\frac{2([ns]-[n\lambda^*])}{n}\frac{1}{\sqrt{n}}\sum_{i=1}^{[n\lambda^*]}\Big(\big(g_1^{(n)}(\xi_i)-\theta_1^{(n)}\big)-\big(g_1(\xi_i)-\theta\big)\Big)\\
&+\frac{2([ns]-[n\lambda^*]-1)}{[ns]-1}\frac{1}{\sqrt{n}}\sum_{i=[n\lambda^*]+1}^{[ns]}\Big(\big(g_1(\xi_i^{(n)}-\theta_1^{(n)}\big)-\big(g_1(\xi_i)-\theta\big)\Big)\\
&+\frac{2([ns]-[n\lambda^*]-1)}{[ns]-1}\frac{1}{\sqrt{n}}\sum_{i=[n\lambda^*]+1}^{[ns]}\Big(\big(g_1^{(n)}(\xi_i^{(n)})-\theta_2^{(n)}\big)-\big(g_1(\xi_i)-\theta\big)\Big)\\
&+\frac{2}{\sqrt{n}([ns]-1)}\sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=1}^{[n\lambda^*]}h_1^{(n)}\big(\xi_i,\xi_j^{(n)}\big)\\
&+\frac{2}{\sqrt{n}([ns]-1)}\sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=[n\lambda^*]+1}^{[ns]}h_2^{(n)}\big(\xi_i^{(n)},\xi_j^{(n)}\big)\\
&-\frac{2}{\sqrt{n}([nt]-1)}\sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=1}^{[n\lambda^*]}h(\xi_i,\xi_j)-\frac{2}{\sqrt{n}([ns]-1)}\sum_{j=[n\lambda^*]+1}^{[ns]}\sum_{i=[n\lambda^*]+1}^{[ns]}h(\xi_i,\xi_j).
\end{aligned}
$$

The right hand side converges to zero uniformly in $s \geq \lambda^*$. For the first four lines on the r.h.s., this follows from our assumptions. Uniform convergence of the terms involving the degenerate kernels h, $h_1^{(n)}$, and $h_2^{(n)}$ follows with arguments similar to those employed in the proof of Theorem 2.5. ■

Returning to the original notation of our paper, we define the U-statistic

$$U_k^{(n)} = \frac{1}{\binom{k}{2}}\sum_{1\leq i<j\leq k} g\big(X_i^{(n)}, X_j^{(n)}\big),\ 0 \leq k \leq n.$$

Using the continuous mapping theorem, we derive the following corollary to the above theorem.

COROLLARY B.4.

$$\frac{1}{2\hat{\sigma}_n}\max_{1\leq k\leq n-1}\frac{k}{\sqrt{n}}\big|U_k^{(n)}-U_n^{(n)}\big| \xrightarrow{d} \sup_{0\leq\lambda\leq 1}\left|B(\lambda)+\frac{1}{2\sigma}\left(\phi_{\lambda^*}(\lambda)-\lambda\phi_{\lambda^*}(1)\right)\right|.$$

C. Proofs of Sections 2, 3, and 4

In this section we prove the theorems of the main text. These are Theorems 2.5 (invariance principle for the sequential U-process), 2.7 (consistency of the long run variance estimator), and Corollary 3.1 (asymptotics of the test statistic under the null), Theorem 4.2 (estimation of the change-point), and Theorem 4.4 (local power analysis).

Proof of Theorem 2.5 (Invariance principle for the sequential U-process)**.** Using the Hoeffding decomposition, we can write the sequential U-process as

$$\frac{[ns]}{\sqrt{n}}\big(U_{[ns]}-U\big) = \frac{2}{\sqrt{n}}\sum_{i=1}^{[ns]}g_1(\boldsymbol{X}_i)+\frac{2}{\sqrt{n}}\frac{1}{[ns]-1}\sum_{1\leq i<j\leq[ns]}g_2(\boldsymbol{X}_i,\boldsymbol{X}_j).$$

For the first summand we find by Assumption 2.2 and Lemma A.2 that the sequence $(g_1(\boldsymbol{X}_n))_{n\in\mathbb{N}}$ is L_2-NED with approximation constants $a_{k,2} = O(k^{-3})$. So

