DEGENERATE U- AND V-STATISTICS UNDER ERGODICITY: ASYMPTOTICS, BOOTSTRAP AND APPLICATIONS IN STATISTICS

Anne Leucht
Technische Universität Braunschweig
Institut für Mathematische Stochastik
Pockelsstr. 14
D-38106 Braunschweig
Germany
E-mail: a.leucht@tu-braunschweig.de

Michael H. Neumann
Friedrich-Schiller-Universität Jena
Institut für Stochastik
Ernst-Abbe-Platz 2
D-07743 Jena
Germany
E-mail: michael.neumann@uni-jena.de

Abstract

We derive the asymptotic distributions of degenerate U- and V-statistics of stationary and ergodic random variables. Statistics of these types naturally appear as approximations of test statistics. Since the limit variables are of complicated structure, quantiles can hardly be obtained directly. Therefore, we prove a general result on the consistency of model-based bootstrap methods for U- and V-statistics under easily verifiable conditions. Three applications to hypothesis testing are presented. Finally, the finite sample behavior of the bootstrap-based tests is illustrated by a simulation study.

2010 Mathematics Subject Classification. Primary 62E20, 62M07; secondary 62F40, 60G10.
Keywords and Phrases. Bootstrap, ergodicity, U-statistic, V-statistic, Cramér-von Mises-type test.
Short title. U- and V-statistics under ergodicity.

version: December 21, 2011

1. Introduction

Many important test statistics can be rewritten as or approximated by degenerate U- or V-statistics. Well-known examples are the Cramér-von Mises statistic, the Anderson-Darling statistic or the χ^2-statistic. In this paper we derive the limit distributions of U- and V-statistics based on random variables from a strictly stationary and ergodic process. While the asymptotics of these statistics have been already derived in the case of mixing random variables, there is no such result under the weaker assumption of ergodicity. The latter condition is partly motivated by processes of interest in statistics which are known to be ergodic but do not satisfy any of the usual mixing conditions. As an example, an L_2-test for the intensity function of a Poisson count model is discussed in Section 5.3 below. According to Neumann (2011), the underlying process is ergodic but not mixing in general. As in the majority of papers in the literature, we employ a spectral decomposition of the kernel to obtain an additive structure where we can use a central limit theorem to proceed to the limit. Most of the existing results have been derived under prerequisites that can hardly be checked in many applications; cf. Section 2 for details. Here, we avoid any of these high-level assumptions. This is achieved by a restriction to positive semidefinite kernels and by a condition slightly stronger than the usual degeneracy property. It can be seen from the examples presented in Section 5 that these conditions are often fulfilled in statistical applications. When composite hypotheses have to be tested, then an estimator of the parameter enters the statistic. It turns out that the effect of estimating the unknown parameter is asymptotically not negligible. We also derive the limit distribution in this case.

Although the limit variables have a simple structure as weighted sums of independent χ^2 variates, we cannot use these results to determine asymptotically correct critical values for tests of hypotheses. This is because the weights in the limiting variables are in most cases only implicitly given as eigenvalues of some integral equation which in turn depends on the distribution of the underlying random variables in a complicated way. Therefore, problems arise as soon as critical values for test statistics of U- and V-type have to be determined. The bootstrap offers a convenient way to circumvent these problems, see Arcones and Giné (1992), Dehling and Mikosch (1994) or Leucht and Neumann (2009) for the i.i.d. case. To the best of our knowledge, with the exception of Leucht (2011), bootstrap validity has not been studied when the observations are dependent. While Leucht imposed non-standard assumptions on the dependence structure of the underlying process, we use classical ergodicity here and prove consistency of general bootstrap methods. Moreover, we use techniques of proof that are completely different from those employed by Leucht (2011) which results from the fact that we do not have covariance inequalities at hand under our weaker assumptions regarding the dependence structure of the underlying process. Furthermore, our Lemma 4.2 is of interest on its own. There we extend well-known results on the convergence of Hilbert-Schmidt operators and the associated eigenvalues.

The paper is organized as follows. In the subsequent Section 2, the limit distributions of degenerate U- and V-statistics are established. Section 3 is dedicated to statistics with estimated parameters. We provide a general result on the consistency of bootstrap methods in Section 4 while three possible applications are presented in Section 5. Afterwards, in Section 6, we report the results of a small simulation study. All proofs are deferred to the concluding Section 7.

2. Asymptotic distributions of U- and V-statistics

Assume that we have observations $X_1, \ldots, X_n$ from a stationary and ergodic process $(X_t)_{t \in T}$ with values in $\mathbb{R}^d$. For convenience of presentation, we assume to have a two-sided sequence, i.e. $T=\mathbb{Z}$. In this section we are concerned with the asymptotic behavior of the U- and V-statistics

$$U_n = \frac{1}{n}\sum_{1\le s,t\le n,\, s\ne t} h(X_s,X_t) \qquad \text{and} \qquad V_n = \frac{1}{n}\sum_{s,t=1}^{n} h(X_s,X_t).$$

We impose the following assumption.

(A1) (i) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process with values in $\mathbb{R}^d$.
(ii) $h\colon \mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ is a symmetric, continuous and positive semidefinite function, i.e., $\forall\, c_1,\dots,c_m\in\mathbb{R}$, $x_1,\dots,x_m\in\mathbb{R}^d$ and $m\in\mathbb{N}$, $\sum_{i,j=1}^m c_i\, c_j\, h(x_i,x_j)\ \ge\ 0$.
(iii) $Eh(X_0,X_0)<\infty$.
(iv) $E(h(x,X_t)\mid X_1,\dots,X_{t-1})=0$ *a.s.* for all $x\in\operatorname{supp}(P^{X_0})$.

Remark 1. *(i)* Concerning the dependence structure we do not assume anything beyond ergodicity of the underlying process. It is well known that strong mixing implies ergodicity; see e.g. Remark 2.6 on page 50 in combination with Proposition 2.8 on page 51 in Bradley (2007). On the other hand, there exist interesting processes which are ergodic but not mixing. Andrews (1984) has shown that a stationary AR(1) process $(X_t)_{t\in\mathbb{Z}}$ given by $X_t=\theta X_{t-1}+\varepsilon_t$ with i.i.d. Bernoulli distributed innovations is not strongly mixing. However, ergodicity is preserved under taking functions of an ergodic process. If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process, $Y_t=g((\dots,\varepsilon_{t-1},\varepsilon_t),(\varepsilon_{t+1},\varepsilon_{t+2},\dots))$ for some Borel-measurable function g, then $(Y_t)_{t\in\mathbb{Z}}$ is also ergodic; see Proposition 2.10 on page 54 in Bradley (2007). Since the above autoregressive process can be represented as a linear process in the ε_t's, it follows that it is also ergodic. Another example of an ergodic and non-mixing process is considered in Section 5.3 below.
(ii) For certain classes of processes, it can be much easier to prove ergodicity rather than mixing. While a verification of mixing properties often requires advanced coupling techniques, the above example shows that one can sometimes get ergodicity almost for free. It is known that any sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ of i.i.d. random variables is ergodic. Hence, it is immediately clear that $(Y_t)_{t\in\mathbb{Z}}$ with $Y_t=g((\dots,\varepsilon_{t-1},\varepsilon_t),(\varepsilon_{t+1},\varepsilon_{t+2},\dots))$ is also ergodic.

Remark 2. *(i)* Within the proofs of the subsequent results, we require $Eh^2(X_0,\widetilde{X}_0)<\infty$, where $\widetilde{X}_0$ denotes an independent copy of X_0. Note that it follows from positive semidefiniteness of h that

$$h(x,x)h(y,y)-h^2(x,y)=\det\left(\begin{pmatrix} h(x,x) & h(x,y)\\ h(x,y) & h(y,y)\end{pmatrix}\right)\ge 0.$$

Therefore, we obtain under (A1) that $Eh^2(X_0,\widetilde{X}_0)\le (Eh(X_0,X_0))^2<\infty$.
(ii) Condition (A1)(iv) implies degeneracy of the kernel, i.e. $Eh(x,X_1)=0\ \forall\, x\in\operatorname{supp}(P^{X_0})$. We show in Section 5 that typical test statistics in time series analysis can be approximated by V-statistics with a kernel satisfying our condition (A1)(iv). This condition is not fulfilled in general when the classical Cramér-von Mises statistic is applied to dependent data.

There are two approaches in the literature to derive the limit distributions of U_n and V_n. If the $X_t,\ t\in\mathbb{Z}$, are i.i.d. real-valued random variables, one can express V_n in terms of the empirical process, that is, $V_n=\iint h(x,y)\,G_n(dx)\,G_n(dy)$, where $G_n(x)=\sqrt{n}(F_n(x)-F(x))$, $F_n(x)=n^{-1}\sum_{t=1}^n \mathbb{1}(X_t\le x)$ and $F(x)=P(X_t\le x)$. Then one can employ empirical process technology to derive the limit distribution that is in this case described as a multiple stochastic integral of the kernel under consideration, with respect to increments of a centered Gaussian process; see e.g. Babbel (1989) and Borisov and Bystrov (2006). Similar techniques have been invoked by Dehling and Taqqu (1991) to obtain the asymptotics of U-statistics under long-range dependence. We do not pursue this approach here since it can get quite difficult to prove the required tightness property of the empirical process if nothing beyond ergodicity is assumed; see Remark 1 in Fokianos and Neumann (2011).

The other approach consists of first approximating the U- and V-statistics by weighted sums of products of partial sums and then applying a central limit theorem (CLT) to these sums. The classical method is based on the spectral theorem for self-adjoint Hilbert-Schmidt operators; see Dunford and Schwartz (1963, p. 1087, Exercise 56). If $Eh^2(X_0, \widetilde{X}_0) < \infty$, we can represent the kernel h as

$$h(x, y) \;=\; \sum_k \lambda_k \, \Phi_k(x) \, \Phi_k(y). \tag{2.1}$$

Here, $(\lambda_k)_k$ is a possibly finite enumeration of the nonzero eigenvalues of the equation

$$E[h(x, X_0) \, \Phi(X_0)] \;=\; \lambda \, \Phi(x), \tag{2.2}$$

repeated according to their multiplicity, and $(\Phi_k)_k$ are associated orthonormal eigenfunctions, i.e. $E[\Phi_j(X_0)\Phi_k(X_0)] = \delta_{j,k}$. In the case of an infinite number of nonzero eigenvalues, convergence of the infinite series (2.1) has to be understood in the L_2-sense, that is,

$$E\left(h(X_0, \widetilde{X}_0) \;-\; \sum_{k=1}^{K} \lambda_k \, \Phi_k(X_0) \, \Phi(\widetilde{X}_0) \right)^2 \underset{K\to\infty}{\longrightarrow} 0.$$

This approach works perfectly well in the case of independent random variables since it is easy to show that U_n can be approximated by $U_n^{(K)}$, which denotes the U-statistic based on the underlying sample and the kernel

$$h^{(K)}(x, y) \;=\; \sum_{k=1}^{K} \lambda_k \, \Phi_k(x) \, \Phi_k(y).$$

The latter statistic can be written as

$$U_n^{(K)} \;=\; \sum_{k=1}^{K} \lambda_k \left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Phi_k(X_t) \right)^2 \;-\; \frac{1}{n} \sum_{t=1}^{n} \Phi^2(X_t) \right]$$

and the limit can be obtained by an application of a CLT and a law of large numbers to the inner sums; see e.g. Gregory (1977) and Serfling (1980).

This method has been adopted for mixing random variables by Eagleson (1979), Carlstein (1988) and Borisov and Volodko (2008) as well as for associated random variables by Dewan and Prakasa Rao (2001) and Huang and Zhang (2006). In the case of dependent random variables, however, this approach requires some care. As pointed out by Borisov and Volodko (2008), approximation (2.1) is valid for almost all $x, y \in \operatorname{supp}(P^{X_0})$. However, it is not guaranteed that the *joint* distribution of X_t and X_{t+h} is absolutely continuous with respect to $P^{X_0} \otimes P^{X_0}$, that is, (2.1) might fail on a set with nonzero measure. Borisov and Volodko (2008) have shown that this problem does not appear if an additional smoothness assumption on the kernel h is imposed what we therefore also do here. Furthermore, while a proof of the fact that

$$\limsup_{n\to\infty} \left\{ E\left(U_n \;-\; U_n^{(K)} \right)^2 \right\} \underset{K\to\infty}{\longrightarrow} 0$$

is very simple in the independent case, it can be much more cumbersome in the case of dependent random variables, particularly if only ergodicity is at our disposal. Mainly for this reason, the authors of the above named papers imposed conditions on the eigenvalues and eigenfunctions whose validity is quite difficult or even impossible to verify for many concrete examples in statistical hypothesis testing. To avoid these problems, Babbel (1989) and Leucht (2011) used a wavelet expansion of the kernel function. They obtained different representations of the limit variables than those appearing from a spectral decomposition and which are not suitable for our purposes. Moreover, we cannot adopt the corresponding approximations because of the lack of the respective covariance inequalities under ergodicity.

It turns out that under assumption (A1) the spectral decomposition of the kernel can be invoked successfully in the stationary and ergodic setting. Indeed, it follows from a version of Mercer's theorem given in Theorem 2 in Sun (2005) that (2.1) holds for all $x, y \in \operatorname{supp}(P^{X_0})$. Although we have now pointwise convergence in (2.1), we still have to show that U_n and V_n can actually be approximated by $U_n^{(K)}$ and $V_n^{(K)}$, respectively. Deviating from the above mentioned papers, we do not prove convergence in L_2 since this would require covariance estimates that are not available under ergodicity alone. At this point our assumption of h being positive semidefinite, together with the degeneracy condition (A1)(iv), proves to be of help. We will need in fact only a few monotonicity arguments and (A1)(iv) to show that

$$\sup_n \left\{ E|V_n - V_n^{(K)}| \right\} \underset{K\to\infty}{\longrightarrow} 0.$$

This approximation together with a CLT for sums of martingale differences allows us to derive the limit distribution of V_n. The limit of the corresponding U-statistic is then obtained by the ergodic theorem since $V_n - U_n = n^{-1} \sum_{t=1}^n h(X_t, X_t) \overset{a.s.}{\longrightarrow} Eh(X_0, X_0)$. The following theorem contains the asymptotic results on the U- and V-statistics.

Theorem 2.1. *Under the assumption (A1),*

$$V_n \overset{d}{\longrightarrow} Z := \sum_k \lambda_k Z_k^2 \qquad \text{and} \qquad U_n \overset{d}{\longrightarrow} Z - Eh(X_0, X_0),$$

as n tends to infinity. Here, $(Z_k)_k$ is a sequence of independent standard normal random variables and $(\lambda_k)_k$ denotes the sequence of nonzero eigenvalues of the equation (2.2), enumerated according to their multiplicity.

Remark 3. It can be seen from Theorem 2.1 that the limit distributions of V_n and U_n are the same as if $X_1, \ldots, X_n$ were independent and identically distributed. This is due to assumption (A1)(iv) which implies that $(\Phi_k(X_t))_{t\in\mathbb{N}}$ is a martingale difference sequence. In the case of i.i.d. random variables, it is well-known that these limits can be described by multiple stochastic integrals; see e.g. Section 1.4 in Dynkin and Mandelbaum (1983). Therefore, we could alternatively represent our limits by such integrals.

3. Approximation of test statistics of Cramér-von Mises type

Statistics of Cramér-von Mises type are an important tool for testing statistical hypotheses. While such a statistic has exactly the form of a V-statistic in the case of a simple null hypothesis, it can often be approximated by a degenerate V-statistic in the more relevant case of a composite null hypothesis. In Proposition 3.1 below we state a quite general approximation result for L_2-type statistics with estimated parameters. We consider kernels of the form

$$h(x, y, \theta) = \int_\Pi [h_1(x, z, \theta)]' h_1(y, z, \theta)\, Q(dz), \tag{3.1}$$

for some vector-valued function $h_1\colon \mathbb{R}^d\times\Pi\times\Theta\to\mathbb{R}^m$, $\Pi\subseteq\mathbb{R}^q$, $\Theta\subseteq\mathbb{R}^p$, and a probability measure Q. (Throughout this paper prime denotes the transposed of a vector.) The corresponding U- and V-statistics based on random variables $X_1,\ldots,X_n$ and an estimator $\widehat{\theta}_n$ of θ_0 are

$$U_n(\widehat{\theta}_n) \;=\; \frac{1}{n}\sum_{1\le s,t\le n,\, s\neq t} h(X_s,X_t,\widehat{\theta}_n) \qquad \text{and} \qquad V_n(\widehat{\theta}_n) \;=\; \frac{1}{n}\sum_{s,t=1}^{n} h(X_s,X_t,\widehat{\theta}_n).$$

Statistics of this type were considered by de Wet and Randles (1987) in the case of i.i.d. random variables. We present three specific applications in the context of stationary and ergodic processes in Section 5 below. We impose the following assumption.

(A2) (i) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process with values in $\mathbb{R}^d$.

