JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **38**: 72–98 (2017)
Published online 9 June 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12200

ORIGINAL ARTICLE

RESIDUAL EMPIRICAL PROCESSES AND WEIGHTED SUMS FOR TIME-VARYING PROCESSES WITH APPLICATIONS TO TESTING FOR HOMOSCEDASTICITY

GABE CHANDLER[a*] AND WOLFGANG POLONIK[b]

[a] *Department of Mathematics, Pomona College, Claremont, CA, USA*
[b] *Department of Statistics, University of California, Davis, CA, USA*

In the context of heteroscedastic time-varying autoregressive (AR)-process we study the estimation of the error/innovation distributions. Our study reveals that the non-parametric estimation of the AR parameter functions has a negligible asymptotic effect on the estimation of the empirical distribution of the residuals even though the AR parameter functions are estimated non-parametrically. The derivation of these results involves the study of both function-indexed sequential residual empirical processes and weighted sum processes. Exponential inequalities and weak convergence results are derived.
As an application of our results we discuss testing for the constancy of the variance function, which in special cases corresponds to testing for stationarity.

Received 12 August 2014; Accepted 20 April 2016

Keywords: Cumulants; empirical process theory; exponential inequality; locally stationary processes; non-stationary processes; tests of stationarity.

1. INTRODUCTION

Consider the following time-varying autoregressive (AR) process that satisfies the system of difference equations

$$Y_t - \sum_{k=1}^{p} \theta_k \left(\frac{t}{n}\right) Y_{t-k} = \sigma \left(\frac{t}{n}\right) \epsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

where θ_k are the AR parameter functions, p is the order of the model, σ is a function controlling the volatility and $\epsilon_t \sim (0, 1)$ denote the i.i.d. errors. Following Dahlhaus (1997), time is rescaled to the interval $[0, 1]$ in order to make a large sample analysis feasible. Observe that this, in particular, means that $Y_t = Y_{t,n}$ satisfying (1) in fact forms a triangular array.

The consideration of non-stationary time series models goes back to Priestley (1965) who considered evolutionary spectra, that is, spectra of time series evolving in time. The time-varying AR process has always been an important special case, either in more methodological and theoretical considerations of non-stationary processes, or in applications such as signal processing and (financial) econometrics (e.g. Subba Rao, 1970; Grenier, 1983; Hall *et al.*, 1983; Rajan and Rayner, 1996; Girault *et al.*, 1998; Eom, 1999; Drees and Stărică, 2002; Orbe *et al.*, 2005; Fryzlewicz *et al.*, 2006; Chandler and Polonik, 2006).

One of our contributions is the estimation of the (average) distribution function of the innovations $\eta_t = \sigma\left(\frac{t}{n}\right)\epsilon_t$. Suppose that we have observed $Y_{1-p}, Y_{2-p}, \ldots, Y_n$, and let $\mathbf{Y}_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})'$, $t = 1, \ldots, n$. Given an estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ and corresponding residuals $\widehat{\eta}_t = Y_t - \widehat{\boldsymbol{\theta}}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}$, we consider the sequential empirical distribution function of the residuals given by

* Correspondence to: Gabe Chandler, Department of Mathematics, Pomona College, Claremont, CA 91711, USA. E-mail: gabriel.chandler@pomona.edu

$$\widehat{F}_n(\alpha, z) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\widehat{\eta}_t \le z\}, \ z \in \mathbb{R}, \ \alpha \in [0, 1].$$

The corresponding distribution function of the true innovations is $F_n(\alpha, z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\eta_t \le z\}$. We show that under appropriate conditions (Theorem 2),

$$\sup_{\alpha \in [0,1], z \in \mathbb{R}} \left| \widehat{F}_n(\alpha, z) - F_n(\alpha, z) \right| = o_P\left(n^{-1/2}\right), \tag{2}$$

even though non-parametric estimation of the parameter functions θ_k is involved. This means that by using appropriate estimators for the parameter functions, the (average) distribution function of the η_t can be estimated just as well as if the parameters were known. In parametric situations, this phenomenon is not new, and it perhaps was first observed in Boldin (1982). We refer to Koul (2002), Ch. 7, for more complex situations. Non-parametric models usually do not allow for such a phenomenon. See, for instance, Akritas and van Keilegom (2001), Koul (2002), Schick and Wefelmeyer (2002), Cheng (2005), Müller *et al.* (2007, 2009a, 2009b, 2012) and van Keilegom *et al.* (2008). In contrast to this work, our model, even though non-parametric in nature, has the crucial structure of being linear in the lagged observations, which also means that the Y_t have mean zero. For more on the role of a zero mean in this context, see Wefelmeyer (1994) and Schick and Wefelmeyer (2002). (Generalized) Autoregressive Conditional Heteroscedastic-type processes are considered in Horváth *et al.* (2001), Stute (2001), Koul (2002), Koul and Ling (2006) and Laïb *et al.* (2008) (cf. Remark (a) given next to Theorem 4).

The proof of the results just mentioned involves the behaviour of two types of stochastic processes, both of which are investigated in this article. One is the residual sequential empirical process, and the other type is a generalized partial sum process or weighted sums process. Both are of independent interest. The *residual sequential empirical process* is defined as

$$\nu_n(\alpha, z, \mathbf{g}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor \alpha n \rfloor} \left[\mathbf{1}\left\{\sigma\left(\frac{t}{n}\right)\epsilon_t \le \mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + z\right\} - F\left(\left(\mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + z\right) / \sigma\left(\frac{t}{n}\right)\right)\right], \tag{3}$$

where $\alpha \in [0, 1]$, $z \in \mathbb{R}$, $\mathbf{g} : [0, 1] \to \mathbb{R}^p$, $\mathbf{g} \in \mathcal{G}^p = \{\mathbf{g} = (g_1, \ldots, g_p)', \, g_i \in \mathcal{G}\}$ with $\mathcal{G}$ an appropriate function class such that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in \mathcal{G}^p$ with probability tending to 1 (see in the succeeding texts) and $F(z)$ denotes the distribution function of the errors ϵ_t. Observe that $\nu_n(\alpha, z, \mathbf{0}) = \sqrt{n}(F_n(\alpha, z) - \mathrm{E}F_n(\alpha, z))$, where, here, $\mathbf{0} = (0, \ldots, 0)'$ denotes the p-vector of null functions. The basic form of ν_n is standard, and residual empirical processes based on non-parametric models have been considered in the literature as indicated earlier. Our contribution here is to study these processes based on a non-stationary Y_t of the form (1).

The second key type of processes in this article is weighted sums of the form

$$Z_n(h) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h\left(\frac{t}{n}\right) Y_t, \quad h \in \mathcal{H},$$

where $\mathcal{H}$ is an appropriate function class. Such processes can be considered as generalizations of partial sum processes. Exponential inequalities and weak convergence results for such processes are derived later. We use properties of $Z_n(h)$ for proving (2), and this motivates its study here.

We also study the empirical distribution of the estimated errors $\widehat{F}_n^*(\alpha, z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}(\widehat{\epsilon}_t \le z)$, where $\widehat{\epsilon}_t = \frac{\widehat{\eta}_t}{\widehat{\sigma}\left(\frac{t}{n}\right)}$ for an appropriate estimator $\widehat{\sigma}\left(\frac{t}{n}\right)$ of the variance function $\sigma\left(\frac{t}{n}\right)$. We will see that under appropriate conditions,

DOI: 10.1111/jtsa.12200

$$\sup_{\alpha\in[0,1],z\in\mathbb{R}} \Big| \widehat{F}_n^*(\alpha,z) - F_n^*(\alpha,z) - z f(z) \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} \Big| = o_P\left(1/\sqrt{n}\right), \tag{4}$$

where $F_n^*(\alpha,z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}(\epsilon_t \le z)$ denotes the empirical distribution of the true errors and f is their density. The derivation of this result involves the study of a residual sequential empirical process slightly more general than (3) in order to accommodate for the estimation of the variance. Notice that, from (4), we can see that in contrast to the estimation of the AR parameter functions, the estimation of the variance function is not negligible, even if the variance was to be assumed constant. See Section 3 for more on this.

Dahlhaus (1997) advanced the formal analysis of time-varying processes by introducing the notion of a locally stationary process. This is a time-varying process with time being rescaled to [0, 1] that satisfies certain regularity assumptions; see (14–16) presented later. We would like to point out, however, that, in our article, local stationarity is only used to calculate the asymptotic covariance function in Theorem 5. All the other results hold under weaker assumptions.

The outline of the article is as follows. In Sections 2, 3 and 4, we analyze the large sample behaviour of the function-indexed residual empirical processes and of function-indexed weighted sums, respectively, under the time-varying model (1), and apply the obtained results to show (2) and (4). As an application of the theoretical results, we discuss in Section 5 a method for testing for homoscedasticity, which in special cases is equivalent to testing for stationarity. Proofs are deferred to Section 6.

Remark on measurability. Suprema of function-indexed processes will enter the theoretical results given in the succeeding texts. We assume throughout the article that such suprema are measurable.

2. RESIDUAL EMPIRICAL PROCESSES UNDER TIME-VARYING AUTOREGRESSIVE MODELS

In order to formulate one of our main results for the residual empirical process $\nu_n(\alpha, z, \mathbf{g})$ defined in (3) earlier, we first introduce some more notations and formulate the underlying assumptions.

Let $\mathcal{H}$ denote a class of functions defined on [0, 1] and let d denote a metric on $\mathcal{H}$. For a given $\delta > 0$, let $N(\delta, \mathcal{H}, d)$ denote the minimal number N of d-balls of radius $\le \delta$ that are needed to cover $\mathcal{H}$. Then $\log N(\delta, \mathcal{H}, d)$ is called the *metric entropy* of $\mathcal{H}$ with respect to d. If the balls $\mathcal{A}_k$ are replaced by brackets $\mathcal{B}_k = \{h \in \mathcal{H} : \underline{g}_k \le h \le \overline{g}_k\}$ for pairs of functions $\underline{g}_k \le \overline{g}_k$, $k = 1, \dots, N$ with $d(\overline{g}_k, \underline{g}_k) \le \delta$, then the minimal number $N = N_B(\delta, \mathcal{H}, d)$ of such brackets with $\mathcal{H} \subset \bigcup_{k=1}^N \mathcal{B}_k$ is called a *bracketing covering number,* and $\log N_B(\delta, \mathcal{H}, d)$ is called the *metric entropy with bracketing* of $\mathcal{H}$ with respect to d. For a function $h : [0,1] \to \mathbb{R}$, we denote $\|h\|_\infty := \sup_{u\in[0,1]} |h(u)|$ and $\|h\|_n^2 := \frac{1}{n}\sum_{t=1}^n h^2\left(\frac{t}{n}\right)$. We further denote by d_n the metric generated by $\|\cdot\|_n$, that is, $d_n(h,g) = \|h - g\|_n$.

Assumptions.

(i) The process $Y_t = Y_{t,n}$ has a Moving Average (MA)-type representation

$$Y_{t,n} = \sum_{j=0}^{\infty} a_{t,n}(j)\epsilon_{t-j}, \tag{5}$$

where $\epsilon_t \sim_{\text{i.i.d.}} (0,1)$. The distribution function F of ϵ_t has a strictly positive Lipschitz continuous Lebesgue density f. The function $\sigma(u)$ in (1) is of bounded variation with $0 < m_* < \sigma(u) < m^* < \infty$ for all u.

(ii) The coefficients $a_{t,n}(\cdot)$ in (5) satisfy

$$\sup_{1\le t\le n} |a_{t,n}(j)| \le \frac{K}{\ell(j)}, \quad j = 0, 1, 2, \dots$$

where, for $j > 1$, $\ell(j) = j\,(\log j)^{1+\kappa}$ for some $K, \kappa > 0$. For $j = 0, 1$, we let $\ell(j) = 1$.

DOI: 10.1111/jtsa.12200

(iii) We have

$$\sup_{\mathbf{g}\in\mathcal{G}^p}\max_{1\le t\le n}\left|\mathbf{g}\left(\tfrac{t}{n}\right)'\mathbf{Y}_{t-1}\right| = O_P(1) \quad \text{as } n\to\infty. \tag{6}$$

(iv) $\sup_{x\in\mathbb{R}}|x|^{1+\beta}f(x)<\infty$ for some $\beta>0$.

