Consistent model-specification tests based on parametric bootstrap

Anne Leucht
Friedrich-Schiller-Universität Jena
Institut für Stochastik
Ernst-Abbe-Platz 2
D-07743 Jena
Germany
E-mail: anne.leucht@uni-jena.de

Abstract

In this paper we establish consistent tests of L_2-type for the parametric functional form of the conditional mean of time series with values in $\mathbb{R}^d$. A recent result on limit distributions of U-statistics of weakly dependent observations is invoked to obtain the asymptotics of the test statistics. Since the limit distributions depend on unknown parameters in a complicated way, we suggest to apply certain parametric bootstrap methods in order to determine critical values of the tests.

2000 Mathematics Subject Classification. 62E20, 62F40.
Keywords and Phrases. Bootstrap; model-specification; U-statistics; time series.
Short title. Consistent model-specification tests.

1. Introduction

An overwhelming amount of statistical and econometrics literature is dedicated to consistent model specification tests. An extensive list of tests in the i.i.d. context is provided by Escanciano [8]. The present article is concerned with a goodness-of-fit test for the parametric conditional mean of a time series model. More precisely, we consider a stationary time series $(Y_t)_t$ with values in $\mathbb{R}^d$ ($d \in \mathbb{N}$) and derive a test for the problem whether the conditional mean of Y_t given some set of information I_t at time t belongs to a specific parametric family, i.e.

$$\begin{aligned}
\mathcal{H}_0 : P(\mathbb{E}(Y_t|I_t) = g(I_t, \theta_0)) \quad &= 1 \quad \text{for some } \theta_0 \in \Theta \subseteq \mathbb{R}^q \quad \text{vs.} \\
\mathcal{H}_1 : P(\mathbb{E}(Y_t|I_t) = g(I_t, \theta)) \quad &< 1 \quad \forall\, \theta \in \Theta \subseteq \mathbb{R}^q.
\end{aligned}$$

If Y_t is integrable, we obtain the tautological expression

$$Y_t = \mathbb{E}(Y_t|I_t) + \epsilon_t,$$

where the conditional expectation of $\epsilon_t = Y_t - \mathbb{E}(Y_t|I_t)$ given I_t is equal to zero almost surely under the null hypothesis.

In the context of time series, the information variable I_t may depend on lagged values of the response process $(Y_t)_t$. Both cases, I_t is finite-dimensional as well as I_t is infinite-dimensional, are treated in the literature. Tests for the latter case with $d = 1$ were derived for example by de Jong [5], Bierens and Ploberger [4] as well as by Escanciano [7]. We restrict ourselves to the finite-dimensional case here. Still, our method leads to tests that are consistent against a broad class of models such as linear and various nonlinear (auto-) regressions.

There are basically two approaches to establish consistent tests. Kernel-based tests with vanishing bandwidths were considered for instance by Fan and Li [10], Hjellvik, Yao, Tjøstheim [14], Fan and Li [11] as well as by Kreiss, Neumann and Yao [16]. While Fan and Li concentrated on the asymptotics under null and alternative hypotheses, the both other papers investigated a parametric bootstrap method and a wild bootstrap procedure, respectively.

The second approach extends the integrated conditional moment test of Bierens [3] towards dependent observations, cf. Koul and Stute [15], Escanciano [8] as well as Escanciano and Jacho Chávez [9]. The first two papers are concerned with the behaviour of the respective test statistics under $\mathcal{H}_0$ and $\mathcal{H}_1$ for real-valued response variables. Both articles rely on the asymptotic behaviour of residual marked empirical processes. Additionally, Escanciano [8] investigated the asymptotics under Pitman local alternatives and justified a wild bootstrap method. Based on the latter work, Escanciano and Jacho Chávez [9] developed a principal components decomposition of Cramér-von Mises types of tests. They approximated the corresponding critical values with the aid of Monte Carlo methods.

A comparative overview of both approaches is provided by Fan and Li [12]. In particular they investigated the behaviour under local alternatives. While the Bierens-type tests are more powerful against Pitman alternatives, the kernel-based method with vanishing bandwidth can detect alternatives characterized by sharp peaks with faster rate.

Here, kernel-based tests of L_2-type with fixed bandwidth are considered. Note that we allow for vector-valued response variables. Alternatively to Escanciano [8], who invoked empirical process theory in the case of real-valued response variables, we employ recent results on degenerate U- and V-type statistics. Let $((Y_t', I_t')')_{t\in\mathbb{Z}}$ be a strictly stationary process, where the marginals of $(Y_t)_t$ have values in $\mathbb{R}^d$. The process

$(Y_t)_t$ is assumed to be nonlinear autoregressive with exogenous terms (NARX), i.e.

$$Y_t := G(Y_{t-p}, \ldots, Y_{t-1}, Z_{t-\bar{p}+1}, \ldots, Z_{t-1}, Z_t) + \epsilon_t \tag{1.1}$$

and $I_t := (Y'_{t-p}, \ldots, Y'_{t-1}, Z'_{t-\bar{p}+1}, \ldots, Z'_{t-1}, Z'_t)'$. Here, $((\epsilon'_t, Z'_t)')_t$ is a sequence of i.i.d. integrable $\mathbb{R}^{d+m}$-valued random variables with independent components ϵ_t and Z_t and $\mathbb{E}\epsilon_1 = 0_d$.

We consider test statistics of U- and V-type with estimated parameters $\widehat{\theta}_n$,

$$\widehat{T}_n^{(u)} := \frac{1}{n} \sum_{\substack{j,k=1 \\ j \neq k}}^{n} [Y_j - g(I_j, \widehat{\theta}_n)]' \, K(I_j, I_k, \widehat{\theta}_n) \, [Y_k - g(I_k, \widehat{\theta}_n)],$$

$$\widehat{T}_n^{(v)} := \frac{1}{n} \sum_{j,k=1}^{n} [Y_j - g(I_j, \widehat{\theta}_n)]' \, K(I_j, I_k, \widehat{\theta}_n) \, [Y_k - g(I_k, \widehat{\theta}_n)]$$

with a diagonal matrix $K = diag(k_1, \ldots, k_d)$ of kernels $k_i : \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$. Both statistics were applied by Bartels [1] to independent observations.

The first statistic can be interpreted as a leave-one-out fixed-kernel estimator of $\mathbb{E}([Y_1 - g(I_1, \theta_0)]\mathbb{E}[Y_1 - g(I_1, \theta_0)|I_1] \, p(I_1))$ multiplied with the sample size, where p denotes the (unknown) density of I_1 and $d = 1$. For details, see Li and Wang [19]. Note that this quantity vanishes under the null hypothesis. Both statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ only differ in view of the diagonal terms. A decision based on the U-type statistic does not take $n^{-1} \sum_{k=1}^{n} [Y_k - g(I_k, \widehat{\theta}_n)]' K(I_k, I_k, \widehat{\theta}_n) [Y_k - g(I_k, \widehat{\theta}_n)]$ into account. If, similar to the kernel-based tests with vanishing bandwidth, $k_i(x, y, \theta) = \bar{k}_i((x-y)/h, \theta)$ for some functions $\bar{k}_i$, $i = 1, \ldots, d$, and $h > 0$, this diagonal expression may be interpreted as an estimator of the weighted variances $\sum_{i=1}^{d} k_i(0, 0, \theta) \, var(\epsilon_{1,i})$ of the components of the innovation vector. In general, this expression has no impact on the validity of $\mathcal{H}_0$.

Statistics of the form of $\widehat{T}_n^{(v)}$ include Bierens-type test statistics, i.e. $\int [\, n^{-1/2} \sum_{j=1}^{n} (Y_j - g(I_j, \widehat{\theta}_n)) \, w(X_j, t) \,]^2 \, \Psi(dt)$; $d = 1$. Here, Ψ denotes an integrating function and w is a weight function. The corresponding kernel of $\widehat{T}_n^{(v)}$ has the representation $K(x, y, \theta) = \int w(x, t) \, w(y, t) \, \Psi(dt)$. Eventually, note that we allow for parametric kernels. This might be of interest if one intends to direct the power towards special alternatives.

The limit distributions of our test statistics are non-normal and depend on the (unknown) parameter θ_0 in a complicated way. Therefore, we suggest to approximate (asymptotic) critical values of the test by a parametric bootstrap procedure, which we prove to be consistent. For Bierens-type tests, Escanciano [8] proposed the application of wild bootstrap procedures in order to derive critical values. Our method may perform better since, in contrast to the wild bootstrap, the bootstrap counterparts of the observed random variables converge in distribution to original ones with probability tending one.

The structure of the paper is as follows. In Section 2, we investigate the behaviour of the test statistics under the null as well as under the alternative hypothesis. A parametric bootstrap method is discussed in Section 3. All proofs a deferred to a final Section 4. In this final part of the paper, we additionally verify validity of parametric bootstrap methods for a wide class of Markovian processes. More precisely, we show that the finite-dimensional distributions of the bootstrap process converge to those of the original one; see Lemma 4.2.

2. Asymptotic behaviour of the test statistics

In order to derive the asymptotic distributions of the test statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ under the null hypothesis, certain smoothness and moment constraints on the conditional mean function $g : \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$ are required. We introduce $\|x\|_l := (\sum_{i=1}^D |x_i|^l)^{1/l},\ x \in \mathbb{R}^D,\ l \in \mathbb{N}$, and assume:

(M1) **(i)** The function g satisfies $\mathbb{E}\|g(x, Z_1, \theta_0)\|_2 < \infty$ for some $x \in \mathbb{R}^{dp+m(\bar{p}-1)}$. Furthermore it admits the estimate

$$\begin{aligned}&\|g(y_1,\ldots,y_p,z_1,\ldots,z_{\bar{p}},\theta) - g(\bar{y}_1,\ldots,\bar{y}_p,\bar{z}_1,\ldots,\bar{z}_{\bar{p}},\theta)\|_2\\ &\le \sum_{j=1}^{P} H_j(z_{\bar{p}},\theta)\|(y_j', z_j')' - (\bar{y}_j', \bar{z}_j')'\|_2 \\ &\quad + H_{P+1}(\bar{y}_1,\ldots \bar{y}_p,\bar{z}_1,\ldots,\bar{z}_{\bar{p}-1},\bar{z}_{\bar{p}},z_{\bar{p}},\theta)\|z_{\bar{p}} - \bar{z}_{\bar{p}}\|_2\end{aligned} \tag{2.1}$$

with $\theta = \theta_0$, $P := \max\{p,\ \bar{p}-1\}$, and continuous functions $(H_j(\cdot,\theta_0))_{j=1}^{P+1}$ such that $\sum_{j=1}^P \mathbb{E}|H_j(Z_1,\theta_0)| < 1$. Moreover, the moments

$$\begin{aligned}&\mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1 + a, I_1, \theta_0)|^{4+\varepsilon},\ \mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1, I_1 + a, \theta_0)|^{4+\varepsilon},\\ &\mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1 + a, \widetilde{I}_1, \theta_0)|^{4+\varepsilon},\ \mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1, \widetilde{I}_1 + a, \theta_0)|^{4+\varepsilon}\end{aligned}$$

of the function $H := \sum_{j=1}^{P+1} H_j$ are finite for some $\varepsilon, A > 0$ and an independent copy $\widetilde{I}_1$ of I_1.

(ii) The function $g(x,\cdot)$ is three times continuously differentiable for all $x \in \mathbb{R}^{dp+m\bar{p}}$. The components of its first two partial derivatives w.r.t. θ satisfy $\sum_{i=1}^d \sum_{\alpha,\beta=1}^q \mathbb{E}\left[|g_{i;\alpha}^{(1)}(I_1,\theta_0)| + |g_{i;\alpha,\beta}^{(2)}(I_1,\theta_0)|\right]^{4+\varepsilon} < \infty$ and

$$\begin{aligned}&\sum_{i=1}^d \sum_{\alpha,\beta=1}^q \left[|g_{i;\alpha}^{(1)}(x,\theta) - g_{i;\alpha}^{(1)}(\bar{x},\theta)| + |g_{i;\alpha,\beta}^{(2)}(x,\theta) - g_{i;\alpha,\beta}^{(2)}(\bar{x},\theta)|\right]\\ &\le f_g(x,\bar{x},\theta)\|x-\bar{x}\|_1\end{aligned} \tag{2.2}$$

with $\theta = \theta_0$, where f_g is continuous and $\sup_{j,k\in\mathbb{N}} \mathbb{E}|f_g(I_j, I_k, \theta_0)|^{4+\varepsilon} + \mathbb{E}|f_g(I_1, \widetilde{I}_1, \theta_0)|^{4+\varepsilon} < \infty$. Moreover, there is a neighbourhood $U(\theta_0) \subseteq \Theta$ such that for all $\theta \in U(\theta_0)$, every element of the third derivative of g w.r.t. θ can be bounded by some function M with $\mathbb{E}M^2(I_1) < \infty$.

The parameter estimator is assumed to satisfy:

(M2) **(i)** The sequence of estimators $\widehat{\theta}_n$ admits the expansion

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{k=1}^n l(X_k,\theta_0) + o_P(n^{-1/2}) \tag{2.3}$$

with $\mathbb{E}l(X_1,\theta_0) = 0,\ \mathbb{E}\|l(X_1,\theta_0)\|_1^{4+\varepsilon} < \infty$, and $X_k = (Y_k', I_k')',\ k \in \mathbb{Z}$.

(ii) Moreover, $\|l(x,\theta_0)-l(\bar{x},\theta_0)\|_1 \le f_l(x,\bar{x},\theta_0)\,\|x-\bar{x}\|_1$, where the function $f_l(\cdot,\cdot,\theta_0)$ is symmetric, continuous and $\sup_{j,k\in\mathbb{N}} \mathbb{E}\max_{a\in[-A,A]^{d(p+1)+m\bar{p}}} |f_l(X_j+a, X_k,\theta_0)|^{4+\varepsilon} < \infty$ as well as $\mathbb{E}\max_{a\in[-A,A]^{d(p+1)+m\bar{p}}} |f_l(X_1+a,\widetilde{X}_1,\theta_0)|^{4+\varepsilon} < \infty$ for an independent copy $\widetilde{X}_1$ of X_1 and some $\varepsilon, A>0$.