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

we can apply Corollary 3.2 of Wooldridge and White (1988), stating that the partial sum process $\left(\frac{2}{\sqrt{n}}\sum_{i=1}^{[ns]} g_1(\boldsymbol{X}_i)\right)_{t\in[0,1]}$ converges weakly to a Brownian motion with variance $4\sigma^2$. For the second summand, we use Lemma D.6, which implies that $\left|\frac{1}{n}\sum_{1\le i<j\le n} g_2(\boldsymbol{X}_i, \boldsymbol{X}_j)\right| \le Cn^{\frac{1}{4}}\log^2(n)$ almost surely and consequently

$$\sup_{s\in[0,1]} \frac{1}{[ns]\sqrt{n}} \sum_{1\le i<j\le[ns]} g_2(\boldsymbol{X}_i, \boldsymbol{X}_j) \le Cn^{-\frac{1}{4}}\log^2(n) \to 0$$

almost surely as $n \to \infty$. Slutsky's theorem completes the proof. ■

Proof of Theorem 2.7 (Consistency of the long run variance estimator)**.** We can write the variance estimator $\hat{\sigma}_n^2$ as

$$\begin{aligned}\hat{\sigma}_n^2 &= \sum_{r=-(n-1)}^{n-1} \kappa\left(\frac{|r|}{b_n}\right)\frac{1}{n}\sum_{i=1}^{n-|r|} \hat{g}_1(\boldsymbol{X}_i)\hat{g}_1(\boldsymbol{X}_{i+|r|}) \\ &= \sum_{r=-(n-1)}^{n-1} \kappa\left(\frac{|r|}{b_n}\right)\frac{1}{n}\sum_{i=1}^{n-|r|} g_1(\boldsymbol{X}_i)g_1(\boldsymbol{X}_{i+|r|}) \\ &\quad + \sum_{r=-(n-1)}^{n-1} \frac{1}{n}\sum_{i=1}^{n-|r|} \left(g_1(\boldsymbol{X}_i)g_1(\boldsymbol{X}_{i+|r|}) - \hat{g}_1(\boldsymbol{X}_i)\hat{g}_1(\boldsymbol{X}_{i+|r|})\right)\kappa(|k|/b_n).\end{aligned}$$

By Theorem 2.1 of de Jong and Davidson (2000), we know that the first summand converges to $\sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left(g_1(\boldsymbol{X}_0), g_1(\boldsymbol{X}_k)\right)$. The second summand converges to 0 by Lemma D.9, and the proof is complete. ■

Proof of Corollary 3.1 (Asymptotic null distribution of change-point test statistic)**.** Corollary 3.1 is a special case of Corollary 2.8 for the specific kernel g given by (5). It remains to show that, under the conditions of Corollary 3.1, the assumptions of Corollary 2.8 are met. Since g is bounded, Assumption 2.3 is satisfied for any $\delta > 0$, and it suffices that Assumption 2.2 is fulfilled for some $\delta > 0$. ■

Proof of Theorem 4.2 (Change-point estimation)**.** Under the assumptions in Theorem 4.2 concerning the value of τ_{FG}, the function $|c(\lambda)|$, $\lambda \in [0,1]$, has a unique maximum at $\lambda = \lambda^*$, where the function $c : [0,1] \to \mathbb{R}$ is given by

$$c(\lambda) = \begin{cases} \left[(1-\lambda^{*2})\tau_F - (1-\lambda^*)^2\tau_G - 2\lambda^*(1-\lambda^*)\tau_{F,G}\right]\lambda & \text{for } 0 \le \lambda < \lambda^*, \\ 2\lambda^*(\tau_{F,G}-\tau_G)(1-\lambda) + \lambda^{*2}(\tau_F+\tau_G-2\tau_{F,G})\left(\frac{1}{\lambda}-\lambda\right) & \text{for } \lambda^* \le \lambda \le 1. \end{cases} \tag{A.3}$$

The proof relies on the fact that

$$(C_n(\lambda))_{0\le\lambda\le1} = \left(\frac{[\lambda n]}{n}\left(\hat{\tau}_{[\lambda n]} - \hat{\tau}_n\right)\right)_{0\le\lambda\le1} \xrightarrow{\mathrm{d}} (c(\lambda))_{0\le\lambda\le1} \tag{A.4}$$

in $D[0,1]$. With the argmax theorem (van der Vaart and Wellner, 1996, Corollary 3.2.3) we have that

$$\hat{\lambda}_n = \operatorname*{arg\,max}_{0\le\lambda\le1} |C_n(\lambda)| \xrightarrow{p} \operatorname*{arg\,max}_{0\le\lambda\le1} |c(\lambda)| = \lambda^*.$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