Assumptions (i) and (ii) have been used in the literature on locally stationary processes before. It is shown in Dahlhaus and Polonik (2005) (see also Dahlhaus and Polonik, 2009) by using a proof similar to Künsch (1995) that Assumption (i) holds for time-varying AR processes (1) if the zeros of the corresponding AR polynomials are bounded away from the unit disk (uniformly in the rescaled time u) and the parameter functions are of bounded variation. In case $p=1$, the functions $a_{t,n}(j)$ are of the form $a_{t,n}(j)=\sigma\left(\frac{t-j}{n}\right)\cdot\prod_{\ell=0}^{j-1}\theta\left(\frac{t-\ell}{n}\right)$, where $\theta(u)=\theta(0)$ for $u<0$ and, similarly, $\sigma(u)=\sigma(0)$ for $u<0$. Assumption (iii) is also not too surprising as it is similar to assumptions used in the analysis of parametric residual empirical processes, [e.g. Koul, 2002, Thm 2.2.3, Ass (2.2.28)]. Assumption (iv) is needed to control the tails of f in the case when the variable z in the process $\nu_n(\alpha,z,\mathbf{g})$ is allowed to range over the entire real line.

Recalling the definition of $\nu_n(\alpha,z,\mathbf{g})$ given in (3), we see that the special case $\mathbf{g}=\underset{\sim}{0}$ (the vector of zero functions) gives a process only based on the innovations, which are independent but not necessarily identically distributed. Our first theorem says that if the index class $\mathcal{G}$ is not too large, then $\nu_n(\alpha,z,\mathbf{g})$ and $\nu(\alpha,z,\mathbf{0})$ behave the same asymptotically:

Theorem 1. Suppose that Assumptions (i) and (ii) hold. Let $\mathcal{G}$ denote a function class such that, for some $c>0$,

$$\int_{c/n}^{1}\sqrt{\log N_B\left(u^2,\mathcal{G},d_n\right)}\,du<\infty \qquad \forall\, n. \tag{7}$$

Then we have for any $0<L<\infty$ and $\delta_n\to 0$ as $n\to\infty$ that

$$\sup_{\substack{\alpha\in[0,1],z\in(-L,L),\\ \mathbf{g}\in\mathcal{G}^p:\sum_{k=1}^p\|g_k\|_n\le\delta_n}}|\nu_n(\alpha,z,\mathbf{g})-\nu_n(\alpha,z,\mathbf{0})|=o_P(1). \tag{8}$$

If, in addition, Assumptions (iii) and (iv) hold, then (8) also holds with $L=\infty$.

The proof of Theorem 1 rests on the crucial technical Lemma 1, which is given in Section 6. This lemma implies asymptotic stochastic equicontinuity of the residual empirical sequential process $\nu_n(\alpha,z,\mathbf{g})$ of which the assertion of Theorem 1 is an immediate consequence.

Statements of type (8) are of typical nature for work on residual empirical processes [e.g. (8.2.32) in Koul, 2002]. Observe, however, that, here, we are dealing with time-varying AR processes and are considering non-parametric index classes. Also keep in mind that we are considering triangular arrays (recall that $Y_t=Y_{t,n}$).

The function class $\mathcal{G}$ is generic in the formulation of Theorem 1. As indicated in Section 1, what we have in mind is function classes $\mathcal{G}$ modelling the differences $\widehat{\theta}_k-\theta_k$ [cf. Assumption (vii)]. A specific example of such a class is given in the discussion next to the formulation of Theorem 2.

Notice that Theorem 1 is considering the difference of two processes where the 'right' centring is used. In contrast to that, we now consider the difference between two sequential empirical distribution functions: one of them based on the residuals and the other on the innovations. As already said earlier, under appropriate assumptions, the difference of these two functions is $o_P\left(1/\sqrt{n}\right)$, so that the non-parametric estimation of the parameter functions has an negligible asymptotic effect on the estimation of the (average) innovation distribution.

DOI: 10.1111/jtsa.12200

Further assumptions.

(v) $|\text{cum}_k(\epsilon_t)| \leq k!C^k$ for $k = 1, 2, \ldots$ for some $C > 0$.
(vi) For $k = 1, \ldots, p$, we have $\|\widehat{\theta}_k - \theta_k\|_n^2 = O_P(m_n^{-2})$ with $m_n \to \infty$ as $n \to \infty$.
(vii) There exists a class $\mathcal{G}$ with $\widehat{\theta}_k(\cdot) - \theta_k(\cdot) \in \mathcal{G}, \quad k = 1, \ldots, p$, with probability tending to 1 as $n \to \infty$ such that $\sup_{g \in \mathcal{G}} \|g\|_\infty < \infty$, and for some $C, c > 0$, we have $\int_{c/n}^1 \log N_B(u, \mathcal{G}, d_n)\, du < C < \infty \; \forall\, n$.

Theorem 2. Assume Assumptions (i) and (ii) and that $\mathcal{G}$ is such that (7) holds. In addition, assume that Assumptions (v)–(vii) hold, with m_n satisfying $\frac{\sqrt{n}\,\beta_n^2}{m_n^2} = o(1)$ as $n \to \infty$, where β_n is such that $\max_{1 \leq t \leq n} Y_t^2 = O_P(\beta_n^2)$. Then we have for $0 < L < \infty$ that

$$\sup_{\alpha \in [0,1], z \in (-L, L)} \left| \widehat{F}_n(\alpha, z) - F_n(\alpha, z) \right| = o_P(1/\sqrt{n}). \tag{9}$$

If further Assumptions (iii) and (iv) hold, then (9) also holds with $L = \infty$.

Discussion of assumptions: Assumption (v) holds, for instance, when $\mathrm{E}|\epsilon_t|^k \leq \left(\frac{C}{2}\right)^k$ for all $k = 1, 2, \ldots$. This is obviously a strong assumption. However, much of the literature on locally stationary processes is using it, in particular those in which rates of convergence for the estimators $\widehat{\theta}$ and/or $\widehat{\sigma}$ are derived. In our work, this assumptions enter the picture through Theorem 5, which is used in the proof of Theorem 2. It is worth pointing out in this context that the assumption on m_n is tied to the large-sample behaviour of the maximum of the Y_t^2, $t = 1, \ldots, T$, and the strong assumption (v) entails 'weak' conditions on the estimators $\widehat{\boldsymbol{\theta}}$. In fact, under Assumption (v), we can choose $\beta_n^2 = \log n$, which follows as in the proof of Lem 5.9 of Dahlhaus and Polonik (2009). Weakening condition (v) would require a stronger condition on β_n, the rate of convergence of the estimator $\widehat{\boldsymbol{\theta}}_n$.

The assumptions on the entropy integral [see Assumptions (vii) and (7)] control the complexity of the class $\mathcal{G}$. Many classes $\mathcal{G}$ are known to satisfy these assumptions – see below for an example. For more examples, we refer to the literature on empirical process theory. A more standard condition on the covering numbers is $\int_{c/n}^1 \sqrt{\log N(u, \mathcal{G}, d_n)}\, du < \infty$ (or similarly with bracketing). Compared with that, the entropy integral in Assumption (iv) does not have a square root, and the latter is similar to condition (7) where the integrand is $\sqrt{\log N_B(u^2, \mathcal{G}, d_n)}$ (notice the u^2). This makes both our entropy conditions stronger than the standard assumption. The reason for this is that the exponential inequality underlying our derivations is not of sub-Gaussian type (Lemma 3), which in turn is caused by the dependence structure of our underlying time-varying process.

A class of non-parametric estimators satisfying conditions (vi) and (vii) is given by the wavelet estimators of Dahlhaus *et al.* (1999). These estimators lie in the Besov smoothness class $B_{p,q}^s(C)$ where the smoothness parameters satisfy the condition $s + \frac{1}{2} - \frac{1}{\max(2,p)} > 1$. The constant $C > 0$ is a uniform bound on the (Besov) norm of the functions in the class. Dahlhaus *et al.* derive conditions under which their estimators converge at rate $\left(\frac{\log n}{n}\right)^{s/(2s+1)}$ in the L_2-norm. For $s \geq 1$, the functions in $B_{p,q}^s(C)$ have uniformly bounded total variation. Assuming that the model parameter functions also posses this property, the rate of convergence in the d_n-distance is the same as the one in L_2, because in this case, the error in approximating the integral by the average over equidistant points is of order $O(n^{-1})$. Consequently, in this case, we have $m_n^{-1} = \left(\frac{\log n}{n}\right)^{s/(2s+1)}$. In order to verify the condition on the bracketing covering numbers from Assumption (iii), we use Nickl and Pötscher (2007). Their Cor. 1, applied with $s = 2$, $p = q = 2$, implies that the bracketing entropy with respect to the L_2-norm can be bounded by $C\,\delta^{-1/2}$. (When applying their Corollary to our situation, choose, in their notation, $\beta = 0$, $\mu = U[0,1]$, $r = 2$ and $\gamma = 2$, say).

3. ESTIMATING THE DISTRIBUTION FUNCTION OF THE ERRORS ϵ_T

Estimating the error distribution can be treated by using similar techniques as used for the innovation distribution, although the estimation of the variance is not negligible – see Theorem 4 and Remark (a) right next to the theorem. For other work involving the estimation of the error distribution and the volatility, see, for instance, Akritas and van Keilegom (2001) and Neumeyer and van Keilegom (2010).

DOI: 10.1111/jtsa.12200

The sequential empirical distribution function of the $\widehat{\epsilon}_t$ can be rewritten as

$$\widehat{F}_n^*(\alpha, z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\widehat{\epsilon}_t \le z\} = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\left\{\epsilon_t \le \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} + z\left(\frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}\right) + z\right\}.$$

Accordingly, we define the modified residual sequential empirical process as

$$\begin{aligned} \nu_n^*(\alpha, z, \mathbf{g}, s) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor \alpha n \rfloor}\Bigg[\mathbf{1}\left\{\sigma\left(\frac{t}{n}\right)\epsilon_t \le \mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + z\left[s\left(\frac{t}{n}\right) + \sigma\left(\frac{t}{n}\right)\right]\right\} \\ -F\left(\left(\mathbf{g}\left(\tfrac{t}{n}\right)' \mathbf{Y}_{t-1} + z\left[s\left(\tfrac{t}{n}\right) + \sigma\left(\tfrac{t}{n}\right)\right]\right)/\sigma\left(\tfrac{t}{n}\right)\right)\Bigg], \end{aligned} \tag{10}$$

where we think of $s \in \mathcal{S}$ with the class $\mathcal{S}$ being a model for the difference $\widehat{\sigma} - \sigma \in \mathcal{S}$. The following is the analogue to Theorem 1.

Theorem 3. Assume Assumptions (i) and (ii) and that $\mathcal{G}$ and $\mathcal{S}$ are classes of sets satisfying the entropy condition (7). Then we have for any $0 < L < \infty$ and $\delta_n \to 0$ as $n \to \infty$ that

$$\sup_{\alpha\in[0,1], z\in(-L,L)} \; \sup_{\{(\mathbf{g},s)\in\mathcal{G}^p\times\mathcal{S}:\, \sum_{k=1}^p \|g_k\|_n + \|s\|_n \le \delta_n,\}} |\nu_n^*(\alpha, z, \mathbf{g}, s) - \nu_n^*(\alpha, z, \mathbf{0}, 0)| = o_P(1). \tag{11}$$

If further Assumptions (iii) and (iv) hold, then assertion (11) also holds with $L = \infty$ and $\|s\|_n$ replaced by $\|s\|_\infty$.

Remarks:

(a) The fact that, for $L = \infty$, we have to replace $\|s\|_n$ by $\|s\|_\infty$ is again reminiscent of assumptions used in the analysis of parametric error processes [e.g. Koul and Ling, 2006, Ass (4.4)].

(b) The residual sequential empirical process from (3) corresponds to not estimating the variance. Formally, this corresponds to using the estimate of the variance being constant equal to 1 (i.e. not dividing the residuals by an estimate of the variance). Since $\mathcal{S}$ models the difference $\widehat{\sigma} - \sigma$, this results in the class $\mathcal{S} = \{1 - \sigma\}$ consisting of only one function. With this choice, the process (10) reduces to the process (3). In fact, we have $\nu_n(\alpha, z, \mathbf{g}) = \nu_n^*(\alpha, z, \mathbf{g}, 1 - \sigma)$.

Further assumptions.

(viii) We have $\|\widehat{\sigma} - \sigma\|_n^2 = o_P(n^{-1/2})$ as $n \to \infty$.

(ix) There exists a function class $\mathcal{S}$ with $\widehat{\sigma} - \sigma \in \mathcal{S}$ with probability tending to 1 as $n \to \infty$, $\sup_{s\in\mathcal{S}} \|s\|_\infty < \infty$, and for some $C, c > 0$, we have $\int_{c/n}^1 \log N_B(u, \mathcal{S}, d_n)\, du < C < \infty \quad \forall\, n$.