Finally, we make the following assumptions concerning the kernel function:

(M3) The entries of $K = diag(k_1,\dots,k_d)$, $k_i : \mathbb{R}^{dp+m\bar{p}} \times \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$, are symmetric in their first two arguments. The functions k_i are three times continuously differentiable w.r.t. θ and $k_i = k_i^{(0)},\dots,k_i^{(3)}$ are bounded and Lipschitz continuous in their first two arguments uniformly for all θ in some neighbourhood $U(\theta_0)\subseteq\Theta$.

We briefly comment on the assumptions. Condition (M1)(i) assures the existence of a unique stationary solution of (1.1) if additionally $\mathbb{E}\|\epsilon_1\|_1 + \mathbb{E}\|Z_1\|_1 < \infty$; see the proof of Lemma 2.1 for details. Moreover, it ensures the process $((Y_t', I_t')')_t$ to satisfy some weak dependence condition, namely the geometric-moment contraction condition GMC(1) of Shao and Wu [21]. For instance, Lipschitz contracting nonlinear AR(1) processes fulfil (M1)(i). Assumption (M1)(ii) is a joint constraint on smoothness and existence of moments of the derivatives of the conditional mean under the null hypothesis; a lack of smoothness has to be compensated by additional moments. In particular, functions g with continuous and bounded partial derivatives up to order three w.r.t. θ satisfy (M1) if their Lipschitz constant is sufficiently small. The condition (M2)(i) is a standard assumption regarding parameter estimators in the field of hypothesis testing, (M2)(ii) states some smoothness restrictions on the corresponding linearizing function. The set of feasible kernels defined in (M3) can be enlarged. However, since the kernel is not model inherent but chosen by hand, we restrict ourselves to this class for sake of technical simplification.

Before the asymptotic distributions of $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ are derived, we verify that these statistics can be approximated by degree-2 degenerate U- and V-statistics of $(X_t)_{t\in\mathbb{Z}}$ with $X_t = (Y_t', I_t')'$ and with fixed parameter θ_0.

Lemma 2.1. *Suppose that $((Y_t', Z_t')')_{t\in\mathbb{Z}}$ is the stationary solution of (1.1) with $G(\cdot) = g(\cdot,\theta_0)$ satisfying $\mathbb{E}\|Z_1\|_1^2 + \mathbb{E}\|Y_1\|_1^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$. Assume further that the assumptions (M1), (M2), and (M3) hold. Then, under $\mathcal{H}_0$,*

$$\widehat{T}_n^{(u)} = T_n^{(u)} + o_P(1) \quad and \quad \widehat{T}_n^{(v)} = T_n^{(v)} + o_P(1)$$

with

$$T_n^{(u)} := \frac{1}{n}\sum_{\substack{j,k=1\\ j\neq k}}^{n} \epsilon_j'[K(I_j,I_k,\theta_0)\,\epsilon_k - 2\,V(I_j,\theta_0)\,l(X_k,\theta_0)] + [l(X_j,\theta_0)]'\,a(\theta_0)\,l(X_k,\theta_0),$$

$$T_n^{(v)} := \frac{1}{n}\sum_{j,k=1}^{n} \epsilon_j'[K(I_j,I_k,\theta_0)\,\epsilon_k - 2\,V(I_j,\theta_0)\,l(X_k,\theta_0)] + [l(X_j,\theta_0)]'\,a(\theta_0)\,l(X_k,\theta_0).$$

Here, $a(\theta_0)$ is a $(q\times q)$-matrix with entries

$$a(\theta_0)_{\alpha,\beta} = \mathbb{E}\left(\left[(g^{(1)}(I_1,\theta_0))'K(I_1,\widetilde{I}_1,\theta_0)g^{(1)}(\widetilde{I}_1,\theta_0)\right]_{\alpha,\beta}\right)$$

and $V(I_j,\theta_0)$ is a $(d\times q)$-matrix with entries

$$V(I_j,\theta_0)_{\alpha,\beta}=\mathbb{E}\left(k_\alpha(I_j,\widetilde{I}_j,\theta_0)g^{(1)}_{\alpha,\beta}(\widetilde{I}_j,\theta_0)|I_j\right),$$

where $\widetilde{I}_j$ denotes an independent copy of $I_j,\ j\in\mathbb{N}$.

Based on this assertion we can employ a recent result by Leucht [18] on degenerate U- and V-type statistics for weakly dependent observations in order to deduce the limiting distributions of $T_n^{(u)}$ and $T_n^{(v)}$, and thus of $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$, respectively.

Theorem 2.1. *Suppose that the prerequisites of Lemma 2.1 are satisfied. Then, there are families of constants $(\lambda^{(c)}_{j;k_1,k_2}(\theta_0))_{c,j,k_1,k_2}$, $(C_{j;k_1,k_2})_{j,k_1,k_2}$ and a sequence $(Z_{j;k})_{j,k}$ of centered, jointly normal random variables such that*

$$\widehat{T}_n^{(u)}\xrightarrow{d}Z_1:=\lim_{c\to\infty}\sum_{j=1}^{\infty}\sum_{k_1,k_2\in\mathbb{Z}^d}\lambda_{j;k_1,k_2}(\theta_0)\left[Z_{j;k_1}Z_{j;k_2}+C_{j;k_1,k_2}\right]\quad and$$

$$\begin{aligned}\widehat{T}_n^{(v)}\xrightarrow{d}Z_2:=Z_1+\mathbb{E}\big\{&\epsilon_1'\,[K(I_1,I_1,\theta_0)\,\epsilon_1-2\,V(I_1,\theta_0)\,l(X_1,\theta_0)]\\&+[l(X_1,\theta_0)]'\,a(\theta_0)\,l(X_1,\theta_0)\big\}\end{aligned}$$

under $\mathcal{H}_0$. Here, the right-hand side converges in the L_2-sense.

Remark 2.1. The limit distributions of the test statistics are of a complicated structure. Therefore, problems arise as soon as (asymptotic) critical values of the tests have to be determined. In the following section we propose the application of certain bootstrap methods in order to circumvent these difficulties.

In what follows, we study the behaviour of the test statistics under the alternative hypothesis $\mathcal{H}_1$. It turns out after some approximation steps that in order to verify asymptotic unboundedness of the statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$, the constraint

$$\Delta(\theta_0)=\mathbb{E}\left([\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]'K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(\widetilde{Y}_1|\widetilde{I}_1)-g(\widetilde{I}_1,\theta_0)]\right)>0$$

has to be satisfied. Here, $(\widetilde{Y}_1',\widetilde{I}_1')'$ denotes an independent copy of $(Y_1',I_1')'$. Three different sufficient conditions concerning the kernel matrix are stated below. All of these constraints are based on the so-called integrated conditional moment approach, that is, they assure equivalence between vanishing conditional moments, $\mathbb{E}(Y_1-g(I_1,\theta_0)|I_1)=0$ a.s., and unconditional moments, $\mathbb{E}[(Y_1-g(I_1,\theta_0))'f(I_1,x)]=0$ for a certain function f and almost all x in a compact subset of $\mathbb{R}^v$. However, this idea does not become apparent from the conditions themselves, which are far from intuitive, but from the proof of the subsequent result.

Lemma 2.2. *Suppose that $(Y_t)_t$ is a stationary solution of (1.1) such that $\mathbb{E}\|G(y_0,\epsilon_t,Z_t)\|_2<\infty$ for some $y_0\in\mathbb{R}^{dp}$ and $\mathbb{E}\|G(y,\epsilon_0,Z_0)-G(\bar{y},\epsilon_0,Z_0)\|_2\le\sum_{k=1}^p a_k\|y_k-\bar{y}_k\|_2$ for all $y=(y_1',\dots,y_p')'$, $\bar{y}=(\bar{y}_1',\dots,\bar{y}_p')'\in\mathbb{R}^{dp}$ and some $\alpha_1,\dots,\alpha_p\ge 0$ with $\sum_{k=1}^p\alpha_k<1$. Moreover, let $\mathbb{E}\|Y_1\|_1^{4+\varepsilon}+\mathbb{E}\|Z_1\|_1^2<\infty$ and $\sqrt{n}(\widehat{\theta}_n-\theta_0)=O_P(1)$ for some $\theta_0\in\Theta$ with $\mathbb{E}\|g(I_1,\theta_0)\|_1^{4+\varepsilon}<\infty$. Furthermore, assume that (M1), (M3), and one of the following conditions hold true with some bounded Lipschitz continuous one-to-one mappings $\zeta_i:\mathbb{R}^{dp+m\bar{p}}\to\mathbb{R}^{dp+m\bar{p}}$, $i=1,\dots,d$:*

(i) *The diagonal terms of $K(\cdot,\cdot,\theta_0)$ are absolutely Lebesgue integrable and admit $k_i(x,y,\theta_0)=\bar{k}_{i,\theta_0}(\zeta_i(x)-\zeta_i(y))$, $i=1,\dots,d$. The Fourier transforms $\mathcal{F}\bar{k}_{i,\theta_0}$ of $\bar{k}_{i,\theta_0}$ are nonnegative and do not vanish in a neighbourhood of the origin.*

(ii) *The diagonal terms of* $K(\cdot,\cdot,\theta_0)$ *are absolutely Lebesgue integrable and admit* $k_i(x,y,\theta_0) = \bar{k}_{i,\theta_0}(x-y)$, $i = 1,\dots,d$. *The Fourier transforms* $\mathcal{F}\bar{k}_{i,\theta_0}$ *of* $\bar{k}_{i,\theta_0}$ *are positive a.e. w.r.t. the Lebesgue measure.*

(iii) *The diagonal elements of* $K(\cdot,\cdot,\theta_0)$ *have the following representation*

$$k_i(x,y,\theta_0) = \int_{\mathbb{R}^{dp+m\bar{p}+1}} W_i((1,[\zeta_i(x)]')'t)\, W_i((1,[\zeta_i(y)]')'t)\, w_i^2(t,\theta_0)\, dt,$$

where $W_i : \mathbb{R} \to \mathbb{R}$ *are analytic, non-polynomial functions. The weight functions* $w_i : \mathbb{R}^{dp+m\bar{p}+1} \times \Theta \to \mathbb{R}$ *are assumed to be measurable and to satisfy*

$$0 < \int_{\mathbb{R}^{dp+m\bar{p}+1}} \sup_{x,y} |W_i((1,[\zeta_i(x)]')'t)\, W_i((1,[\zeta_i(y)]')'t)| w_i^2(t,\theta_0)\, dt < \infty.$$

Then, under any fixed alternative in $\mathcal{H}_1$,

$$P\left(\widehat{T}_n^{(u)} > K\right) \underset{n\to\infty}{\longrightarrow} 1 \quad \text{and} \quad P\left(\widehat{T}_n^{(v)} > K\right) \underset{n\to\infty}{\longrightarrow} 1, \quad \forall\, K \in \mathbb{R}.$$

Since the conditions are rather technical, we give some examples. Let $\zeta(x) = (\arctan(x_1),\dots,\arctan(x_{dp+m\bar{p}}))'$. Assumptions (i) and (ii) of the foregoing lemma are satisfied for instance by the following frequently used kernels:

- Gauss kernel $\bar{k}_i(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$,
- Cauchy kernel $\bar{k}_i(x) = \frac{1}{\pi(1+x^2)}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}} e^{-|t|}$,
- Triangular kernel $\bar{k}_i(x) = (1-|x|)\mathbb{1}_{[-1,1]}(x)$ with $\mathcal{F}(t) = \frac{2(1-\cos t)}{\sqrt{2\pi}t^2}$,
- Picard kernel $\bar{k}_i(x) = \frac{1}{2} e^{-|x|}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}(1+t^2)}$.

They are violated when the uniform or the Epanechnikov kernel are applied.

The third constraint holds true for square integrable functions w with bounded support if e.g. $W(y) = e^y$, which was applied by Bierens [3], or the logistic function $W(y) = 1/[1+e^{c-y}]$, $c \neq 0$, are used.

3. Bootstrapping critical values of the tests

According to Theorem 2.1 the limit distributions of both test statistics are basically infinite weighted sums of products of correlated normal variables. The weights depend on the unknown parameter θ_0 in a complicated way. Thus, the (asymptotic) critical values of these tests can hardly be determined analytically or be tabulated. In a similar context, Escanciano [8] proposed to circumvent these difficulties by approximating the critical values with the aid of a wild bootstrap method. In order to prove consistency of this approach, he makes use of the additional assumption that the linearizing function l in (M2) has a product structure, i.e. $l(Y_k, I_k, \theta_0) = \epsilon_k\, \bar{l}(I_k, \theta_0)$ for some function $\bar{l}$. We suggest a parametric bootstrap algorithm instead, where l does not have to factorize as above. Recall that the random variables ϵ_k and Z_k, $k \in \mathbb{Z}$, are independent.

(1) Determine $\widehat{\theta}_n$.
(2) Calculate $\epsilon_k^{(n)} = Y_k - g(I_k, \widehat{\theta}_n)$ and $\bar{\epsilon} = n^{-1} \sum_{k=1}^n \epsilon_k^{(n)}$.
(3) Draw $\epsilon_1^*,\dots,\epsilon_n^*$ via Efron's bootstrap from $(\epsilon_k^{(n)} - \bar{\epsilon})_k$.
(4) Draw $Z_{1-\bar{p}}^*,\dots,Z_n^*$ via Efron's bootstrap from $Z_{1-\bar{p}},\dots,Z_n$ (independently of $\epsilon_1^*,\dots,\epsilon_n^*$).
(5) Determine an initial vector $(Y_0^{*\prime},\dots,Y_{1-p}^{*\prime})'$ independently of $(\epsilon_k^*)_k$ and $(Z_k^*)_k$.