(ii) The parameter estimator admits the expansion

$$\widehat{\theta}_n \,-\, \theta_0 \;=\; \frac{1}{n}\sum_{t=1}^{n} l_t \,+\, o_P(n^{-1/2}),$$

where $l_t = L(X_t, X_{t-1},\ldots)$ for some measurable function L, $E_{\theta_0}(l_t \mid X_{t-1}, X_{t-2},\ldots) = 0_p$ *a.s.* with 0_p denoting the p-dimensional vector of zeros, and $E_{\theta_0}\|l_t\|_2^2<\infty$.

(iii) The function h_1 satisfies $\int_\Pi \|h_1(x,z,\theta_0)\|_2^2\,Q(dz) < \infty$ for all $x\in\mathbb{R}^d$ and it holds $\int_\Pi \|h_1(x,z,\theta_0)-h_1(\bar{x},z,\theta_0)\|_2^2\,Q(dz) \longrightarrow_{\bar{x}\to x} 0$. Moreover, $E_{\theta_0}(h_1(X_t,z,\theta_0)\mid X_t,X_{t-1},\ldots) = 0_m$ *a.s.* $\forall z\in\Pi$ and $\int_\Pi E_{\theta_0}\|h_1(X_0,z,\theta_0)\|_2^2\,Q(dz)<\infty$.

(iv) The function h_1 is continuously differentiable w.r.t. θ in a neighborhood $\mathcal{U} = \{\theta\in\Theta\colon \|\theta-\theta_0\|_2<\delta\}$, $\delta>0$, of θ_0 for all $(x',z')'\in \operatorname{supp}(P^{X_0})\times\Pi$. Additionally $E_{\theta_0}\int_\Pi \|\dot{h}_1(X_0,z,\theta_0)\|_F^2\,Q(dz)<\infty$ and

$$E_{\theta_0}\left[\int_\Pi \sup_{\theta\colon \|\theta-\theta_0\|_2<\delta}\left\{\left\|\dot{h}_1(X_0,z,\theta) \,-\, \dot{h}_1(X_0,z,\theta_0)\right\|_F^2\right\} Q(dz)\right] \underset{\delta\to 0}{\longrightarrow} 0,$$

where $\|A\|_F := (\sum_{i=1}^m\sum_{j=1}^p |a_{i,j}|^2)^{1/2}$ denotes the Frobenius norm of a matrix $A=(a_{i,j})_{i=1,\ldots,m;\, j=1,\ldots,p}$.

The following proposition states that U- and V-statistics with kernels involving a parameter estimator can be approximated by statistics with appropriate fixed kernels.

Proposition 3.1. *Suppose that (A2) is fulfilled. Then*

$$U_n(\widehat{\theta}_n) \;=\; \widehat{U}_n \,+\, o_P(1) \quad \textit{with} \quad \widehat{U}_n \;=\; \frac{1}{n}\sum_{1\le s,t\le n,\, s\neq t}\widehat{h}(\bar{X}_s,\bar{X}_t)$$

and

$$V_n(\widehat{\theta}_n) \;=\; \widehat{V}_n \,+\, o_P(1) \quad \textit{with} \quad \widehat{V}_n \;=\; \frac{1}{n}\sum_{s,t=1}^{n}\widehat{h}(\bar{X}_s,\bar{X}_t),$$

where $\bar{X}_t = (X_t', l_t')'$ *and*

$$\begin{aligned}\widehat{h}(x,y) \;&=\; \int_\Pi \big(h_1(x_1,z,\theta_0) \,+\, E_{\theta_0}[\dot{h}_1(X_1,z,\theta_0)]\, x_2\big)' \\ &\qquad\times \big(h_1(y_1,z,\theta_0) \,+\, E_{\theta_0}[\dot{h}_1(X_1,z,\theta_0)]\, y_2\big)\, Q(dz).\end{aligned}$$

Remark 4. Due to the latter proposition, the asymptotic distributions of $U_n(\widehat{\theta}_n)$ and $V_n(\widehat{\theta}_n)$ coincide with those of $\widehat{U}_n$ and $\widehat{V}_n$. Their limit in turn can be deduced immediately from Theorem 2.1.

4. Bootstrap consistency

The limit distributions of degenerate U- and V-statistics depend on the eigenvalues of equation (2.1). The explicit derivation of these quantities is usually very complicated or even impossible. Moreover, they depend on an unknown parameter θ_0 in the setting of Section 3. Thus, quantiles of the (asymptotic) distributions of degenerate U- and V-statistics can hardly be determined directly. The bootstrap offers a suitable tool to circumvent these difficulties. Denote by $(X_t^*)_{t\in\mathbb{Z}}$ a bootstrap process which is constructed on the basis of the sample $X_1,\dots,X_n$ and introduced to mimic the unknown stochastic properties of $(X_t)_{t\in\mathbb{Z}}$. As usual, starred symbols such as P^* and E^* refer to the distribution associated with X_t^*, conditioned on $X_1,\dots,X_n$. In order to verify bootstrap consistency, we assume:

(B1) (i) The bootstrap process $(X_t^*)_{t\in\mathbb{Z}}$ is strictly stationary with probability tending to one and takes its values in $\mathbb{R}^d$. Additionally,

$$P^*\left(\sup_{\omega:\,\|\omega\|_2\le K}\left|\frac{1}{n}\sum_{t=1}^n e^{i\omega' X_t^*} - E_{\theta_0} e^{i\omega' X_t}\right| > \epsilon\right) \xrightarrow{P} 0 \qquad \forall K<\infty, \epsilon>0, \tag{4.1}$$

i.e. the empirical bootstrap measure converges weakly to P^{X_0} in probability.

(ii) The kernels of the bootstrap statistics $h^*\colon \mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ are symmetric, positive semidefinite, and equicontinuous on compacta in probability, i.e. $\forall\, K<\infty, \varepsilon>0, x_0,y_0\in\mathbb{R}^d$, $\exists\,\delta>0$ such that

$$P\left(\sup_{x_0,y_0:\,\|x_0\|_2,\|y_0\|_2\le K}\ \sup_{x,y:\,\|x-x_0\|_2,\|y-y_0\|_2\le\delta} |h^*(x,y) - h^*(x_0,y_0)| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

(iii) $h^*(x,y) \xrightarrow{P} h(x,y) \quad \forall\, x,y\in \operatorname{supp}(P^{X_0})$.

(iv) $E^* h^*(X_0^*,X_0^*) \xrightarrow{P} E h(X_0,X_0)$.

(v) $E^*(h^*(x,X_t^*) \mid X_{t-1}^*,\dots,X_1^*) = 0\ a.s. \quad \forall\, x\in\operatorname{supp}(P^{*X_0^*})$.

Remark 5. A verification of (4.1) seems to be difficult at first glance. However, suppose that, based on the underlying sample $\mathbb{X}_n=(X_1',\dots,X_n')'$, we can construct on an appropriate probability space $(\Omega,\mathcal{A},Q)$ versions $(\widetilde{X}_t^{(n)})_{t=1,\dots,n}$ and $(\widetilde{X}_t^{*(n)})_{t=1,\dots,n}$ of the processes $(X_t)_{t=1,\dots,n}$ and $(X_t^*)_{t=1,\dots,n}$ such that $Q^{(\widetilde{X}_t)_t}=P^{(X_t)_t}$, $Q^{(\widetilde{X}_t^*)_t}=P^{*(X_t^*)_t}$ and

$$Q\left(\frac{1}{n}\sum_{t=1}^n \min\left\{\|\widetilde{X}_t^* - \widetilde{X}_t\|_2, 1\right\} > \varepsilon\right) \xrightarrow{P} 0 \quad \forall\,\varepsilon>0 \ \text{ as } \ n\to\infty. \tag{4.2}$$

Then (4.1) is an immediate consequence. We show for the example of Poisson count processes in Section 5.3 below how such a coupling can actually be constructed. We think that this can also be done for other model-based bootstrap schemes when the processes satisfy a certain contractive condition. Furthermore, it might be expected that $X_t^* \xrightarrow{d} X_t$ in probability is some sort of minimal requirement for bootstrap consistency. It can be seen that this also follows from (4.2). The following lemma clarifies the connection between (4.1), (4.2), and the latter convergence result.

Lemma 4.1. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process on a probability space $(\Omega, \mathcal{A}, P)$.*

(i) *If $(X_t^{(n)})_{t\in\mathbb{Z}}$, $n \in \mathbb{N}$, are processes on $(\Omega, \mathcal{A}, P)$ with*

$$\frac{1}{n}\sum_{t=1}^{n} \min\left\{\|X_t^{(n)} \,-\, X_t\|_2, 1\right\} \stackrel{P}{\longrightarrow} 0,$$

then

$$P\left(\sup_{\omega:\ \|\omega\|_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n} e^{i\omega' X_t^{(n)}} \,-\, E e^{i\omega' X_0}\right| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall K<\infty, \epsilon>0. \tag{4.3}$$

(ii) *The relation (4.3) implies $X_0^{(n)} \stackrel{d}{\longrightarrow} X_0$.*

We introduce the bootstrap counterparts of the statistics U_n and V_n:

$$U_n^* \,=\, \frac{1}{n}\sum_{1\le s,t\le n,\, s\ne t} h^*(X_s^*, X_t^*) \qquad \text{and} \qquad V_n^* \,=\, \frac{1}{n}\sum_{s,t=1}^{n} h^*(X_s^*, X_t^*).$$

To derive the limit distributions of U_n^* and V_n^* we will again use a spectral decomposition of the corresponding kernel function h^*. We denote by $(\lambda_k^*)_k$ the sequence of nonzero eigenvalues of the equation $E^*[h^*(x, X_0^*)\,\Phi(X_0^*)] = \lambda\,\Phi(x)$, arranged in non-increasing order and according to multiplicity. Moreover let the eigenvalues $(\lambda_k)_k$ of $E[h(x, X_0)\,\Phi(X_0)] = \lambda\,\Phi(x)$ also be arranged in non-increasing order and according to multiplicity. It is well-known from functional analysis that the eigenvalues of two Hilbert-Schmidt operators converge if the corresponding kernels converge in $L_2(\Omega_X, \mathcal{A}_X, P_X)$. However, according to our knowledge, there is no such result if additionally the underlying Hilbert spaces vary, i.e., if we have $L_2(\Omega_X, \mathcal{A}_X, P_X^{(n)})$ instead of $L_2(\Omega_X, \mathcal{A}_X, P_X)$. The following result turns out to be crucial for proving bootstrap consistency.

Lemma 4.2. *Suppose that (A1) and (B1) are fulfilled. Then*

$$\sup_k |\lambda_k^* \,-\, \lambda_k| \stackrel{P}{\longrightarrow} 0.$$

Distributional convergence of the bootstrap statistics towards the limits of V_n and U_n implies bootstrap consistency if the limit distribution function is continuous. The latter property is ensured if $P(h(X_0, \widetilde{X}_0) \ne 0) > 0$. The following theorem summarizes the results concerning bootstrap consistency.

Theorem 4.1. *Under the assumptions (A1) and (B1),*

$$V_n^* \stackrel{d}{\longrightarrow} Z \qquad \text{and} \qquad U_n^* \stackrel{d}{\longrightarrow} Z - Eh(X_0, X_0) \quad \text{in probability.}$$

If additionally $P(h(X_0, \widetilde{X}_0) \ne 0) > 0$, then

$$\sup_{x\in\mathbb{R}} |P^*(U_n^* \le x) - P(U_n \le x)| \stackrel{P}{\longrightarrow} 0 \quad \text{and} \quad \sup_{x\in\mathbb{R}} |P^*(V_n^* \le x) - P(V_n \le x)| \stackrel{P}{\longrightarrow} 0.$$

As a last result in this section, we derive consistency for the bootstrap counterparts of the statistics with estimated parameters that have been investigated in Proposition 3.1 and Remark 4. We write $R_n^* = o_{P^*}(a_n)$ if $P^*(\|R_n^*\|_2/|a_n| > \varepsilon) \stackrel{P}{\longrightarrow} 0$, $\forall\,\varepsilon > 0$, and assume:

(B2) (i) The parameter estimator $\widehat{\theta}_n^* \in \Theta \subseteq \mathbb{R}^p$ admits the expansion

$$\widehat{\theta}_n^* - \widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^n l_t^* + o_{P^*}\left(n^{-1/2}\right),$$

where $l_t^* = L_{\widehat{\theta}_n}(X_t^*, X_{t-1}^*, \dots)$ for some measurable function $L_{\widehat{\theta}_n}$, $E^*(l_t^* \mid X_{t-1}^*, X_{t-2}^*, \dots) = 0_p$ *a.s.* and $E^*\|l_t^*\|_2^2 \xrightarrow{P} E\|l_t\|_2^2$.

(ii) (B1)(i) holds true for $(\widetilde{X}_t^*)_t$ with $\widetilde{X}_t^* = ((X_t^*)', (l_t^*)')'$.

(iii) The function h_1 satisfies $\sup_{\theta\in\mathcal{U}} \int_\Pi \|h_1(x,z,\theta)\|_2^2\, Q(dz) < \infty$ $\forall x \in \mathbb{R}^d$ and $\mathcal{U} = \{\theta \in \Theta\colon \|\theta-\theta_0\|_2 < \delta\}$, $\delta > 0$. It holds $\int_\Pi \|h_1(x,z,\theta) - h_1(\bar{x},z,\bar{\theta})\|_2^2\, Q(dz) \longrightarrow 0$ as $\bar{x} \to x$ and $\bar{\theta} \to \theta \in \mathcal{U}$. Moreover, $E^*(h_1(X_t^*, z, \widehat{\theta}_n) \mid X_{t-1}^*, \dots, X_1^*) = 0_m$, *a.s.* $\forall z \in \Pi$ and

$$\int_\Pi E^*\|h_1(X_1^*, z, \widehat{\theta}_n)\|_2^2\, Q(dz) \xrightarrow{P} \int_\Pi E_{\theta_0}\|h_1(X_1, z, \theta_0)\|_2^2\, Q(dz).$$

(iv) The function h_1 is continuously differentiable w.r.t. θ in $\mathcal{U}$ and

$$E^*\left[\int_\Pi \sup_{\theta\colon \|\theta-\widehat{\theta}_n\|_2<\delta} \left\{\left\|\dot{h}_1(X_1^*, z, \theta) - \dot{h}_1(X_1^*, z, \widehat{\theta}_n)\right\|_F^2\right\} Q(dz)\right] \xrightarrow[\delta\to 0]{P} 0,$$

$$\int_\Pi E^*\|\dot{h}_1(X_0^*, z, \widehat{\theta}_n)\|_F^2\, Q(dz) \xrightarrow{P} \int_\Pi E_{\theta_0}\|\dot{h}_1(X_0, z, \theta_0)\|_F^2\, Q(dz).$$

The function $\dot{h}_1$ satisfies $\sup_{\theta\in\mathcal{U}} \int_\Pi \|\dot{h}_1(x,z,\theta)\|_F^2\, Q(dz) < \infty$, $\forall x \in \mathbb{R}^d$. It holds $\int_\Pi \|\dot{h}_1(x,z,\theta) - h_1(\bar{x},z,\bar{\theta})\|_F^2\, Q(dz) \longrightarrow 0$ as $\bar{x} \to x$ and $\bar{\theta} \to \theta \in \mathcal{U}$.

Defining the bootstrap counterparts of $U_n(\widehat{\theta}_n)$ and $V_n(\widehat{\theta}_n)$ as

$$U_n^*(\widehat{\theta}_n^*) = \frac{1}{n}\sum_{1\le s,t\le n,\, s\ne t} h(X_s^*, X_t^*, \widehat{\theta}_n) \qquad \text{and} \qquad V_n^*(\widehat{\theta}_n^*) = \frac{1}{n}\sum_{s,t=1}^n h(X_s^*, X_t^*, \widehat{\theta}_n)$$

we obtain the following result.

Proposition 4.1. *Under the assumptions (A2) and (B2),*

$$V_n^*(\widehat{\theta}_n^*) \xrightarrow{d} \widehat{Z} := \sum_k \widehat{\lambda}_k Z_k^2 \qquad \textit{and} \qquad U_n^*(\widehat{\theta}_n^*) \xrightarrow{d} \widehat{Z} - E_{\theta_0}\widehat{h}(X_0, X_0) \qquad \textit{in probability,}$$

where $\widehat{h}$ is defined as in Proposition 3.1, $(Z_k)_k$ is a sequence of independent standard normal random variables and $(\widehat{\lambda}_k)_k$ denotes the sequence of nonzero eigenvalues of the equation $E_{\theta_0}[\widehat{h}(x, X_0)\Phi(X_0)] = \lambda\Phi(x)$, enumerated according to their multiplicity.
If additionally $P_{\theta_0}(\widehat{h}(X_0, \widetilde{X}_0) \ne 0) > 0$, then

$$\sup_{x\in\mathbb{R}} \left|P^*\big(U_n^*(\widehat{\theta}_n^*) \le x\big) - P_{\theta_0}\big(U_n(\widehat{\theta}_n) \le x\big)\right| \xrightarrow{P} 0$$

and

$$\sup_{x\in\mathbb{R}} |P^*(V_n^*(\widehat{\theta}_n^*) \le x) - P_{\theta_0}(V_n(\widehat{\theta}_n) \le x)| \xrightarrow{P} 0.$$

5. Applications

In this section we present three different goodness-of-fit tests with test statistics that can be approximated by V-statistics.