Theorem 4. Assume Assumptions (i), (ii), (v)–(ix) and that $\mathcal{G}$ and $\mathcal{S}$ are classes of sets satisfying the entropy condition (7). In addition, assume that f is differentiable with bounded derivative and that the sequence $\{m_n\}$ in Assumption (vi) is satisfying $\frac{\sqrt{n}\,\beta_n^2}{m_n^2} = o(1)$ as $n \to \infty$, where β_n is such that $\max_{1\le t\le n} Y_t^2 = O_P(\beta_n^2)$. Then we have for $0 < L < \infty$ that

$$\sup_{\alpha\in[0,1], z\in(-L,L)} \left| \widehat{F}_n^*(\alpha, z) - F_n^*(\alpha, z) - z f(z)\, \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} \right| = o_P(1/\sqrt{n}). \tag{12}$$

If further Assumptions (iii) and (iv) hold and if $\frac{1}{\|\widehat{\sigma}\|_\infty} = O_P(1)$, then (12) also holds with $L = \infty$.

DOI: 10.1111/jtsa.12200

Remarks.

(a) The theorem shows that, in contrast to the estimation of the AR parameters, the estimation of the variance function is not negligible, and this even holds in the i.i.d. case when the variance is constant. That the estimation of the AR parameters $\theta_k\left(\frac{t}{n}\right)$ is negligible is due to the technical fact that, in the Taylor expansion of $E\left(\widehat{F}_n^* - F_n^*\right)(\alpha, z)$, the estimates of the AR parameters appear in sums of the form $\frac{1}{n}\sum_{t=1}^{n}(\widehat{\theta}_k - \theta)\left(\frac{t}{n}\right) Y_{t-k}$. Now, if $(\widehat{\theta}_k - \theta)\left(\frac{t}{n}\right)$ lies in a class of functions $\mathcal{G}$, then this sum can be absolutely bounded by $\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{t=1}^{n} g\left(\frac{t}{n}\right) Y_{t-k}\right|$, and because the Y_{t-k} have mean zero and the functions $g\left(\frac{t}{n}\right)$ have small norm, the latter sum is $o_P(n^{-1/2})$ under appropriate assumptions. In contrast to that, the sum of the differences $(\widehat{\sigma} - \sigma)\left(\frac{t}{n}\right)$ does not involve the zero mean Y_t's, so that the same 'trick' does not apply.

(b) The term $z f(z)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}\frac{\widehat{\sigma}\left(\frac{t}{n}\right)-\sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}$ in (12) is the first-order stochastic approximation to the difference of the appropriate centrings $F_{\frac{t}{n}}\left((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + z\,\widehat{\sigma}\left(\frac{t}{n}\right)\right) - F(z)$ [see (49) and (51) shown later]. The factor $z f(z)$ also appears in other work that involves the estimation of variance functions such as Horváth *et al.* (2001) and Koul and Ling (2006).

(c) The wavelet-based estimator of the variance function of Dahlhaus and Neumann (2001) satisfies the necessary assumptions on $\widehat{\sigma}$ for $0 < L < \infty$.

(d) In order to be able to utilize the earlier results to develop statistical methodology, finer knowledge about the asymptotic behaviour of the term $\frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}\frac{\widehat{\sigma}\left(\frac{t}{n}\right)-\sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}$ is needed. It can be expected that, under appropriate assumptions, this term becomes asymptotically normal for reasonable estimators, but to the best of our knowledge, such a result is not available in the literature for estimators of the variance function under local stationarity. While it might be possible to derive such a result for certain estimators under appropriate assumptions, exploring this question goes beyond the scope of this work. It is also unclear, whether the asymptotic normality, if it can be derived, could be easily used for statistical purposes, for what really is needed is the knowledge about the asymptotic distribution of the sum $F_n^*(\alpha, z) + z f(z)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}\frac{\widehat{\sigma}\left(\frac{t}{n}\right)-\sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}$. For this, a stochastic expansion of the sum $\frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}\frac{\widehat{\sigma}\left(\frac{t}{n}\right)-\sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}$ would be ideal. Such an expansion will of course heavily depend on the specific estimator considered. Moreover, one then has to be able to estimate the resulting asymptotic variance, which might not be straightforward either, because the asymptotic variance might have a complex form. We would like to point out, however, that one of the basic ideas underlying our testing approach discussed in Section 5 is to avoid estimating the variance function altogether and extract information of the shape of the variance functions by other means.

4. WEIGHTED SUMS UNDER LOCAL STATIONARITY

The second type of processes of importance in our context is weighted partial sums of locally stationary processes given by

$$Z_n(h) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} h\left(\frac{t}{n}\right) Y_t, \quad h \in \mathcal{H}. \tag{13}$$

In the i.i.d. case, weighted sums have received some attention in the literature. For functional central limit theorems and exponential inequalities, see, for instance, Alexander and Pyke (1986) and van de Geer (2000) and the references therein.

We will show in the succeeding discussion that, under appropriate assumptions, $Z_n(h)$ converges weakly to a Gaussian process. In order to calculate the covariance function of the limit, we assume that the process Y_t is locally stationary as in Dahlhaus and Polonik (2009). Recalling Assumption (i), we assume the existence of functions $a(\cdot, j) : (0, 1] \to \mathbb{R}$ with

DOI: 10.1111/jtsa.12200

$$\sup_u |a(u,j)| \leq \frac{K}{\ell(j)}, \tag{14}$$

$$\sup_j \sum_{t=1}^{n} \left| a_{t,n}(j) - a\left(\frac{t}{n}, j\right) \right| \leq K, \tag{15}$$

$$TV(a(\cdot, j)) \leq \frac{K}{\ell(j)}, \tag{16}$$

where for a function $g : [0,1] \to \mathbb{R}$, we denote by $TV(g)$ the total variation of g on [0, 1]. Conditions (14)–(16) hold if the zeros of the corresponding AR polynomials are bounded away from the unit disk (uniformly in the rescaled time u) and the parameter functions are of bounded variation (see Dahlhaus and Polonik, 2006). Further, we define the time-varying spectral density as the function

$$f(u,\lambda) := \frac{1}{2\pi} |A(u,\lambda)|^2$$

with

$$A(u,\lambda) := \sum_{j=-\infty}^{\infty} a(u,j) \exp(-i\lambda j),$$

and

$$c(u,k) := \int_{-\pi}^{\pi} f(u,\lambda) \exp(i\lambda k) d\lambda = \sum_{j=-\infty}^{\infty} a(u,k+j)\, a(u,j)$$

is the time-varying covariance of lag k at rescaled time $u \in [0,1]$. We also denote by $\mathrm{cum}_k(X)$, the kth-order cumulant of a random variable X.

Theorem 5. Let $\mathcal{H}$ denote a class of uniformly bounded, real-valued functions of bounded variation defined on [0, 1]. Assume further that for some $C, c > 0$,

$$\int_{c/n}^{1} \log N(u, \mathcal{H}, d_n)\, du < C < \infty \quad \forall\, n. \tag{17}$$

Then we have under Assumptions (i), (ii) and (v) that, as $n \to \infty$, the process $Z_n(h),\ h \in \mathcal{H}$, converges weakly to a tight, mean zero Gaussian process $\{G(h), h \in \mathcal{H}\}$. If, in addition, (14)–(16) hold, then the variance–covariance function of $G(h)$ can be calculated as $C(h_1, h_2) = \int_0^1 h_1(u)\, h_2(u)\, S(u)\, du$, where $S(u) = \sum_{k=-\infty}^{\infty} c(u,k)$.

Remarks.

(a) Here, weak convergence is meant in the sense of Hoffman–Jørgensen – see van der Vaart and Wellner (1996) for more details.
(b) Weighted partial sums of the form

$$Z_n(\alpha, h) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor \alpha n \rfloor} g\left(\frac{t}{n}\right) Y_t$$

DOI: 10.1111/jtsa.12200

are in fact a special case of processes considered in the theorem. Here, $h(u) = h_{g,\alpha}(u) = \mathbf{1}_{[0,\alpha]}(u)\, g(u)$. Note that if $g \in \mathcal{G}$ and $\mathcal{G}$ satisfies the assumptions on the covering integral from the aforementioned theorem, then so does the class $\{h_{g,\alpha}(u) : g \in \mathcal{G}, \alpha \in [0,1]\}$. In this case, the limit covariance can then be written as $C(h_{g_1,\alpha_1}, h_{g_2,\alpha_2}) = \int_0^{\alpha_1 \wedge \alpha_2} g_1(u)\, g_2(u)\, S(u)\, du$.

(c) Assumptions (14)–(16) are only used for calculating the covariance function of the limit process.

The main ingredients to the proof of Theorem 5 are presented in the following results. These results are of independent interest.

Theorem 6. Let $\{Y_t, t = 1, \ldots, n\}$ satisfy Assumptions (i), (ii) and (v), and let $\mathcal{H} = \{h : [0,1] \to \mathbb{R}\}$ be totally bounded with respect to d_n. Further, let $\mathbf{A}_n = \left\{ \frac{1}{n} \sum_{t=1}^n Y_t^2 \le M^2 \right\}$, where $M > 0$. There exist constants $c_0, c_1, c_2 > 0$ such that, for all $\eta > 0$ satisfying

$$\eta < 16\, M\, \sqrt{n}\, \tau \tag{18}$$

and

$$\eta > c_0 \left(\int_{\frac{\eta}{8 M \sqrt{n}}}^{\tau} \log N(u, \mathcal{H}, d_n)\, du \;\vee\; \tau \right), \tag{19}$$

we have

$$P\left[\sup_{h \in \mathcal{H},\, \|h\|_n \le \tau} |Z_n(h)| > \eta,\, \mathbf{A}_n \right] \le c_1 \, \exp\left\{ -\frac{\eta}{c_2\, \tau} \right\}.$$

5. APPLICATIONS

We consider an application that motivates the aforementioned study of the two types of processes. The application consists of testing for the constancy of the variance function, that is, testing for homoscedasticity. We note that, for our test statistics, this particular application does not require estimation of the variance function $\sigma^2(u)$; rather, the question is about the constancy of this function. This allows us to use the results of Theorem 2, avoiding the need to handle the ‘bias’ term arising in Theorem 4 [see Remark (i) following this theorem]. As we are primarily concerned with questions of homoscedasticity, as well as lacking methodology dealing with distributional tests regarding the error terms ϵ_t in this setting on which to compare, we do not explore questions regarding F^* here. We note that such applications would be more difficult because of this additional term. However, when exploring questions of homoscedasticity and related questions about stationarity, the current methodology is significantly easier to implement than the methods to which we compare; see succeeding discussion. Additionally, more general questions of homoscedasticity are not able to be handled by those methodologies. In the case of a time-varying AR model, the homoscedastic and heteroscedastic model both live in the alternative space for the tests of stationarity. A second testing problem, closely related although more involved than the first, is a test for determining the modality of the variance function, details of which can be found in Chandler and Polonik (2012).

5.1. A test of homoscedasticity

We are interested in testing the null hypothesis $H : \sigma(u) = \sigma_0$ for all $u \in [0,1]$. (Note that if we additionally assume that the AR parameters do not depend on time, then under mild conditions on the θ_k, this is also a test for

DOI: 10.1111/jtsa.12200

weak stationarity.) To this end, consider the process

$$\widehat{G}_{n,\gamma}(\alpha) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\left(\widehat{\eta}_t^2 \geq \widehat{q}_\gamma^2\right), \quad \alpha \in [0,1], \tag{20}$$

where $\widehat{q}_\gamma$ is the empirical quantile of the squared residuals. Notice that $\widehat{G}_{n,\gamma}(\alpha)$ counts the number of large (squared) residuals within the first $(100 \times \alpha)\%$ of the observations $Y_1, \ldots, Y_n$, where 'large' is determined by the choice of γ. If the variance function is constant, then, since one can expect to have a total of $\lfloor n\gamma \rfloor$ large residuals, the expected value of $\widehat{G}_{n,\gamma}(\alpha)$ approximately equals $\alpha\gamma$. This motivates the form of our test statistic,

$$T_n = \sup_{\alpha \in [0,1]} \sqrt{\frac{n}{\gamma(1-\gamma)}} \left| \widehat{G}_{n,\gamma}(\alpha) - \alpha\gamma \right| .$$

Following the discussion of our application, we argue that the large-sample behaviour of this test statistic under the null is that of the supremum of a Brownian bridge (Section 5.2).

While there seems to be an obvious robust component to this methodology, it is not obvious that it should be very powerful, and perhaps, too much information is being lost in considering the data in this way. As we are unaware of any competing tests for constancy of the variance function in an otherwise non-stationary setting, we attempt to allay fears about a lack of power by considering the following test of stationarity via our methodology.

We consider the univariate locally stationary model used in Puchstein and Preuß (2016),

$$X_{t,n} = \left(1 + \frac{t}{n}\right) Z_t, \tag{21}$$

where the $Z_t \sim N(0,1)$ i.i.d. Thus, we run our test on simply the squared observations.