(6) Generate the bootstrap sample $Y_k^* = g(I_k^*, \widehat{\theta}_n) + \epsilon_k^*$, where $I_k^* = (Y_{k-p}^{*\prime}, \ldots, Y_{k-1}^{*\prime}, Z_{k-\bar{p}+1}^{*\prime}, \ldots, Z_k^{*\prime})'$, $k = 1, \ldots, n$.

(7) Compute the bootstrap parameter estimator $\widehat{\theta}_n^*$.

(8) Calculate the bootstrap versions of the test statistics

$$\widehat{T}_n^{(u)*} := \frac{1}{n} \sum_{\substack{j,k=1 \\ j \neq k}}^{n} [Y_j^* - g(I_j^*, \widehat{\theta}_n^*)]' \, K(I_j^*, I_k^*, \widehat{\theta}_n^*) \, [Y_k^* - g(I_k^*, \widehat{\theta}_n^*)],$$

$$\widehat{T}_n^{(v)*} := \frac{1}{n} \sum_{j,k=1}^{n} [Y_j^* - g(I_j^*, \widehat{\theta}_n^*)]' \, K(I_j^*, I_k^*, \widehat{\theta}_n^*) \, [Y_k^* - g(I_k^*, \widehat{\theta}_n^*)].$$

(9) Define the critical values $t_\alpha^{(u)*}$ and $t_\alpha^{(v)*}$ as the $(1-\alpha)$-quantiles of the (conditional) distributions of $\widehat{T}_n^{(u)*}$ and $\widehat{T}_n^{(v)*}$, respectively.
Reject $\mathcal{H}_0$ if $\widehat{T}_n^{(u)} > t_\alpha^{(u)*}$ and $\widehat{T}_n^{(v)} > t_\alpha^{(v)*}$, respectively.

Validity of the bootstrap algorithm can be verified if we assume besides (M1) to (M3):

(M5) **(i)** The function g satisfies $P(\mathbb{E}^* \| g(x^*, Z_1^*, \widehat{\theta}_n) \|_2 \leq K) \longrightarrow_{n \to \infty} 1$ for some $x^* \in \mathbb{R}^{dp+m\bar{p}}$ and a $K < \infty$. It admits (2.1) with $\theta = \widehat{\theta}_n$ and continuous functions $(H_j(\cdot, \widehat{\theta}_n))_{j=1}^{P+1}$ with $P(\sum_{j=1}^{P} \mathbb{E}^* |H_j(Z_1^*, \widehat{\theta}_n)| \leq 1 - \delta) \longrightarrow_{n \to \infty} 1$ for some $\delta \in (0,1)$. Moreover, the moments

$$\mathbb{E}^* \max_{a \in [-A,A]^{dp+m\bar{p}}} |H(I_1^* + a, I_1^*, \widehat{\theta}_n)|^{4+\varepsilon}, \ \mathbb{E}^* \max_{a \in [-A,A]^{dp+m\bar{p}}} |H(I_1^*, I_1^* + a, \widehat{\theta}_n)|^{4+\varepsilon},$$
$$\mathbb{E}^* \max_{a \in [-A,A]^{dp+m\bar{p}}} |H(I_1^* + a, \widetilde{I}_1^*, \widehat{\theta}_n)|^{4+\varepsilon}, \ \mathbb{E}^* \max_{a \in [-A,A]^{dp+m\bar{p}}} |H(I_1^*, \widetilde{I}_1^* + a, \widehat{\theta}_n)|^{4+\varepsilon}$$

of the function $H := \sum_{j=1}^{P+1} H_j$ are of order $O_P(1)$ for some $\varepsilon, A > 0$ and an independent copy $\widetilde{I}_1^*$ of I_1^*, conditionally on $X_1, \ldots, X_n$.

(ii) The function $g(x, \cdot)$ is three times continuously differentiable for all $x \in \mathbb{R}^{dp+m\bar{p}}$. The components of the first two partial derivatives of g w.r.t. θ satisfy $\sum_{i=1}^{d} \sum_{\alpha,\beta=1}^{q} \mathbb{E}^*[|g_i(I_1^*, \widehat{\theta}_n)| + |g_{i;\alpha}^{(1)}(I_1^*, \widehat{\theta}_n)| + |g_{i;\alpha,\beta}^{(2)}(I_1^*, \widehat{\theta}_n)|]^{4+\varepsilon} = O_P(1)$. Moreover, (2.2) holds true with $\theta = \widehat{\theta}_n$, $\mathbb{E}|f_g(I_1^*, \widetilde{I}_1^*, \widehat{\theta}_n)|^{4+\varepsilon} = O_P(1)$, and $P(\sup_{j,k \in \mathbb{N}} \mathbb{E}^* |f_g(I_j^*, I_k^*, \widehat{\theta}_n)|^{4+\varepsilon} \leq K) \longrightarrow_{n \to \infty} 1$. The function M, defined in (M1), satisfies $P(\mathbb{E}^*[M(I_1^*)]^2 \leq K) \longrightarrow_{n \to \infty} 1$.

(iii) The function g fulfils $\|g(y, \theta_1) - g(y, \theta_2)\|_1 \leq L(y) \, \|\theta_1 - \theta_2\|_1$ where $L : \mathbb{R}^{dp+m\bar{p}} \to \mathbb{R}_+$ is a locally Lipschitz continuous function with $\mathbb{E}\|L(I_1)\|_1^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$. The random vector $Y_1 - g(I_1, \theta_0)$ has continuous marginal distribution functions.

The bootstrap estimator is assumed to satisfy:

(M6) **(i)** The sequence of estimators $\widehat{\theta}_n^*$ admits the expansion

$$\widehat{\theta}_n^* - \widehat{\theta}_n = \frac{1}{n} \sum_{k=1}^{n} l(X_k^*, \widehat{\theta}_n) + o_{P^*}(n^{-1/2}) \tag{3.1}$$

with $X_k^* = (Y_k^{*\prime}, I_k^{*\prime})'$ and $\mathbb{E}^* l(X_1^*, \widehat{\theta}_n) = 0$, $\mathbb{E}^* \|l(X_1^*, \widehat{\theta}_n)\|_1^{4+\varepsilon} = O_P(1)$.

(ii) Moreover, $\|l(x,\widehat{\theta}_n) - l(\bar{x},\widehat{\theta}_n)\|_1 \le f_l(x,\bar{x},\widehat{\theta}_n)\|x-\bar{x}\|_1$, where f_l is a symmetric and continuous function with

$$P\left(\sup_{j,k\in\mathbb{N}} \mathbb{E}^* \max_{a\in[-A,A]^{pd+\bar{p}m}} |f_l(X_j^*+a, X_k^*, \widehat{\theta}_n)|^{4+\varepsilon} \le K\right) \underset{n\to\infty}{\longrightarrow} 1$$

and $\mathbb{E}^* \max_{a\in[-A,A]^{pd+\bar{p}m}} |f_l(X_1^*+a, \widetilde{X}_1^*, \widehat{\theta}_n)|^{4+\varepsilon} = O_P(1)$ for any independent copy $\widetilde{X}_1^*$ of X_1^*, some $\varepsilon, A > 0$, and a $K < \infty$.

The assumptions (M5)(i),(ii) and (M6) can be interpreted as the bootstrap counterparts to (M1) and (M2). Note that the condition (M5) holds for instance when the function g is linear with coefficients whose absolute values sum up to some constant that is less than one and if the corresponding parameter estimators converge. The additional assumption (M5)(iii) is required to prove that the finite-dimensional distributions of the bootstrap process converge to those of the original process.

Applying the result of Leucht [18] concerning bootstrap consistency of degenerate U- and V-statistics under weak dependence, we obtain bootstrap consistency for our test statistics.

Proposition 3.1. *Let the assumptions of Lemma 2.1 hold true. Suppose that the conditions* $(M5)$ *and* $(M6)$ *are fulfilled and that* $\widehat{T}_n^{(u)*}$ *as well as* $\widehat{T}_n^{(v)*}$ *are generated via the aforementioned algorithm, where the initial response vector* Y_0^* *is drawn from the stationary bootstrap distribution if it exists. Then, under* $\mathcal{H}_0$,

$$\widehat{T}_n^{(u)*} \stackrel{d}{\longrightarrow} Z_1 \qquad \text{and} \qquad \widehat{T}_n^{(v)*} \stackrel{d}{\longrightarrow} Z_2,$$

in probability as $n\to\infty$, *where* Z_1 *and* Z_2 *are defined as in Theorem 2.1. Moreover, if* $var(Z_1) > 0$,

$$\sup_{-\infty<x<\infty} \left| P(\widehat{T}_n^{(u)*} \le x | X_1,\dots,X_n) - P(\widehat{T}_n^{(u)} \le x)\right| \stackrel{P}{\longrightarrow} 0,$$
$$\sup_{-\infty<x<\infty} \left| P(\widehat{T}_n^{(v)*} \le x | X_1,\dots,X_n) - P(\widehat{T}_n^{(v)} \le x)\right| \stackrel{P}{\longrightarrow} 0.$$

Providing that the initial vector exhibits the stationary bootstrap distribution, which exists with probability tending to one, we obtain bootstrap consistency. However, this is often unknown. Still, since the L_1-distance of the m^{th} variable of a reasonably started bootstrap process and the process started with the bootstrap stationary distribution tends to zero rapidly with increasing m (cf. Shao and Wu [21], Theorem 5.1), we conjecture that the assertion above remains valid if we take an "arbitrary" initial vector. In practice one would drop the first generated values.

Under (M4) to (M6), the proposed bootstrap method imitates a null situation under $\mathcal{H}_0$ as well as under $\mathcal{H}_1$. Consequently, the bootstrap statistic is of order $O_P(1)$. In accordance with Lemma 2.2 this immediately implies consistency of the bootstrap-based test. The asymptotic behaviour of the bootstrap-aided test can be summarized as follows:

Lemma 3.1. *Suppose that the conditions (M5) and (M6) are fulfilled and let* $\alpha \in (0,1)$. *Moreover, assume that the prerequisites of Lemma 2.1 and Lemma 2.2 hold. The proposed bootstrap test based on the algorithm above satisfies*

$$\lim_{n\to\infty} P(\widehat{T}_n^{(u)} > t_\alpha^{(u)*}) = \lim_{n\to\infty} P(\widehat{T}_n^{(v)} > t_\alpha^{(v)*}) = \begin{cases} \alpha & \text{if } \mathcal{H}_0 \text{ is true and } var(Z_1) > 0, \\ 1 & \text{if } \mathcal{H}_1 \text{ is true.} \end{cases}$$

4. Proofs and auxiliary results

Throughout this section, C denotes a generic finite constant that may change its value even within a single calculation.

4.1. **Proofs of the main results.**
Based on Lemma 4.1 of Subsection 4.2 and the corresponding techniques of proof, we can verify the results of the previous sections.

Proof of 2.1. First, note that $\mathbb{E}\|\varepsilon_1\|_1^{4+\varepsilon} = \mathbb{E}\|Y_1 - E(Y_1|I_1)\|_1^{4+\varepsilon} \leq C\,\mathbb{E}\|Y_1\|_1^{4+\varepsilon} < \infty$ under the assumptions of the lemma. According to Shao and Wu [21], Theorem 5.1 with $X_k = (Y_k', Z_k')'$ and $\varepsilon_k = (\epsilon_k', Z_k')$, there exists a unique stationary solution $((Y_k', Z_k')')_{k\in\mathbb{Z}}$ to (1.1) with $G(\cdot) = g(\cdot, \theta_0)$ under $\mathcal{H}_0$ and (M1)(i). Moreover, their result implies that for all $n \in \mathbb{N}$ there is a random vector $(\widetilde{Y}_n', \widetilde{Z}_n')' \stackrel{d}{=} (Y_n', Z_n')'$ that is independent of $((Y_k', Z_k')')_{k\leq 0}$ and that satisfies their GMC(1) condition, i.e., $\mathbb{E}\|(\widetilde{Y}_n', \widetilde{Z}_n')' - (Y_n', Z_n')'\|_2 \leq K\rho^n$ for some $K < \infty$ and $\rho \in (0,1)$. Moreover, the definition of the GMC(1) condition implies the existence of a copy $(\widetilde{Y}_n', \ldots, \widetilde{Y}_{n-P}', \widetilde{Z}_n', \ldots, \widetilde{Z}_{n-P}')'$ of the random vector $(Y_n', \ldots, Y_{n-P}', Z_n', \ldots, Z_{n-P}')'$ that is independent of $((Y_k', Z_k')')_{k\leq 0}$ for $n - P - 1 > 0$ and such that

$$\begin{aligned}&\mathbb{E}\|(Y_n', \ldots, Y_{n-P}', Z_n', \ldots, Z_{n-P}')' - (\widetilde{Y}_n', \ldots, \widetilde{Y}_{n-P}', \widetilde{Z}_n', \ldots, \widetilde{Z}_{n-P}')'\|_2 \\ &\leq K\,(P+1)\,\rho^{-P-1}\,\rho^n.\end{aligned}$$

Due to the equivalence of norms on $\mathbb{R}^D$, the latter inequality leads to

$$\mathbb{E}\|X_n - \widetilde{X}_n\|_1 \leq \sqrt{(d+m)(P+1)}\,K\,(P+1)\,\rho^{-P-1}\,\rho^n, \quad n \in \mathbb{N}, \tag{4.1}$$

where $X_n = (Y_n', I_n')'$ and $\widetilde{X}_n = (\widetilde{Y}_n', \widetilde{I}_n')'$ with $\widetilde{I}_n := (\widetilde{Y}_{n-p}', \ldots, \widetilde{Y}_{n-1}', Z_{n-\bar{p}+1}', \ldots, Z_n')'$.