5.1. **A goodness-of-fit test for the conditional mean function of a time series.** Let $(X_t)_{t\in\mathbb{Z}}$ with $X_t = (Y_t', I_{t-1}')'$ be a strictly stationary and ergodic process with values in $\mathbb{R}^d \times \mathbb{R}^m$. In this part we are concerned with a test for the following problem:

$$\begin{aligned}
&\mathcal{H}_0\colon \quad E(Y_t|I_{t-1}) = f(I_{t-1},\theta_0) \qquad a.s. \text{ for some } \theta_0 \in \Theta \subseteq \mathbb{R}^p \quad \text{vs.}\\
&\mathcal{H}_1\colon \quad P(E(Y_t|I_{t-1}) \neq f(I_{t-1},\theta)) > 0 \quad \forall\,\theta \in \Theta.
\end{aligned}$$

There is already a great variety of specification tests regarding the conditional mean in the literature. For a comprehensive overview of approaches in the i.i.d. as well as in the time series case, we refer the reader to Escanciano (2007). Here, we consider the test statistic

$$\widehat{T}_n^{(1)} \;=\; \int_\Pi \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[Y_t - f(I_{t-1},\widehat{\theta}_n)\right] w(I_{t-1},z) \right\|_2^2 Q(dz),$$

where w denotes a certain weight function and $\widehat{\theta}_n$ is an estimator for the unknown parameter θ_0. This type of statistic has also been investigated by Escanciano (2007) in the case of real-valued response variables Y_t. However, since he invoked empirical process theory in order to establish the asymptotic null distribution, he had to impose a condition on the conditional distributions which is not fulfilled e.g. by Poisson count processes considered in Section 5.3 below. We avoid such a condition with our V-statistics approach which does not require a tightness proof for the underlying process. Fan and Li (1999) proposed a test based on a similar statistic. While their test statistic is based on a non-parametric estimator with vanishing bandwidth of the conditional mean function, we use the corresponding fixed-kernel estimator instead. In comparison to our work, the method of Fan and Li (1999) is more suitable to detect local alternatives with sharp peaks while it suffers from a loss of power against so-called Pitman alternatives, see also Fan and Li (2000) for a comparative overview.

(G1) (i) $(X_t)_{t\in\mathbb{Z}}$ with $X_t = (Y_t', I_{t-1}')'$ is a strictly stationary and ergodic process.
(ii) The sequence of parameter estimators $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ satisfies (A2)(ii) with $l_t = l(X_t,\theta_0)$.
(iii) The function f is continuous and continuously differentiable w.r.t. its second argument in a neighborhood $\mathcal{U}(\theta_0)$ of θ_0 and $E_{\theta_0}\|\dot{f}(I_0,\theta_0)\|_F^2 < \infty$. Moreover,

$$E_{\theta_0}\left[\sup_{\theta\colon \|\theta-\theta_0\|_2<\delta} \left\|\dot{f}(I_0,\theta) \;-\; \dot{f}(I_0,\theta_0)\right\|_F^2\right] \underset{\delta\to 0}{\longrightarrow} 0,$$

(iv) $E_{\theta_0}(Y_t - f(I_{t-1},\theta_0) \mid I_{t-1}, X_{t-1}, X_{t-2},\dots) = 0$ $a.s.$ and $E_{\theta_0}\|Y_t - f(I_{t-1},\theta_0)\|_2^2 < \infty$.
(v) The weight function $w\colon \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}$ is bounded and $\int_\Pi w(x,z)\,w(y,z)Q(dz)$ is continuous in x and y.

Note that the first moment assumption of (G1)(iv) implies (A1)(iv). It is equivalent to $\mathcal{H}_0$ if $E_{\theta_0}(Y_{t-1}|I_{t-1},X_{t-1},X_{t-2},\dots) = E_{\theta_0}(Y_{t-1}|I_{t-1})$. The subsequent result follows immediately from Remark 4.

Corollary 5.1. *Under (G1),*

$$\widehat{T}_n^{(1)} \xrightarrow{d} \sum_k \lambda_k^{(1)}\, Z_k^2,$$

where $(Z_k)_k$ is a sequence of i.i.d. standard normal random variables and $(\lambda_k^{(1)})_k$ is the sequence of nonzero eigenvalues of the equation $E_{\theta_0}[h^{(1)}(x,X_0)\,\Phi(X_0)] = \lambda\,\Phi(x)$ with $x =$

$(x_1', x_2')'$, $y = (y_1', y_2')' \in \mathbb{R}^d \times \mathbb{R}^m$ *and kernel*

$$h^{(1)}(x,y) \;=\; \int_\Pi \big\{(x_1 - f(x_2,\theta_0))\, w(x_2,z) - E_{\theta_0}[\dot f(I_0,\theta_0) w(I_0,z)]\, l(x,\theta_0)\big\}' \times \big\{(y_1 - f(y_2,\theta_0))\, w(y_2,z) - E_{\theta_0}[\dot f(I_0,\theta_0)\, w(I_0,z)]\, l(y,\theta_0)\big\}\, Q(dz).$$

It can be seen easily that under the alternative $\mathcal{H}_1$

$$\frac{1}{n}\,\widehat{T}_n^{(1)} \;=\; \int_\Pi \big\| E[(Y_1 - f(I_0,\theta_0))\, w(I_0,z)] \big\|_2^2\, Q(dz) + o_P(1).$$

if $E\|Y_1\|_2^2 < \infty$. Under additional conditions concerning the weight function, the leading term is strictly positive; see e.g. Bierens and Ploberger (1997) or Stinchcombe and White (1998). Thus under $\mathcal{H}_1$, we get

$$P\left(\widehat{T}_n^{(1)} > K\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad \forall\, K < \infty.$$

While Escanciano (2007) proposed a wild bootstrap method to determine critical values of the test, one can alternatively employ certain model-based procedures in view of our Proposition 4.1. The latter approach may perform better since the bootstrap counterparts of the observed process converge to the original ones which does not hold true for the wild bootstrap. For instance, the algorithm proposed by Leucht (2010) is applicable for some special cases of the present framework. In particular, the author verifies $X_1^* \xrightarrow{d} X_1$ in probability. In conjunction with a weak law of large numbers (Lemma 5.1 in Leucht (2011)), this implies the validity of (B1)(i).

5.2. **A goodness-of-fit test for the conditional distribution of Markovian time series.** It might happen that the conditional mean functions of two models coincide whereas their conditional distribution functions are essentially different. In what follows, we are concerned with the question whether the conditional distribution of a time series belongs to a certain parametric class. A review of the literature on this topic is given by Bierens and Wang (2011). Inspired by Neumann and Paparoditis (2008), who established a Kolmogorov-Smirnov-type test for the conditional distributions of AR(p) and ARCH(p) processes, we consider a test of Cramér-von Mises-type for the validity of certain Markovian time series models of order m. More precisely, based on

$$\mathcal{M} := \big\{P_{(X_t)_t} \mid P\left(X_t \in B \mid \sigma(X_s, s<t)\right) = P\left(X_t \in B \mid \sigma(X_s, t-m \le s < t)\right) \forall\, B \in \mathcal{B}, t \in \mathbb{Z}\big\}$$

and

$$\mathcal{M}_0 := \big\{P_{(X_t)_t} \mid X_t = G(\mathbb{X}_{t-1}, \varepsilon_t, \theta) \iff \varepsilon_t = H(\mathbb{X}_{t-1}, X_t, \theta) \sim F_\varepsilon \text{ i.i.d.}, \quad \theta \in \Theta \subseteq \mathbb{R}^p,\ \mathbb{X}_{t-1} := (X_{t-1}', \ldots, X_{t-m}')'\big\}$$

with known measurable functions G and H, G monotonically increasing in its second argument, the following problem is illuminated:

$$\mathcal{H}_0\colon\ P_{(X_t)_t} \in \mathcal{M}_0 \qquad \text{versus} \qquad \mathcal{H}_1\colon\ P_{(X_t)_t} \in \mathcal{M}\backslash\mathcal{M}_0.$$

Here, we suggest a test statistic of L_2-type based on the empirical cumulative distribution function:

$$\widehat{T}_n^{(2)} := \int_{\mathbb{R}^d\times\Pi} \Big\{\frac{1}{\sqrt{n}} \sum_{t=1}^n \Big[\mathbb{1}(X_t \preceq z_1) - F_\varepsilon(H(\mathbb{X}_{t-1}, z_1, \widehat{\theta}_n))\Big]\, w(\mathbb{X}_{t-1}, z_2)\Big\}^2 Q(dz),$$

where $z = (z_1', z_2')' \in \mathbb{R}^d \times \Pi$, $\Pi \subseteq \mathbb{R}^{dm}$. $\widehat{\theta}_n$ is an estimator for the unknown parameter θ_0 and Q is a probability measure on $\mathbb{R}^d \times \Pi$. ($x \preceq y$ means that $x_i \le y_i\ \forall i$.) In order to derive the asymptotics of the test statistic under the null, we assume:

(G2) (i) $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary, ergodic process in $\mathcal{M}$ with values in $\mathbb{R}^d$.
(ii) The sequence of parameter estimators satisfies (A2)(ii) with $l_t = l(X_t, \mathbb{X}_{t-1}, \theta_0)$.
(iii) F_ε is continuously differentiable.
(iv) H is continuous and partially continuously differentiable w.r.t. its third component in some neighborhood $\mathcal{U}(\theta_0)$ of θ_0. Additionally, $\int_{\mathbb{R}^d\times\Pi} E_{\theta_0}\|\dot F_\varepsilon(H(\mathbb{X}_0,z_1,\theta_0))\dot H(\mathbb{X}_0,z_1,\theta_0)\, w(\mathbb{X}_0,z_2)\|_F^2\, Q(dz) < \infty$ and

$$E_{\theta_0}\left[\int_{\mathbb{R}^d\times\Pi} \sup_{\theta:\,\|\theta-\theta_0\|_2<\delta} \left\{\Big\| \dot F_\varepsilon(H(\mathbb{X}_0,z_1,\theta))\dot H(\mathbb{X}_0,z_1,\theta) - \dot F_\varepsilon(H(\mathbb{X}_0,z_1,\theta_0))\dot H(\mathbb{X}_0,z_1,\theta_0)\Big\|_F^2 |w(\mathbb{X}_0,z_2)|\right\} Q(dz)\right] \underset{\delta\to 0}{\longrightarrow} 0.$$

(v) The weight function $w\colon \mathbb{R}^{dm}\times\Pi\to\mathbb{R}$ is continuous and bounded. The probability measure Q is absolutely continuous w.r.t. the Lebesgue measure.

Again by Remark 4, we obtain the limit distribution of $\widehat{T}_n^{(2)}$.

Corollary 5.2. *Suppose that (G2) holds. Then, under $\mathcal{H}_0$,*

$$\widehat{T}_n^{(2)} \xrightarrow{d} \sum_k \lambda_k^{(2)} Z_k^2,$$

where $(Z_k)_k$ is a sequence of i.i.d. standard normal random variables and $(\lambda_k^{(2)})_k$ is the sequence of nonzero eigenvalues of the equation $E_{\theta_0}[h^{(2)}(x,X_0)\,\Phi(X_0)] = \lambda\,\Phi(x)$ for $x = (x_1', x_2')'$, $y = (y_1', y_2')' \in \mathbb{R}^d\times\mathbb{R}^{dm}$ and with kernel

$$\begin{aligned} h^{(2)}(x,y) := \int_{\mathbb{R}^d\times\Pi} &\big\{[\mathbb{1}(x_1\preceq z_1) - F_\varepsilon(H(x_2,z_1,\theta_0))]\, w(x_2,z_2) \\ &\quad - E_{\theta_0}[\dot F_\varepsilon(H(\mathbb{X}_0,z_1,\theta_0))\,\dot H(\mathbb{X}_0,z_1,\theta_0)\, w(\mathbb{X}_0,z_2)]\, l(x,\theta_0)\big\} \\ &\times \big\{[\mathbb{1}(y_1\preceq z_1) - F_\varepsilon(H(y_2,z_1,\theta_0))]\, w(y_2,z_2) \\ &\quad - E_{\theta_0}[\dot F_\varepsilon(H(\mathbb{X}_0,z_1,\theta_0))\,\dot H(\mathbb{X}_0,z_1,\theta_0)\, w(\mathbb{X}_0,z_2)]\, l(y,\theta_0)\big\}\, Q(dz). \end{aligned}$$

Concerning the behavior of the test statistic under the alternative hypothesis, we have

$$\frac{1}{n}\widehat{T}_n^{(2)} = \int_{\mathbb{R}^d\times\Pi} \Big\{E\Big[\big(\mathbb{1}(X_t\preceq z_1) - F_\varepsilon(H(\mathbb{X}_{t-1},z_1,\theta_0)\big)w(\mathbb{X}_{t-1},z_2)\Big]\Big\}^2 Q(dz) + o_P(1).$$

Thus for suitably chosen weight functions w and probability measures Q, one obtains

$$P\left(\widehat{T}_n^{(2)} > K\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad \forall\, K<\infty.$$

To approximate critical values of the test, model-based bootstrap methods can be employed. Naturally, one would draw bootstrap innovations $(\varepsilon_t^*)_{t=1,\ldots,n}$ according to an estimator of F_ε. After determining an initial value $\mathbb{X}_0^*$, the bootstrap sample can be generated iteratively, $X_t^* = G(\mathbb{X}_{t-1}^*, \varepsilon_t^*, \widehat{\theta}_n)$. Then quantiles of the empirical distribution of

$$\widehat{T}_n^{(2)*} = \int_{\mathbb{R}^d\times\Pi} \Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^n \Big[\mathbb{1}(X_t^*\preceq z_1) - F_\varepsilon(H(\mathbb{X}_{t-1}^*, z_1, \widehat{\theta}_n^*))\Big]\, w(\mathbb{X}_{t-1}^*, z_2)\Big\}^2 Q(dz)$$

are used to estimate critical values, where $\widehat{\theta}_n^*$ denotes the bootstrap parameter estimator. Again Proposition 4.1 can be invoked to verify the validity of this algorithm under certain regularity conditions on the function G and the innovation distribution function F_ε. As two concrete examples we mention AR(p) and ARCH(p) bootstrap methods and refer the reader to Neumann and Paparoditis (2008) for details.

5.3. **A goodness-of-fit test for Poisson count processes.** Assume that observations $Y_0, \dots, Y_n$ are available, where $((Y_t, \lambda_t)')_{t\in\mathbb{Z}}$ is a strictly stationary process with

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t),$$

$\mathcal{F}_t = \sigma(Y_t, \lambda_t, Y_{t-1}, \lambda_{t-1}, \dots)$. We assume that

$$\lambda_t = f(\lambda_{t-1}, Y_{t-1}),$$

with a function $f\colon [0,\infty) \times \mathbb{N}_0 \to (0,\infty)$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Models of this type have been considered before e.g. by Rydberg and Shephard (2000), Streett (2000), Davis, Dunsmuir and Streett (2003), Ferland, Latour and Oraichi (2006), Fokianos, Rahbek and Tjøstheim (2009), Gao, King, Lu and Tjøstheim (2009), Fokianos and Tjøstheim (2010), Neumann (2011), and Fokianos and Neumann (2011). According to Theorem 2.1(i) and Theorem 3.1(iii) in Neumann (2011), the contractive condition (G3)(i) below ensures the existence of a unique strictly stationary and ergodic solution $((Y_t, \lambda_t)')_{t\in\mathbb{Z}}$ to the system of model equations above. However, these processes are not mixing in general; see Remark 3 of Neumann (2011) for a counterexample.

A first goodness-of-fit test for the problem

$$\mathcal{H}_0 : f \in \{f_\theta \mid \theta \in \Theta\} \qquad \text{versus} \qquad \mathcal{H}_1 : f \notin \{f_\theta \mid \theta \in \Theta\}$$

with $\Theta \subseteq \mathbb{R}^p$ based on the statistic $G_n = n^{-1/2} \sum_{t=1}^n [(Y_t - \widehat{\lambda}_t)^2 - Y_t]$ was discussed by Neumann (2011). Here, $\widehat{\theta}_n$ was any $\sqrt{n}$-consistent estimator of θ, $\widehat{\lambda}_1$ an arbitrary initial value, and, for $t = 2, \dots, n$, $\widehat{\lambda}_t = f_{\widehat{\theta}_n}(\widehat{\lambda}_{t-1}, Y_{t-1})$. Fokianos and Neumann (2011) proposed a Kolmogorov-Smirnov-type test based on the statistic $\sup_z \big| \sum_{t=1}^n ((Y_t - \widehat{\lambda}_t)/\sqrt{\widehat{\lambda}_t}) w(z - \widehat{I}_{t-1}) \big|$, where $\widehat{I}_s = (Y_s, \widehat{\lambda}_s)'$. They investigated the asymptotics of their test statistic employing empirical process theory and assumed the involved weight function w to be Lipschitz continuous in order to be able to prove tightness of the corresponding process. In particular, they failed to include the natural case of indicator weight functions; see Remark 1 in their paper.