As compared with our test, most tests for stationarity in the literature (e.g. von Sachs and Neumann, 2000; Paparoditis, 2009, 2010; Preuß *et al.*, 2013; Puchstein and Preuß, 2016) work in the spectral domain, comparing an estimate of the time-varying spectrum with the stationary estimate. We compare the power behaviour of our test based on T_n to three of these test. To this end, we use results presented in Puchstein and Preuß (2016), which compare the methods of Paparoditis (2010) and Preuß *et al.* (2013) with their proposed method. This latest method has the benefit of not requiring a window size to be chosen, while the other two methods were run at two different window sizes to optimize the procedures. For our proposed time domain method, we use a $\gamma = 0.8$, although this does not prove to be too critical. For example, with $n = 256$ and under Model (21), the power was 0.870, 0.885 and 0.861 for $\gamma = (0.7, 0.8, 0.9)$. The proposed model is competitive (Table I) and much simpler both conceptually and implementation-wise.

According to our theory, the estimation of the parameter functions should not have an effect on the test, asymptotically. To see how far this also holds for finite samples, we simulated from a tvAR(1) of the form

$$X_{t,n} = 0.8 \sin\left(2\pi \frac{t}{n}\right) X_{t-1,n} + Z_t$$

Table I. Simulation results for model (21)

n	T_n	Paparoditis (2010)	Preuß *et al.* (2013)	Puchstein and Preuß (2016)
64	0.311	0.054	0.322	0.341
128	0.549	0.150	0.686	0.698
256	0.878	0.234	0.958	0.958

DOI: 10.1111/jtsa.12200

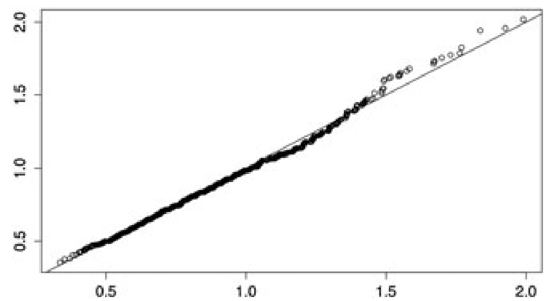

Figure 1. $T_n = \sqrt{\frac{n}{q(1-q)}}||\hat{G} - G||$ (horizontal) vs $\sqrt{n}||\hat{F} - F||$ (vertical)

and then estimated the parameter function $\theta(u) = 0.8\sin(2\pi u)$ using local least squares followed by a kernel smooth. Note that this particular form of the estimator does not satisfy the assumptions underlying our theory, but the discussion later demonstrates the robustness of the result to these assumptions. We generated 1000 data sets according to this model with $n = 256$, and the test statistic regarding the residuals was computed for each. We then generated a sample from a uniform distribution and computed the Kolmogorov–Smirnov statistic, properly normalized for $n = 1000$ as a sample from the limiting null distribution. A QQ plot of the two test statistics is provided in Figure 1. We then simulated from the model

$$X_{t,n} = 0.8\sin\left(2\pi\frac{t}{n}\right) X_{t-1,n} + \left(1 + \frac{t}{n}\right) Z_t,$$

which has the same volatility function as the modulated white noise process (21). With $n = 256$, 884 of 1000 time series from this model resulted in a rejection of the null of constant variance using an $\alpha = 0.05$. We note that this is very close to the estimated power (0.878) found for Model (21), as expected. Based on the results comparing this idea with multiple methods in the better studied question about stationarity, we are led to believe that these results are promising, despite lacking a second method on which to compare.

5.2. Asymptotic distribution of a test statistic T_n

Notice that $\widehat{G}_{n,\gamma}(\alpha)$ is closely related to the sequential residual empirical process, and as can be seen from the proof of Theorem 7, weighted empirical processes enter the analysis of $\widehat{G}_{n,\gamma}(\alpha)$ through handling the estimation of q_γ. Theorem 7 given in the succeeding texts provides an approximation of the test statistic T_n by independent (but not necessarily identically distributed) random variables. This result crucially enters the proofs in Chandler and Polonik (2012), where a similar test statistic is used to test for the modality of the variance function. In particular, it implies that the large-sample behaviour of the test statistic T_n under the null hypothesis is not influenced by the, in general, non-parametric estimation of the parameter functions, as long as the rate of convergence of these estimators is sufficiently fast.

First, we introduce some additional notation. Let f_u denote the pdf of $\sigma(u)\,\epsilon_t$, that is, $f_u(z) = \frac{1}{\sigma(u)}\,f\left(\frac{z}{\sigma(u)}\right)$, and

$$G_{n,\gamma}(\alpha) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n\rfloor} \mathbf{1}\left(\epsilon_t^2\sigma^2\left(\frac{t}{n}\right) \geq q_\gamma^2\right),$$

where q_γ is defined by means of

$$\Psi(z) = \int_0^1 F\left(\frac{z}{\sigma(u)}\right) du - \int_0^1 F\left(\frac{-z}{\sigma(u)}\right) du, \quad z \geq 0, \tag{22}$$

DOI: 10.1111/jtsa.12200

as the solution to the equation

$$\Psi(q_\gamma) = 1 - \gamma. \tag{23}$$

Notice that this solution is unique since we have assumed F to be strictly monotonic, and if $\sigma^2(u) = \sigma_0^2$ is constant for all $u \in [a, b]$, then q_γ^2 equals the upper γ-quantile of the squared innovations $\eta_t^2 = \sigma_0^2 \epsilon_t^2$. The approximation result that follows does not assume that the variance is constant, however.

Theorem 7. Let $\gamma \in [0, 1]$ and suppose that $0 \le a < b \le 1$ are non-random. Then, under Assumptions (i)–(v), with $\frac{n^{1/2} \log n}{m_n^2} = o(1)$, we have as $n \to \infty$ that

$$\sqrt{n} \sup_{\alpha \in [0,1]} \left| \widehat{G}_{n,\gamma}(\alpha) - G_{n,\gamma}(\alpha) + c(\alpha)\left(G_{n,\gamma}(1) - \mathrm{E}G_{n,\gamma}(1)\right) \right| = o_p(1), \tag{24}$$

where

$$c(\alpha) = \frac{\int_0^\alpha [\, f_u(q_\gamma) + f_u(-q_\gamma)\,]\, du}{\int_0^1 [f_u(q_\gamma) + f_u(-q_\gamma)]\, du}.$$

Under the null hypothesis $\sigma(u) \equiv \sigma_0 > 0$ for $u \in [0, 1]$, we have $c(\alpha) = \alpha$. Moreover, in case the AR parameter in Model (1) is constant and $\sqrt{n}$-consistent estimators are used, then the moment assumptions on the innovations can be significantly relaxed to $\mathrm{E}\,\epsilon_t^2 < \infty$.

Under the null hypothesis, $\sigma^2(u) = \sigma_0^2$ for all $u \in [0, 1]$, the innovations are i.i.d. Using that in this case $c(\alpha) = \alpha$ and $\mathrm{E}G_{n,\gamma}(\alpha) = \alpha\gamma$, we see that the aforementioned result implies that, in this case, $(\gamma(1-\gamma))^{-1/2}\sqrt{n}(\widehat{G}_{n,\gamma}(\alpha) - \alpha\gamma)$ converges weakly to a standard Brownian Bridge (cf. Chandler and Polonik, 2012).

6. PROOFS

Throughout the proofs, we use the notation $F_u(z) = P(\sigma(u)\epsilon_t \le z)$.

6.1. A crucial maximal inequality

The proofs of Theorems 1 and 3 rest on the following lemma, which is of independent interest. It is modelled after a similar result for empirical processes (van de Geer, 2000, Thm 5.11). Let $\mathcal{H}_L^* = [0, 1] \times (-L, L) \times \mathcal{G}^p \times \mathcal{S}$ denote the index space of the process ν_n^* defined in (10), where $0 < L \le \infty$. Define a metric on $\mathcal{H}_L^*$ as

$$d_{n,W}(h_1, h_2) = |\alpha_1 - \alpha_2| + |W(z_1) - W(z_2)| + \sum_{k=1}^{p} \|g_{1,k} - g_{2,k}\|_n + \|s_1 - s_2\|_n, \tag{25}$$

where $W(x) = \int_{-\infty}^x w(y)\mathrm{d}y$ with $w(y) > 0$ integrable such that, with $\beta > 0$ from Assumption (iv), we have $\lim_{|y|\to\infty} y^{1+\beta} w(y) = 1$. Observe that W is a strictly increasing, positive, bounded function on $\mathbb{R}$ and the tails of f are not heavier than the ones of w, assuming that Assumption (iv) holds.

Lemma 1. Assume Assumptions (i) and (ii). Further assume that both $\mathcal{G}$ and $\mathcal{S}$ are totally bounded with respect to d_n. Let $\mathbf{A}_n = \left\{\frac{1}{n}\sum_{s=-p+1}^n Y_s^2 \le C_0^2\right\} \cap \left\{\sup_{g\in\mathcal{G}} |\mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}| < C_0^*\right\}$. Define $K^* = \frac{\|f\|_\infty}{m_*}\left(\sup_{s\in\mathcal{S}} \|s\|_n + m^* + L + \sqrt{p}\, C_0\right)$. Suppose that $C_1, \eta, \tau > 0$ are such that

DOI: 10.1111/jtsa.12200

$$\eta \geq \frac{2^6 \, K^*}{\sqrt{n}}, \tag{26}$$

$$\eta \leq \frac{1}{2} \, K^* \sqrt{n} \, (\tau^2 \wedge \tau), \tag{27}$$

$$\eta \geq C_1 \left(\int_{\eta/2^8 \, K^* \sqrt{n}}^{\tau} \sqrt{\log N_B \left(u^2, \mathcal{H}_L^*, d_{n,W}\right)} \, du \ \vee \ \tau \right). \tag{28}$$

Then for $0 < L < \infty$ and $C_1 \geq 2^6 \sqrt{10\,K^*}$, we have with $C_2 = \left(\frac{2^6(2^6+1)K^*}{C_1^2} + 2\right)$ that

$$P\left[\sup_{h_1,h_2 \in \mathcal{H}_L^* :\, d_{n,W}(h_1,h_2) \leq \tau^2} |\, \nu_n^*(h_1) - \nu_n^*(h_2)| \geq \eta, \, \mathbf{A}_n \right] \leq C_2 \, \exp\left(-\frac{\eta^2}{2^6(2^6+1)\,K^*\,\tau^2} \right).$$

If, in addition, Assumption (iv) holds and, for some $c_* > 0$, we have $\inf_{s \in \mathcal{S}} \inf_{u \in [0,1]} [(s+\sigma)(u)] > c_*$, then the assertion also holds for $L = \infty$ with a modified K^* (see Proof).

Remarks:

(a) The assertion also holds (with a slightly modified K^*) on the simplified set $\mathbf{A}_n = \left\{ \frac{1}{n} \sum_{s=-p+1}^{n} Y_s^2 \leq C_0^2 \right\}$ if we, in addition, assume that $\sup_{g \in \mathcal{G}} \|g\|_\infty < \infty$.

(b) This lemma of course also applies to $\mathcal{H}_L = [0,1] \times (-L, L) \times \mathcal{G}^p$ replacing $\mathcal{H}_L^*$ [with the appropriately modified (simplified) $d_{n,W}$]. Since this case formally corresponds to putting $\mathcal{S} = \{1 - \sigma\}$, all the assumptions involving $\mathcal{S}$ are trivially satisfied. Moreover, in this case, the assumptions $\sup_{g \in \mathcal{G}} \|g\|_\infty < \infty$ and $\sup_{x \in \mathbb{R}} |x|^{1+\beta} f(x) < \infty$ for some $\beta > 0$ can both be dropped (see also Proof).

(c) The assumption $\inf_{s \in \mathcal{S}} \inf_{u \in [0,1]} [(s+\sigma)(u)] > c_*$ for some $c_* > 0$ is fulfilled if $\sup_{s \in \mathcal{S}} \|s\|_\infty$ is arbitrarily small (recall that we assumed σ to be bounded away from zero). In our application, we think of s modelling the difference $\widehat{\sigma} - \sigma$, so that (uniform) consistency of the estimator $\widehat{\sigma}$ will ensure that, with high probability, $\widehat{\sigma} - \sigma$ will lie in $\mathcal{S}$, satisfying the assumption.