Based on these L_1-coupling inequalities, the assertions of our lemma can be derived. To this end, we restrict ourselves to the approximation of the statistic $\widehat{T}_n^{(v)}$ here. The calculations for $\widehat{T}_n^{(u)}$ can be carried out in complete analogy. Let

$$\begin{aligned}\widetilde{T}_n^{(v)} := \frac{1}{n}\sum_{j,k=1}^{n} \big\{&\epsilon_j'\,K(I_j, I_k, \theta_0)\,\epsilon_k - 2\,\epsilon_j'\,K(I_j, I_k, \theta_0)\,g^{(1)}(I_k, \theta_0)\,(\widehat{\theta}_n - \theta_0) \\ &+ (\widehat{\theta}_n - \theta_0)'\left[g^{(1)}(I_j, \theta_0)\right]' K(I_j, I_k, \theta_0)\,g^{(1)}(I_k, \theta_0)\,(\widehat{\theta}_n - \theta_0)\big\}.\end{aligned}$$

<u>Step 1:</u> $\widehat{T}_n^{(v)} = \widetilde{T}_n^{(v)} + o_P(1)$.
Elementwise Taylor expansion of $g(\cdot, \widehat{\theta}_n)$ and $K(\cdot, \cdot, \widehat{\theta}_n)$ results in

$$
\begin{aligned}
&\widehat{T}_n^{(v)} - \widetilde{T}_n^{(v)} = \\
&\frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i} \left[(\widehat{\theta}_n - \theta_0)' \, k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k}) \right] \epsilon_{k,i} && (4.2)\\
&-\frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i} \, k_i(I_j, I_k, \theta_0) \left[(\widehat{\theta}_n - \theta_0)' \, g_i^{(2)}(X_k, \bar{\theta}_{n,i,k}) \, (\widehat{\theta}_n - \theta_0) \right] && (4.3)\\
&-\frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i} \left[k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k}) \, (\widehat{\theta}_n - \theta_0) \right] && (4.4)\\
&\quad \times (\widehat{\theta}_n - \theta_0)' \left[2 \, g_i^{(1)}(I_k, \theta_0) + g_i^{(2)}(I_k, \bar{\theta}_{n,i,k}) \, (\widehat{\theta}_n - \theta_0) \right] \\
&+\frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} (\widehat{\theta}_n - \theta_0)' \left[g_i^{(1)}(I_j, \theta_0) + \frac{1}{2} g_i^{(2)}(I_j, \bar{\theta}_{n,i,j}) \, (\widehat{\theta}_n - \theta_0) \right] && (4.5)\\
&\quad \times k_i(I_j, I_k, \theta_0) \left[(\widehat{\theta}_n - \theta_0)' \, g_i^{(2)}(I_k, \bar{\theta}_{n,i,k}) \, (\widehat{\theta}_n - \theta_0) \right] \\
&+\frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} (\widehat{\theta}_n - \theta_0)' \left[g_i^{(1)}(I_j, \theta_0) + \frac{1}{2} g_i^{(2)}(I_j, \bar{\theta}_{n,i,j}) \, (\widehat{\theta}_n - \theta_0) \right] && (4.6)\\
&\quad \times \left[(\widehat{\theta}_n - \theta_0)' \, k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k}) \right] (\widehat{\theta}_n - \theta_0)' \left[g_i^{(1)}(I_k, \theta_0) + \frac{1}{2} g_i^{(2)}(I_k, \bar{\theta}_{n,i,k}) \, (\widehat{\theta}_n - \theta_0) \right]
\end{aligned}
$$

for some random $\bar{\theta}_{n,i,j}$ and $\bar{\theta}_{n,i,j,k}$ between $\widehat{\theta}_n$ and θ_0. For sake of notational simplicity let $q = 1$ in the sequel. The extension of the approximations to higher dimensions is straightforward.

To verify asymptotic negligibility of the term (4.2), we expand $k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k}) = k_i^{(1)}(I_j, I_k, \theta_0) + k_i^{(2)}(I_j, I_k, \theta_0)(\bar{\theta}_{n,i,j,k} - \theta_0) + k_i^{(3)}(I_j, I_k, \bar{\bar{\theta}}_{n,i,j,k})(\bar{\theta}_{n,i,j,k} - \theta_0)^2/2$ for some random $\bar{\bar{\theta}}_{n,i,j,k}$ with $|\bar{\bar{\theta}}_{n,i,j,k} - \theta_0| \leq |\widehat{\theta}_n - \theta_0|$. Thus, the absolute value of (4.2) can be bounded from above by

$$
o_P(1) \sum_{i=1}^{d} \left[\left| \frac{1}{n} \sum_{1 \leq j < k \leq n} \epsilon_{j,i} \, k_i^{(1)}(I_j, I_k, \theta_0) \, \epsilon_{k,i} \right| + \left| \frac{1}{n} \sum_{1 \leq j < k \leq n} \epsilon_{j,i} \, k_i^{(2)}(I_j, I_k, \theta_0) \, \epsilon_{k,i} \right| \right] + o_P(1)
$$

as $\widehat{\theta}_n - \theta_0 = o_P(1)$ in view of its expansion (2.3) and Lemma 4.1 of Leucht [17]. For every i the expressions within the absolute-value signs form degenerate U-statistics of the random variables $(X_k)_k$. Both U-statistics have finite second moments, which implies that they are of order $O_P(1)$, since the prerequisites of Lemma 4.1 of this paper are fulfilled due to assumption (M1)(i), (M3), and (4.1).

In order to approximate the expression (4.3), we use a Taylor expansion of $g_i^{(2)}$ for $i = 1, \ldots, d$. This yields

$$|(4.3)| = O_P(1) \left| \frac{1}{n^2} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0) \left[g_i^{(2)}(I_k, \theta_0) + g_i^{(3)}(I_k, \bar{\bar{\theta}}_{n,i,k})\, (\bar{\theta}_{n,i,k} - \theta_0) \right] \right|$$
$$= o_P(1) + O_P(n^{-1/2}) \left| \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0) \right|$$

for some $\bar{\bar{\theta}}_{n,i,k}$ between $\widehat{\theta}_n$ and θ_0. It is now sufficient to show that the remaining multiple sum is of order $O_P(1)$. For this purpose, we consider

$$\frac{1}{n^3} \sum_{j,k,l,m=1}^{n} \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\}$$
$$\leq O(1) + \frac{2}{n^3} \sum_{1 \leq j,k,m < l \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right|$$
$$+ \frac{2}{n^3} \sum_{1 \leq j < k,l,m \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right|$$
$$+ \frac{2}{n^3} \sum_{1 \leq k < j < l < m \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right|$$
$$+ \frac{2}{n^3} \sum_{1 \leq m < j < l < k \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right|.$$

It can easily be seen that the first two sums are uniformly bounded in n. To this end, we use a coupling with appropriate copies $\widetilde{X}_l$ and $(\widetilde{X}_k', \widetilde{X}_l', \widetilde{X}_m')'$ of X_l and $(X_k', X_l', X_m')'$ that are independent of $(X_j', X_k', X_m')'$ and X_j, respectively, if $l - (P+1) - \max\{j,k,m\} > 0$ and $\min\{k,l,m\} - (P+1) - j > 0$, respectively. Note that in these cases

$$\mathbb{E} \left\{ \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0) \right\} \left\{ [\widetilde{Y}_{l,i} - g_i(\widetilde{I}_l, \theta_0)]\, k_i(\widetilde{I}_l, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0) \right\} = 0$$

and

$$\mathbb{E} \left\{ \epsilon_{j,i}\, k_i(I_j, \widetilde{I}_k, \theta_0)\, g_i^{(2)}(\widetilde{I}_k, \theta_0) \right\} \left\{ [\widetilde{Y}_{l,i} - g_i(\widetilde{I}_l, \theta_0)]\, k_i(\widetilde{I}_l, \widetilde{I}_m, \theta_0)\, g_i^{(2)}(\widetilde{I}_m, \theta_0) \right\} = 0.$$

The desired order of the corresponding summands can then be obtained under (M1) and (M3) invoking the similar arguments as in the proof of Lemma 4.1.

In order to show uniform boundedness of the third sum, we introduce a vector $(\widetilde{X}_l', \widetilde{X}_m')' \stackrel{d}{=} (X_l', X_m')'$ that is independent of $(X_k', X_j')'$ and that satisfies the condition (4.1) with $n = l - j$ as long as $l - (P+1) - j > 0$. This leads to

$$\frac{2}{n^3} \sum_{1\le k<j<l<m\le n} \Big| \mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)\right\} \left\{\epsilon_{l,i}\, k_i(I_l, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0)\right\}\Big|$$
$$\le \frac{2}{n^3} \sum_{1\le k<j<l<m\le n} \Big| \mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)\right\} \Big\{\epsilon_{l,i}\, k_i(I_l, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0)$$
$$- \widetilde{\epsilon}_{l,i}\, k_i(\widetilde{I}_l, \widetilde{I}_m, \theta_0)\, g_i^{(2)}(\widetilde{I}_m, \theta_0)\Big\}\Big| + \frac{1}{n}\left| \frac{1}{n} \sum_{j,k=1;\, j\ne k}^{n} \mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)\right|^2,$$

where $\widetilde{\epsilon}_{l,i} := \widetilde{Y}_l - g(\widetilde{I}_l, \theta_0)$. In the same way as in the proof of Lemma 4.1, we obtain boundedness of the first summand on the r.h.s. uniformly in n. The second one is of order $O(1/n)$ since

$$\left|\mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)\right| \le C\, \mathbb{E}|\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0) - \widetilde{\epsilon}_{j,i}\, k_i(\widetilde{I}_j, I_k, \theta_0)|\, |g_i^{(2)}(I_k, \theta_0)|$$
$$\le C\, \rho^{(j-k)\delta}$$

for a $\delta > 0$, $j > k + P + 1$ and some appropriate $\widetilde{X}_j \stackrel{d}{=} X_j$ that is independent of X_k. Similarly, one gets $|\mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)| \le C\, \rho^{(k-j)\delta}$ when $k > j + P + 1$ by introducing an suitable copy $\widetilde{I}_k$ of I_k that is independent of X_j. Eventually, we have to verify

$$\frac{2}{n^3} \sum_{1\le m<j<l<k\le n} \Big| \mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0)\right\} \left\{\epsilon_{l,i}\, k_i(I_l, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0)\right\}\Big| = O(1).$$

Under (M1) and (M3) it suffices to consider those summands such that the minimal gap between the corresponding indices is larger than $P+1$. Denote this set of indices by J. Invoking L_1-coupling techniques and the continuity properties of the involved functions in the same manner as before, the introduction of copies $\widetilde{I}_k$ of I_k that are independent of $(X_m', X_j', X_l')'$ imply that the above sum can be bounded by

$$\frac{2}{n^3} \sum_{\substack{1\le m<j<l<k\le n\\ (m,j,l,k)\in J}} \Big| \mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, \widetilde{I}_k, \theta_0)\, g_i^{(2)}(\widetilde{I}_k, \theta_0)\right\} \left\{\epsilon_{l,i}\, k_i(I_l, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0)\right\}\Big| + O(1).$$

Proceeding iteratively in complete analogy with the remaining indices l, j and m, i.e. introducing a coupling of the variable with the largest index, we end up with the upper estimate

$$O(1) + \frac{2}{n} \sum_{\substack{1\le m<j\le n\\ (j,m)\in J}} \left| \mathbb{E} \int_{\mathbb{R}^{dp+m\bar{p}}} \epsilon_{j,i}\, k_i(I_j, y, \theta_0)\, g_i^{(2)}(y, \theta_0) P_{I_1}(dy)\right|$$
$$\times \left| \mathbb{E} \int_{\mathbb{R}^{d(p+1)+m\bar{p}}} [z_1 - g_i(z_2, \theta_0)]\, k_i(z_2, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0) P_{X_1}(dz)\right|,$$

where $z = (z_1', z_2')'$ with $z_1 \in \mathbb{R}^d$ and $z_2 \in \mathbb{R}^{dp+m\bar{p}}$. Note that all summands of the remaining sum are equal to zero. Consequently, we finally achieve the desired order of (4.3).

Similar arguments as in the considerations of (4.2) and (4.3) lead to (4.4) + (4.5) + (4.6) $= o_P(1)$ due to the moment constraints concerning $g_i^{(1)}, g_i^{(2)}$, and $g_i^{(3)}$, $i = 1, \ldots, d$. Thus, $\widehat{T}_n^{(v)} - \widetilde{T}_n^{(v)} = o_P(1)$.

Step 2: $\widetilde{T}_n^{(v)} = T_n^{(v)} + o_P(1)$.