Below we derive the limit distribution of an L_2-type statistic by means of our results on V-statistics, where we allow for a more general class of weight functions. We consider the following test statistic:

$$\widehat{T}_n^{(3)} = \int_\Pi \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n (Y_t - \widehat{\lambda}_t)\, w(z - \widehat{I}_{t-1}) \right\}^2 Q(dz),$$

where Q is a probability measure on $\Pi = \mathbb{N}_0 \times [0,\infty)$.

Proposition 3.1 and Theorem 2.1 cannot be used directly for deriving the limit distribution of $\widehat{T}_n^{(3)}$ since this statistic has not the structure required in Proposition 3.1. The estimated intensities $\widehat{\lambda}_t$ do not form a stationary process and since we do not want to assume that w is differentiable, we cannot simply treat the effect of estimating θ_0 by $\widehat{\theta}_n$ in $w(z - \widehat{I}_{t-1})$ by a direct Taylor expansion. It can be shown by backward iterations that, for given $(Y_t)_{t\in\mathbb{Z}}$ and $\theta \in \Theta$, the system of equations $\lambda_t = f_\theta(\lambda_{t-1}, Y_{t-1})$ $(t \in \mathbb{Z})$ has a unique stationary solution $(\lambda_t(\theta))_{t\in\mathbb{Z}}$, where $\lambda_t(\theta) = g_\theta(Y_{t-1}, Y_{t-2}, \dots)$ for some measurable function g_θ; see also the proof of Theorem 3.1 in Neumann (2011). In the technical part below we will use $\lambda_t(\widehat{\theta}_n)$ as a substitute for $\widehat{\lambda}_t$. Of course, we have that $\lambda_t = \lambda_t(\theta_0)$.

In order to show that the test statistic behaves asymptotically as a degenerate V-statistic and to derive its limit distribution on the basis of Theorem 2.1, we assume:

(G3) (i) $|f_\theta(\lambda, y) - f_\theta(\bar{\lambda}, \bar{y})| \leq \kappa_1|\lambda - \bar{\lambda}| + \kappa_2|y - \bar{y}|,\ \forall \lambda, \bar{\lambda} \geq 0,\ y, \bar{y} \in \mathbb{N}_0, \theta \in \Theta$ with $\|\theta - \theta_0\|_2 \leq \delta$ for some $\delta > 0$, $\kappa_1, \kappa_2 \geq 0$ and $\kappa := \kappa_1 + \kappa_2 < 1$.

(ii) $|f_\theta(\lambda, y) - f_{\theta_0}(\lambda, y)| \leq C\|\theta - \theta_0\|_2(\lambda + y + 1),\ \forall \lambda \geq 0, y \in \mathbb{N}_0, \theta \in \Theta$ with $\|\theta - \theta_0\|_2 \leq \delta$ for some $\delta > 0$.

(iii) $\lambda_t(\theta)$ is continuously differentiable w.r.t. θ in some neighborhood of θ_0 such that $E_{\theta_0}\dot{\lambda}_1^2(\theta_0) < \infty$ and $E_{\theta_0}[\sup_{\theta:\, \|\theta-\theta_0\|_2 \leq \delta} |\dot{\lambda}_1(\theta) - \dot{\lambda}_1(\theta_0)|^2] \longrightarrow 0$ as $\delta \to 0$.

(iv) The weight function w is non-negative, bounded, and satisfies

$$\sup_{y \in \mathbb{N}_0} \int_\Pi |w(z - (y, \lambda_1)') - w(z - (y, \lambda_2)')|^2\, Q(dz) = O(|\lambda_1 - \lambda_2|).$$

(v) The parameter estimator admits the expansion

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{t=1}^n l_t + o_P(n^{-1/2}),$$

where $l_t = L_{\theta_0}(Y_t, Y_{t-1}, \ldots)$ with some measurable function L_{θ_0}. Moreover, $E_{\theta_0}(l_t|\mathcal{F}_{t-1}) = 0_p$ *a.s.* and $E_{\theta_0}\|l_t\|_2^2 < \infty$.

Remark 6. (i) Assumptions (G3)(i) to (iii) are satisfied for instance in the case of linear Poisson autoregressions; see Fokianos and Neumann (2011).

(ii) (G3)(iv) is obviously satisfied if w is Lipschitz continuous. It is also satisfied if w is of bounded squared variation in its second argument, that is,

$$\sup_{y \in \mathbb{N}_0} \sup_{-\infty < \lambda_0 < \lambda_1 < \cdots < \lambda_M < \infty,\, M \in \mathbb{N}} \sum_{j=1}^M |w(y, \lambda_{j-1}) - w(y, \lambda_j)|^2 < \infty.$$

This includes the case where $w(z - I_{t-1}) = \mathbb{1}(I_{t-1} \preceq z)$.

(iii) (G3)(v) is satisfied for the conditional maximum likelihood estimator in the case of linear Poisson autoregressions; see e.g. Fokianos, Rahbek and Tjøstheim (2009). It is also fulfilled for the least squares estimator in the model $\lambda_t = \theta_1 - \theta_2 Y_{t-1}$; for more details see below.

The following lemma shows that the test statistic can be approximated by a statistic as required in Theorem 2.1.

Lemma 5.1. *Suppose that (G3) is satisfied. Then*

$$\widehat{T}_n^{(3)} = \frac{1}{n^2}\sum_{s,t=1}^n h^{(3)}(X_s, X_t) + o_P(1),$$

where $X_t = (Y_t, \lambda_t, I_{t-1}', l_t')', t \in \mathbb{N}$ *and*

$$h^{(3)}(x, y) = \int_\Pi \Big\{(x_1 - x_2)\, w(z - x_3) - [E_{\theta_0}(\dot{\lambda}_1(\theta_0)\, w(z - I_0))]' x_4\Big\} \times \Big\{(y_1 - y_2)\, w(z - y_3) - [E_{\theta_0}(\dot{\lambda}_1(\theta_0)\, w(z - I_0))]' y_4\Big\}\, Q(dz).$$

Corollary 5.3. *Suppose that (G3) is satisfied. Then, under $\mathcal{H}_0$,*

$$\widehat{T}_n^{(3)} \xrightarrow{d} \sum_k \lambda_k^{(3)} Z_k^2,$$

where $(Z_k)_k$ is a sequence of i.i.d. standard normal variables and $(\lambda_k^{(3)})_k$ is the sequence of nonzero eigenvalues of the equation $E_{\theta_0}[h^{(3)}(x, X_1)\, \Phi(X_1)] = \lambda\, \Phi(x)$.

Under $\mathcal{H}_1$, we obtain

$$\frac{1}{n}\, \widehat{T}_n^{(3)} = \int_\Pi \Big[E(Y_t - \lambda_t(\theta_0)) w(z - I_{t-1}(\theta_0)) \Big]^2 Q(dz) \,+\, o_P(1)$$

which in turn implies

$$P\left(\widehat{T}_n^{(3)} > K\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad \forall\, K < \infty$$

under suitable assumptions on w and Q.

Finally, we suggest a parametric bootstrap method to determine critical values of the test:

(1) Determine $\widehat{\theta}_n$.
(2) Generate an initial intensity λ_0^* and, for $t = 0, \ldots, n$, generate conditioned on $\lambda_0^*, \ldots, \lambda_t^*, Y_0^*, \ldots, Y_{t-1}^*$ counts $Y_t^* \sim \text{Poisson}(\lambda_t^*)$, where $\lambda_t^* = f_{\widehat{\theta}_n}(\lambda_{t-1}^*, Y_{t-1}^*)$.
(3) Determine $\widehat{\theta}_n^*$ such that $\widehat{\theta}_n^* - \widehat{\theta}_n = n^{-1} \sum_{t=1}^n l_t^* + o_{P^*}(1)$.
(4) Generate $\widehat{\lambda}_t^* = f_{\widehat{\theta}_n^*}(\widehat{\lambda}_{t-1}^*, Y_{t-1}^*),\ t \in \mathbb{N}$, with an arbitrary $\widehat{\lambda}_0^*$.
(5) Compute the bootstrap test statistic

$$\widehat{T}_n^{(3)*} \,=\, \int_\Pi \Big[\frac{1}{\sqrt{n}} \sum_{t=1}^n (Y_t^* - \widehat{\lambda}_t^*)\, w(z - \widehat{I}_{t-1}^*) \Big]^2 Q(dz).$$

We can again approximate the bootstrap version of the test statistic by a bootstrap V-statistic. Consistency will then follow from Theorem 4.1. In our simulations presented in the next section, we restricted our attention to a null hypothesis with

$$\lambda_t \,=\, \theta_1 \,+\, \theta_2 Y_{t-1},$$

where $\theta \in \Theta := \{(\theta_1, \theta_2)' \colon \theta_1 > 0, 0 \le \theta_2 < 1\}$. The model equation can be rewritten in form of a linear autoregressive model as

$$Y_t \,=\, \theta_1 \,+\, \theta_2 Y_{t-1} \,+\, \varepsilon_t,$$

where $\varepsilon_t = Y_t - \lambda_t$ satisfies $E(\epsilon_t \mid \mathcal{F}_{t-1}) = 0$ and $E(\epsilon_t^2 \mid \mathcal{F}_{t-1}) = \lambda_t$. Here, the unknown parameter $\theta = (\theta_1, \theta_2)'$ can be most easily estimated by least squares, i.e.,

$$\widehat{\theta}_n \,=\, \arg\min_{\theta\in\Theta} \sum_{t=1}^n \left(Y_t \,-\, \theta_1 \,-\, \theta_2 Y_{t-1}\right)^2.$$

It will be shown in the course of the proof of Lemma 5.2 below that

$$\widehat{\theta}_n^* \,-\, \widehat{\theta}_n \,=\, \frac{1}{n} \sum_{t=1}^n l_t^* \,+\, o_{P^*}(n^{-1/2}),$$

where $l_t^* = M^{-1}(Y_t^* - \lambda_t^*, Y_{t-1}^*(Y_t^* - \lambda_t^*))'$, $M = \begin{pmatrix} 1 & EY_0^* \\ EY_0^* & E(Y_0^*)^2 \end{pmatrix}$. Among the conditions summarized in (B1), (4.1) seems to be the most difficult one to check. We therefore conclude this section with an assertion that allows us to apply Lemma 4.1 which yields that (4.1) holds true.

Lemma 5.2. *Suppose that the conditions imposed above hold true. Then there exists a coupling of* $X_1, \ldots, X_n$ *and* $X_1^*, \ldots, X_n^*$ *on a probability space* $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ *such that*

$$\frac{1}{n}\sum_{t=1}^{n} \min\{\|X_t - X_t^*\|_2, 1\} \stackrel{\widetilde{P}}{\longrightarrow} 0.$$

Hence, we obtain bootstrap consistency as indicated above.

6. Simulations

We explored the finite sample behavior of our bootstrap-based tests by a few numerical examples. We considered the goodness-of-fit test for Poisson count models of Section 5.3 and we assumed that observations $Y_0, \ldots, Y_n$ from a strictly stationary process $((Y_t, \lambda_t))_{t\in\mathbb{Z}}$ are available, where

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t),$$

$\mathcal{F}_t = \sigma(Y_t, \lambda_t, Y_{t-1}, \lambda_{t-1}, \ldots)$. To keep the computational effort at a reasonable size, we restricted our attention to the simple case of $\lambda_t = f(Y_{t-1})$. The null model was a linear specification for f, that is,

$$H_0\colon \qquad f \in \{g\colon \ \ g(y) = \theta_1 + \theta_2 y \quad \text{for } (\theta_1, \theta_2)' \in \Theta\},$$

where $\Theta = \{(\theta_1, \theta_2)'\colon \ \ \theta_1 > 0, 0 \le \theta_2 < 1\}$. The parameter $\theta = (\theta_1, \theta_2)'$ was estimated by least squares and we used the test statistic

$$\widehat{T}_n^{(3)} = \int_{\mathbb{N}_0\times[0,\infty)} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - \widehat{\lambda}_t)\, w(z - (Y_{t-1}, \widehat{\lambda}_{t-1})') \right\}^2 Q(dz).$$

Our choice of $w(x_1, x_2) = \mathbb{1}_{(-1,1)}(x_1)\mathbb{1}_{(-1,1)}(x_2)$ and $Q = Q_1 \otimes Q_2$ with $Q_1 = \text{Poisson}(1)$ and $Q_2 = \text{Exp}(1)$ led to

$$\begin{aligned} \widehat{T}_n^{(3)} &= \frac{1}{n}\sum_{s,t=1}^{n} (Y_s - \widehat{\lambda}_s)(Y_t - \widehat{\lambda}_t) \sum_{k=0}^{\infty} \mathbb{1}_{(-1,1)}(k - Y_{s-1})\mathbb{1}_{(-1,1)}(k - Y_{t-1}) \frac{e^{-1}}{k!} \\ &\qquad\qquad \times \int_0^\infty \mathbb{1}_{(-1,1)}(\lambda - \lambda_{s-1})\mathbb{1}_{(-1,1)}(\lambda - \lambda_{t-1}) e^{-\lambda}\, d\lambda \\ &= \frac{1}{n}\sum_{s,t=1}^{n} (Y_s - \widehat{\lambda}_s)(Y_t - \widehat{\lambda}_t)\, \frac{\mathbb{1}(Y_{s-1} = Y_{t-1})}{e\, Y_{s-1}!} \\ &\qquad\qquad \times \max\left\{0, e^{-\max\{\widehat{\lambda}_{s-1}-1, \widehat{\lambda}_{t-1}-1, 0\}} - e^{-\min\{\widehat{\lambda}_{s-1}+1, \widehat{\lambda}_{t-1}+1\}}\right\}. \end{aligned}$$

To get some impression about the power, we considered three different alternatives,

$$\begin{aligned} H_1^{(1)}\colon &\qquad f(y) = \theta_1 + \theta_2\, y\, e^{-y^2}, \\ H_1^{(2)}\colon &\qquad f(y) = \theta_1 + \theta_2\,(1 - e^{-y}), \\ H_1^{(3)}\colon &\qquad f(y) = \theta_1 + \theta_2\, e^{-y}; \end{aligned}$$

see Figure 1. In all four cases, we have chosen $\theta_1 = 0.5$ and $\theta_2 = 0.7$. To obtain critical values, we generated 500 bootstrap samples according to the model-based method described in Section 5.3. Size and power were estimated on the basis of 500 simulation runs. The implementations were carried out with the aid of the statistical software package *R*; see R Development Core Team (2007). The results for nominal significance levels $\alpha = 0.05$ and 0.1 and sample sizes $n = 100, 200$ and 300 are shown in Table 1. It can be seen that the prescribed size is kept fairly well. The power behavior is very satisfactory for all of our

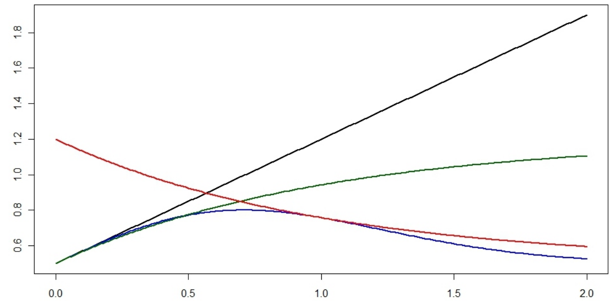

FIGURE 1. Intensity functions under H_0 (black), $H_1^{(1)}$ (blue), $H_1^{(2)}$ (green), and $H_1^{(3)}$ (red).

Table 1. Rejection frequencies

	$\alpha = 0.05$				$\alpha = 0.1$			
n	H_0	$H_1^{(1)}$	$H_1^{(2)}$	$H_1^{(3)}$	H_0	$H_1^{(1)}$	$H_1^{(2)}$	$H_1^{(3)}$
100	0.080	0.212	0.274	0.700	0.121	0.310	0.414	0.798
200	0.070	0.438	0.480	0.922	0.126	0.572	0.608	0.964
300	0.056	0.600	0.566	0.992	0.102	0.708	0.678	0.996

alternatives. Having a particular alternative in mind, the power can even be increased by a tailor-made choice of the weights w and Q; cf. Anderson and Darling (1954) in the case of generalized Cramér-von Mises statistics.