Proof

We only present an outline of the proof. Let $h = (\alpha, z, \mathbf{g}, s) \in \mathcal{H}_L^*$. First notice that $\nu_n^*(h) = \nu_n^*(\alpha, z, \mathbf{g}, s)$ is a sum of bounded martingale differences. To see this, let $\xi_t^{z,\mathbf{g},s} = \mathbf{1}\left(\sigma\left(\frac{t}{n}\right)\epsilon_t \leq \mathbf{g}'\left(\frac{t}{n}\right)\mathbf{Y}_{t-1} + z\,s\left(\frac{t}{n}\right) + z\sigma\left(\frac{t}{n}\right)\right)$ and $\tilde{\xi}_t^{z,\mathbf{g},s} = \xi_t^{z,\mathbf{g},s} - \mathrm{E}\left(\xi_t^{z,\mathbf{g},s} | \mathcal{F}_{t-1}\right)$, where $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ denotes the σ-algebra generated by $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$, then

$$\nu_n^*(\alpha, z, \mathbf{g}, s) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\xi}_t^{z,\mathbf{g},s} \mathbf{1}\left(\frac{t}{n} \leq \alpha\right).$$

Obviously, also $\nu_n^*(\alpha_1, z_1, \mathbf{g}_1, s_1) - \nu_n^*(\alpha_2, z_2, \mathbf{g}_2, s_2)$ are sums of martingale differences. The proof of the lemma is based on the basic chaining device that is well known in empirical process theory, utilizing the following exponential inequality for sums of bounded martingale differences from Freedman (1975). □

Lemma (Freedman 1975)**.** Let $Z_1, \ldots, Z_n$ denote martingale differences with respect to a filtration $\{\mathcal{F}_t,\, t = 0, \ldots, n-1\}$ with $|Z_t| \leq C$ for all $t = 1, \ldots, n$. Let further $S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t$ and $V_n = V_n(S_n) =$

DOI: 10.1111/jtsa.12200

$\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left(Z_t^2 | \mathcal{F}_{t-1}\right)$. Then we have for all $\epsilon, \tau^2 > 0$ that

$$P\left(S_n \geq \epsilon,\ V_n \leq \tau^2\right) \leq \exp\left(-\frac{\epsilon^2}{2\tau^2 + \frac{2\epsilon C}{\sqrt{n}}}\right). \tag{29}$$

In order to be able to apply (29) to our problem, it is crucial to control the quadratic variation V_n. We now indicate how to do this. Let

$$\eta_t^{\alpha,z,\mathbf{g},s} = \tilde{\xi}_t^{z,\mathbf{g},s}\ \mathbf{1}\left(\frac{t}{n} \leq \alpha\right).$$

We have for $h_1 = (\alpha_1, z_1, \mathbf{g}_1, s_1), h_2 = (\alpha_2, z_2, \mathbf{g}_2, s_2) \in \mathcal{H}$ with $d_{n,W}(h_1, h_2) \leq \epsilon$ that

$$\begin{aligned}
V_n &= V_n(\nu_n(h_1) - \nu_n(h_2)) \\
&= \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[\left(\widetilde{\xi}_t^{\alpha_1,z_1,\mathbf{g}_1,s_1}\ \mathbf{1}\left(\frac{t}{n} \leq \alpha_1\right) - \widetilde{\xi}_t^{\alpha_2,z_2,\mathbf{g}_2,s_2}\right)^2 \mathbf{1}\left(\frac{t}{n} \leq \alpha_2\right) | \mathcal{F}_{t-1}\right] \\
&\leq \frac{1}{n}\sum_{t=1}^{n} \left| 1\left(\frac{t}{n} \leq \alpha_1\right) - 1\left(\frac{t}{n} \leq \alpha_2\right) \right| \mathrm{E}\left(\tilde{\xi}_t^{z_1,\mathbf{g}_1,s_1} | \mathcal{F}_{t-1}\right) + \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[\left|\tilde{\xi}_t^{z_1,\mathbf{g}_1,s_1} - \tilde{\xi}_t^{z_2,\mathbf{g}_2,s_2}\right| |\mathcal{F}_{t-1}\right] \\
&\leq |\alpha_1 - \alpha_2| + \frac{1}{n} + \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[\left|\xi_t^{z_1,\mathbf{g}_1,s_1} - \xi_t^{z_2,\mathbf{g}_2,s_2}\right| |\mathcal{F}_{t-1}\right].
\end{aligned}$$

For the last sum, we obtain by telescoping

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left(\left|\xi_t^{z_1,\mathbf{g}_1,s_1} - \xi_t^{z_2,\mathbf{g}_2,s_2}\right| \mathcal{F}_{t-1}\right)$$

$$\leq \frac{1}{n}\sum_{t=1}^{n} \left| F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_1+\sigma)\left(\tfrac{t}{n}\right) z_1\right) - F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_1+\sigma)\left(\tfrac{t}{n}\right) z_2\right)\right| \tag{30}$$

$$+ \frac{1}{n}\sum_{t=1}^{n} \left| F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_1+\sigma)\left(\tfrac{t}{n}\right) z_2\right) - F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_2+\sigma)\left(\tfrac{t}{n}\right) z_2\right)\right| \tag{31}$$

$$+ \frac{1}{n}\sum_{t=1}^{n} \left| F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_2+\sigma)\left(\tfrac{t}{n}\right) z_2\right) - F_{\frac{t}{n}}\left(\mathbf{g}_2\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_2+\sigma)\left(\tfrac{t}{n}\right) z_2\right)\right| \tag{32}$$

We estimate the three terms on the right separately. To control (30), let $V_{\frac{t}{n}}(y) = F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_1+\sigma)\left(\frac{t}{n}\right) W^{-1}(y)\right)$, where W is introduced in (25) earlier. Then

$$\begin{aligned}
(30) &= \frac{1}{n}\sum_{t=1}^{n} \left| V_{\frac{t}{n}}(W(z_1)) - V_{\frac{t}{n}}(W(z_2))\right| \\
&= \frac{1}{n}\sum_{t=1}^{n} \left| \frac{f_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + (s_1+\sigma)\left(\frac{t}{n}\right) W^{-1}(\eta)\right)}{w\left(W^{-1}(\eta)\right)}\right| \left|W(z_1) - W(z_2)\right| \left|(s_1+\sigma)\left(\tfrac{t}{n}\right)\right|
\end{aligned} \tag{33}$$

with η between $W(z_1)$ and $W(z_2)$, and thus, by monotonicity of W, we have $W^{-1}(\eta) \in (-L, L)$ (since $z_1, z_2 \in (-L, L)$). In fact, for $L < \infty$, we can choose $w(x) = 1$ for $x \in (-L, L)$ and $w(x) = 0$ else, and the term (33)

DOI: 10.1111/jtsa.12200

(i.e. (30)) can therefore be bounded by

$$\frac{\|f\|_\infty\ (\sup_{s\in\mathcal{S}}\|s\|_n + m^*)}{m_*}\ \epsilon.$$

For $L=\infty$, we also need to consider cases in which $W^{-1}(\eta)$ is large. We fix L_0 large enough so that, for $W^{-1}(\eta) > L_0$, we have on $\mathbf{A}_n$ that $\left|\mathbf{g}_1\left(\frac{t}{n}\right)' Y_{t-1} + (s_1+\sigma)\left(\frac{t}{n}\right) W^{-1}(\eta)\right| > |W^{-1}(\eta)/2|$ and that $\sup_{|x|>L_0}|x|^{1+\beta}w(x) > 1/2$. The latter can be achieved by our assumption on w (given right before the formulation of the lemma). To see the former, assume that (a) $s_1\left(\frac{t}{n}\right)+\sigma\left(\frac{t}{n}\right) > c_*$ and (b) $\sup_{g\in\mathcal{G}}\left|\mathbf{g}\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}\right| < C_0^*$. Then, if $|z| \geq \frac{2C_0^*}{c_*}$, then it is easy to see that $|\eta_{\frac{t}{n}}(z_2)| \leq \frac{|z|}{2}$, which is the assertion. (a) holds by assumption and (b) holds on $\mathbf{A}_n$. Using these properties, we obtain

$$\begin{aligned}
\frac{f_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}+(s_1+\sigma)\left(\frac{t}{n}\right)W^{-1}(\eta)\right)}{w\left(W^{-1}(\eta)\right)} &\leq \frac{M\sigma^\beta\left(\frac{t}{n}\right)}{\left|\left(\mathbf{g}_1\left(\frac{t}{n}\right)\right)'\mathbf{Y}_{t-1}+(s_1+\sigma)\left(\frac{t}{n}\right)W^{-1}(\eta)\right|^{1+\beta} w\left(W^{-1}(\eta)\right)}\\
&\leq \frac{2^{1+\beta}\,(m^*)^\beta}{\left|\left(W^{-1}(\eta)\right)^{1+\beta}\right| w(W^{-1}(\eta))} \leq 2^{2+\beta}\,(m^*)^\beta\,M.
\end{aligned}$$

Thus, for $L=\infty$, we have the bound

$$(30) \leq \max\left(\frac{\|f\|_\infty}{m_*},\ 2^{2+\beta}\,(m^*)^\beta \sup_{x\in\mathbb{R}}\left|x^{1+\beta}f(x)\right|\right)\left(\sup_{s\in\mathcal{S}}\|s\|_n + m^*\right)\ \epsilon.$$

Next, we consider (31). By an application of a one-term Taylor expansion, there exists $\xi\left(\frac{t}{n}\right)$ lying between $s_1\left(\frac{t}{n}\right)$ and $s_2\left(\frac{t}{n}\right)$ such that, with the shorthand notation $\eta_{\frac{t}{n}}(z_2) = \mathbf{g}_1\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1} + \left(\xi\left(\frac{t}{n}\right)+\sigma\left(\frac{t}{n}\right)\right) z_2$, we have that

$$F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}+(s_1+\sigma)\left(\frac{t}{n}\right)z_2\right)-F_{\frac{t}{n}}\left(\mathbf{g}_1\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}+(s_2+\sigma)\left(\frac{t}{n}\right)z_2\right)=f_{\frac{t}{n}}\left(\eta_{\frac{t}{n}}(z_2)\right)(s_1-s_2)\left(\frac{t}{n}\right)z_2.$$

Consequently, for $L<\infty$,

$$(31) = \frac{1}{n}\sum_{t=1}^n f_{\frac{t}{n}}\left(\eta_{\frac{t}{n}}(z_2)\right)\left|(s_1-s_2)\left(\frac{t}{n}\right)z_2\right| \leq \frac{\|f\|_\infty}{m_*}\,L\,\epsilon.$$

For $L=\infty$, we argue as follows. Since $\xi\left(\frac{t}{n}\right)$ lies between $s_1\left(\frac{t}{n}\right)$ and $s_2\left(\frac{t}{n}\right)$, we have from our assumption that $\xi\left(\frac{t}{n}\right)+\sigma\left(\frac{t}{n}\right) > c_*$. As mentioned earlier in the estimation of (33), it now follows that if $|z| \geq \frac{2C_0^*}{c_*}$, then $|\eta_{\frac{t}{n}}(z_2)| \leq \frac{|z|}{2}$, and we have for $|z| \geq \frac{2C_0^*}{c_*}$ that

$$\begin{aligned}
(31) &= \frac{1}{n}\sum_{t=1}^n f_{\frac{t}{n}}\left(\eta_{\frac{t}{n}}(z_2)\right)\left|(s_1-s_2)\left(\frac{t}{n}\right)z_2\right|\\
&\leq \frac{1}{n}\sum_{t=1}^n \frac{1}{\sigma\left(\frac{t}{n}\right)} f\left(\frac{\eta_{\frac{t}{n}}(z_2)}{\sigma\left(\frac{t}{n}\right)}\right)\left|(s_1-s_2)\left(\frac{t}{n}\right)\right|\left|\frac{\eta_{\frac{t}{n}}(z_2)}{\sigma\left(\frac{t}{n}\right)}\right|\sigma\left(\frac{t}{n}\right)\ \frac{|z_2|}{|\eta_{\frac{t}{n}}(z_2)|}\\
&\leq \frac{m^*\,\sup_{x\in\mathbb{R}}|xf(x)|}{2m_*}\ \frac{1}{n}\sum_{t=1}^n\left|(s_1-s_2)\left(\frac{t}{n}\right)\right|\\
&\leq \frac{m^*\,\sup_{x\in\mathbb{R}}|xf(x)|}{2m_*}\ \epsilon.
\end{aligned} \tag{34}$$

DOI: 10.1111/jtsa.12200

Consequently, we have uniformly over $z \in \mathbb{R}$

$$(31) \leq \frac{1}{m_*} \max\left(\frac{2C_0^* \|f\|_\infty}{c_*}, \frac{m^* \sup_{x\in\mathbb{R}} |xf(x)|}{2} \right) \epsilon.$$

As for (32), we have on $\mathbf{A}_n$ (and note that we do not need the event $\left\{\sup_{g\in\mathcal{G}} |\mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}| < C_0^*\right\}$ to hold here)

$$\begin{aligned}(32) &\leq \sup_{u,x} f_u(x) \frac{1}{n}\sum_{t=1}^{n} \left| (\mathbf{g}_1 - \mathbf{g}_2)\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} \right| \\ &\leq \frac{\|f\|_\infty}{m_*} \sum_{k=1}^{p} \|g_{1k} - g_{2k}\|_n \sqrt{\frac{p}{n}\sum_{t=-p}^{n} Y_t^2} \leq \sqrt{p}\, C_0 \frac{\|f\|_\infty}{m_*} \epsilon.\end{aligned} \tag{35}$$

Putting everything together, we obtain that, for $h_1, h_2 \in \mathcal{H}_L^*$ with $d_{n,W}(h_1, h_2) \leq \epsilon$ and $\epsilon \geq \frac{1}{n}$, we have on $\mathbf{A}_n$ for $0 < L < \infty$ that

$$V_n(\nu_n(h_1) - \nu_n(h_2)) \leq K^* \epsilon, \tag{36}$$

with

$$K^* = \frac{\|f\|_\infty}{m_*} \left(\sup_{s\in\mathcal{S}} \|s\|_n + m^* + L + \sqrt{p}\, C_0 \right) \epsilon,$$

as defined in the formulation of the lemma. If $L = \infty$, then the formula just shown shows that (36) holds where the corresponding K^* can immediately be obtained from the aforementioned estimates.