For sake of notational simplicity all calculations are only stated for $q = 1$ again. According to the representation (2.3) of $\widehat{\theta}_n - \theta$, we obtain

$$\frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, (\widehat{\theta}_n - \theta_0)$$

$$= \frac{1}{n^2} \sum_{j,k,m=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i} \, k_i(I_j, I_k, \theta_0) \, g_i^{(1)}(I_k, \theta_0) \, l(X_m, \theta_0) + o_P(1).$$

Moreover,

$$\frac{1}{n^2} \sum_{j,k,m=1}^{n} \epsilon_j' \, [K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) - V(I_j, \theta_0)] \, l(X_m, \theta_0)$$

$$= O_P(1) \, \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \big\{ \epsilon_j' \, [K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) - V(I_j, \theta_0)] + \epsilon_k' \, [K(I_k, I_j, \theta_0) \, g^{(1)}(I_j, \theta_0) - V(I_k, \theta_0)] \big\}.$$

The double sum builds a degenerate V-statistic multiplied with $\sqrt{n}$. According to the continuity assumptions on the involved functions and Lemma 4.1, this quantity vanishes asymptotically. Summing up, we obtain

$$\frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, (\widehat{\theta}_n - \theta_0) = \frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j' \, V(I_j, \theta_0) \, l(X_k, \theta_0) + o_P(1).$$

It remains to consider

$$\frac{1}{n} \sum_{j,k=1}^{n} (\widehat{\theta}_n - \theta_0) \, \big[g^{(1)}(I_j, \theta_0)\big]' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, (\widehat{\theta}_n - \theta_0)$$

$$= \frac{1}{n^3} \sum_{i,j,k,m=1}^{n} l(X_i, \theta_0) \, \big[g^{(1)}(I_j, \theta_0)\big]' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, l(X_m, \theta_0)$$

$$+ o_P(1) \left[\frac{1}{n^2} \sum_{j,k=1}^{n} \big[g^{(1)}(I_j, \theta_0)\big]' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \right] \left[\frac{1}{\sqrt{n}} \sum_{m=1}^{n} l(X_m, \theta_0) \right] + o_P(1),$$

$$= \frac{1}{n^3} \sum_{i,j,k,m=1}^{n} l(X_i, \theta_0) \, \big[g^{(1)}(I_j, \theta_0)\big]' \, K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, l(X_m, \theta_0) + o_P(1).$$

The latter equality is obtained by applying Lemma 4.1 of [17] to $n^{-1/2} \sum_{m=1}^{n} l(X_m, \theta_0)$ and by virtue of the assumptions (M1) to (M3). Thus, we still have to prove that

$T_n := n^{-2}\sum_{j,k=1}^n [g^{(1)}(I_j,\theta_0)]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0) - a(\theta_0) \xrightarrow{P} 0$. To this end, a Hoeffding decomposition of the kernel associated with the above V-statistic is invoked as follows:

$$T_n = \frac{2}{n}\sum_{j=1}^n h_1(I_j) + \frac{1}{n^2}\sum_{j,k=1}^n h_2(I_j,I_k),$$

where $h_1(x) := \mathbb{E}[g^{(1)}(x,\theta_0)]'K(x,I_1,\theta_0)g^{(1)}(I_1,\theta_0) + a(\theta_0)$ and

$$h_2(x,y) = [g^{(1)}(x,\theta_0)]'K(x,y,\theta_0)g^{(1)}(y,\theta_0) + a(\theta_0) - h_1(x) - h_1(y), \quad x,y \in \mathbb{R}^{dp+m\bar{p}}.$$

Note that h_2 is degenerate. Due to Lemma 4.1 of this paper the corresponding double sum tends to zero in probability. Again by Lemma 4.1 of [17] $\mathbb{E}(n^{-1}\sum_{j=1}^n h_1(I_j))^2 = o(1)$. This eventually completes the proof. □

Proof of Theorem 2.1. According to Lemma 2.1 it suffices to consider the expressions $T_n^{(u)}$ and $T_n^{(v)}$. Under the null hypothesis, these are degenerate U- and V-statistics multiplied with the sample size and whose kernel is given by

$$\begin{aligned} h(x,y,\theta_0) :=& [x_1 - g(x_2,\theta_0)]'\,K(x_2,y_2,\theta_0)\,[y_1 - g(y_2,\theta_0)] + [l(x,\theta_0)]'\,a(\theta_0)\,l(y,\theta_0) \\ & - [V(y_1,\theta_0)\,l(x,\theta_0)]'\,[y_1 - g(y_2,\theta_0)] - [x_1 - g(x_2,\theta_0)]'\,V(x_2,\theta_0)\,l(y,\theta_0) \end{aligned}$$

with $x = (x_1', x_2')'$, $y = (y_1', y_2')'$, $x_1, y_1 \in \mathbb{R}^d$ and $x_2, y_2 \in \mathbb{R}^{dp+m\bar{p}}$. We obtain their limits from Theorem 2.2 in Leucht [18] if her assumptions (A1), (A2) and (A4) are satisfied. As verified already at the beginning of the previous proof, the underlying sample satisfies the GMC(1) condition of Shao and Wu [21], which implies that (A1) holds true with $\tau_r = C\rho^r$. The function h that satisfies $\sup_{k\in\mathbb{N}}\mathbb{E}|h(X_1,X_k)|^{2+\varepsilon} + \mathbb{E}|h(X_1,\widetilde{X}_1)|^{2+\varepsilon} < \infty$, where $X_k = (Y_k', I_k')', k \in \mathbb{Z}$ and $\widetilde{X}_1$ denotes an i.i.d. copy of X_1. Moreover in view of (M1) to (M3), there is a continuous function f such that

$$|h(x,y) - h(\bar{x},\bar{y})| \le f(x,\bar{x},y,\bar{y})[\|x-\bar{x}\|_1 + |y-\bar{y}\|_1]$$

with

$$\sup_{k_1,\dots,k_5\in\mathbb{N}} \mathbb{E}\left\{\max_{a_1,a_2\in[-A,A]^{(p+1)d+\bar{p}m)}} f(\bar{X}_{k_1}, \bar{X}_{k_2}+a_1, \bar{X}_{k_3}, \bar{X}_{k_4}+a_2)^{1+\varepsilon}\|\bar{X}_{k_1}\|_1\right\} < \infty$$

for any any $(\bar{X}_{k_1}',\dots,\bar{X}_{k_5}')'$ consisting of independent subvectors $(\bar{X}_{k_{j_1(m)}}',\dots,\bar{X}_{k_{j_l(m)}}')' \stackrel{d}{=} (X_{k_{j_1(m)}}',\dots,X_{k_{j_l(m)}}')'$, $l,m = 1,\dots,5$. Thus, all prerequisites of this result are satisfied and the construction of $(\lambda_{j;k_1,k_2}^{(c)})$ and $(Z_{j;k_1,k_2})$ can be carried out as follows: Let ϕ and ψ denote $\mathbb{R}$-valued, Lipschitz continuous, compactly supported scale and wavelet functions associated with an one-dimensional multiresolution analysis such that $\int_{-\infty}^{\infty}\phi(x)dx = 1$ and $\int_{-\infty}^{\infty}\psi(x)dx = 0$. Set $D = d(p+1) + m\bar{p}$,

$$\varphi^{(i)} := \begin{cases} \phi & \text{for } i = 0, \\ \psi & \text{for } i = 1 \end{cases}$$

and $\Phi^{(e)}_{j;k}(x) = 2^{jD/2}\prod_{i=1}^{D}\varphi^{(e_i)}(2^j x_i - k_i)$ for $x \in \mathbb{R}^D,\ k \in \mathbb{Z}^D,\ e \in \{0,1\}^D$. Let $h^{(c)}$ denote the truncated version of h, given by the degenerate counterpart of $\widetilde{h}^{(c)}$

$$\widetilde{h}^{(c)}(x,y,\theta_0) := \begin{cases} h(x,y,\theta_0) & \text{for } |h(x,y,\theta_0)| \le c_h, \\ -c_h & \text{for } h(x,y,\theta_0) < -c_h, \\ c_h & \text{for } h(x,y,\theta_0) > c_h, \end{cases}$$

where $c_h = \max_{x,y\in[-c,c]^D} |h(x,y,\theta_0)|$. The limit distributions of Leucht [18] and the present theorem coincide if we define

$$\lambda^{(c)}_{j;k_1+e_1,k_2+e_2}(\theta_0) := \begin{cases} \iint_{\mathbb{R}^D\times\mathbb{R}^D} h^{(c)}(x,y,\theta_0)\Phi^{(e_1)}_{j;k_1/2}(x)\Phi^{(e_2)}_{j;k_2/2}(y)dx\,dy, & \\ \qquad \text{for } \frac{k_1}{2},\frac{k_2}{2} \in \mathbb{Z}^D, (e_1',e_2')' \in \{0,1\}^{2D}\backslash\{0_{2D}\} & \\ \qquad \text{or } \frac{k_1}{2},\frac{k_2}{2} \in \mathbb{Z}^D, (e_1',e_2')' = 0_{2D}, j=0, & \\ 0, \qquad \text{else} & \end{cases}$$

and if

$$\begin{aligned} cov(Z_{j_1;k_1+e_1}, Z_{j_2;k_2+e_2}) =\ & cov\left(\Phi^{(e_1)}_{j_1;k_1/2}(X_1), \Phi^{(e_2)}_{j_1;k_2/2}(X_1)\right) \\ + \sum_{k=2}^{\infty}\Big[& cov\left(\Phi^{(e_1)}_{j_1;k_1/2}(X_1), \Phi^{(e_2)}_{j_1;k_2/2}(X_k)\right) + cov\left(\Phi^{(e_1)}_{j_1;k_1/2}(X_k), \Phi^{(e_2)}_{j_1;k_2/2}(X_1)\right)\Big]. \end{aligned}$$

□

Proof of Lemma 2.2. Step 1:
$n^{-1}\,\widehat{T}^{(v)}_n \stackrel{P}{\longrightarrow} \mathbb{E}([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]'K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(Y_1|\widetilde{I}_1) - g(\widetilde{I}_1,\theta_0)])$.
According to the assumptions on the set of functions g under $\mathcal{H}_0$, on the sequence of parameter estimators $(\widehat{\theta}_n)_n$ and on the kernel function k, we obtain

$$\begin{aligned} \frac{1}{n}\,\widehat{T}^{(v)}_n =\ & \frac{1}{n^2}\sum_{j,k=1}^{n}[Y_j - \mathbb{E}(Y_j|I_j)]'K(I_j,I_k,\theta_0)[Y_k - \mathbb{E}(Y_k|I_k)] \\ & + \frac{2}{n^2}\sum_{j,k=1}^{n}[Y_j - \mathbb{E}(Y_j|I_j)]'K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] \\ & + \frac{1}{n^2}\sum_{j,k=1}^{n}[\mathbb{E}(Y_j|I_j) - g(I_j,\theta_0)]'K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] + o_P(1). \end{aligned}$$

In view of Theorem 5.1 of Shao and Wu [21] the process $(X_k)_k$ has a Bernoulli shift representation with innovations $((\epsilon_k', Z_k')')_k$ and satisfies the GMC(1) condition. Since the first sum is a degenerate V-statistic in these variables, we obtain its asymptotic negligibility in analogy to the proof of Lemma 4.1. Even though the function G, determining $\mathbb{E}(Y_k|I_k) = G(I_k)$, does not satisfy the continuity assumptions of this result, its proof remains valid since the relation $\mathbb{E}\|G(I_k) - G(\widetilde{I}_k)\|_1 \le C\rho^k$, for some $\rho \in (0,1)$ and a copy $\widetilde{I}_k$ of I_k that is independent of $(I_l)_{l\le 0}$ for $k > P+1$, can be employed instead. The latter inequality results from the contraction condition on G and the fact that the vector $\widetilde{I}_k$ can be chosen such that its last component, $\widetilde{Z}_k$, coincides with the last component, Z_k, of I_k. Coupling methods, similar to those used to investigate the quantity (4.3) in the proof of Lemma 2.1, yield that the middle term is asymptotically negligible. Finally, a Hoeffding decomposition of the kernel of the remaining summand leads to

$$\begin{aligned}&\frac{1}{n^2}\sum_{j,k=1}^{n}[\mathbb{E}(Y_j|I_j)-g(I_j,\theta_0)]'K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k)-g(I_k,\theta_0)]\\&=\Delta(\theta_0)+\frac{2}{n}\sum_{j=1}^{n}h_1(I_j)+\frac{1}{n^2}\sum_{j,k=1}^{n}h_2(I_j,I_k)\end{aligned}\tag{4.7}$$

with

$$\begin{aligned}\Delta(\theta_0)&=\mathbb{E}([\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]'K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(\widetilde{Y}_1|\widetilde{I}_1)-g(\widetilde{I}_1,\theta_0)]),\\h_1(I_j)&=\mathbb{E}([\mathbb{E}(Y_j|I_j)-g(I_j,\theta_0)]'\,K(I_j,\widetilde{I}_j,\theta_0)\,[\mathbb{E}(\widetilde{Y}_j|\widetilde{I}_j)-g(\widetilde{I}_j,\theta_0)]\mid I_j)-\Delta(\theta_0),\\h_2(I_j,I_k)&=[\mathbb{E}(Y_j|I_j)-g(I_j,\theta_0)]'\,K(I_j,I_k,\theta_0)\,[\mathbb{E}(Y_k|I_k)-g(I_k,\theta_0)]\\&\quad-h_1(I_k)-h_1(I_j)-\Delta(\theta_0),\end{aligned}$$

where $(\widetilde{Y}_j',\widetilde{I}_j')'$ is an independent copy of $(Y_j',I_j')'$. Analogous arguments as at the end of proof of Lemma 2.1 imply that the middle summand of the r.h.s. of equation (4.7) is of order $o_P(1)$. The same order can be derived for the last sum which is a degenerate U-statistic in $(I_k)_k$. This finally completes the first step. Obviously, we also have $n^{-1}\,\widehat{T}_n^{(u)}\xrightarrow{P}\Delta(\theta_0)$.

<u>*Step 2:*</u> $\Delta(\theta_0)>0$ *if (i) holds true.*

By inverse Fourier transform we obtain $\bar{k}_{j,\theta_0}(x)=(2\pi)^{-(dp+m\bar{p})/2}\int_{\mathbb{R}^{dp+m\bar{p}}}\mathcal{F}\bar{k}_{j,\theta_0}(t)e^{it'x}dt$, $j=1,\ldots,d$. Thus, the application of Fubini's theorem leads to

$$\begin{aligned}\Delta(\theta_0)&=(2\pi)^{-(dp+m\bar{p})/2}\sum_{j=1}^{d}\int_{\mathbb{R}^{dp+m\bar{p}}}\mathbb{E}\left([\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]_j\,e^{i\zeta_j(I_1)'t}\right)\\&\qquad\times\mathbb{E}\left([\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]_j\,e^{-i\zeta_j(I_1)'t}\right)\mathcal{F}\bar{k}_{i,\theta_0}(t)\,dt\\&=(2\pi)^{-(dp+m\bar{p})/2}\sum_{j=1}^{d}\int_{\mathbb{R}^{dp+m\bar{p}}}\left|\mathbb{E}\left([\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]_j\,e^{i\zeta_j(I_1)'t}\right)\right|^2\mathcal{F}\bar{k}_{i,\theta_0}(t)\,dt,\end{aligned}$$

where the subscript j denotes the j^{th} element of the corresponding vector. By Theorem 1(II) of Bierens [2], the expression $\Delta(\theta_0)$ is strictly positive since $\mathbb{E}(Y_1-g(I_1,\theta_0)|I_1)=\mathbb{E}(Y_1-g(I_1,\theta_0)|\zeta_j(I_1))$ almost surely.