7. Proofs

Proof of Theorem 2.1. We denote by $(\lambda_k)_k$ an enumeration of the positive eigenvalues of (2.1) in decreasing order and according to their multiplicity and by $(\Phi_k)_k$ the corresponding eigenfunctions with $E[\Phi_j(X_0)\Phi_k(X_0)] = \delta_{jk}$. To avoid an explicit distinction of the cases whether the number of nonzero eigenvalues of (2.1) is finite or not, we set $\lambda_k := 0$ and $\Phi_k \equiv 0$, $\forall\, k > L$ if the number L of nonzero eigenvalues is finite. It follows from a version of Mercer's theorem (see Theorem 2 of Sun (2005) with $\mathcal{X} = \operatorname{supp}(P^{X_0})$) that

$$h^{(K)}(x,y) \;=\; \sum_{k=1}^{K} \lambda_k\; \Phi_k(x)\; \Phi_k(y) \;\underset{K\to\infty}{\longrightarrow}\; h(x,y) \qquad \forall\, x,y \in \operatorname{supp}(P^{X_0}). \tag{7.1}$$

The convergence of the series in (7.1) is absolute and uniform on compact subsets of $\operatorname{supp}(P^{X_0}) := \{x \in \mathbb{R}^d \mid \forall \text{open } O : x \in O \Rightarrow P^{X_0}(O) > 0\}$. The prerequisites of this result can be checked fairly easily here. Of course, P^{X_0} is nondegenerate on $\operatorname{supp}(P^{X_0})$ and there are compact sets $A_1 \subseteq A_2 \subseteq \ldots$ such that $\operatorname{supp}(P^{X_0}) = \cup_{n=1}^{\infty} A_n$. Assumption 1 of Sun (2005) is a consequence of $E|h(X_0,X_0)| < \infty$. Moreover, (A1) implies his Assumptions 2 and 3 in view of the Propositions 1 - 3 in that paper.

We define

$$V_n^{(K)} \;=\; \frac{1}{n} \sum_{s,t=1}^{n} h^{(K)}(X_s, X_t).$$

It follows from the non-negativity of the eigenvalues λ_k that

$$V_n \,-\, V_n^{(K)} \;=\; \frac{1}{n} \sum_{k=K+1}^{\infty} \lambda_k \left(\sum_{s=1}^{n} \Phi_k(X_s) \right)^2 \;\geq\; 0.$$

Therefore, from $EV_n = Eh(X_0, X_0) < \infty$ and (7.1) we obtain by majorized convergence that

$$E\big|V_n - V_n^{(K)}\big| = E\left[h(X_0,X_0) - h^{(K)}(X_0,X_0)\right] \underset{K\to\infty}{\longrightarrow} 0. \tag{7.2}$$

This means that we can actually approximate V_n by $V_n^{(K)}$. We rewrite this quantity as

$$V_n^{(K)} = \sum_{k=1}^{K} \lambda_k \, Z_{n,k}^2,$$

where $Z_{n,k} = n^{-1/2}\sum_{t=1}^n \Phi_k(X_t)$. Next we will show that

$$\begin{pmatrix} Z_{n,1} \\ \vdots \\ Z_{n,K} \end{pmatrix} \stackrel{d}{\longrightarrow} \begin{pmatrix} Z_1 \\ \vdots \\ Z_K \end{pmatrix} \sim \mathcal{N}(0_K, I_K), \tag{7.3}$$

where I_K denotes the $K \times K$ identity matrix and $K \le L$. By the Cramér-Wold device it suffices to show that, for arbitrary real $c_1, \ldots, c_K$,

$$\sum_{k=1}^{K} c_k \, Z_{n,k} \stackrel{d}{\longrightarrow} \sum_{k=1}^{K} c_k \, Z_k \sim \mathcal{N}\left(0, \sum_{k=1}^{K} c_k^2\right). \tag{7.4}$$

Let $Y_t = \sum_{k=1}^K c_k \, \Phi_k(X_t)$. We will show that the Y_t's satisfy the conditions of a CLT from McLeish (1974). It is clear that the process $(Y_t)_{t\in\mathbb{Z}}$ inherits the properties of stationarity and ergodicity from $(X_t)_{t\in\mathbb{Z}}$. Let $\mathcal{F}_t = \sigma(X_t, \ldots, X_1)$. It follows from (iv) of assumption (A1) that

$$\lambda_k \, E(\Phi_k(X_t) \mid \mathcal{F}_{t-1}) = \int E(h(x, X_t) \mid \mathcal{F}_{t-1}) \, \Phi_k(x) \, P^{X_0}(dx) = 0,$$

which implies that

$$E(Y_t \mid \mathcal{F}_{t-1}) = 0. \tag{7.5}$$

Furthermore, we have that $EY_t^2 = \sum_{j,k=1}^K c_j \, c_k \, E[\Phi_j(X_t)\Phi_k(X_t)] = \sum_{k=1}^K c_k^2$. Since $(Y_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, we obtain by the ergodic theorem (see e.g. Theorem 2.3 on page 48 in Bradley (2007)) that

$$\frac{1}{n}\sum_{t=1}^{n} Y_t^2 \stackrel{a.s.}{\longrightarrow} \sum_{k=1}^{K} c_k^2. \tag{7.6}$$

Finally, since $EY_t^2 < \infty$, we conclude that the Lindeberg condition is fulfilled, that is,

$$\frac{1}{n}\sum_{t=1}^{n} E\big[Y_t^2 \, \mathbb{1}(|Y_t| > \epsilon\sqrt{n})\big] \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall \epsilon > 0. \tag{7.7}$$

From (7.5) to (7.7) we see that the conditions of the CLT of McLeish (1974, Theorem 2.3) are fulfilled, see also his comments after this theorem. Therefore, (7.4) and also (7.3) hold true which implies by the continuous mapping theorem that

$$V_n^{(K)} \stackrel{d}{\longrightarrow} Z^{(K)} = \sum_{k=1}^{K} \lambda_k Z_k^2. \tag{7.8}$$

Finally, since $\sum_{k=1}^\infty \lambda_k = Eh(X_0, X_0) < \infty$, we obtain

$$E|Z - Z^{(K)}| = \sum_{k=K}^{\infty} \lambda_k \underset{K\to\infty}{\longrightarrow} 0. \tag{7.9}$$

Applying Theorem 3.2 from Billingsley (1999), we see that (7.2), (7.8) and (7.9) imply that

$$V_n \stackrel{d}{\longrightarrow} Z.$$

Using once more the ergodic theorem, we obtain

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t, X_t) \stackrel{a.s.}{\longrightarrow} Eh(X_0, X_0),$$

which proves that

$$U_n \stackrel{d}{\longrightarrow} Z \,-\, Eh(X_0, X_0).$$

□

Proof of Proposition 3.1. We consider the V-statistic first and decompose the integrand in $V_n(\widehat{\theta}_n)$ as follows:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h_1(X_t, z, \widehat{\theta}_n) \\
&= \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{h_1(X_t, z, \theta_0) \,+\, E_{\theta_0}[\dot{h}_1(X_1, z, \theta_0)]\; l_t\Big\} \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\dot{h}_1(X_t, z, \theta_0) \,-\, E_{\theta_0}[\dot{h}_1(X_1, z, \theta_0)]\Big)\; \frac{1}{n}\sum_{s=1}^{n} l_s \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(h_1(X_t, z, \widehat{\theta}_n) \,-\, h_1(X_t, z, \theta_0) \,-\, \dot{h}_1(X_t, z, \theta_0)(\widehat{\theta}_n \,-\, \theta_0)\Big) \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dot{h}_1(X_t, z, \theta_0)\;\left(\widehat{\theta}_n \,-\, \theta_0 \,-\, \frac{1}{n}\sum_{s=1}^{n} l_s\right) \\
&=: \quad S_n(z) \,+\, R_{n,1}(z) \,+\, R_{n,2}(z) \,+\, R_{n,3}(z).
\end{aligned} \tag{7.10}$$

Since

$$E_{\theta_0}\left[\int \|S_n(z)\|_2^2\, Q(dz)\right] \;=\; E_{\theta_0}\left[\int \Big\|h_1(X_1, z, \theta_0) \,+\, E_{\theta_0}[\dot{h}_1(X_1, z, \theta_0)]\; l_1\Big\|_2^2\; Q(dz)\right] \;<\; \infty,$$

it suffices to show that

$$T_{n,i} \;:=\; \int \|R_{n,i}(z)\|_2^2\, Q(dz) \;=\; o_P(1) \qquad \text{for } i = 1, 2, 3.$$

To prove negligibility of $T_{n,1}$, we will first show that

$$\frac{1}{n^2}\sum_{s,t=1}^{n} \widetilde{h}(X_s, X_t) \stackrel{P}{\longrightarrow} 0_{p\times p}, \tag{7.11}$$

where $0_{p\times p}$ denotes the $(p \times p)$ null matrix and

$$\widetilde{h}(x, y) \;=\; \int \Big(\dot{h}_1(x, z, \theta_0) \,-\, E_{\theta_0}[\dot{h}_1(X_1, z, \theta_0)]\Big)' \Big(\dot{h}_1(y, z, \theta_0) \,-\, E_{\theta_0}[\dot{h}_1(X_1, z, \theta_0)]\Big)\; Q(dz).$$

Denote by $\widetilde{h}^{(i,j)}(x, y)$ the (i, j)th entry of $\widetilde{h}(x, y)$ and let $\widetilde{h}_M^{(i,j)}(x, y) = (\widetilde{h}^{(i,j)}(x, y) \wedge M) \vee (-M)$. It follows from the ergodic theorem that

$$P_n \;:=\; n^{-1}\sum_{t=1}^{n} \delta_{X_t} \Longrightarrow P^{X_0} \quad a.s.,$$

which implies that $P_n \otimes P_n \Longrightarrow P^{X_0} \otimes P^{X_0}$ holds almost surely. Therefore, and since $\widetilde{h}_M^{(i,j)}$ is a bounded and continuous function we obtain that

$$\frac{1}{n^2} \sum_{s,t=1}^n \widetilde{h}_M^{(i,j)}(X_s, X_t) \;=\; \int \widetilde{h}_M^{(i,j)}\, dP_n \otimes P_n \;\overset{a.s.}{\longrightarrow}\; \int \widetilde{h}_M^{(i,j)}\, dP^{X_0} \otimes P^{X_0} \;=\; E_{\theta_0} \widetilde{h}_M^{(i,j)}(X_0, \widetilde{X}_0) \tag{7.12}$$

holds for all $M < \infty$. Furthermore, it follows from the Cauchy-Schwarz inequality that

$$|\widetilde{h}^{(i,j)}(x,y)| \;\leq\; \sqrt{\widetilde{h}^{(i,i)}(x,x)}\, \sqrt{\widetilde{h}^{(j,j)}(y,y)} \;\leq\; \widetilde{h}^{(i,i)}(x,x) \vee \widetilde{h}^{(j,j)}(y,y),$$

which implies by (A2)(iv) that

$$\begin{aligned}
&\sup_n \left\{ E_{\theta_0} \left| \frac{1}{n^2} \sum_{s,t=1}^n \widetilde{h}_M^{(i,j)}(X_s, X_t) \;-\; \widetilde{h}^{(i,j)}(X_s, X_t) \right| \right\} \\
&\leq \quad \sup_n \left\{ \frac{1}{n^2} \sum_{s,t=1}^n E_{\theta_0} \left[\left| \widetilde{h}^{(i,j)}(X_s, X_t) \right| \mathbb{1}(|\widetilde{h}^{(i,j)}(X_s, X_t)| > M) \right] \right\} \\
&\leq \quad E_{\theta_0} \left[\widetilde{h}^{(i,i)}(X_1, X_1)\; \mathbb{1}(\widetilde{h}^{(i,i)}(X_1, X_1) > M) \right] \\
&\qquad + E_{\theta_0} \left[\widetilde{h}^{(j,j)}(X_1, X_1)\; \mathbb{1}(\widetilde{h}^{(j,j)}(X_1, X_1) > M) \right] \\
&\underset{M \to \infty}{\longrightarrow} \quad 0.
\end{aligned}$$

This yields

$$E_{\theta_0} \widetilde{h}_M^{(i,j)}(X_0, \widetilde{X}_0) \underset{M \to \infty}{\longrightarrow} E_{\theta_0} \widetilde{h}^{(i,j)}(X_0, \widetilde{X}_0) \;=\; 0,$$

which implies in conjunction with (7.12) that (7.11) holds true. Since $n^{-1/2} \sum_{s=1}^n l_s = O_P(1)$ we obtain from (7.11) that

$$T_{n,1} \;=\; o_P(1). \tag{7.13}$$

From (A2)(iv) and $\widehat{\theta}_n - \theta_0 = O_P(n^{-1/2})$ we conclude

$$\begin{aligned}
T_{n,2} \;&=\; \frac{\|\widehat{\theta}_n \,-\, \theta_0\|_2^2}{n} \int \left(\sum_{t=1}^n \sup_{\theta\colon \|\theta - \theta_0\|_2 \leq \|\widehat{\theta}_n - \theta_0\|_2} \left\| \dot{h}_1(X_t, z, \theta) \,-\, \dot{h}_1(X_t, z, \theta_0) \right\|_F \right)^2 Q(dz) \\
&=\; o_P(1).
\end{aligned} \tag{7.14}$$

Moreover, we get from (A2)(ii) and (iv)

$$\begin{aligned}
T_{n,3} \;&\leq\; \frac{1}{n} \int \Big\| \sum_{t=1}^n \dot{h}_1(X_t, z, \theta_0) \Big\|_F^2 Q(dz) \;\times\; \Big\| \widehat{\theta}_n \,-\, \theta_0 \,-\, \frac{1}{n} \sum_{s=1}^n l_s \Big\|_2^2 \\
&=\; o_P(1).
\end{aligned} \tag{7.15}$$

Finally, (7.10) and (7.13) to (7.15) yield the approximation result for the V-statistic. The result for the U-statistic can be obtained in a similar manner. □

Proof of Lemma 4.1. Since $|e^{ix} - e^{iy}| \leq \min\{|x-y|, 2\}\ \forall x, y \in \mathbb{R}$, we obtain

$$\sup_{\omega\colon \|\omega\|_2 \leq K} \left| \frac{1}{n} \sum_{t=1}^n e^{i\omega' X_t^{(n)}} \;-\; \frac{1}{n} \sum_{t=1}^n e^{i\omega' X_t} \right| \;\leq\; \frac{1}{n} \sum_{t=1}^n \min\left\{ K\, \|X_t^{(n)} \,-\, X_t\|_2, 2 \right\} \;\overset{P}{\longrightarrow}\; 0.$$

Furthermore, since $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic, we obtain by Corollary 11.3.4 in Dudley (1989) that

$$\sup_{\omega:\,\|\omega\|_2\le K}\left|\frac{1}{n}\sum_{t=1}^n e^{i\omega' X_t} \,-\, Ee^{i\omega' X_t}\right| \stackrel{a.s.}{\longrightarrow} 0,$$

which completes the proof of (i).

It follows from (4.3) that we can choose for arbitrary $\epsilon>0$ a sequence $(K_n)_{n\in\mathbb{N}}$ with $K_n\to\infty$ as $n\to\infty$ such that

$$P\left(\sup_{\omega:\,\|\omega\|_2\le K_n}\left|\frac{1}{n}\sum_{t=1}^n e^{i\omega' X_t^{(n)}} \,-\, Ee^{i\omega' X_0}\right| > \epsilon\right) =:\, \delta_n \underset{n\to\infty}{\longrightarrow} 0$$

holds. Therefore we obtain by Jensen's inequality, for fixed $\omega\in\mathbb{R}^d$,

$$\begin{aligned}\left|Ee^{i\omega' X_0^{(n)}} \,-\, Ee^{i\omega' X_0}\right| \;&\le\; E\left|\frac{1}{n}\sum_{t=1}^n e^{i\omega' X_t^{(n)}} \,-\, Ee^{i\omega' X_0}\right| \\ &\le\; \epsilon \,+\, 2\,\delta_n \qquad \forall\, n \text{ with } \|\omega\|_2\le K_n.\end{aligned}$$

This implies that $Ee^{i\omega' X_0^{(n)}} \longrightarrow_{n\to\infty} Ee^{i\omega' X_0}$ $\forall\,\omega$ and hence $P^{X_0^{(n)}} \Longrightarrow P^{X_0}$. □

Proof of Lemma 4.2. We intend to invoke Corollary XI.9.4(a) from Dunford and Schwartz (1963, page 1090) to derive the desired estimate for the eigenvalues. In the case of two eigenvalue problems,

$$\int g(x,y)\,\Phi(x)\,\tau(dy) \;=\; \lambda\,\Phi(x) \qquad \text{and} \qquad \int g^*(x,y)\,\Phi(x)\,\tau(dy) \;=\; \lambda\,\Phi(x),$$

with symmetric, positive semi-definite functions g and g^*, denote by $(\mu_k)_{k\in\mathbb{N}}$ and $(\mu_k^*)_{k\in\mathbb{N}}$ the corresponding eigenvalues, arranged in decreasing order and repeated according to multiplicity. According to symmetry of the functions g and g^*, the associated Hilbert-Schmidt operators T and T^* are self-adjoint. Moreover, they are positive in view of the positive semi-definiteness of the kernels g and g^*. This implies that $(T_{adj}T)^{1/2} = |T| = T$ and $(T^*_{adj}T^*)^{1/2} = T^*$. Therefore, $(\mu_k)_{k\in\mathbb{N}}$ and $(\mu_k^*)_{k\in\mathbb{N}}$ coincide with the so-called characteristic numbers, i.e. the eigenvalues of $(T_{adj}T)^{1/2}$ and $(T^*_{adj}T^*)^{1/2}$. Now it follows from the corollary mentioned above and Lemma XI.6.2 of Dunford and Schwartz (1963, page 1010) that

$$\sup_k |\mu_k^* \,-\, \mu_k| \;\le\; \|T^* \,-\, T\|,$$

where $\|\cdot\|$ denotes the Hilbert-Schmidt norm satisfying

$$\|T^* \,-\, T\| \;=\; \sqrt{\int\!\!\int (g^*(x,y) \,-\, g(x,y))^2\,\tau(dx)\,\tau(dy)};$$

see Exercise XI.8.44 in Dunford and Schwartz (1963, page 1083).