Proofs of statements in Remarks given after the lemma

(a) If we modify $\mathbf{A}_n$ by dropping the event $\left\{\sup_{g\in\mathcal{G}} |\mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}| < C_0^*\right\}$ and instead assume that $\sup_{g\in\mathcal{G}} \|g\|_\infty < \infty$, then we can estimate the last average in (34) as follows:

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} \left|(s_1 - s_2)\left(\tfrac{t}{n}\right) \left(\mathbf{g}_1\left(\tfrac{t}{n}\right)\right)' \mathbf{Y}_{t-1}\right| &\leq \sqrt{\frac{1}{n}\sum_{t=1}^{n}(s_1 - s_2)^2\left(\tfrac{t}{n}\right) \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{g}_1\left(\tfrac{t}{n}\right)\right)' \mathbf{Y}_{t-1})^2} \\ &\leq \epsilon \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{g}_1\left(\tfrac{t}{n}\right)\right)' \mathbf{Y}_{t-1})^2} \\ &\leq \epsilon \sqrt{\sup_{g\in\mathcal{G}} \|g\|_\infty\, p\, \frac{1}{n}\sum_{t=-p+1}^{n} Y_t^2} \\ &\leq C_0 \left(\sup_{g\in\mathcal{G}} \|g\|_\infty\, p\right)^{1/2} \epsilon.\end{aligned}$$

This then leads to a slightly modified constant K^*, which can easily be determined.

(b) If $\mathcal{H}_L^*$ is replaced by $\mathcal{H}_L$ then, as has been discussed earlier, this formally corresponds to setting $\mathcal{S} = \{1 - \sigma\}$, which is a class consists of just one function, implying that, in the aforementioned proof, $s_1 - s_2 = 0$. In other words, all the assumptions used to bound (31) become obsolete.

DOI: 10.1111/jtsa.12200

The just shown control of the quadratic variation in conjunction with Freedman's exponential bound for martingales now enables us to apply the (restricted) chaining argument in a way similar to the proof of Thm 5.11 in van de Geer (2000). Details are omitted. □

6.2. Proofs of Theorems 1 and 3

We only present the proof of Theorem 3. The proof of Theorem 1 is exactly the same. Recall that the metric $d_{n,W}(h_1, h_2)$ on $\mathcal{H}_L^*$ is defined in (25), and observe that, for $h_1 = (\alpha, z, \mathbf{g}, s)$ and $h_2 = (\alpha, z, \mathbf{0}, 0)$, we have $d_{n,W}(h_1, h_2) = \sum_{k=1}^{p} \|\mathbf{g}_k - \mathbf{0}\|_n + \|s\|_n$. We thus obtain for $\eta > 0$ and $0 < L < \infty$ and $C > 0$ that

$$P\left(\sup_{\substack{\alpha\in[0,1], z\in(-L,L), \mathbf{g}\in\mathcal{G}^p \\ \sum_{k=1}^{p}\|g_k\|_n+\|s\|_n\le\delta_n}} |\nu_n(\alpha, z, \mathbf{g}, s) - \nu_n(\alpha, z, \mathbf{0}, 0)| \ge \eta\right)$$
$$\le P\left(\mathbf{A}_n^{\complement}\right) + P\left(\sup_{\substack{d_{n,W}(h_1,h_2)\le\delta_n \\ h_1,h_2\in\mathcal{H}_L^*}} |\nu_n(h_1) - \nu_n(h_2)| \ge \eta, \mathbf{A}_n\right),$$

where as in Lemma 1, $\mathbf{A}_n = \left\{\frac{1}{n}\sum_{s=-p+1}^{n} Y_s^2 \le C_0^2\right\} \cap \left\{\sup_{g\in\mathcal{G}} |\mathbf{g}\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}| < C_0^*\right\}$ for some $C_0, C_0^* > 0$. An application of Lemma 1 gives the assertion once we have shown that $P\left(\mathbf{A}_n^{\complement}\right)$ can be made arbitrarily small (for sufficiently large n) and that $\int_{\frac{c}{n}}^{1} \sqrt{\log N_B\left(u^2, \mathcal{H}_L^*, d_{n,W}\right)}\, du < \infty$.

To see this, notice that, by our assumptions, we have both $\int_{\frac{c}{n}}^{1} \sqrt{\log N_B(u^2, \mathcal{G}, d_n)}\, du < \infty$ and $\int_{\frac{c}{n}}^{1} \sqrt{\log N_B\left(u^2, \mathcal{S}, d_n\right)}\, du < \infty$. This implies $\int_{\frac{c}{n}}^{1} \sqrt{\log N_B\left(u^2, \mathcal{H}_L^*, d_{n,W}\right)}\, du < \infty$, because for $h_i = (\alpha_i, z_i, \mathbf{g}_i, s_i), i = 1, 2$ with $|\alpha_1 - \alpha_2| < \frac{\epsilon}{4}$, $|W(z_1) - W(z_2)| < \frac{\epsilon}{4}$, $\|g_{k1} - g_{k2}\|_n < \frac{\epsilon}{4p}, k = 1, \ldots, p$ and $\|s\|_n < \frac{\epsilon}{4}$, we obviously have $d_{n,W}(h_1, h_2) < \epsilon$, and thus, by using standard arguments, it is not difficult to see that

$$\log N_B\left(\epsilon, \mathcal{H}_L^*, d_{n,W}\right) \le C_0 \log\tfrac{4}{\epsilon} + p\, \log N_B\left(\tfrac{\epsilon}{4p}, \mathcal{G}, d_n\right) + \log N_B\left(\tfrac{\epsilon}{4}, \mathcal{S}, d_n\right) \tag{37}$$

for some $C_0 > 0$. It remains to show that $P(\mathbf{A}_n^{\complement})$ can be made arbitrarily small (for n large) by choosing C_0, C sufficiently large. Because of Assumption (iii), this follows from $\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[Y_t^2\right] < \infty$, which in turn follows easily from

$$\mathrm{E}Y_t^2 = \mathrm{E}\left[\sum_{j=-\infty}^{t} a_{t,n}(t-j)\epsilon_j\right]^2 = \mathrm{E}\sum_{j=-\infty}^{t}\sum_{k=-\infty}^{t} a_{t,n}(t-j)a_{t,n}(t-k)\epsilon_j\epsilon_k$$
$$= \sum_{j=-\infty}^{t} a_{t,n}^2(t-j) \le \sum_{j=0}^{\infty}\left(\frac{K}{\ell(j)}\right)^2 < C < \infty,$$

for some $C > 0$. Here, we are using Assumption (ii).

6.3. Proof of Theorem 5

First, we formulate and prove a lemma that is needed in the proof of Theorem 5. For random variables $X_1, \ldots, X_k$, we denote by $\mathrm{cum}(X_1, \ldots, X_k)$ their joint cumulant, and if $X_i = X$ for all $i = 1, \ldots, k$, then $\mathrm{cum}(X_1, \ldots, X_k) = \mathrm{cum}(X, \ldots, X) = \mathrm{cum}_k(X)$, the kth-order cumulant of X.

DOI: 10.1111/jtsa.12200

Lemma 2. Let $\{Y_t,\ t = 1, \dots, n\}$ satisfy Assumptions (i) and (ii). For $j = 1, 2, \dots$, let h_j be functions defined on $[a, b]$ with $\|h_j\|_n < \infty$. Then there exists a constant $1 \le K_0 < \infty$ such that for all $k \ge 1$ such that

$$\left|\operatorname{cum}(Z_n(h_1), \dots, Z_n(h_k))\right| \le K_0^{k-1} \left|\operatorname{cum}_k(\epsilon_1)\right| \prod_{j=1}^{k} \|h_j\|_n.$$

If, in addition, $\|h_j\|_\infty \le M < \infty,\ j = 1, \dots, k$, then for $k \ge 3$,

$$\left|\operatorname{cum}(Z_n(h_1), \dots, Z_n(h_k))\right| \le (K_0)^{k-2} M^k \left|\operatorname{cum}_k(\epsilon_1)\right| n^{-\frac{k-2}{2}}.$$

Proof

We have $\operatorname{cum}(Z_n(h_1), \dots, Z_n(h_k)) = n^{-\frac{k}{2}} \sum_{t_1, t_2, \dots, t_k = 1}^{n} h_1\left(\frac{t_1}{n}\right) \dots h_k\left(\frac{t_k}{n}\right) \operatorname{cum}(Y_{t_1}, \dots, Y_{t_k})$ by utilizing multilinearity of cumulants. In order to estimate $\operatorname{cum}(Y_{t_1}, \dots, Y_{t_k})$, we utilize the special structure of the Y_t-variables. Since the ϵ_j are independent and have mean zero, we have that $\operatorname{cum}(\epsilon_{j_1}, \dots, \epsilon_{j_k}) = 0$ unless all the $j_\ell,\ \ell = 1, \dots, k$ are equal, and by again using multilinearity of the cumulants, we obtain

$$\operatorname{cum}(Y_{t_1}, \dots, Y_{t_k}) = \operatorname{cum}_k(\epsilon_1) \sum_{j=0}^{\min\{t_1, \dots, t_k\}} a_{t_1,n}(t_1 - j) \dots a_{t_k,n}(t_k - j). \tag{38}$$

Thus, $\left|\operatorname{cum}(Y_{t_1}, \dots, Y_{t_k})\right| \le \left|\operatorname{cum}_k(\epsilon_1)\right| \sum_{j=0}^{\infty} \prod_{i=1}^{k} \frac{K}{\ell(|t_i - j|)}$, and consequently,

$$\begin{aligned}
\left|\operatorname{cum}(Z_n(h_1), \dots, Z_n(h_k))\right| &\le n^{-\frac{k}{2}} \left|\operatorname{cum}_k(\epsilon_1)\right| \sum_{j=-\infty}^{\infty} \prod_{i=1}^{k} \left[\sum_{t_i=0}^{n} \left|h_i\left(\frac{t_i}{n}\right)\right| \frac{K}{\ell(|t_i - j|)}\right] \\
&= n^{-\frac{k}{2}} \left|\operatorname{cum}_k(\epsilon_1)\right| \sum_{j=-\infty}^{\infty} \prod_{i=1}^{2} \left[\sum_{t_i=0}^{n} \left|h_i\left(\frac{t_i}{n}\right)\right| \frac{K}{\ell(|t_i - j|)}\right] \\
&\quad \times \prod_{i=3}^{k} \left[\sum_{t_i=0}^{n} \left|h_i\left(\frac{t_i}{n}\right)\right| \frac{K}{\ell(|t_i - j|)}\right].
\end{aligned}$$

Utilizing the Cauchy–Schwarz inequality, we have for the last product

$$\begin{aligned}
\prod_{i=3}^{k} \left[\sum_{t_i=0}^{n} \left|h_i\left(\tfrac{t_i}{n}\right)\right| \frac{1}{\ell(|t_i - j|)}\right] &\le \prod_{i=3}^{k} \sqrt{\sum_{t_i=0}^{n} h_i\left(\frac{t_i}{n}\right)^2} \sqrt{\sum_{t_i=0}^{n} \left(\frac{K}{\ell(|t_i - j|)}\right)^2} \\
&\le n^{\frac{k}{2}-1} \prod_{i=3}^{k} \|h_i\|_n \sqrt{\sum_{t=-\infty}^{\infty} \left(\frac{K}{\ell(|t|)}\right)^2} \\
&\le K_0^{k-2} n^{\frac{k}{2}-1} \prod_{i=3}^{k} \|h_i\|_n,
\end{aligned} \tag{39}$$

DOI: 10.1111/jtsa.12200

where we used the fact that $\sqrt{\sum_{t=-\infty}^{\infty}\left(\frac{K}{\ell(|t|)}\right)^2} \leq \sum_{t=-\infty}^{\infty}\frac{K}{\ell(|t|)} \leq K_0$ for some $K_0 < \infty$. Notice that the bound (39) does not depend on the index j anymore, so that

$$\left|\mathrm{cum}(Z_n(h_1),\ldots,Z_n(h_k))\right| \leq K_0^{k-2}n^{-1}\prod_{i=3}^{k}\|h_i\|_n\left|\mathrm{cum}_k(\epsilon_1)\right|\sum_{j=-\infty}^{\infty}\prod_{i=1}^{2}\left[\sum_{t_i=0}^{n}\left|h_i\left(\tfrac{t_i}{n}\right)\right|\frac{K}{\ell(|t_i-j|)}\right].$$