It remains to prove (A.4). To simplify notation, we let $m = [\lambda^* n]$, write $\boldsymbol{Z}_i^{(n)}$ short for $\big(X_i^{(n)}, Y_i^{(n)}\big)$ and further suppress the subscript n. Assume for an instant that the $\boldsymbol{Z}_i$, $i = 1, \ldots, n$, are independent. Then we have

$$t_n(k) = E(\hat{\tau}_k) = \begin{cases} \tau_F & \text{for } k \le m, \\ \frac{m(m-1)}{k(k-1)}\tau_F + \frac{(k-m)(k-m-1)}{k(k-1)}\tau_G + \frac{2m(k-m)}{k(k-1)}\tau_{F,G} & \text{for } k \ge m+1, \end{cases}$$

from where we derive the mean function $c_n(\lambda) = E[C_n(\lambda)] = [\lambda n] n^{-1}(t_n([\lambda n]) - t_n(n))$, $\lambda \in [0,1]$, of the process of $(C_n(\lambda))_{0 \le \lambda \le 1}$ and observe that it converges to the function c. Thus it remains to show that $\max_{0 \le \lambda \le 1} |C_n(\lambda) - c_n(\lambda)|$ converges to zero in probability also under the short-range dependence assumption of Model 4.1. In the following, let the $\boldsymbol{Z}_i$, $i = 1, \ldots, n$ be weakly dependent as specified by Model 4.1. We will prove that

$$\max_{m < k \le n} \frac{k}{n} \left| \hat{\tau}_k - t_n(k) \right| \xrightarrow{p} 0. \tag{A.5}$$

The convergence of $\max_{1 < k \le m} \frac{k}{n} \left| \hat{\tau}_k - t_n(k) \right|$ follows along the same lines. Hence the maximum in (A.5) can be extended to the range $k = 1, \ldots, n$. We split the difference $\frac{k}{n} \left| \hat{\tau}_k - t_n(k) \right|$ into three parts: two one-sample U-statistics and one two-sample U-statistic with the kernel

$$g((x_1, y_1), (x_2, y_2)) = \mathbf{1}_{(0,\infty)}((x_2 - x_1)(y_2 - y_1)) - \mathbf{1}_{(-\infty,0)}((x_2 - x_1)(y_2 - y_1))$$

as in Section 3. By the triangle inequality we get the following:

$$\begin{aligned} \frac{k}{n} \left| \hat{\tau}_k - t_n(k) \right| \le & \left| \frac{2}{n(k-1)} \sum_{1 \le i < j \le m} \big(g(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - \tau_F \big) \right| \\ & + \left| \frac{2}{n(k-1)} \sum_{m+1 \le i < j \le k} \big(g(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - \tau_G \big) \right| \\ & + \left| \frac{2}{n(k-1)} \sum_{1 \le i \le m < j \le k} \big(g(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - \tau_{F,G} \big) \right|. \end{aligned} \tag{A.6}$$

For the first summand on the right-hand side, we have by Theorem 2.5:

$$\max_{m < k \le n} \left| \tfrac{2}{n(k-1)} \sum_{1 \le i < j \le m} \big(g(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - \tau_F \big) \right| \le \left| \tfrac{2}{n(m-1)} \sum_{1 \le i < j \le m} \big(g(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - \tau_F \big) \right| \xrightarrow{p} 0.$$

Due to our assumptions, $\big(\boldsymbol{Z}_i^{(n)}\big)_{m+1 \le i \le n}$ is a stationary process which satisfies Condition 2.2, so we also treat the second summand by Theorem 2.5. For the third summand, we apply a two-sample Hoeffding decomposition $g(\boldsymbol{z}_1, \boldsymbol{z}_2) = \tau_{F,G} + \tilde{g}_1(\boldsymbol{z}_1) + \tilde{g}_2(\boldsymbol{z}_2) + \tilde{g}_3(\boldsymbol{z}_1, \boldsymbol{z}_2)$ with

$$\tilde{g}_1(\boldsymbol{z}_1) = E\big(g(\boldsymbol{z}_1, \boldsymbol{Z}_{m+1})\big) - \tau_{F,G}, \quad \tilde{g}_2(\boldsymbol{z}_1) = E\left(g(\boldsymbol{Z}_1, \boldsymbol{z}_2)\right) - \tau_{F,G},$$

$$\tilde{g}_3(\boldsymbol{z}_1, \boldsymbol{z}_2) = g(\boldsymbol{z}_1, \boldsymbol{z}_2) - \tilde{g}_1(\boldsymbol{z}_1) - \tilde{g}_2(\boldsymbol{z}_2) - \tau_{F,G},$$