To prove the assertion of the lemma, we have to compare the eigenvalues of the eigenvalue problems

$$\int h(x,y)\,\Phi(y)\,P^{X_0}(dy) \;=\; \lambda\,\Phi(x) \tag{7.16}$$

and

$$\int h^*(x,y)\,\Phi(y)\,P^{*X_0^*}(dy) \;=\; \lambda\,\Phi(x). \tag{7.17}$$

Unfortunately, the integrating measures, P^{X_0} and $P^{*X_0^*}$, are not the same and the above corollary cannot be applied directly. In what follows we replace the eigenvalue problems

(7.16) and (7.17) by equivalent ones with one and the same integrating measure. It follows from (B1)(i) that $X_0^* \stackrel{d}{\longrightarrow} X_0$, which implies $(X_0^*, \widetilde{X}_0^*) \stackrel{d}{\longrightarrow} (X_0, \widetilde{X}_0)$ in probability. Given the underlying sample $\mathbb{X}_n = (X_1', \ldots, X_n')'$, we can construct appropriate probability spaces $(\Omega, \mathcal{A}, Q)$ with random elements $(X, \widetilde{X})$ and $(X^*, \widetilde{X}^*)$ such that

$$Q^{(X,\widetilde{X})} = P^{(X_0,\widetilde{X}_0)}, \qquad\qquad Q^{(X^*,\widetilde{X}^*)} = P^{*(X_0^*,\widetilde{X}_0^*)}$$

and, because of $(X_0^*, \widetilde{X}_0^*) \stackrel{d}{\longrightarrow} (X_0, \widetilde{X}_0)$ in probability,

$$(X^*, \widetilde{X}^*) \stackrel{Q}{\longrightarrow} (X, \widetilde{X}) \quad \text{in } P\text{-probability} \tag{7.18}$$

according to the Skorohod representation theorem (Theorem 6.7 in Billingsley (1999, p. 70)). Here we can choose the canonical space, i.e. $\Omega = \operatorname{supp}\{P^{X_0}\} \times \operatorname{supp}\{P^{X_0}\} \times \operatorname{supp}\{P^{*X_0^*}\} \times \operatorname{supp}\{P^{*X_0^*}\}$ and $\mathcal{A} = \mathcal{B}^{4d}$. The convergence in (7.18) means in particular that the distributions $(Q^{(X^*,\widetilde{X}^*)})$ are tight in probability. Therefore, we conclude from (B1)(ii) that

$$Q\left(|h^*(X^*, \widetilde{X}^*) \,-\, h^*(X, \widetilde{X})| \,>\, \varepsilon\right) \stackrel{P}{\longrightarrow} 0 \qquad \forall \varepsilon > 0.$$

Furthermore, from (B1)(ii) and (iii),

$$Q\left(|h^*(X, \widetilde{X}) \,-\, h(X, \widetilde{X})| \,>\, \varepsilon\right) \stackrel{P}{\longrightarrow} 0 \qquad \forall \varepsilon > 0,$$

which yields

$$h^*(X^*, \widetilde{X}^*) \stackrel{Q}{\longrightarrow} h(X, \widetilde{X}) \quad \text{in } P\text{-probability}. \tag{7.19}$$

In view of (i) of Remark 2 and (B1)(iv), $([h^*(X^*, \widetilde{X}^*)]^2)$ is uniformly integrable in probability. This implies in conjunction with (7.19) that

$$E_Q\left(h^*(X^*, \widetilde{X}^*) \,-\, h(X, \widetilde{X})\right)^2 \stackrel{P}{\longrightarrow} 0. \tag{7.20}$$

Instead of (7.16) and (7.17), we consider the eigenvalue problems

$$\int h_Q(\omega, \nu)\, \Psi(\nu)\, Q(d\nu) \;=\; \lambda\, \Psi(\omega) \tag{7.21}$$

and

$$\int h_Q^*(\omega, \nu)\, \Psi(\nu)\, Q(d\nu) \;=\; \lambda\, \Psi(\omega), \tag{7.22}$$

where $h_Q(\omega, \nu) = h(X(\omega), \widetilde{X}(\nu))$ and $h_Q^*(\omega, \nu) = h^*(X^*(\omega), \widetilde{X}^*(\nu))$. It can be easily verified that (7.21) and (7.22) have the same eigenvalues with the same multiplicities as (7.16) and (7.17), respectively. To see this, suppose first that Φ is an eigenfunction of (7.16) to the eigenvalue λ, that is,

$$\int h(x, y)\, \Phi(y)\, P^{X_0}(dy) \;=\; \lambda\, \Phi(x).$$

Define $\Psi(\omega) = \Phi(X(\omega))$. Then

$$\begin{aligned}\int h_Q(\omega, \nu)\Psi(\nu)\, Q(d\nu) \;&=\; \int h(X(\omega), \widetilde{X}(\nu))\Phi(\widetilde{X}(\nu))\, Q(d\nu) \\ &=\; \int h(X(\omega), y)\Phi(y)\, P^{X_0}(dy) \;=\; \lambda\, \Phi(X(\omega)) \;=\; \lambda\, \Psi(\omega),\end{aligned}$$

that is, λ is also an eigenvalue of (7.21) and Ψ is a corresponding eigenfunction. Vice versa, suppose that Ψ is an eigenfunction of (7.21) to the eigenvalue λ, that is,

$$\int h_Q(\omega, \nu)\Psi(\nu)\, Q(d\nu) \;=\; \lambda\, \Psi(\omega).$$

Since $h_Q(\omega,\nu) = h_Q(\omega',\nu)$ if $X(\omega) = X(\omega')$, we obtain from the eigenvalue equation that $\Psi(\omega) = \Psi(\omega')$ if $X(\omega) = X(\omega')$. Therefore, we can define Φ on the range of X as

$$\Phi(x) \,=\, \Psi(\omega) \qquad \text{if } x = X(\omega).$$

Now it follows, for $x = X(\omega)$, that

$$\begin{aligned}\int h(x,y)\Phi(y)\, P^{X_0}(dy) \;&=\; \int h_Q(\omega,\nu)\Psi(\nu)\, Q(d\nu)\\ &=\; \lambda\, \Psi(\omega) \,=\, \lambda\, \Phi(x),\end{aligned}$$

that is, λ is also an eigenvalue of (7.16) and Φ is a corresponding eigenfunction. Hence, we obtain from Corollary XI.9.4(a) and Exercise XI.8.44 in Dunford and Schwartz (1963) that

$$\begin{aligned}\sup_k |\lambda_k^* \,-\, \lambda_k| \;&\leq\; \|T^* \,-\, T\|\\ &=\; \sqrt{\iint \Big(h_Q(\omega,\nu) \,-\, h_Q^*(\omega,\nu)\Big)^2\, Q(d\omega)\, Q(d\nu)}\\ &=\; \sqrt{E_Q\left(h^*(X^*,\widetilde{X}^*) \,-\, h(X,\widetilde{X})\right)^2},\end{aligned}$$

which yields in conjunction with (7.20) the assertion. □

Proof of Theorem 4.1. We denote by $(\lambda_k^*)_k$ an enumeration of the positive eigenvalues of $E^*[h^*(x,X_1^*)\Phi(X_1^*)] = \lambda\,\Phi(x)$ in nonincreasing order and according to their multiplicity and by $(\Phi_k^*)_k$ the corresponding orthonormal eigenfunctions. For sake of notational simplicity, we set $\lambda_k^* := 0$ and $\Phi_k^* \equiv 0$, $\forall\, k > L^*$ if the number L^* of nonzero eigenvalues is finite. Again Mercer's theorem (see Theorem 2 of Sun (2005), this time with $\mathcal{X} = \operatorname{supp}(P^{*X_0^*})$) yields

$$h^*(x,y) \,=\, \sum_{k=1}^{\infty} \lambda_k^*\, \Phi_k^*(x)\, \Phi_k^*(y), \quad \forall\, x,\, y \in \operatorname{supp}(P^{*X_0^*}).$$

We approximate this infinite series again by a finite one,

$$h^{*(K)}(x,y) \,=\, \sum_{k=1}^{K} \lambda_k^*\, \Phi_k^*(x)\, \Phi_k^*(y).$$

As an appropriate approximation to V_n^* we use

$$V_n^{*(K)} \,=\, \frac{1}{n} \sum_{s,t=1}^{n} h^{*(K)}(X_s^*, X_t^*).$$

First we show that the error of this approximation can be made arbitrarily small if K is sufficiently large. It follows from Lemma 4.2 that

$$\begin{aligned}E^* V_n^{*(K)} \;&=\; \sum_{k=1}^{K} \lambda_k^*\, E^*(\Phi_k^*(X_0^*))^2\\ &=\; \sum_{k=1}^{K} \lambda_k^* \;\stackrel{P}{\longrightarrow}\; \sum_{k=1}^{K} \lambda_k \,=\, EV_n^{(K)}.\end{aligned}$$

Furthermore, we obtain from assumption (B1)(iv) that

$$E^* V_n^* \,=\, E^* h^*(X_0^*, X_0^*) \;\stackrel{P}{\longrightarrow}\; Eh(X_0,X_0) \,=\, EV_n,$$

which implies in conjunction with $V_n^* \geq V_n^{*(K)}$ that

$$E^*|V_n^* - V_n^{*(K)}| = E^*[V_n^* - V_n^{*(K)}] \xrightarrow{P} \delta_K, \tag{7.23}$$

where $\delta_K := E|V_n - V_n^{(K)}| \longrightarrow_{K\to\infty} 0$; see the proof of Theorem 2.1. Hence, it suffices to study the asymptotics of $V_n^{*(K)}$. We will prove that

$$V_n^{*(K)} \xrightarrow{d} \sum_{k=1}^{K} \lambda_k Z_k^2 \qquad \text{in probability}, \tag{7.24}$$

where $Z_1, \ldots, Z_K$ are independent standard normal variables and $K \leq L$. By Lemma 4.2, the continuous mapping theorem and the Cramér-Wold device, (7.24) will follow from

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{K} Y_t^* \xrightarrow{d} \mathcal{N}\left(0, \sum_{k=1}^{K} c_k^2\right) \qquad \text{in probability}, \tag{7.25}$$

where $Y_t^* = \sum_{k=1}^{K} c_k \, \Phi_k^*(X_t^*)$ and $c_1, \ldots, c_K$ are arbitrary real numbers. In order to prove this, we verify that the conditions of the CLT of McLeish (1974) are fulfilled in probability.

Let $\mathcal{F}_t^* = \sigma(X_t^*, \ldots, X_1^*)$. Since

$$\lambda_k^* \, E^*(\Phi_k^*(X_t^*) \mid \mathcal{F}_{t-1}^*) = \int E^*[h^*(x, X_t^*) \mid \mathcal{F}_{t-1}^*] \Phi_k^*(x) \, P^{X_0^*}(dx) = 0 \quad a.s.,$$

we get

$$E(Y_t^* \mid \mathcal{F}_{t-1}^*) = 0 \qquad a.s.\,. \tag{7.26}$$

Next,

$$\frac{1}{n} \sum_{t=1}^{n} (Y_t^*)^2 - EY_0^2 = o_{P^*}(1) \tag{7.27}$$

has to be verified, which is equivalent to $n^{-1} \sum_{t=1}^{n} (Y_t^*)^2 - E^*[Y_0^*]^2 = o_{P^*}(1)$ since $EY_0^2 = E^*[Y_0^*]^2 = \sum_{k=1}^{K} c_k^2$. Using the representation $h^*(x,x) = \sum_k \lambda_k^* (\Phi_k^*(x))^2$ we conclude from the uniform integrability of $h^*(X_0^*, X_0^*)$ that $((\Phi_k^*(X_0^*))^2)_n$ is also uniformly integrable. Since additionally the bootstrap distributions and the empirical bootstrap distributions are tight with probability tending to one, it suffices to prove

$$\frac{1}{n} \sum_{t=1}^{n} \Phi_{k,l,M}^*(X_t^*) - E^* \Phi_{k,l,M}^*(X_0^*) = o_{P^*}(1) \quad \forall\, k, l = 1, \ldots, K, \; M \in \mathbb{N}, \tag{7.28}$$

where $\Phi_{k,l,M}^*(x) = [(\Phi_k^*(x)\Phi_l^*(x) \wedge M) \vee (-M)] \, \mathbb{1}(x \in [-M, M]^d)$. The basic idea of the proof is to establish a coupling between variables that obey the bootstrap and the empirical bootstrap law, respectively, such that their difference tends to zero. In order to verify (7.28), we then employ the equicontinuity (with probability tending to one) of $(\Phi_{k,l,M}^*)$ on $\operatorname{supp}(P^{*X_0^*}) \cap [-M, M]^d$. The latter property can be concluded from (B1)(ii) and the inequality

$$\begin{aligned}\lambda_k^* \, (\Phi_k^*(x) - \Phi_k^*(y))^2 &\leq \sum_{j=1}^{\infty} \lambda_j^* \, \left(\Phi_j^*(x) - \Phi_j^*(y)\right)^2 \\ &= h^*(x,x) - h^*(x,y) - h^*(y,x) - h^*(y,y), \quad x, y \in \operatorname{supp}(P^{*X_0^*}).\end{aligned}$$

Before constructing our coupling, we point out that the bootstrap distribution is a random measure depending on $\mathbb{X}_n = (X_1', \ldots, X_n')'$ and the empirical bootstrap distribution is random as well, depending on $\mathbb{X}_n^*(\mathbb{X}_n) = ((X_1^*(\mathbb{X}_n))', \ldots, (X_n^*(\mathbb{X}_n))')'$. For sake of clarity, we therefore introduce sequences of "favorable events" as follows: We choose a sequence of sets $(\mathfrak{X}_n)_{n\in\mathbb{N}}$ such that (h^*) is equicontinuous on $[-M,M]^{2d}\cap\text{supp}(P^{X_0^*|\mathbb{X}_n=x_n})\times\text{supp}(P^{X_0^*|\mathbb{X}_n=x_n})$ and $P^{X_0^*|\mathbb{X}_n=x_n}\Longrightarrow P^{X_0}$ uniformly for any sequence $(x_n)_n$ with $x_n\in\mathfrak{X}_n,\ n\in\mathbb{N}$. Similarly, $(\mathfrak{X}_n^*)_n$ are defined such that $P^{\bar{X}_{n,0}^*|\mathbb{X}_n^*(\mathbb{X}_n)=x_n^*(x_n)}\Longrightarrow P^{X_0}$ uniformly for all sequences $(x_n^*(x_n))_n$ and $(x_n)_n$ with $x_n^*\in\mathfrak{X}_n^*$ and $x_n\in\mathfrak{X}_n$. Here $\bar{X}_{n,0}^*$ is distributed according to the empirical bootstrap distribution conditionally on $\mathbb{X}_n^*(\mathbb{X}_n)$. According to (B1)(i),(ii) the sequences of sets $(\mathfrak{X}_n)_n$ and $(\mathfrak{X}_n^*)_n$ can be chosen such that $P(\mathbb{X}_n\in\mathfrak{X}_n)\longrightarrow 1$ and $P(\mathbb{X}_n^*\in\mathfrak{X}_n^*\mid\mathbb{X}_n=x_n)\longrightarrow 1$ as $n\to\infty$ uniformly for all $(x_n)_n$ with $x_n\in\mathfrak{X}_n$. Now we consider arbitrary but fixed sequences $(x_n)_n$ and $(x_n^*(x_n))_n$ as above. The coupling can be established following the lines of the proof of the Skorohod representation theorem. However, this result can not be applied directly since it is dedicated to derive *a.s.* convergence from weak convergence in the case of a fixed limit measure. In contrast we intend to construct a probability space $(\Omega,\mathcal{A},Q)$ such that there exist processes $(Y_n)_n$ and $(\bar{Y}_n)_n$ with $Q^{Y_n}=P^{X_0^*|\mathbb{X}_n=x_n}$, $Q^{\bar{Y}_n}=P^{\bar{X}_{n,0}^*|\mathbb{X}_n^*(\mathbb{X}_n)=x_n^*(x_n)}$ and $\bar{Y}_n(\omega)-Y_n(\omega)\longrightarrow 0,\ \forall\,\omega\in\Omega$. Proceeding as in the proof of Skorohod's representation theorem (see e.g. Theorem 6.7 in Billingsley (1999)), one can define $(\Omega,\mathcal{A},Q)$ as well as $(Y_n)_n$ and $(\bar{Y}_n)_n$ with the desired marginals such that $Y_n(\omega)\longrightarrow Y(\omega)$ and $\bar{Y}_n(\omega)\longrightarrow Y(\omega)$, $\forall\,\omega\in\Omega$, where $Q^Y=P^{X_1}$. This in turn yields the aforementioned convergence of $(\bar{Y}_n-Y_n)_n$. The latter relation finally implies

$$E_Q[\Phi_{k,l,M}^{(x_n)}(\bar{Y}_n)-\Phi_{k,l,M}^{(x_n)}(Y_n)]\le\varepsilon\,Q(\|\bar{Y}_n-Y_n\|_2\le\delta)+2M\,Q(\|\bar{Y}_n-Y_n\|_2>\delta)\underset{n\to\infty}{\longrightarrow}\varepsilon$$

for any $\varepsilon>0$ and suitably chosen $\delta=\delta(\varepsilon)>0$. Here, $\Phi_{k,l,M}^{(x_n)}$ denotes the version of $\Phi_{k,l,M}^*$ given $\mathbb{X}_n=x_n$. Summing up, (7.28) and thus (7.27) can be deduced.