By using $\sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)}\frac{K}{\ell(|t_2-j|)} = \sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)}\frac{K}{\ell(|(t_1-t_2)+t_1-j|)} \leq \frac{K}{\ell(|t_1-t_2|)}\sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)} \leq \frac{K^*}{\ell(|t_1-t_2|)}$ for some $K^* > 0$ and Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
\sum_{j=-\infty}^{\infty}\prod_{i=1}^{2}\left[\sum_{t_i=0}^{n}\left|h_i\left(\frac{t_i}{n}\right)\right|\frac{K}{\ell(|t_i-j|)}\right] &= \sum_{t_1=0}^{n}\sum_{t_2=0}^{n}h_1\left(\frac{t_1}{n}\right)h_2\left(\frac{t_2}{n}\right)\sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)}\frac{K}{\ell(|t_2-j|)}\\
&\leq \sum_{t_1=0}^{n}\sum_{t_2=0}^{n}h_1\left(\frac{t_1}{n}\right)h_2\left(\frac{t_2}{n}\right)\frac{K^*}{\ell(|t_1-t_2|)}\\
&\leq K^*\sqrt{\sum_{t_1=0}^{n}h_1\left(\frac{t_1}{n}\right)^2\sum_{t_2=0}^{n}\frac{1}{\ell(|t_1-t_2|)}}\sqrt{\sum_{t_1=0}^{n}h_1\left(\frac{t_2}{n}\right)^2\sum_{t_2=0}^{n}\frac{1}{\ell(|t_1-t_2|)}}\\
&\leq K_0\, n\,\|h_1\|_n\,\|h_2\|_n.
\end{aligned}$$

This completes the proof of the first part of the lemma. The second part follows similar to the aforementioned by observing that if $\|h_i\|_\infty < M$ for all $i = 1,\ldots,k$, then, instead of the estimate (39), we have with $K_0 = \sum_{t=-\infty}^{\infty}\frac{1}{\ell(|t|)}$ that

$$\prod_{i=3}^{k}\left[\sum_{t_i=0}^{n}\left|h_i\left(\tfrac{t_i}{n}\right)\right|\frac{1}{\ell(|t_i-j|)}\right] \leq M^{k-2}\prod_{i=3}^{k}\sum_{t_i=0}^{n}\frac{1}{\ell(|t_i-j|)} \leq (MK_0)^{k-2}.$$

□

Now, we continue with the proof of Theorem 5.

Showing weak convergence of $Z_n(h)$ means proving asymptotic tightness and convergence of the finite dimensional distribution (e.g. van der Vaart and Wellner, 1996). Tightness follows from Theorem 6. It remains to show the convergence of the finite dimensional distributions. To this end, we will utilize the Cramér–Wold device in conjunction with the method of cumulants. It follows from Lemma 2 that all the cumulants of $Z_n(h)$ of order $k \geq 3$ converge to zero as $n \to \infty$. Using the linearity of the cumulants, the same holds for any linear combination of $Z_n(h_i)$, $i = 1,\ldots,K$. The mean of all the $Z_n(h)$ equals zero. It remains to show that convergence of the covariances $\mathrm{cov}(Z_n(h_1), Z_n(h_2))$. The range of the summation indices shown next are such that the indices of the Y-variables are between 1 and n. For ease of notation, we achieve this by formally setting $h_i(u) = 0$ for $u \leq 0$ and $u > 1$, $i = 1, 2$. We have

$$\begin{aligned}
\mathrm{cov}(Z_n(h_1), Z_n(h_2)) &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}h_1\left(\frac{t}{n}\right)\cdot h_2\left(\tfrac{s}{n}\right)\mathrm{cov}(Y_s, Y_t)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\leq\sqrt{n}}h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t-k}{n}\right)\mathrm{cov}(Y_t, Y_{t-k}) \quad + \quad R_{1n},
\end{aligned} \tag{40}$$

DOI: 10.1111/jtsa.12200

where for n sufficiently large,

$$|R_{1n}| \leq \frac{1}{n}\sum_{t=1}^{n}\sum_{|k|>\sqrt{n}}\left|h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t-k}{n}\right)\right|\,\left|\operatorname{cov}(Y_t, Y_{t-k})\right|.$$

From Proposition 5.4 of Dahlhaus and Polonik (2009), we obtain that $\sup_t |\operatorname{cov}(Y_t,\, Y_{t-k})| \leq \frac{K}{\ell(k)}$ for some constant K. Since both h_1 and h_2 are bounded and $\sum_{k=-\infty}^{\infty}\frac{1}{\ell(k)} < \infty$, we can conclude that $R_{1n} = o(1)$. The main term in (40) can be approximated as

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\leq\sqrt{n}} h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t-k}{n}\right)\; c\left(\frac{t}{n},\, k\right) + R_{2n}, \tag{41}$$

where

$$|R_{2,n}| \leq \frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\leq\sqrt{n}}\left|h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t-k}{n}\right)\right|\,\left|\operatorname{cov}(Y_t, Y_{t-k}) - c\left(\frac{t}{n},\, k\right)\right|.$$

Proposition 5.4 of Dahlhaus and Polonik (2009) also gives us that, for $|k| \leq \sqrt{n}$, we have $\sum_{t=0}^{n}\left|\operatorname{cov}(Y_t, Y_{t-k}) - c\left(\frac{t}{n},\, k\right)\right| \leq K\left(1+\frac{|k|}{n}\right)$ for some $K > 0$. Using this, we obtain

$$\begin{aligned}|R_{2,n}| &\leq \frac{1}{n}\sum_{t=1}^{n}\sum_{k=-\sqrt{n}}^{\sqrt{n}}\left|h_1\left(\frac{t}{n}\right)\cdot h_2\left(\tfrac{t-k}{n}\right)\right|\,\left|\operatorname{cov}(Y_t, Y_{t-k}) - c\left(\frac{t}{n},\, k\right)\right|\\ &\leq K_1\,\frac{1}{n}\sum_{k=-\sqrt{n}}^{\sqrt{n}}\sum_{t=1}^{n}\left|\operatorname{cov}(Y_t, Y_{t-k}) - c\left(\frac{t}{n},\, k\right)\right| \leq K_1\,\frac{1}{n}\sum_{k=-\sqrt{n}}^{\sqrt{n}}\left(1+\frac{|k|}{\ell(|k|)}\right) = o(1)\end{aligned}$$

as $n \to \infty$. Next, we replace $h_2(\frac{t-k}{n})$ in the main term of (41) by $h_2\left(\frac{t}{n}\right)$. The approximation error can be bounded by

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\leq\sqrt{n}}\left|h_1\left(\frac{t}{n}\right)\right|\cdot\left|h_2\left(\frac{t-k}{n}\right) - h_2\left(\frac{t}{n}\right)\right|\,\frac{K}{\ell(|k|)} = o(1).$$

Here, we are using the fact that $\sup_u |c(u,k)| \leq \frac{K}{\ell(|k|)}$ (see Propn 5.4 in Dahlhaus and Polonik, 2009) together with the assumed (uniform) continuity of h_2, the boundedness of h_1 and the boundedness of $\sum_{k=-\infty}^{\infty}\frac{1}{\ell(|k|)}$. We have seen that

$$\operatorname{cov}(Z_n(h_1),\, Z_n(h_2)) = \frac{1}{n}\sum_{t=1}^{n} h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t}{n}\right)\sum_{k\leq\sqrt{n}} c\left(\frac{t}{n},\, k\right) + o(1).$$

Since $S(u) = \sum_{k=-\infty}^{\infty} c\left(\frac{t}{n},\, k\right) < \infty$, we also have

$$\operatorname{cov}(Z_n(h_1),\, Z_n(h_2)) = \sum_{k=-\infty}^{\infty}\frac{1}{n}\sum_{t=1}^{n} h_1\left(\frac{t}{n}\right)\cdot h_2\left(\frac{t}{n}\right)\; c\left(\frac{t}{n},\, k\right) + o(1).$$

DOI: 10.1111/jtsa.12200

Finally, we utilize the fact that $TV(c(\cdot,k)) \leq \frac{K}{\ell(|k|)}$, which is another result from Propn 5.4 of Dahlhaus and Polonik (2009). This result, together with the assumed bounded variation of both h_1 and h_2, allows us to replace the average over t by the integral.

6.4. A crucial exponential inequality

The following exponential inequality is a crucial ingredient to the proof of Theorem 6. It relies on the control of the cumulants, which is provided in Lemma 2 shown earlier.

Lemma 3. Let $\{Y_t,\ t = 1,\dots,n\}$ satisfy Assumptions (i) and (ii). Let h be a function with $\|h\|_n < \infty$. Assume that there exists a constant $C > 0$ such that, for all $k = 1,2,\dots,$ we have $|\mathrm{cum}_k(\epsilon_t)| \leq k!\,C^k$. Then there exist constants $c_1, c_2 > 0$ such that, for any $\eta > 0$, we have

$$P\left[\,|Z_n(h)| > \eta\,\right] \leq c_1 \exp\left\{-\frac{\eta}{c_2\,\|h\|_n}\right\}. \tag{42}$$

Proof
Using Lemma 2, we have the assumptions on the cumulants implying that

$$\Psi_{Z_n(h)}(t) = \log \mathrm{E}e^{t\,Z_n(h)} \leq \frac{1}{K_0}\sum_{k=1}^{\infty}(tCK_0\,\|h\|_n)^k,$$

assuming that $t > 0$ is such that the infinite sum exists and is finite. We obtain

$$P\left[|Z_n(h)| > \eta\right] \leq 2\,e^{-t\eta}\,\mathrm{E}\left(e^{Z_n(h)}\right) \leq 2\,\exp\left\{-t\eta + K_0^{-1}\sum_{k=1}^{\infty}(tCK_0\,\|h\|_n)^k\right\}.$$

Choosing $t = \frac{1}{2CK_0\|h\|_n}$ gives the assertion with $c_1 = 2\,e^{1/K_0}$ and $c_2 = 2CK_0$. The fact that

$$\left|\mathrm{cum}_j(\epsilon_t)\right| \leq j!\,C^j,\quad j = 1,2,\dots \tag{43}$$

holds if $\mathrm{E}|\epsilon_t|^k \leq \left(\frac{C}{2}\right)^k,\ k = 1,2,\dots,$ can be seen by induction. Details are omitted. □

6.5. Proof of Theorem 6

Using Lemma 3, we can mimic the proof of Lem 3.2 from van de Geer (2000). As compared with that of van de Geer, our exponential bound is of the form $c_1 \exp\left\{-c_2\frac{\eta}{\|h\|_n}\right\}$ rather than $c_1 \exp\left\{-c_2\left(\frac{\eta}{\|h\|_n}\right)^2\right\}$. It is well known that this type of inequality leads to the covering integral being the integral of the metric entropy rather than the square root of the metric entropy. (See, for instance, Thm 2.2.4 in van der Vaart and Wellner, 1996.) This indicates the necessary modifications to the proof in van de Geer. Details are omitted.