where $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^2$. We get

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

$$\frac{2}{n(k-1)} \sum_{1\le i\le m<j\le k} \big(g(\boldsymbol{Z}_i,\boldsymbol{Z}_j)-\tau_{F,G}\big)$$

$$= \frac{2(k-m)}{n(k-1)}\sum_{i=1}^{m}\tilde{g}_1(\boldsymbol{Z}_i)+\frac{2m}{n(k-1)}\sum_{j=m+1}^{k}\tilde{g}_2(\boldsymbol{Z}_j)+\frac{2}{n(k-1)}\sum_{1\le i\le m<j\le k}\tilde{g}_3(\boldsymbol{Z}_i,\boldsymbol{Z}_j). \quad \textbf{(A.7)}$$

To obtain a maximal inequality, we use Theorem 2.4.1 of Stout (1974): For random variables $R_1,\dots,R_n$ with $E(\sum_{j=k}^{k+l-1} R_j)^2 \le C_1 l$ for a constant C_1, we have that

$$E\left[\max_{1\le l\le n}\left(\sum\nolimits_{j=1}^{l} R_j\right)^2\right] \le C_1 n\,(\log(2n)/\log 2)^2\,. \quad \textbf{(A.8)}$$

We define the random variables $R_j=\sum_{i=1}^{m}\tilde{g}_3(\boldsymbol{Z}_i,\boldsymbol{Z}_{j+m})$. Without loss of generality, we can assume that the random variables $\boldsymbol{Z}_i$ are bounded and thus the process is L_1-NED by Lemma A.1(ii). Furthermore, the kernel $\tilde{g}_3$ is degenerate, so we can apply Proposition A.2 of Dehling and Fried (2012) to obtain the moment bound $E(\sum_{j=1}^{k+l-1} R_j)^2 = E\big[\sum_{j=1}^{k+l-1}\sum_{i=1}^{m}\tilde{g}_3(\boldsymbol{Z}_i,\boldsymbol{Z}_{j+m})\big]^2 \le Cml$. Applying (A.8), we find

$$E\left(\max_{m<k\le n}\left|\frac{2}{n(k-1)}\sum_{1\le i\le m<j\le k}\tilde{g}_3(\boldsymbol{Z}_i,\boldsymbol{Z}_j)\right|\right)^2 \le m^{-2}E\left(\max_{1<k\le n-m}\left|\frac{2}{n}\sum_{j=1}^{k}R_j\right|\right)^2 \le \frac{C\log^2(2n)}{mn\log^2 2}$$

converges to zero as $n\to 0$. Thus the third summand in (A.7) converges to zero in probability. As for the first two summands, we have that $E\big[\sum_{j=k}^{k+l-1}\tilde{g}_2(\boldsymbol{Z}_j)\big]^2 \le Cl$, since $l^{-1}\mathrm{Var}\Big[\sum_{j=k}^{k+l-1}\tilde{g}_2(\boldsymbol{Z}_j)\Big]$ converges to a finite limit as $l\to\infty$. Hence, (A.8) applied to $R_j=\tilde{g}_2(\boldsymbol{Z}_{j+m})$ leads to

$$\max_{m<k\le n}\left|\frac{2}{n}\sum\nolimits_{j=m+1}^{k}\tilde{g}_2(\boldsymbol{Z}_j)\right| \xrightarrow{p} 0$$

as $n\to\infty$. Finally, $2/n\sum_{i=1}^{m}\tilde{g}_1(\boldsymbol{Z}_i)\xrightarrow{p}0$, as its variance converges to zero. We have thus shown (A.5), which completes the proof. ■

Proof of Theorem 4.4. It remains to calculate the quantities c_1, c_2 introduced above for the change-point test based on Kendall's τ, under the local alternative as specified in Model 4.3. Observe that, in this case, $\theta=\tau_F$, $\theta_1^{(n)}=\tau_{F,F_n}$, and $\theta_2^{(n)}=\tau_{F_n,F_n}$, where F denotes the joint distribution of (X_i,Y_i) before the change, and F_n denotes the joint distribution of $\big(X_i^{(n)},Y_i^{(n)}\big)$ after the change.