Again from uniform integrability of $((\Phi_k^*(X_0^*))^2)_n$, we obtain that

$$\frac{1}{n}\sum_{t=1}^{n}E^*\left[(Y_t^*)^2\,\mathbb{1}(|Y_t^*|>\epsilon\sqrt{n})\right]\overset{P}{\longrightarrow}0\qquad\forall\epsilon>0,\tag{7.29}$$

that is, the Lindeberg condition is fulfilled in probability. From (7.26), (7.27) and (7.29) we conclude that we can apply Theorem 2.4 from McLeish (1974), which proves (7.25) and therefore also (7.24). From (7.23), (7.24) and (7.9) we obtain by Theorem 3.2 in Billingsley (1999) the first assertion,

$$V_n^*\overset{d}{\longrightarrow}Z\qquad\text{in probability.}$$

The latter relation also implies

$$U_n^*\overset{d}{\longrightarrow}Z-Eh(X_0,X_0)\qquad\text{in probability.}\tag{7.30}$$

if additionally $n^{-1}\sum_{t=1}^n h^*(X_t^*,X_t^*)-Eh(X_0,X_0)=o_{P^*}(1)$. Assumptions (B1)(i), (iv) yield that $(h^*(X_1^*,X_1^*))$ is uniformly integrable in probability. In view of (B1)(ii), we obtain $n^{-1}\sum_{t=1}^n[h^*(X_t^*,X_t^*)\wedge M]\vee(-M)-E[(h(X_0,X_0)\wedge M)\vee(-M)]=o_{P^*}(1)$ by invoking Corollary 11.3.4 of Dudley (1989). Combining both results, we get $n^{-1}\sum_{t=1}^n h^*(X_t^*,X_t^*)-Eh(X_0,X_0)=o_{P^*}(1)$ and eventually (7.30).

Finally, if additionally $P(h(X_0,\widetilde{X}_0)\neq 0)>0$, it follows from $\text{var}(Z)=2\sum_k\lambda_k^2>0$ that the random variable Z has a continuous distribution. In this case, the last assertion of Theorem 4.1 is an immediate consequence of the first one and of Theorem 2.1. □

Proof of Proposition 4.1. We consider the V-type statistics only since the corresponding results on U-statistics can be established in a similar manner. To show that

$$V_n^*(\widehat{\theta}_n^*) \;=\; \int \Big\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big\{ h_1(X_t^*, z, \widehat{\theta}_n) \;+\; E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, l_t^* \Big\} \Big\|_2^2 Q(dz) \;+\; o_{P^*}(1),$$

we invoke the bootstrap counterpart of the decomposition (7.10). Denote the bootstrap counterpart of S_n by S_n^*. Then, $E^* \int \|S_n^*(z)\|_2^2\, Q(dz)$ is bounded with probability tending to one due to our moment assumptions. In view of the continuity assumption in (B2)(iv), we obtain the counterpart of (7.12) by Corollary 11.3.4 from Dudley (1989). (Note that instead of $a.s.$ convergence, we have stochastic convergence here.) Now the verification of the analogues to (7.13), (7.14) and (7.15) is straightforward under (B2).

Therefore, we obtain the limits of $U_n^*(\widehat{\theta}_n)$ and $V_n^*(\widehat{\theta}_n)$ by Theorem 4.1 if

$$\begin{aligned} h^*(x,y) = \int & \Big\{ h_1(x_1, z, \widehat{\theta}_n) \;+\; E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, x_2 \Big\}' \\ & \times \Big\{ h_1(y_1, z, \widehat{\theta}_n) \;+\; E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, y_2 \Big\}\; Q(dz) \end{aligned}$$

satisfies (B1)(ii) to (v). Invoking the Cauchy-Schwarz inequality, (B1)(ii) results from (B2)(iii),(iv). Next, we have to show that $h^*(x,y) \xrightarrow{P} \widehat{h}(x,y)$, where $\widehat{h}$ is defined as in Proposition 3.1. It follows from (B2)(iii) and $\widehat{\theta}_n \xrightarrow{P} \theta_0$, that

$$\int \Big\{ h_1(x_1, z, \widehat{\theta}_n) \Big\}' \Big\{ h_1(y_1, z, \widehat{\theta}_n) \Big\}\; Q(dz) \xrightarrow{P} \int \{ h_1(x_1, z, \theta_0) \}' \{ h_1(y_1, z, \theta_0) \}\; Q(dz).$$

The desired convergence of $\int \{E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, x_2\}' \{E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, y_2\}\, Q(dz)$ follows from $(X_1^*, \widehat{\theta}_n) \xrightarrow{d} (X_1, \theta_0)$ and (B2)(iv). To this end, note that
$\int \{\dot{h}_1(x_1, z, \theta)\}' \{\dot{h}_1(x_2, z, \theta)\}\, Q(dz)$ is continuous in $(x_1', x_2', \theta')'$ under (B2)(iv). It remains to verify

$$\begin{aligned} & \int \Big\{ h_1(x_1, z, \widehat{\theta}_n) \Big\}' \Big\{ E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, y_2 \Big\}\; Q(dz) \\ & \xrightarrow{P} \int \{ h_1(x_1, z, \theta_0) \}' \Big\{ E_{\theta_0}[\dot{h}_1(X_0, z, \theta_0)]\, y_2 \Big\}\; Q(dz) \end{aligned}$$

On the one hand,

$$\int \Big\{ h_1(x_1, z, \widehat{\theta}_n) - h_1(x_1, z, \theta_0) \Big\}' \Big\{ E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)]\, y_2 \Big\}\; Q(dz) = o_P(1).$$

On the other hand, $\int \{h_1(x_1, z, \theta_0)\}' \{[E^*[\dot{h}_1(X_0^*, z, \widehat{\theta}_n)] \,-\, E_{\theta_0}[\dot{h}_1(X_0, z, \theta_0)]]\, y_2\}\, Q(dz)$ is asymptotically negligible under the moment assumptions of (B2)(iv) since $(X_1^*, \widehat{\theta}_n) \xrightarrow{d} (X_1, \theta_0)$ and because of the continuity of $\int \{h_1(x_1, z, \theta_0)\}' \{\dot{h}_1(y_1, z, \theta)\}\, Q(dz)$ in $(y_1', \theta')'$. Similarly we obtain (B1)(iv) from (B2)(iii),(iv) and the relation $(X_1^*, \widehat{\theta}_n) \xrightarrow{d} (X_1, \theta_0)$. The validity of (B1)(v) is obvious, which finally allows for the application of Theorem 4.1. □

Proof of Lemma 5.1. First, we show that the effect of choosing an arbitrary initial intensity $\widehat{\lambda}_0$ is asymptotically negligible. Let $\widetilde{T}_n^{(3)}$ be the statistic based on the stationary counterparts $\lambda_t(\widehat{\theta}_n)$ of $\widehat{\lambda}_t$, i.e.

$$\widetilde{T}_n^{(3)} \;=\; \int \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t \,-\, \lambda_t(\widehat{\theta}_n))\; w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \right)^2\; Q(dz),$$

where $I_t(\theta) = (Y_t, \lambda_t(\theta))'$. It follows from the contractive property (G3)(i) that

$$|\widehat{\lambda}_t \,-\, \lambda_t(\widehat{\theta}_n)| \,\leq\, \kappa_1^t \, |\widehat{\lambda}_0 \,-\, \lambda_0(\widehat{\theta}_n)|.$$

Therefore we obtain that

$$\widehat{T}_n^{(3)} \,=\, \widetilde{T}_n^{(3)} \,+\, o_P(1). \tag{7.31}$$

Now we decompose the integrand in $\widetilde{T}_n^{(3)}$ as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^n (Y_t \,-\, \lambda_t(\widehat{\theta}_n))\, w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \\
&= \quad \frac{1}{\sqrt{n}}\sum_{t=1}^n (Y_t \,-\, \lambda_t(\theta_0))\, w(z \,-\, I_{t-1}) - E_{\theta_0}[\dot{\lambda}_1(\theta_0) w(z - I_0)]\, l_t \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{t=1}^n (Y_t \,-\, \lambda_t(\theta_0)) \Big[w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \,-\, w(z \,-\, I_{t-1}) \Big] \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{t=1}^n (\lambda_t(\theta_0) \,-\, \lambda_t(\widehat{\theta}_n)) \Big[w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \,-\, w(z \,-\, I_{t-1}) \Big] \\
&\qquad - \frac{1}{\sqrt{n}}\sum_{t=1}^n \Big(\dot{\lambda}_t(\theta_0)\, w(z \,-\, I_{t-1}) \,-\, E_{\theta_0}[\dot{\lambda}_1(\theta_0)\, w(z \,-\, I_0)] \Big)\, \frac{1}{n}\sum_{s=1}^n l_s \\
&\qquad - \frac{1}{\sqrt{n}}\sum_{t=1}^n \Big(\lambda_t(\widehat{\theta}_n) \,-\, \lambda_t(\theta_0) \,-\, \dot{\lambda}_t(\theta_0)(\widehat{\theta}_n \,-\, \theta_0) \Big)\, w(z \,-\, I_{t-1}) \\
&\qquad - \frac{1}{\sqrt{n}}\sum_{t=1}^n \dot{\lambda}_t(\theta_0)\, \big(\widehat{\theta}_n \,-\, \theta_0 \,-\, \frac{1}{n}\sum_{s=1}^n l_s\big)\, w(z \,-\, I_{t-1}) \\
&=: \quad S_n^{(3)}(z) \,+\, R_{n,1}(z) \,+\, R_{n,2}(z) \,-\, R_{n,3}(z) \,-\, R_{n,4}(z) \,-\, R_{n,5}(z),
\end{aligned} \tag{7.32}$$

say. Since $\int (S_n^{(3)}(z))^2\, Q(dz) = O_P(1)$ by Theorem 2.1, it remains to show that

$$\int R_{n,i}^2(z)\, Q(dz) \,=\, o_P(1), \qquad \text{for } i = 1, \ldots, 5. \tag{7.33}$$

The main tool to estimate $\int R_{n,1}^2(z)\, Q(dz)$ will be the Bernstein-type inequality for martingales given in Proposition 2.1 in Freedman (1975). Since this inequality requires bounded random variables, we have to truncate $Y_t - \lambda_t$ and define $\xi_{n,t} = (Y_t - \lambda_t)\mathbb{1}(|Y_t - \lambda_t| \leq \sqrt{n}) - E_{\theta_0}((Y_t - \lambda_t)\mathbb{1}(|Y_t - \lambda_t| \leq \sqrt{n}) \mid \mathcal{F}_{t-1})$ and $\bar{\xi}_{n,t} = (Y_t - \lambda_t)\mathbb{1}(|Y_t - \lambda_t| > \sqrt{n})$. Then, with

$$W_{s,t}(\theta) \,=\, \int [w(z - I_{s-1}(\theta)) - w(z - I_{s-1})][w(z - I_{t-1}(\theta)) - w(z - I_{t-1})]\, Q(dz),$$

$$
\begin{aligned}
\int R_{n,1}^2(z)\, Q(dz) \quad \leq \quad & \frac{3}{n} \sum_{s,t=1}^n E_{\theta_0}((Y_s-\lambda_s)\mathbb{1}(|Y_s-\lambda_s| \leq \sqrt{n}) \mid \mathcal{F}_{s-1}) \\
& \qquad \times\; E_{\theta_0}((Y_t-\lambda_t)\mathbb{1}(|Y_t-\lambda_t| \leq \sqrt{n}) \mid \mathcal{F}_{t-1}) W_{s,t}(\widehat{\theta}_n) \\
& + \frac{3}{n} \sum_{s,t=1}^n \bar{\xi}_{n,s}\, \bar{\xi}_{n,t}\, W_{s,t}(\widehat{\theta}_n) \\
& + \frac{3}{n} \sum_{t=1}^n \xi_{n,t}^2\, W_{t,t}(\widehat{\theta}_n) \\
& + \frac{6}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s}\, W_{s,t}(\widehat{\theta}_n) \right\} \\
=: \quad & T_{n,1} \,+\, \cdots \,+\, T_{n,4}.
\end{aligned} \tag{7.34}
$$

Since $E_{\theta_0}((Y_t-\lambda_t)\mathbb{1}(|Y_t-\lambda_t| \leq \sqrt{n}) \mid \mathcal{F}_{t-1}) = -E_{\theta_0}((Y_t-\lambda_t)\mathbb{1}(|Y_t-\lambda_t| > \sqrt{n}) \mid \mathcal{F}_{t-1})$, we obtain that

$$
\begin{aligned}
& E_{\theta_0}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^n \left| E_{\theta_0}((Y_t-\lambda_t)\mathbb{1}(|Y_t-\lambda_t| \leq \sqrt{n}) \mid \mathcal{F}_{t-1}) \right| \right] \\
\leq \quad & \frac{1}{\sqrt{n}} \sum_{t=1}^n E_{\theta_0}\left[|Y_t-\lambda_t| \mathbb{1}(|Y_t-\lambda_t| > \sqrt{n}) \right] \\
\leq \quad & E_{\theta_0}\left[(Y_1-\lambda_1)^2 \mathbb{1}(|Y_t-\lambda_t| > \sqrt{n}) \right] \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

This implies that

$$
T_{n,1} \;\leq\; \frac{3\, \|w\|_\infty^2}{n} \left(\sum_{t=1}^n \left| E_{\theta_0}((Y_t-\lambda_t)\mathbb{1}(|Y_t-\lambda_t| \leq \sqrt{n}) \mid \mathcal{F}_{t-1}) \right| \right)^2 \;=\; o_P(1). \tag{7.35}
$$

Furthermore, we have

$$
\begin{aligned}
& P\left(|Y_t-\lambda_t| \;>\; \sqrt{n} \quad \text{ for some } t \in \{1,\ldots,n\} \right) \\
\leq \quad & \sum_{t=1}^n P\left(|Y_t-\lambda_t| \;>\; \sqrt{n} \right) \\
\leq \quad & E_{\theta_0}\left[(Y_1-\lambda_1)^2 \mathbb{1}(|Y_1-\lambda_1| > \sqrt{n}) \right] \underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

which leads to

$$
T_{n,2} \;=\; o_P(1). \tag{7.36}
$$

It follows from a repeated application of (G3)(i),(ii) that, if $\|\theta-\theta_0\|_2 \leq \delta$,

$$
|\lambda_t(\theta) \,-\, \lambda_t(\theta_0)| \;\leq\; C\, \|\theta \,-\, \theta_0\|_2 \sum_{k=1}^\infty \kappa_1^{k-1} (Y_{t-k} \,+\, \lambda_{t-k} \,+\, 1); \tag{7.37}
$$

see also inequality (5.13) in Neumann (2011). Since

$$
W_{s,t}(\theta) \;=\; O\left(\sqrt{|\lambda_{s-1}(\theta)-\lambda_{s-1}(\theta_0)|}\, \sqrt{|\lambda_{t-1}(\theta)-\lambda_{t-1}(\theta_0)|} \right), \tag{7.38}
$$

we obtain that

$$
T_{n,3} \;=\; O_P\left(\|\widehat{\theta}_n \,-\, \theta_0\|_2\, \frac{1}{n} \sum_{t=1}^n \left(\xi_{n,t}^2 \sum_{k=1}^\infty \kappa_1^{k-1} (Y_{t-k} \,+\, \lambda_{t-k} \,+\, 1) \right) \right) \;=\; o_P(1). \tag{7.39}
$$

The estimation of the term $T_{n,4}$ turns out to be much more delicate. The fact that $W_{s,t}(\widehat{\theta}_n)$ is of order $O_P(\|\widehat{\theta}_n - \theta_0\|_2)$ might suggest that $T_{n,4}$ is negligible, however, these weights depend via $\widehat{\theta}_n$ on the whole sample and we cannot use any standard inequality for sums of martingale differences directly. To proceed, we choose a sequence of increasingly fine grids $\Theta_n = \{\theta_{n,1}, \ldots, \theta_{n,M_n}\}$ on $\Theta \cap \{\theta : \|\theta - \theta_0\| \leq \gamma_n n^{-1/2}\}$, where $\gamma_n \longrightarrow_{n\to\infty} \infty$ and $\gamma_n = O(n^\gamma)$, for some $\gamma < 1/2$. Terms such as $\sum_{t=2}^n \xi_{n,t} \left\{\sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i})\right\}$ have now the desired martingale structure and we will show that

$$\max_{1\leq i\leq M_n} \left| \frac{1}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| = o_P(1). \tag{7.40}$$