<u>*Step 3:*</u> $\Delta(\theta_0)>0$ *if (ii) holds true.*

One proceeds as in the previous step but employs Theorem 1(I) of Bierens [2].

<u>*Step 4:*</u> $\Delta(\theta_0)>0$ *if (iii) holds true.*

Fubini's theorem yields

$$\Delta(\theta_0)=\sum_{j=1}^{d}\int_{\mathbb{R}^{dp+m\bar{p}+1}}|\mathbb{E}\{[\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]_j\,W_j((1,[\zeta_j(I_1)]')'t)\,w_j(t,\theta_0)\}|^2\,dt.$$

Thus, it remains to verify that there exists a $j\in\{1,\ldots,d\}$ such that $\mathbb{E}\{[\mathbb{E}(Y_1|I_1)-g(I_1,\theta_0)]_j\,W_j((1,[\zeta_j(I_1)]')'t)\,w_j(t,\theta_0)\}\neq 0$ on a set of positive Lebesgue measure, which follows from Theorem 2.3 of Stinchcombe and White [22]. □

Proof of Proposition 3.1. The proof is based on Theorem 3.1 of Leucht [18] and carried out in two steps. First, we verify the bootstrap sample $(X_k^*)_k$ to satisfy her condition (A1*), i.e. convergence of the bivariate bootstrap distributions to the corresponding bivariate distribution of the original process as well as stationarity and weak dependence of the bootstrap variables with probability tending to one. Afterwards, the assertions of the present theorem is proved.
Step 1: Verification of (A1*) of [18].
We apply Lemma 4.2. To this end, the notation

$$\bar{X}_t = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} g(Y_{t-p}, \dots, Y_{t-1}, Z_{t-\bar{p}+1}, \dots, Z_t, \theta_0) + \epsilon_t \\ Z_t \end{pmatrix} =: G(\bar{X}_{t-1}, \dots, \bar{X}_{t-P}, \varepsilon_t; \theta_0)$$

is introduced with $\varepsilon_t = (\epsilon_t', Z_t')'$. Thus, our bootstrap procedure is equivalent to the Algorithm [B1] in Subsection 4.2 if $\varepsilon_1^* := (\epsilon_t^{*\prime}, Z_t^{*\prime})' \xrightarrow{d} \varepsilon_1$, in probability. The validity of (A1*) then follows from Lemma 4.2 if its prerequisites are fulfilled. Under (M1) the process $(X_t)_t$ satisfies the condition (4.10) of Subsection 4.2. Moreover, the prerequisites (a) and (b) of Lemma 4.2 are satisfied due to (M5)(i). Therefore, in order to obtain (A1*), it remains to verify $\varepsilon_1^* \xrightarrow{d} \varepsilon_1$, in probability. First, note that $Z_1^* \xrightarrow{d} Z_1$, in probability, since the variables Z_k^* are drawn via Efron's bootstrap from $Z_{1-\bar{p}}, \dots, Z_n$. Next, we prove $\epsilon_1^* \xrightarrow{d} \epsilon_1$, in probability, more precisely, we show that $P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) \leq \delta$ for any $\eta, \delta > 0$ and all $n \geq n_0(\eta, \delta)$ and arbitrary fixed $x \in \mathbb{R}^d$. To this end, one splits up

$$\begin{aligned} P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) \leq P\Big(\sum_{r=1}^d \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{x_r - \|g(I_j,\widehat{\theta}_n) - g(I_j,\theta_0)\|_1 - \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i,\widehat{\theta}_n)]\|_1 \leq} \\ {}_{\epsilon_{j,r} \leq x_r + \|g(I_j,\widehat{\theta}_n) + g(I_j,\theta_0)\|_1 + \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i,\widehat{\theta}_n)]\|_1} > \frac{\eta}{2}\Big) \\ + P\left(\left| \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\epsilon_j \preceq x} - F_{\epsilon_1}(x) \right| > \frac{\eta}{2} \right). \end{aligned}$$

The first summand of the r.h.s. can be bounded by employing Markov's inequality:

$$\begin{aligned} &\frac{C}{\eta}\Big[\sum_{r=1}^d P\big(\epsilon_{1,r} \in [x_r - 3\kappa, x_r + 3\kappa]\big) + P\Big(\sum_{r=1}^d \Big| \frac{1}{n} \sum_{i=1}^n \epsilon_{i,r} \Big| > \kappa \Big) \\ &\quad + P\Big(\frac{1}{n} \sum_{i=1}^n \Big\| g(I_i, \widehat{\theta}_n) - g(I_i, \theta) \Big\|_1 > \kappa \Big) + \frac{1}{n} \sum_{j=1}^n P\Big(\Big\| g(I_j, \widehat{\theta}_n) - g(I_j, \theta) \Big\|_1 > \kappa \Big) \Big] \\ &\leq \frac{C}{\eta}\Big[\sum_{r=1}^d P\big(\epsilon_{1,r} \in [x_r - 3\kappa, x_r + 3\kappa]\big) + n^{-1}\kappa^{-2} + P(L(I_1) > 1/\kappa) \\ &\quad + P\Big(\|\widehat{\theta}_n - \theta_0\|_1 > \kappa^2 \Big) + P\Big(n^{-1} \sum_{i=1}^n L(I_i) > 1/\kappa \Big) \Big], \end{aligned}$$

which is less than $\delta/2$ for sufficiently small $\kappa > 0$ and all $n \geq n_0$ for some $n_0(\eta, \delta, \kappa)$. Concerning the second summand, Chebychev's inequality yields

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\epsilon_j\preceq x} - F_{\epsilon_1}(x)\right| > \eta\right)\\
&\leq \frac{1}{n\,\eta^2} + \frac{2}{(n\,\eta)^2}\sum_{1\leq j<k\leq n}\mathbb{E}\big\{[\mathbb{1}_{\epsilon_j\preceq x} - F_{\epsilon_1}(x)][\mathbb{1}_{\epsilon_k\preceq x} - F_{\epsilon_1}(x)]\big\}\\
&\leq \frac{C}{n\,\eta^2}
\end{aligned}$$

due to the independence of the innovations. Hence, for all $\delta > 0$ and any sufficiently small $\kappa > 0$ there exists an $n_0(\kappa, \delta)$ with $P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) < \delta,\ \forall\, n \geq n_0$. This completes the proof of the convergence of the (conditional) distribution of the bootstrap innovations to P_{ϵ_1}. Due to the independence of Z_1 and ϵ_1 as well as of Z_1^* and ϵ_1^*, we have $\varepsilon_1^* \stackrel{d}{\longrightarrow} \varepsilon_1$, in probability. Therefore, all prerequisites of Lemma 4.2 are fulfilled which assures the validity of (A1*) of [18].

Step 2: Convergence of $\widehat{T}_n^{(u)*}$ and $\widehat{T}_n^{(v)*}$.

First, we have $\mathbb{E}^*|Z_k^*|^2 \stackrel{P}{\longrightarrow} \mathbb{E}|Z_1|^2$. Moreover, $\mathbb{E}^*|Y_k^*|^{4+\varepsilon} = O_P(1)$ for some $\varepsilon > 0$ as, on the one hand, $\mathbb{E}^*|g(I_k^*, \widehat{\theta}_n)|^{4+\varepsilon} = O_P(1)$ by assumption and, on the other hand,

$$\begin{aligned}
\mathbb{E}^*|\epsilon_1^*|^{4+\varepsilon} &\leq C\left[\mathbb{E}|\epsilon_1|^{4+\varepsilon} + \frac{1}{n}\sum_{k=1}^{n}|g(I_k, \widehat{\theta}_n) - g(I_k, \theta_0)|^{4+\varepsilon}\right]\\
&\leq C + o_P(1)\,\frac{1}{n}\sum_{k=1}^{n}|L(I_k)|^{4+\varepsilon}\\
&= O_P(1).
\end{aligned}$$

W.l.o.g. we consider $\widehat{T}_n^{(v)*}$ only. Note that for all $\alpha, \beta \in \{1, \ldots, q\}$

$$|\mathbb{E}^*([g^{(1)}(I_1^*, \widehat{\theta}_n)'K(I_1^*, \widetilde{I}_1^*, \widehat{\theta}_n)g^{(1)}(\widetilde{I}_1^*, \widehat{\theta}_n)]_{\alpha,\beta}) - a(\theta_0)| \stackrel{P}{\longrightarrow} 0,\ \forall\, \alpha, \beta \in \{1, \ldots, q\},$$

where $\widetilde{I}_1^*$ denotes and independent copy of I_1^*, conditionally on $\mathbb{X}_n$. Following the lines of the proof of Lemma 2.1, one obtains $\widehat{T}_n^{(v)*} - T_n^{(v)*} = o_{P^*}(1)$ with

$$\begin{aligned}
T_n^{(v)*} := \frac{1}{n}\sum_{j,k=1}^{n}\big\{&\epsilon_j^{*\prime}\left[K(I_j^*, I_k^*, \widehat{\theta}_n)\epsilon_k^* - 2\,V(I_j^*, \theta_0)\,l(X_k^*, \widehat{\theta}_n)\right]\\
&+ [l(X_j^*, \widehat{\theta}_n)]'\,a(\theta_0)\,l(X_k^*, \widehat{\theta}_n)\big\}
\end{aligned}$$

if for all $\alpha, \beta \in \{1, \ldots, q\}$,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j^{*\prime}\,[\mathbb{E}^*(k_\alpha(I_j^*, \widetilde{I}_j^*, \widehat{\theta}_n)\,g_{\alpha,\beta}^{(1)}(\widetilde{I}_j^*, \widehat{\theta}_n)|I_j^*) - V_{\alpha,\beta}(I_j^*, \theta_0)] = o_{P^*}(1). \tag{4.8}$$

In order to verify this relation, we prove asymptotic negligibility of

$$\mathbb{E}^*\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j^{*\prime}\left[\int_{\mathbb{R}^{dp+m\bar{p}}}k_\alpha(I_j^*, x, \widehat{\theta}_n)\,g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n)P_{I_1^*|\mathbb{X}_n} - V_{\alpha,\beta}(I_j^*, \theta_0)\right]\right\}^2.$$

The introduction of a copy $\widetilde{X}_k^*$ of $X_k^*,\ k \in \mathbb{N}$, that is independent of X_j^*, conditionally on $X_1, \dots, X_n$, and such that the bootstrap counterpart of inequality (4.1) holds for $n = k - j > P + 1$ with probability tending to one, leads to

$$\begin{aligned}
&\mathbb{E}^* \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n \epsilon_j^{*\prime} \left[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \widehat{\theta}_n) \right] \right\}^2 \\
&\le O_P(1) \left(\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_1^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_1^*, \theta_0) \right|^2 \right)^{1/2} \\
&\quad + \frac{2}{n} \sum_{\substack{j,k=1 \\ k-j>P}}^n \mathbb{E}^* \Big(\epsilon_j^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \theta_0) \Big] \\
&\quad \times \Big\{ \epsilon_k^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_k^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n} - V_{\alpha,\beta}(I_k^*, \theta_0) \Big] \\
&\quad - \widetilde{\epsilon}_k^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(\widetilde{I}_k^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(\widetilde{I}_k^*, \theta_0) \Big] \Big\} \Big) \\
&\le O_P(1) \left(\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_1^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_1^*, \theta_0) \right|^{1/\delta} \right)^\delta
\end{aligned}$$

for some $\delta \in (0, 1/2]$. It remains to verify that the quantity in round brackets can be bounded with probability tending to one by any given $\varepsilon > 0$ if n is sufficiently large. For this purpose we introduce a compact set D such that

$$\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n) g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \theta_0) \right|^{1/\delta} \mathbb{1}_{I_1^* \notin D} \le C P(I_1^* \notin D) \le \frac{\varepsilon}{2}$$

holds with probability tending to one. Moreover, according to the convergence of the finite-dimensional distributions and due to Lipschitz continuity of the elements of K, we have

$$\max_{z \in D} \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(z, x, \widehat{\theta}_n) g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_\alpha(z, \theta_0) \right|^{1/\delta} \le \frac{\varepsilon}{2}$$

with probability tending to one, which finally yields (4.8).