We now provide a formula for $\tau_{F,G}$ in the case when both distributions F,G are absolutely continuous, with densities $f(x,y)$ and $g(x,y)$. Let (X_1,Y_1) and (X_2,Y_2) be random variables with densities $f(x,y)$, and $g(x,y)$, respectively, and let (X_1,Y_1) be independent of (X_2,Y_2). Then we have

$$\tau_{F,G}=2\Big(\tau_{F,G}^{(I)}+\tau_{F,G}^{(II)}\Big)-1, \quad \textbf{(A.9)}$$

where

$$\tau_{F,G}^{(I)}=P(X_1\le X_2,Y_1\le Y_2)=\int_{-\infty}^{\infty}\int_{x_1}^{\infty}\int_{-\infty}^{\infty}\int_{y_1}^{\infty} f(x_1,y_1)g(x_2,y_2)dy_2dy_1dx_2dx_1,$$

$$\tau_{F,G}^{(II)}=P(X_1\ge X_2,Y_1\ge Y_2)=\int_{-\infty}^{\infty}\int_{x_2}^{\infty}\int_{-\infty}^{\infty}\int_{y_2}^{\infty} f(x_1,y_1)g(x_2,y_2)dy_2dy_1dx_2dx_1.$$

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press

Noting that $\tau_F = \tau_{F,F}$, we obtain $\tau_F = 2\big(\tau_{F,F}^{(I)} + \tau_{F,F}^{(II)}\big) - 1$, where $\tau_{F,F}^{(I)}$ and $\tau_{F,F}^{(II)}$ are given by the above integrals with $f = g$. Finally, we obtain

$$\tau_{F,G} - \tau_F = 2\Big(\big(\tau_{F,G}^{(I)} - \tau_{F,F}^{(I)}\big) + \big(\tau_{F,G}^{(II)} - \tau_{F,F}^{(II)}\big)\Big).$$

Under Model 4.3, we have $G = F_n$, where F_n is the distribution of $\big(X_1 + \frac{\Delta}{\sqrt{n}} Y_1, Y_1\big)$. By the transformation formula for densities, F_n has the density

$$f_n(x_2, y_2) = f\left(x_2 - \frac{\Delta}{\sqrt{n}} y_2, y_2\right).$$

Thus, we obtain

$$\tau_{F,F_n}^{(I)} - \tau_{F,F}^{(I)} = \int_{-\infty}^{\infty}\int_{x_1}^{\infty}\int_{-\infty}^{\infty}\int_{y_1}^{\infty} f(x_1, y_1)\left(f\left(x_2 - \frac{\Delta}{\sqrt{n}} y_2, y_2\right) - f(x_2, y_2)\right) dy_2 dy_1 dx_2 dx_1.$$

Under the assumptions made in Model 4.3 on the densities, we then obtain

$$\lim_{n\to\infty} \sqrt{n}\left(\tau_{F,F_n}^{(I)} - \tau_{F,F}^{(I)}\right) = -\Delta \int_{-\infty}^{\infty}\int_{x_1}^{\infty}\int_{-\infty}^{\infty}\int_{y_1}^{\infty} f(x_1, y_1) y_2 f_1(x_2, y_2) dy_2 dy_1 dx_2 dx_1,$$

where $f_1(x, y) = \frac{\partial}{\partial x} f(x, y)$. Furthermore,

$$\int_{x_1}^{\infty} f_1(x_2, y_2) dx_2 = -f(x_1, y_2),$$

and thus

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{n}\left(\tau_{F,F_n}^{(I)} - \tau_{F,F}^{(I)}\right) &= \Delta \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{y_1}^{\infty} f(x_1, y_1) y_2 f(x_1, y_2) dy_2 dy_1 dx_1 \\
&= \Delta \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{y_2} y_2 f(y_1|x) f(y_2|x) dy_1 dy_2\right) f_X^2(x) dx \\
&= \Delta \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y_2 F(y_2|x) f(y_2|x) dy_2\right) f_X^2(x) dx \\
&= \Delta \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y_2 F(y_2|x) F(dy_2|x)\right) f_X^2(x) dx,
\end{aligned}$$

where f_X denotes the marginal density of X, and where $f(y|x)$ and $F(y|x)$ denote the conditional density, respectively, the conditional distribution function of Y given $X = x$.

With similar calculations, we obtain

$$\lim_{n\to\infty} \sqrt{n}\left(\tau_{F,F_n}^{(II)} - \tau_{F,F}^{(II)}\right) = \Delta \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y_2 (F(y_2|x) - 1) F(dy_2|x)\right) f_X^2(x) dx.$$

Finally, one can show that $\lim_{n\to\infty} \sqrt{n}(\tau_{F_n,F_n} - \tau_{F,F}) = 2\lim_{n\to\infty} \sqrt{n}(\tau_{F,F_n} - \tau_{F,F})$. ■

https://doi.org/10.1017/S026646661600044X Published online by Cambridge University Press