In order to get a meaningful result, we will choose the grids sufficiently fine such that

$$\min_{1\leq i\leq M_n} \left| \frac{1}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\widehat{\theta}_n) \right\} - \frac{1}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| = o_P(1). \tag{7.41}$$

Then (7.40) and (7.41) eventually yield that

$$T_{n,4} = o_P(1). \tag{7.42}$$

To prove (7.40), we consider the sum of conditional variances,

$$V_n^2(\theta) = \frac{1}{n^2} \sum_{t=2}^n E\left(\left(\xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta) \right\} \right)^2 \middle| \mathcal{F}_{t-1} \right).$$

The Bernstein-type inequality given in Proposition 2.1 in Freedman (1975) yields that

$$\begin{aligned} &P\left(\left| \frac{1}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| > \nu_n\, a_n, \quad V_n(\theta_{n,i}) \leq \nu_n, \quad \max_{1\leq i\leq M_n} \{K_{n,i}\} \leq \mu_n \right) \\ &\leq\; 2\, \exp\left\{ -\frac{\nu_n^2\, a_n^2}{2\,(\mu_n\, \nu_n\, a_n + \nu_n^2)} \right\} = 2\, \exp(-a_n^2/4), \end{aligned} \tag{7.43}$$

where $K_{n,i} = \text{ess sup} \max_{1\leq t\leq n}\{|\xi_{n,t}|\; |n^{-1} \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i})|\}$ and $\mu_n = \nu_n/a_n$. Therefore, with the choice of $a_n = c\sqrt{\log n}$ with an appropriate c we obtain that

$$\begin{aligned} &\sum_{i=1}^{M_n} P\left(\left| \frac{1}{n} \sum_{t=2}^n \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| > \nu_n\, a_n, \quad V_n(\theta_{n,i}) \leq \nu_n, \quad \max_{1\leq i\leq M_n} \{K_{n,i}\} \leq \mu_n \right) \\ &= o(1). \end{aligned} \tag{7.44}$$

Next we have to choose ν_n such that $\nu_n\, a_n \to 0$ as $n \to \infty$ and

$$P\left(V_n(\theta_{n,i}) > \nu_n \quad \text{for some } i \in \{1, \ldots, M_n\}\right) = o(1). \tag{7.45}$$

Since $E(\xi_{n,t}^2 \mid \mathcal{F}_{t-1}) \leq E((Y_t - \lambda_t)^2 \mid \mathcal{F}_{t-1}) = \lambda_t$ we see that

$$\begin{aligned} V_n^2(\theta) &= \frac{1}{n^2} \sum_{t=2}^n \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta) \right\}^2 E\left(\xi_{n,t}^2 \mid \mathcal{F}_{t-1}\right) \\ &\leq \frac{1}{n^2} \sum_{t=2}^n \lambda_t \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta) \right\}^2. \end{aligned}$$

The terms in curly braces are not sums of martingale differences since the weights $W_{s,t}(\theta)$ depend on $I_{t-1}(\theta)$ and $I_{t-1}(\theta_0)$. Again by an approximation on a sequence of sufficiently fine grids, this time for $I_{t-1}(\theta_{n,i})$ and $I_{t-1}(\theta_0)$, we can show that with the choice of $\nu_n = n^{-\delta}$ for any $\delta \in (0, 1/2)$ (7.45) holds true. Finally, we get

$$P\left(\max_{1\le i\le M_n}\{K_{n,i}\} > \mu_n\right) = o(1).$$

This, (7.44) and (7.45) imply (7.42). (7.35), (7.36), (7.39) and (7.42) yield

$$\int R_{n,1}^2(z)\, Q(dz) = o_P(1), \tag{7.46}$$

The estimation of the remaining terms is much easier. It follows from (7.37) and (7.38) that

$$\int R_{n,2}^2(z)\, Q(dz) = o_P(1). \tag{7.47}$$

We obtain analogously to the proof of (7.13) that

$$\int R_{n,3}^2(z)\, Q(dz) = o_P(1). \tag{7.48}$$

Since $\lambda_t(\widehat{\theta}_n) - \lambda_t(\theta_0) = \dot{\lambda}_t(\widetilde{\theta}_{n,t})(\widehat{\theta}_n - \theta_0)$ for some $\widetilde{\theta}_{n,t}$ between θ_0 and $\widehat{\theta}_n$ we obtain that

$$\begin{aligned}
&\int R_{n,4}^2(z)\, Q(dz) \\
&\le \frac{\|\widehat{\theta}_n - \theta_0\|_2^2\, \|w\|_\infty}{n} \int \left(\sum_{t=1}^n \sup_{\theta:\, \|\theta-\theta_0\|_2 \le \|\widehat{\theta}_n-\theta_0\|_2} \|\dot{\lambda}_t(\theta) - \dot{\lambda}_t(\theta_0)\|_2\right)^2 Q(dz) \\
&= o_P(1).
\end{aligned} \tag{7.49}$$

Finally,

$$\begin{aligned}
\int R_{n,5}^2(z)\, Q(dz) &\le \frac{\|\widehat{\theta}_n - \theta_0 - n^{-1}\sum_{s=1}^n l_s\|_2^2\, \|w\|_\infty}{n} \int \left\|\sum_{t=1}^n \dot{\lambda}_t(\theta_0)\right\|_2^2 Q(dz) \\
&= o_P(1).
\end{aligned} \tag{7.50}$$

We see from (7.46) to (7.50) that (7.33) actually holds true, which completes the proof. □

Proof of Lemma 5.2. It suffices to construct a coupling such that, for arbitrary $\epsilon > 0$,

$$P\left(n^{-1}\sum_{t=1}^n \min\{\|X_t^* - X_t\|_2, 1\} > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{7.51}$$

The main difficulty arises from the fact that the parameter controlling the process $((Y_t^*, \lambda_t^*)')_{t\in\mathbb{Z}}$, $\widehat{\theta}_n$, is random. In this sense, we have to deal with a triangular scheme of processes and we cannot use, for example, the ergodic theorem here. To circumvent this problem, we construct a coupling of $((Y_t, \lambda_t)')_{t\in\mathbb{Z}}$ with a whole family of processes, with parameters θ in a neighborhood of θ_0.

As in Fokianos, Rahbek and Tjøstheim (2009), we draw all Poisson random variables from a family of independent Poisson processes $(N_t(\lambda))_{\lambda\in[0,\infty)}$, $t \in \mathbb{Z}$, with intensity functions equal to 1. For any parameter $\theta \in \Theta$, we contruct the corresponding process as follows. For any fixed $K \in \mathbb{Z}$, we begin the construction by setting $\lambda_K^{(K)} = 0$ and $Y_K^{(K)} = N_K(\lambda_K^{(K)}) = 0$. Provided $\lambda_K^{(K)}, Y_K^{(K)}, \ldots, \lambda_{t-1}^{(K)}, Y_{t-1}^{(K)}$ are defined, we define the next intensity according to the model equation, $\lambda_t^{(K)} = \theta_1 + \theta_2 Y_{t-1}^{(K)}$ and we set $Y_t^{(K)} = N_t(\lambda_t^{(K)})$. Finally, we define, for all $t \in \mathbb{Z}$, $\lambda_t = \lim_{K\to-\infty} \lambda_t^{(K)}$ and $Y_t = \lim_{K\to-\infty} Y_t^{(K)}$. Since the contractive property is fulfilled for $\theta \in \Theta$, these limits exist and $((Y_t, \lambda_t)')_{t\in\mathbb{Z}}$ is a stationary and ergodic version of the bivariate process with parameter θ. Since $\widehat{\theta}_n$ is a consistent estimator of θ_0, the bootstrap process $((Y_t^*, \lambda_t^*)')_{t\in\mathbb{Z}}$ can be sandwiched by two processes $((Y_t^+, \lambda_t^+)')_{t\in\mathbb{Z}}$ and $((Y_t^-, \lambda_t^-)')_{t\in\mathbb{Z}}$ that are sufficiently close to the original process. For some $\delta > 0$, we set $\theta_1^+ = \theta_{0,1} + \delta$, $\theta_2^+ = \theta_{0,2} + \delta$ and $\theta_1^- = \max\{\theta_{0,1} - \delta, 0\}$, $\theta_2^- = \max\{\theta_{0,2} - \delta, 0\}$. We denote by $((Y_t^+, \lambda_t^+)')_{t\in\mathbb{Z}}$ and $((Y_t^-, \lambda_t^-)')_{t\in\mathbb{Z}}$ those processes generated as decribed above, with parameters $(\theta_1^+, \theta_2^+)'$ and $(\theta_1^-, \theta_2^-)'$, respectively.

Let $K(\delta) = \max\{E|\lambda_t^+ - \lambda_t|, E|\lambda_t^- - \lambda_t|\}$. We can show that

$$K(\delta) \underset{\delta\to 0}{\longrightarrow} 0. \tag{7.52}$$

Now we obtain by the ergodic theorem that

$$P\left(n^{-1}\sum_{t=1}^n |\lambda_t^o - \lambda_t| \,>\, 2K(\delta)\right) \,+\, P\left(n^{-1}\sum_{t=0}^n |Y_t^o - Y_t| \,>\, 2K(\delta)\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{7.53}$$

where the index o stands for + or -.

Now we can make use of the fact that the random variables $\lambda_t = \lambda_t(\theta)$ and $Y_t = Y_t(\theta)$ are monotone in both components of θ. If $\theta_1^- \le \widehat{\theta}_{n,1} \le \theta_1^+$ and $\theta_2^- \le \widehat{\theta}_{n,2} \le \theta_2^+$, then we obtain from the above construction that $\lambda_t^- \le \lambda_t^* \le \lambda_t^+$ and $Y_t^- \le Y_t^* \le Y_t^+$ for all $t \in \mathbb{Z}$. Since the least squares estimator is consistent, we have

$$P\left(\theta_1^- \le \widehat{\theta}_{n,1} \le \theta_1^+ \quad \text{and} \quad \theta_2^- \le \widehat{\theta}_{n,2} \le \theta_2^+\right) \underset{n\to\infty}{\longrightarrow} 1,$$

which yields that

$$P\left(n^{-1}\sum_{t=1}^n |\lambda_t^* - \lambda_t| \,>\, 2K(\delta)\right) \,+\, P\left(n^{-1}\sum_{t=0}^n |Y_t^* - Y_t| \,>\, 2K(\delta)\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{7.54}$$

As for the l_t^*, note first that it follows from the above calculations that $M^* = \begin{pmatrix} 1 & n^{-1}\sum_{t=1}^n Y_{t-1}^* \\ n^{-1}\sum_{t=1}^n Y_{t-1}^* & n^{-1}\sum_{t=1}^n (Y_{t-1}^*)^2 \end{pmatrix}$ converges in probability to the matrix M. Therefore, with a probability tending to 1, $\widehat{\theta}_n^*$ has an explicit representation as

$$\widehat{\theta}_n^* \,=\, (M^*)^{-1}\,\frac{1}{n}\sum_{t=1}^n \begin{pmatrix} Y_t^* \\ Y_{t-1}^* Y_t^* \end{pmatrix},$$

which implies that

$$\widehat{\theta}_n^* \,-\, \widehat{\theta}_n \,=\, (M^*)^{-1}\,\frac{1}{n}\sum_{t=1}^n \begin{pmatrix} Y_t^* - \lambda_t^* \\ Y_{t-1}^*(Y_t^* - \lambda_t^*) \end{pmatrix}.$$

It follows from Proposition 6 of Ferland, Latour and Oraichi (2006) that all moments of Y_t^* are bounded in probability. Therefore, we obtain that $E^*(n^{-1/2}\sum_{t=1}^n (Y_t^* - \lambda_t^*))^2 = E^*(Y_1^* - \lambda_1^*)^2 = O_P(1)$ and $E^*(n^{-1/2}\sum_{t=1}^n Y_{t-1}^*(Y_t^* - \lambda_t^*))^2 = E^*[(Y_0^*)^2(Y_1^* - \lambda_1^*)^2] = O_P(1)$. Hence, we can replace the random matrix M^* by its nonrandom limit and obtain

$$\widehat{\theta}_n^* \,-\, \widehat{\theta}_n \,=\, \frac{1}{n}\sum_{t=1}^n l_t^* \,+\, o_{P^*}(n^{-1/2}),$$

with $l_t^* = M^{-1}\begin{pmatrix} Y_t^* - \lambda_t^* \\ Y_{t-1}^*(Y_t^* - \lambda_t^*) \end{pmatrix}$. Now we can deduce from (7.54) that

$$P\left(n^{-1}\sum_{t=1}^n \min\{\|l_t^* - l_t\|_2, 1\} \,>\, \widetilde{K}(\delta)\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{7.55}$$

for some $\widetilde{K}(\delta) \underset{\delta\to\infty}{\longrightarrow} 0$. (7.54) and (7.55) imply (7.51), which completes the proof. □

Acknowledgment . This research was funded by the German Research Foundation DFG, projects NE 606/2-1 and NE 606/2-2.

References

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *J. Amer. Statist. Assoc.* **49**, 765–769.

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934.

Arcones, M. A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20**, 655–674.

Babbel, B. (1989). Invariance principles for U-statistics and von Mises functionals. *J. Statist. Plann. Inference* **22**, 337–354.

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129–1151.

Bierens, H. J. and Wang, L. (2011). Integrated conditional moment tests for parametric conditional distributions. *Econometric Theory* (to appear) DOI: 10.1017/S0266466611000168.

Billingsley, P. (1999). *Convergence of Probability Measures.* 2nd Edition. New York: Wiley.

Borisov, I. S. and Bystrov, A. A. (2006). Limit theorems for the canonical von Mises statistics with dependent data. *Sibirsk. Mat. Zh.* **47**, 1205–1217. [English translation: *Siberian Math. J.* **47**, 980–989.]

Borisov, I. S. and Volodko, N. V. (2008) Orthogonal series and limit theorems for canonical U- and V-statistics of stationary connected observations. *Siberian Advances in Mathematics* **18**, 242–257.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions, Volume I.* Kendrick Press.

Carlstein, E. (1988). Degenerate U-statistics based on non-independent observations. *Calcutta Statist. Assoc. Bull.* **37**, 55–65.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90**, 777–790.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51**, 392–413.

Dehling, H. and Taqqu, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference* **28**, 153–165.

Dewan, I. and Prakasa Rao, B. L. S. (2001). Asymptotic normality for U-statistics of associated random variables. *J. Statist. Plann. Inference* **97**, 201–225.

de Wet, T. and Randles, R. H. (1987). On the effect of substituting parameter estimators in limiting χ^2 U and V statistics. *Ann. Statist.* **15**, 398–412.

Dudley, R. M. (1989). *Real Analysis and Probability.* New York: Chapman and Hall.

Dunford, N. and Schwartz, J. T. (1963). *Linear Operators, Part II, Spectral Theory. Self Adjoint Operators in Hilbert Space.* New York: John Wiley and Sons.

Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11**, 739–745.

Eagleson, G. K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* **21**, 221–237.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica* **17**, 115–138.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Stat.* **10**, 245–271.

Fan, Y. and Li, Q. (2000). Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16**, 1016–1041.

Fokianos, K. and Neumann, M. H. (2011). A goodness-of-fit test for Poisson count processes. Preprint.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis* **27**, 923–942.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104**, 1430–1439.

Fokianos, K. and Tjøstheim, D. (2010). Log-linear Poisson autoregression. *J. Multivariate Anal.* **102**, 563–578.

Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3**, 100–118.

Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009). Specification testing in nonlinear and nonstationary time series regression. *Annals of Statistics* **37**, 3893–3928.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5**, 110–123.

Huang, W. and Zhang, L.-X. (2006). Asymptotic normality for U-statistics of negatively associated random variables. *Statist. Probab. Lett.* **76**, 1125–1131.

Leucht, A. (2010). Consistent model-specification tests based on parametric bootstrap. *Reports of the Department of Mathematics and Computer Science, Friedrich Schiller University Jena* **10-07**.

Leucht, A. (2011). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli*, forthcoming.

Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U- and V-type statistics. *J. Multivariate Anal.* **100**, 1622–1633.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 620–628.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17**, 1268–1284.

Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap approximations. *Bernoulli* **14**, 14–46.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rydberg, T. H. and Shephard, N. (2000). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. World Conference Econometric Society, 2000, Seattle, Contributed Paper 0740.

Serfling, R. S. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Stinchcombe, M. B. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14**, 295–325.

Streett, S. (2000). Some observation driven models for time series of counts. Ph.D. Thesis, Colorado State University, Department of Statistics.

Sun, H. (2005). Mercer theorem for RKHS on noncompact sets. *J. Complexity* **214**, 337–349.