Thus, it suffices to show the desired convergence for $T_n^{(v)*}$ instead of $\widehat{T}_n^{(v)*}$. The statistic $T_n^{(v)*}$ is the bootstrap counterpart of the degenerate V-statistic $T_n^{(v)}$ of Lemma 2.1. It remains to show that its kernel satisfies (A2*) and (A3*) of Leucht [18]. Define h and f_θ as in the proof of Theorem 2.1. Then we get

$$P \left(\sup_{k \in \mathbb{N}} \mathbb{E}^* |h(X_1^*, X_{1+k}^*, \widehat{\theta}_n)|^{2+\varepsilon} + \mathbb{E}^* |h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|^{2+\varepsilon} \le K \right) \xrightarrow{P} 1$$

as well as

$$P \left(\sup_{k_1, \dots, k_5 \in \mathbb{N}} \mathbb{E}^* \left(\max_{a_1, a_2 \in [-A, A]^{d(p+1)+m\bar{p}}} [f_{\widehat{\theta}_n}(\bar{X}_{k_1}^*, \bar{X}_{k_2}^* + a_1, \bar{X}_{k_3}^*, \bar{X}_{k_4}^* + a_2)]^\eta \|\bar{X}_{k_5}^*\|_1 \right) \le K \right)$$
$$\xrightarrow{P} 1$$

for some $\varepsilon > 0,\ K < \infty$ and any $(\bar{X}_{k_1}^{*\prime}, \ldots, \bar{X}_{k_5}^{*\prime})'$ consisting of independent subvectors $(\bar{X}_{k_{j_1(m)}}^{*\prime}, \ldots, \bar{X}_{k_{j_l(m)}}^{*\prime})' \stackrel{d}{=} (X_{k_{j_1(m)}}^{*\prime}, \ldots, X_{k_{j_l(m)}}^{*\prime})',\ l, m = 1, \ldots, 5$. Hence, the conditions (A2*) and (A3*) of [18] are fulfilled. Moreover, according to the first step of the proof, the bootstrap sample satisfies (A1*). Consequently, we obtain bootstrap consistency by Theorem 3.1 of her paper. □

Proof of Lemma 3.1. *Step 1: Behaviour under $\mathcal{H}_0$.*
These results follow from Proposition 3.1.
Step 2: Behaviour under $\mathcal{H}_1$.
Under our assumptions one can prove that the proposed bootstrap method imitates a null situation, namely the (unique) stationary solution of $\bar{Y}_k = g(\bar{Y}_{k-p}, \ldots, \bar{Y}_{k-1}, Z_{k-\bar{p}+1}, \ldots, Z_k, \theta_0) + \bar{\epsilon}_k$. Here, $((Z_k', \bar{\epsilon}_k')')_k$ is a sequence of i.i.d. random variables, where Z_1 and $\bar{\epsilon}_1$ are independent and $\bar{\epsilon}_1 \stackrel{d}{=} Y_1 - g(I_1, \theta_0) - \mathbb{E}Y_1 + \mathbb{E}g(I_1, \theta_0)$. To this end, the arguments of step 1 of the proof of Proposition 3.1 are invoked. Thus, it remains to show that $\epsilon_1^* \stackrel{d}{\longrightarrow} \bar{\epsilon}_1$, in probability. We have

$$
\begin{aligned}
&P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) \\
&\le P\Big(\sum_{r=1}^{d} \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{x_r - \|g(I_j, \widehat{\theta}_n) - g(I_j, \theta_0)\|_1 - \|n^{-1} \sum_{i=1}^{n} [Y_i - g(I_i, \widehat{\theta}_n) - \mathbb{E}Y_1 + \mathbb{E}g(I_1, \theta_0)]\|_1 \le \bar{\epsilon}_{j,r}} \\
&\qquad\qquad {}_{\le x_r + \|g(I_j, \widehat{\theta}_n) + g(I_j, \theta_0)\|_1 + \|n^{-1} \sum_{i=1}^{n} [Y_i - g(I_i, \widehat{\theta}_n) - \mathbb{E}Y_1 + \mathbb{E}g(I_1, \theta_0)]\|_1} > \frac{\eta}{2} \Big) \\
&\quad + P\left(\left| \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x) \right| > \frac{\eta}{2} \right).
\end{aligned}
$$

An upper bound for the first summand on the r.h.s. is obtained similarly as in the proof of Proposition 3.1. However, the approximation of the second summand has to be modified since the underlying variables are no longer independent here. Thus, we have to use certain coupling arguments to verify asymptotic negligibility of

$$
\begin{aligned}
&P\left(\left| \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x) \right| > \eta \right) \\
&\le \frac{1}{n\,\eta^2} + \frac{2}{(n\,\eta)^2} \sum_{1 \le j < k \le n} \mathbb{E}\big\{ [\mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)][\mathbb{1}_{\bar{\epsilon}_k \preceq x} - F_{\bar{\epsilon}_1}(x)] \big\}.
\end{aligned}
$$

Unfortunately, the involved terms are not continuous and thus the coupling techniques cannot be invoked directly. Therefore, the introduction of a smoothing function is helpful in order to approximate the expectations above. For $a > 0$ we define

$$
G_a(z) := \int_{\mathbb{R}^d} w_a(u) \mathbb{1}_{z \preceq x + u}\, du - F_{\bar{\epsilon}_1}(x), \quad z \in \mathbb{R}^d.
$$

Here, $(w_a)_{a>0}$ is a family of nonnegative functions with $\operatorname{supp}(w_a) \subseteq \{u \in \mathbb{R}^d \mid u \succeq 0, \|u\|_1 \le a\}$, $\int_{\mathbb{R}^d} w_a(u)\, du = 1$ and $\|w_a\|_\infty \le C a^{-d}$. Thus, G_a is Lipschitz continuous with $\operatorname{Lip}(G_a) \le C a^{-1}$ as

$$
|G_a(z) - G_a(\bar{z})| \le \sum_{i=1}^{d} \int_{\mathbb{R}^d} w_a(u)\, [\mathbb{1}_{z_i - x_i \le u_i \le \bar{z}_i - x_i} + \mathbb{1}_{\bar{z}_i - x_i \le u_i \le z_i - x_i}]\, du \le C a^{-1} \|z - \bar{z}\|_1.
$$

Let $\widetilde{\epsilon}_k$ be a copy of $\bar{\epsilon}_k$ that is independent of $\bar{\epsilon}_j$ such that (4.1) holds with $n = k - j$ for $k - j > P + 1$. Then,

$$\begin{aligned}
&\frac{2}{(n\,\eta)^2} \sum_{1 \le j < k \le n} \mathbb{E}\big\{[\mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)][\mathbb{1}_{\bar{\epsilon}_k \preceq x} - F_{\bar{\epsilon}_1}(x)]\big\} \\
&= \frac{2}{(n\,\eta)^2} \sum_{1 \le j < k \le n} \mathbb{E}\Big\{[\mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)]\Big[\mathbb{1}_{\bar{\epsilon}_k \preceq x} - \int_{\mathbb{R}^d} w_a(u) \mathbb{1}_{\bar{\epsilon}_k \preceq x+u} du\Big]\Big\} \\
&\quad + \frac{2}{(n\,\eta)^2} \sum_{1 \le j < k \le n} \mathbb{E}\big\{[\mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)] G_a(\bar{\epsilon}_k)]\big\} \\
&\le \frac{2}{\eta^2} \sum_{r=1}^d P(x_r \le \bar{\epsilon}_{1,r} \le x_r + a) + \frac{C}{\eta^2\, n} + \frac{2}{(n\,\eta)^2} \sum_{\substack{1 \le j < k \le n \\ k-j > P+1}} \mathbb{E}|G_a(\bar{\epsilon}_k) - G_a(\widetilde{\epsilon}_k)| \\
&\le \frac{2}{\eta^2} \sum_{r=1}^d P(x_r \le \bar{\epsilon}_{1,r} \le x_r + a) + \frac{C}{\eta^2\, n} + \frac{C}{\eta^2\, a\, n} \sum_{k=1}^\infty \rho^k
\end{aligned}$$

for some $\rho \in (0,1)$. Thus, the latter expression can be bounded by any $\delta > 0$ whenever a is sufficiently small and $n \ge n_0(\delta, a)$. This finally implies the validity of (A1*) in this case. Step 2 of the proof of Proposition 3.1 remains valid under $\mathcal{H}_1$. Consequently, the quantiles of the bootstrap statistics are bounded with probability tending to one. In accordance with Lemma 2.2, we obtain the desired assertions. □

4.2. **Auxiliary results.** Let V_n be a degenerate V-statistic, i.e. with a symmetric kernel function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $V_n = n^{-2} \sum_{j,k=1}^n h(X_j, X_k)$ and $\int_{\mathbb{R}^d} h(x,y) P_{X_0}(dx) = 0,\ \forall\, y \in \mathbb{R}^d$.

Lemma 4.1. *Let $(X_k)_{k \in \mathbb{Z}}$ be a stationary process with values in $\mathbb{R}^d$ and $E\|X_0\|_1 < \infty$ such that for all $n_1 \le n_2 \le n_3 \in \mathbb{N}$, there exists a copy $(\widetilde{X}'_{n_1}, \widetilde{X}'_{n_2}, \widetilde{X}'_{n_3})'$ of $(X'_{n_1}, X'_{n_2}, X'_{n_3})'$ that is independent of $\sigma(X_k,\ k \le 0)$ and such that $\sum_{k=1}^3 \mathbb{E}\|\widetilde{X}_{n_k} - X'_{n_k}\|_1 \le K \rho^{n_1}$ for some $K < \infty$ and $\rho \in (0,1)$. Assume that $\sup_{k \in \mathbb{Z}} \mathbb{E}|h(X_0, X_k)|^{2+\varepsilon} + \mathbb{E}|h(X_0, \widetilde{X}_0)|^{2+\varepsilon} < \infty$, where $\widetilde{X}_0$ denotes an independent copy of X_0. Moreover, suppose that there exists a function $f : \mathbb{R}^{4d} \to \mathbb{R}$ such that*

$$|h(x,y) - h(\bar{x}, \bar{y})| \le f(x, \bar{x}, y, \bar{y})\, [\|x - \bar{x}\|_1 + \|y - \bar{y}\|_1], \quad \forall\, x, \bar{x}, y, \bar{y} \in \mathbb{R}^d,$$

with $\sup_{k_1, \ldots, k_5 \in \mathbb{Z}} \mathbb{E}\left(f(Y_{k_1}, Y_{k_2} + c_1, Y_{k_3}, Y_{k_4} + c_2)^{1+\varepsilon} \|Y_{k_5}\|_1\right) < \infty$, for some $\varepsilon > 0$ and any $(Y'_{k_1}, \ldots, Y'_{k_5})'$ consisting of independent subvectors $(Y'_{k_{j_1(m)}}, \ldots, Y'_{k_{j_l(m)}})' \overset{d}{=} (X'_{k_{j_1(m)}}, \ldots, X'_{k_{j_l(m)}})'$, $l, m = 1, \ldots, 5$. Then, there exists a finite constant C such that

$$\limsup_{n \to \infty} \mathbb{E}\,(n\, V_n)^2 \le C.$$

Proof. We have

$$\mathbb{E}\,(n\, V_n)^2 \le \frac{8}{n^2} \sum_{i \le j; k \le l; i \le k}^{n} |\mathbb{E} h(X_i, X_j) h(X_k, X_l)| \le 8 \sup_{1 \le k \le n} \mathbb{E}|h(X_1, X_k)|^2 + \frac{8}{n^2} \sum_{r=1}^{n-1} \sum_{t=1}^{4} Z_{n,r}^{(t)}$$

with

$$Z_{n,r}^{(1)} = \sum_{\substack{i\leq j,k\leq l \\ r:=\min\{j,k\}-i\geq l-\max\{j,k\}}}^{n} |\mathbb{E}h(X_i,X_j)h(X_k,X_l) - \mathbb{E}h(X_i,\widetilde{X}_j)h(\widetilde{X}_k,\widetilde{X}_l)|,$$

$$Z_{n,r}^{(2)} = \sum_{\substack{i\leq j,k\leq l \\ r:=l-\max\{j,k\}>\min\{j,k\}-i}}^{n} |\mathbb{E}h(X_i,X_j)h(X_k,X_l) - \mathbb{E}h(X_i,X_j)h(X_k,\widetilde{X}_l)|,$$

$$Z_{n,r}^{(3)} = \sum_{\substack{i\leq k\leq l<j \\ r:=i-k\geq j-l}}^{n} |\mathbb{E}h(X_i,X_j)h(X_k,X_l) - \mathbb{E}h(X_i,\widetilde{X}_j)h(\widetilde{X}_k,\widetilde{X}_l)|,$$

$$Z_{n,r}^{(4)} = \sum_{\substack{i\leq k\leq l<j \\ r:=j-l>i-k}}^{n} |\mathbb{E}h(X_i,X_j)h(X_k,X_l) - \mathbb{E}h(X_i,\widetilde{X}_j)h(X_k,X_l)|.$$

Here, in each summand of $Z_{n,r}^{(1)}$ and $Z_{n,r}^{(3)}$ the vector $(\widetilde{X}_j',\widetilde{X}_k',\widetilde{X}_l')'$ is chosen such that it is independent of X_i, $(\widetilde{X}_j',\widetilde{X}_k',\widetilde{X}_l')' \stackrel{d}{=} (X_j',X_k',X_l')'$ and $\mathbb{E}\|(\widetilde{X}_j',\widetilde{X}_k',\widetilde{X}_l')' - (X_j',X_k',X_l')'\|_1 \leq C\rho^r$ holds. Within $Z_{n,r}^{(2)}$ (respectively $Z_{n,r}^{(4)}$) the random variable $\widetilde{X}_l$ (respectively $\widetilde{X}_j$) is chosen to be independent of $(X_i',X_j',X_k')'$ (respectively $(X_i',X_k',X_l')'$) such that $\widetilde{X}_l \stackrel{d}{=} X_l$ (respectively $\widetilde{X}_j \stackrel{d}{=} X_j$) and $\mathbb{E}\|\widetilde{X}_j - X_j\|_1 \leq C\rho^r$ ($\mathbb{E}\|\widetilde{X}_j - X_j\|_1 \leq C\rho^r$) hold. Thus, by degeneracy the subtrahends of these expressions vanish. Moreover, note that the number of summands of $Z_{n,r}^{(t)},\ t=1,\ldots,4$, is bounded by $(r+1)n^2$. For further approximations we exemplarily take $Z_{n,r}^{(2)}$ and $\sup_{1\leq k\leq n}\mathbb{E}|h(X_1,X_k)|^2$ into further considerations. Obviously, the latter term is bounded by assumption. Applying Hölder's inequality iteratively, the first expression can be bounded from above as follows:

$$\begin{aligned}
Z_{n,r}^{(2)} \leq & \sum_{\substack{i\leq j,k\leq l \\ r:=l-\max\{j,k\}>\min\{j,k\}-i}}^{n} |(\mathbb{E}|h(X_i,X_j)|^{1+\varepsilon/2}|h(X_k,X_l)-h(X_k,\widetilde{X}_l)|)^{1/(1+\varepsilon/2)} \\
& \qquad \times (\mathbb{E}|h(X_k,X_l)-h(X_k,\widetilde{X}_l)|)^{\varepsilon/(2+\varepsilon)} \\
\leq & \, C \sum_{\substack{i\leq j,k\leq l \\ r>\min\{j,k\}-i}}^{n} (\mathbb{E}[f(X_k,X_l,X_k,\widetilde{X}_l)^{1+\varepsilon/2}(\|X_l\|_1+\|\widetilde{X}_l\|_1)])^{2\varepsilon/(1+\varepsilon/2)^2}\rho^{r\varepsilon^2/(2+\varepsilon)^2} \\
\leq & \, Cn^2\sum_{r=1}^{\infty}(r+1)\rho^{r\varepsilon^2/(2+\varepsilon)^2}
\end{aligned}$$

We can proceed similarly for the remaining summands and end up with $n^{-2}\sum_{r=1}^{n-1}\sum_{t=1}^{4} Z_{n,r}^{(t)} \leq n^{-2}\sum_{r=1}^{\infty}(r+1)\rho^{r\varepsilon^2/(2+\varepsilon)^2} \leq C$ which in turn implies the assertion of the lemma. $\square$

As a final result we verify the validity of parametric bootstrap for contractive iterative random functions. More precisely, we show that the finite dimensional distributions of the bootstrap counterpart of a Markov process with

$$X_t = G(X_{t-1}, \ldots, X_{t-p}, \varepsilon_t), \quad t \in \mathbb{Z}, \tag{4.9}$$

converge to those of the original process. To this end, we impose the following contraction condition:

$$\exists y_0 \in \mathbb{R}^{dp} \text{ with } \mathbb{E}\|G(y_0, \varepsilon_0)\|_2 < \infty \text{ and } \exists \alpha_1, \ldots \alpha_p \geq 0 \text{ with } \sum_{k=1}^{p} a_k < 1 \text{ and}$$

$$\mathbb{E}\|G(y, \varepsilon_0) - G(\bar{y}, \varepsilon_0)\|_2 \leq \sum_{k=1}^{p} a_k \|y_k - \bar{y}_k\|_2, \quad \forall\, y = (y_1', \ldots, y_p')',\ \bar{y} = (\bar{y}_1', \ldots, \bar{y}_p')'. \tag{4.10}$$

In particular, this assures stationarity and the GMC(1) dependence condition of Shao and Wu [21]. The following nodel-based bootstrap method is considered:
[B1] Algorithm.

(1) Calculate the parameter estimator $\widehat{\theta}_n(X_1, \ldots, X_n)$ such that $\widehat{\theta}_n \xrightarrow{P} \theta_0$.
(2) Draw i.i.d. bootstrap innovations $\varepsilon_k^*,\ k \geq 1$, such that $\varepsilon_1^* \xrightarrow{d} \varepsilon_1$, in probability.
(3) Choose some initial vector $(X_0^{*\prime}, \ldots, X_{1-p}^{*\prime})'$.
(4) Generate $X_k^* = G(X_{k-1}^*, \ldots, X_{k-p}^*, \varepsilon_k^*; \widehat{\theta}_n),\ k = 1, 2, \ldots$.

Lemma 4.2. *Let $(X_k)_{k\in\mathbb{Z}}$ be the stationary solution of (4.9), satisfying the condition (4.10) Moreover, assume the function $G : \mathbb{R}^{dp+q} \times \Theta \to \mathbb{R}^d$ to be continuous, where Θ is an open subset of $\mathbb{R}^m$. Suppose that the following conditions are fulfilled:*

(a) *There exists a $y_0^* \in \mathbb{R}^d$ and some $K_1 \in \mathbb{R}$ such that $\mathbb{E}^*\|G(y_0^*, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \leq K_1$, with probability tending to one.*
(b) *There are a constant $\delta \in (0, 1)$ and nonnegative random variables $a_{n,1}, \ldots, a_{n,p}$ with $P(\sum_{i=1}^p a_{n,i} \leq 1 - \delta) \longrightarrow_{n\to\infty} 1$ and such that*

$$\mathbb{E}^*\|G(y, \varepsilon_0^*; \widehat{\theta}_n) - G(\bar{y}, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \leq \sum_{k=1}^{p} a_{n,k} \|y_k - \bar{y}_k\|_2, \quad \forall\, y, \bar{y} \in \mathbb{R}^{dp}.$$

Then the bootstrap process generated by Algorithm [B1] satisfies $(X_{t_1}^*, \ldots, X_{t_k}^*) \xrightarrow{d} (X_{t_1}, \ldots, X_{t_k}),\ t_1, \ldots, t_k \in \mathbb{Z},\ k \in \mathbb{N}$, if the initial vector is drawn from the stationary bootstrap distribution, which exists with probability tending to one. Moreover, the bootstrap process satisfies the GMC(1) condition of [21] with probability tending to one.

Proof. Step 1: Stationarity.
Define

$$\mathfrak{X}_n := \Big\{ \mathbb{X}_n = x_n \Big| \sum_{i=1}^{p} |a_{n,i}| \leq 1 - \delta,\ \mathbb{E}\left(\|G(y_0^*, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \mid \mathbb{X}_n = x_n \right) \leq K_1,$$

$$\|\widehat{\theta}_n - \theta_0\|_2 \leq \delta_n,\ \widehat{\theta}_n \in \Theta \Big\},\ n \in \mathbb{N},$$

where the null sequence $(\delta_n)_n$ can be chosen such that $P(\mathbb{X}_n \in \mathfrak{X}_n) \longrightarrow_{n\to\infty} 1$. Thus, there exists a stationary law to the bootstrap model provided $\mathbb{X}_n \in \mathfrak{X}_n$, cf. Theorem 5.1 of Shao and Wu [21].

Step 2: GMC(1) condition.

According to the proofs of Proposition 6.3.22 and Lemma 6.2.10 of Duflo [6], the GMC(1) condition is fulfilled whenever the the spectral radii of the matrices

$$A_n = \begin{pmatrix} a_{n,1} & a_{n,2} & \cdots & a_{n,p-1} & a_{n,p} \\ 1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad n \in \mathbb{N},$$

are less than a constant that is strictly smaller than one uniformly on $(\mathfrak{X}_n)_n$. By virtue of Lemma 4.1.1 of Duflo [6], the spectral radius is the absolute largest inverse of the zeros of the polynomial $1 - \sum_{k=1}^n a_{n,k}\, z^k,\ z \in \mathbb{C}$. On $\mathfrak{X}_n$, a necessary condition for $z \in \mathbb{C}$ being a zero of the considered polynomial is given by $|z| \geq 1/(1-\delta)$. This leads to an upper bound $(1-\delta)$ for the spectral radius of $A_n,\ n \in \mathbb{N}$.

Step 3: Convergence of the finite-dimensional distributions.

Here, Lemma 4.3 of Neumann and Paparoditis [20] is invoked. Thus, we have to check whether its prerequisites (i) to (iv) hold in probability. Condition (i) is satisfied if for each $\delta > 0$ and each compact subset K of $\mathbb{R}^{dp}$,

$$P\left(\sup_{y\in K} d\left(P_{G(y_1,\dots,y_p,\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n}, P_{G(y_1,\dots,y_p,\varepsilon_1;\theta_0)}\right) > \delta\right) \xrightarrow{P} 0.$$

To this end, suppose that the bootstrap process is started with its stationary distribution if it exists and define $\widetilde{\mathfrak{X}}_n \subseteq \mathfrak{X}_n$ such that moreover $P_{\varepsilon_1^*|\mathbb{X}_n=x_n} \Longrightarrow P_{\varepsilon_1}$ uniformly for all sequences $(x_n)_n$ with $x_n \in \widetilde{\mathfrak{X}}_n,\ n \in \mathbb{N}$, and $P(\mathbb{X}_n \in \widetilde{\mathfrak{X}}_n) \longrightarrow_{n\to\infty} 1$. We construct a grid $y^{(1)}, \dots, y^{(M)}$ in K such that $\sup_{y\in K} \min\{\|y - y^{(i)}\|_2 \mid i \in \{1,\dots,M\}\} \leq \delta/3$. Let $(x_n)_n$ be an arbitrary sequence with $x_n \in \widetilde{\mathfrak{X}}_n,\ n \in \mathbb{N}$. The application of the triangular inequality yields

$$\begin{aligned} &\sup_{y\in K} d\left(P_{G(y_1,\dots,y_p,\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n=x_n}, P_{G(y_1,\dots,y_p,\varepsilon;\theta_0)}\right) \\ &\leq \frac{\delta}{3}\sum_{i=1}^{p}[a_k + a_{n,k}(x_n)] + \sum_{i=1}^{M} d\left(P_{G(y_1^{(i)},\dots,y_p^{(i)},\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n=x_n}, P_{G(y_1^{(i)},\dots,y_p^{(i)},\varepsilon_1;\theta_0)}\right), \end{aligned}$$

where the first summand can be bounded from above by $2\,\delta/3$. The second one is less than $\delta/3$ for all $n \geq n_0$ for some $n_0 \in \mathbb{N}$ since G is continuous and $P_{(\varepsilon_1^{*\prime},\widehat{\theta}_n')'|\mathbb{X}_n=x_n} \Longrightarrow P_{(\varepsilon_1',\theta_0')'}$ for all $(x_n)_n$ with $x_n \in \widetilde{\mathfrak{X}}_n,\ n \in \mathbb{N}$. This implies the condition (i) of Lemma 4.3 of [20].

Regarding the constraint (ii) we obtain

$$\begin{aligned} &\sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\leq\delta} d\left(P_{Y_t|Y_{t-1}=y_1,\dots,Y_{t-p}=y_p}, P_{Y_t|Y_{t-1}=\widetilde{y}_1,\dots,Y_{t-p}=\widetilde{y}_p}\right) \\ &\leq \sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\leq\delta} \mathbb{E}\|G(y,\varepsilon_0;\theta_0) - G(\widetilde{y},\varepsilon_0;\theta_0)\|_2 \\ &\leq \sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\leq\delta} \sum_{k=1}^{p} a_k\|y_k - \bar{y}_k\|_2, \end{aligned}$$

where the latter expression tends to zero as $\delta \to 0$.

Next, tightness of the bootstrap process has to be verified. To this end, we consider again an arbitrary sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$. Based on $\mathbb{X}_n = x_n$, let $(X_k^*)_k$ be the bootstrap process that is started with its stationary distribution. In order to prove tightness of the bootstrap process, it now suffices to show that $\sup_n \mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n) \leq K_2$ for some finite constant K_2 and all sequences $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$. Since $\sup_{k\in\mathbb{Z}} \mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n)$ is finite, we obtain

$$\begin{aligned}
&\mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n) \\
&\leq \mathbb{E}(\|G(X_{k-1}^*, \ldots, X_{k-p}^*, \varepsilon_k^*; \widehat{\theta}_n) - G(y_0^*, \varepsilon_k^*; \widehat{\theta}_n)\|_2 \mid \mathbb{X}_n = x_n) + K_1 \\
&\leq \sum_{j=1}^{p} a_{n,j} \mathbb{E}(\|X_{k-1}^*\|_2 \mid \mathbb{X}_n = x_n) + \|y_0^*\|_2 + K_1.
\end{aligned}$$

By stationarity this implies

$$\mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n) \leq \frac{\|y_0^*\|_2 + K_1}{1 - \sum_{j=1}^{p} a_{n,j}} \leq \frac{\|y_0^*\|_2 + K_1}{\delta}, \quad k \in \mathbb{N},$$

for any sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$.

Finally, the condition (iv) of Lemma 4.3 of Neumann and Paparoditis [20] follows from Theorem 5.1 of Shao and Wu [21] since they show existence of a stationary distribution and convergence of every arbitrarily started process towards this limit. This in turn yields its uniqueness. Consequently, for all $k \in \mathbb{N}$, $P_{X_1^*,\ldots,X_k^*} \Longrightarrow P_{X_1,\ldots,X_k}$, in probability. □

Acknowledgment . This research was funded by the German Research Foundation DFG (project: NE 606/2-1).

References

[1] Bartels, K. (1999). Tests zur Modellspezifikation in der nichtlinearen Regression. *Dissertation*.

[2] Bierens, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20**, 105–134.

[3] Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443-1458.

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

[5] de Jong, R M. (1996). The Bierens' tests under data dependence. *J. Econometrics* **72**, 1–32.

[6] Duflo, M. (1997). Random Iterative Models. *Applications of Mathematics* **34**, *Springer, Berlin*

[7] Escanciano, J. C. (2006). Goodness-of-fit Tests for Linear and Nonlinear Time Series Models. *Journal of the American Statistical Association* **101**, 531–541.

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

[9] Escanciano, J. C. and Jacho Chavéz, D. T. (2010). Approximating the Critical Values of Cramér-von Mises Tests in General Parametric Conditional Specifications. *Computational Statistics & Data Analysis* **54**, 625-636.

[10] Fan, Y. and Li, Q. (1997). A consistent nonparametric test for linearity of AR(p) models. *Economics Letters* **55**, 53-59.

[11] 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.

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

[13] Ghosh, B. K. and Huang, W.-M. (1991). The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit. *Ann. Statist.* **19**, 999–1009.

[14] Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference* **68**, 295–321.

[15] Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27**, 204–236.

[16] Kreiss, J.-P., Neumann, M. H. and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and its Interface* **1**, 367–380.

[17] Leucht, A. (2010a). Characteristic function-based goodness-of-fit tests under weak dependence. *Reports of the Department of Mathematics and Computer Science, Friedrich Schiller University Jena* **10-06**.

[18] Leucht, A. (2010b). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Reports of the Department of Mathematics and Computer Science, Friedrich Schiller University Jena* **10-05**.

[19] Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87**, 145–165.

[20] 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.

[21] Shao, X. and Wu, W. B. (2007), Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35**, 1773-1801.

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