6.6. Proofs of Theorems 2 and 4

Proof of Theorem 4
Recall that $\widehat{F}^*_n(\alpha,z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n\rfloor}\mathbf{1}\left\{\sigma\left(\frac{t}{n}\right)\epsilon_t \leq (\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1} + (\widehat{\sigma}-\sigma)\left(\frac{t}{n}\right)z + \sigma\left(\frac{t}{n}\right)z\right\}$ and that, by

DOI: 10.1111/jtsa.12200

assumption, $P\left(\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta},\, \widehat{\sigma} - \sigma\right) \in \mathcal{G}^p \times \mathcal{S}\right) \to 1$ as $n \to \infty$. Let

$$F_n^*(\alpha, z, \mathbf{g}, s) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\left\{\sigma\left(\frac{t}{n}\right) \epsilon_t \leq \mathbf{g}'\left(\frac{t}{n}\right) \mathbf{Y}_{t-1} + (s+\sigma)\left(\tfrac{t}{n}\right) z\right\}. \tag{44}$$

Then we have with $F_n^*(\alpha, z, \mathbf{0}, 0) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\epsilon_t \leq z\}$ that $F_n^*(\alpha, z) = F_n^*(\alpha, z, \mathbf{0}, 0)$ and $\widehat{F}_n^*(\alpha, z) = F_n(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma)$. Further, let

$$E_n^*(\alpha, z, \mathbf{g}, s) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} F_{\frac{t}{n}}\left(\left(\mathbf{g}\left(\frac{t}{n}\right)\right)' \mathbf{Y}_{t-1} + (s+\sigma)\left(\tfrac{t}{n}\right) z\right) \tag{45}$$

denote the conditional expectation of $F_n^*(\alpha, z, \mathbf{g}, s)$ given $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, \ldots)$. The purpose of introducing E_n^* is to make $F_n^* - E_n^*$ a martingale difference (see in the following texts). We now have

$$\begin{aligned}
\widehat{F}_n^*(\alpha, z) - F_n^*(\alpha, z) &= F_n^*(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma) - F_n^*(\alpha, z, \mathbf{0}, 0) \\
&= \left[(F_n^* - E_n^*)(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma) - (F_n^* - E_n^*)(\alpha, z, \mathbf{0}, 0)\right] \\
&\quad + \left[E_n^*(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma) - E_n^*(\alpha, z, \mathbf{0}, 0)\right] \\
&=: T_{n1}(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma) + T_{n2}(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma).
\end{aligned} \tag{46}$$

Using this decomposition, we will show the assertion by proving the following two properties:

$$\sup_{\alpha \in [0,1],\, z \in (-L, L)} \left|\sqrt{n}\, T_{n1}(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma)\right| = o_P(1) \tag{47}$$

$$\sup_{\alpha \in [0,1],\, z \in (-L, L)} \sqrt{n}\left[T_{n2}(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma) - z f(z) \frac{1}{n} \sum_{t=1}^{n} \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}\right] = o_P(1). \tag{48}$$

Verification of (47). Let $\nu_n^*(\alpha, z, \mathbf{g}, s) = \sqrt{n}\left(F_n^* - E_n^*\right)(\alpha, z, \mathbf{g}, s)$ denote the error sequential empirical process. By taking into account that, by assumption, $\sum_{k=1}^{p} \|\widehat{\theta}_k - \theta_k\|_n = O_P(m_n^{-1})$ and $\|\widehat{\sigma} - \sigma\|_n = O_P(n^{-1/4})$, we have that, for any $\epsilon > 0$, we can find a $C > 0$ such that, with probability at least $1 - \epsilon$ for large enough n,

$$\sup_{\alpha \in [0,1],\, z \in (-L, L)} \left|\sqrt{n}\, T_{n1}(\alpha, z, \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \widehat{\sigma} - \sigma)\right| \leq \sup_{\substack{\alpha \in [0,1],\, z \in (-L, L), \\ \mathbf{g} \in \mathcal{G}^p,\, \sum_{k=1}^{p} \|g_k\|_n \leq C m_n^{-1} \\ s \in \mathcal{S},\, \|s\|_n \leq C^{-1} n^{-1/4}}} |\sqrt{n}\, T_{n1}(\alpha, z, \mathbf{g}, s)|.$$

That the right-hand side is $o_P(1)$ is an immediate application of Theorem 3, and (47) is verified.

Verification of (48). We have by a simple one-term expansion that

DOI: 10.1111/jtsa.12200

$$\sqrt{n}\,T_{n2}(\alpha, z, \widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}, \widehat{\sigma}-\sigma) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \left[F\left(\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1} + \widehat{\sigma}\left(\frac{t}{n}\right) z}{\sigma\left(\frac{t}{n}\right)} \right) - F(z) \right]$$

$$= f(z) \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \left[\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} + \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} z \right] \tag{49}$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} f'(\eta_t(z)) \left[\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} + \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} z \right]^2 \tag{50}$$

with $\eta_t(z)$ between z and $\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} + \frac{\widehat{\sigma}\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} z$. The term (50) is a remainder term that will be treated as shown later. The sum in (49) can be written as a sum of two terms

$$f(z) \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} + z f(z) \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)}. \tag{51}$$

The first terms are (finite sum of) weighted sums of the Y_t's, and we will now use Theorem 6 to show that this term is $o_P(1)$. By recalling that, by assumption, the probability of $\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta} \in \mathcal{G}^p\right\}$ tends to 1 as $n \to \infty$, we can assume that $(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})_k \in \mathcal{G}$ for all $k = 1, \ldots, p$. For $h \in \mathcal{H} = \{h_{\alpha,g} = \mathbf{1}_\alpha(u) \frac{g(u)}{\sigma(u)};\ \alpha \in [0,1],\ g \in \mathcal{G}\}$ with the shorthand notation $\mathbf{1}_\alpha(u) = \mathbf{1}(u \le \alpha)$, let

$$Z_{k,n}(h) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h\left(\frac{t}{n}\right) Y_{t-k}, \quad k = 1, \ldots, p. \tag{52}$$

Then we can write the first term in (51) as $f(z) \sum_{k=1}^p Z_{k,n}\left(\mathbf{1}_\alpha \frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})_k}{\sigma}\right)$. With this notation, we have (on the event $\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta} \in \mathcal{G}^p\right\}$) that

$$\left| f(z) \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n\alpha \rceil} \frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} \right| \le \sup_x f(x) \sum_{k=1}^{p} \sup_{h \in \mathcal{H}} |Z_{k,n}(h)|.$$

It thus remains to show that $\sup_{h\in\mathcal{H}} |Z_{k,n}(h)| = o_P(1)$ for all $k = 1, \ldots, p$. To this end, we will apply Theorem 6. Clearly, $\sup_{h\in\mathcal{H}} \|h\|_n^2 \le \frac{1}{m_*} \sup_{g\in\mathcal{G}} \|g\|_n^2 \le \frac{1}{m_*} m_n^{-2} = o(1)$. Since $\mathrm{E}Z_{k,n}(h) = 0$ for all $h \in \mathcal{H}$, an application of Theorem 6 now gives the assertion once we have verified that the class $\mathcal{H}$ satisfies the condition on the covering numbers. This, however, follows easily because by assumption, $\mathcal{G}$ has a finite covering integral with respect to d_n, $\frac{1}{\sigma(u)}$ is just a single function that is bounded above and below and the class of functions $\{\mathbf{1}_\alpha(u), \alpha \in [0,1]\}$ is a Vapnik-Chervonenkis (VC)-subgraph class (and thus has a finite covering integral as well).

Now, we show that the term in (50) is $o_P(1)$. We have

$$(50) \le \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \left|f'(\eta_t(z))\right| \left[\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)' \mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)} \right]^2 \tag{53}$$

$$+ \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \left|f'(\eta_t(z))\right| z^2 \left[\frac{\widehat{\sigma}\left(\frac{t}{n}\right) - \sigma\left(\frac{t}{n}\right)}{\sigma\left(\frac{t}{n}\right)} \right]^2, \tag{54}$$

DOI: 10.1111/jtsa.12200

and under our assumptions, both (53) and (54) are $o_P(1)$ uniformly over z. In fact, for (53), we have

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left|f'(\eta_t(z))\right|\left[\frac{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}}{\sigma\left(\frac{t}{n}\right)}\right]^2$$
$$\leq \frac{2\sup_x\left|f'(x)\right|}{m_*}\;\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\left((\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)\right)'\mathbf{Y}_{t-1}\right]^2,$$

where for the last sum we have (recall that β_n is such that $\max_{1\leq t\leq n}Y_t^2=O_P\left(\beta_n^2\right)$)

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\frac{t}{n}\right)'\mathbf{Y}_{t-1}\right]^2\leq\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\sum_{k=1}^{p}(\widehat{\theta}-\theta)_k^2\left(\frac{t}{n}\right)\sum_{j=0}^{p}Y_{t-j}^2\right)$$
$$\leq p\max_{-p\leq t\leq n}Y_t^2\;\sqrt{n}\,O_P\left(\frac{1}{m_n^2}\right)=p\,\frac{\sqrt{n}}{m_n^2}\,\beta_n^2\;O_P(1)=o_P(1),$$

where the last equality uses our assumptions. The fact that (54) it $o_P(1)$ follows immediately once we have shown that

$$\sup_{z\in\mathbb{R}}\left|f'(\eta_t(z))\right|z^2=O_P(1). \tag{55}$$

To see that (55) holds, first observe that, by assumption, we have $\lim_{|x|\to\infty}|xf(x)|=0$, and an application of l'Hospital's Rule gives $\lim_{|x|\to\infty}|x^2f'(x)|=0$. Since f' is continuous, we conclude that f' is bounded, and thus, it is clear that (55) holds for $|z|<L<\infty$. For $L=\infty$, the argument is as follows. The previous arguments show that $\sup_{x\in\mathbb{R}}|f'(x)x^2|<\infty$. Since $\sup_{z\in\mathbb{R}}\max_{1\leq t\leq n}|f'(\eta_t(z))z^2|=\sup_{z\in\mathbb{R}}\max_{1\leq t\leq n}|f'(\eta_t(z))\eta_t^2(z)|\,|\frac{z^2}{\eta_t^2(z)}|$, we see that it is sufficient to show that

$$\sup_{z\in\mathbb{R}}\max_{1\leq t\leq n}\left|\frac{z}{\eta_t(z)}\right|=O_P(1). \tag{56}$$

This follows by an argument along the lines used to bound (33) in the proof of Lemma 3. Further details are omitted. Treating residual empirical process, i.e. *proving Theorem* 2, is very similar. We omit the details. □

Proof of Theorem 7
First recall that $\widehat{F}_n(\alpha,z)=\frac{1}{n}\sum_{t=1}^{\lfloor\alpha n\rfloor}\mathbf{1}(\widehat{\eta}_t\leq z)$ and $F_n(\alpha,z)=\frac{1}{n}\sum_{t=1}^{\lfloor\alpha n\rfloor}\mathbf{1}\left(\sigma\left(\frac{t}{n}\right)\epsilon_t\leq z\right)$. With this notation, we can write

$$\begin{aligned}
&\sqrt{n}\;\left(\widehat{G}_{n,\gamma}(\alpha)-G_{n,\gamma}(\alpha)\right)\\
&=\sqrt{n}\left[\left(1-\widehat{F}_n(\alpha,\widehat{q}_\gamma)+\widehat{F}_n(\alpha,-\widehat{q}_\gamma)\right)-(1-F_n(\alpha,q_\gamma)+F_n(\alpha,-q_\gamma))\right]\\
&=\left[\sqrt{n}\left(F_n(\alpha,\widehat{q}_\gamma)-\widehat{F}_n(\alpha,\widehat{q}_\gamma)\right)-\sqrt{n}\left(F_n(\alpha,-\widehat{q}_\gamma)-\widehat{F}_n(\alpha,-\widehat{q}_\gamma)\right)\right]\\
&\quad-\left[\sqrt{n}\;(F_n(\alpha,\widehat{q}_\gamma)-F_n(\alpha,q_\gamma))-\sqrt{n}\;(F_n(\alpha,-\widehat{q}_\gamma)-F_n(\alpha,-q_\gamma))\right]\\
&=:I_n(\alpha)-II_n(\alpha)
\end{aligned} \tag{57}$$

DOI: 10.1111/jtsa.12200

with $I_n(\alpha)$ and $II_n(\alpha)$, respectively, denoting the two quantities inside the two [·]-brackets. The assertion of Theorem 7 follows from

$$\sup_{\alpha\in[0,1]} |I_n(\alpha)| = o_P(1) \qquad \text{and} \tag{58}$$

$$\sup_{\alpha\in[0,1]} |II_n(\alpha) - c(\alpha)\ (G_{n,\gamma}(1) - \mathrm{E}G_{n,\gamma}(1))\,| = o_P(1). \tag{59}$$

Clearly, $\sup_{\alpha\in[0,1]} |I_n(\alpha)| \le 2 \sup_{\alpha\in[0,1],z\in\mathbb{R}} |\widehat{F}_n(\alpha,z) - F_n(\alpha,z)|$, and thus, property (58) follows from Theorem 2. Property (59) will now be shown by using empirical process theory based on independent, but not identically distributed, random variables. □

Proof of (59)
Define

$$\overline{F}_n(\alpha,z) := EF_n(\alpha,z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n\rfloor} F\left(\frac{z}{\sigma\left(\frac{t}{n}\right)}\right).$$

We can write

$$\begin{aligned} II_n(\alpha) = {} & \sqrt{n}\,\big((F_n - \overline{F}_n)(\alpha,\widehat{q}_\gamma) - (F_n - \overline{F}_n)(\alpha,q_\gamma)\big) && (60)\\ & + \sqrt{n}\,\big((F_n - \overline{F}_n)(\alpha,-\widehat{q}_\gamma)\ - (F_n - \overline{F}_n)(\alpha,-q_\gamma)\big) && (61)\\ & + \sqrt{n}\,(\overline{F}_n(\alpha,\widehat{q}_\gamma) - \overline{F}_n(\alpha,q_\gamma))\ -\ \sqrt{n}\,(\overline{F}_n(\alpha,-\widehat{q}_\gamma) - \overline{F}_n(\alpha,-q_\gamma)). && (62)\end{aligned}$$

The process $\nu_n(\alpha,z,\mathbf{0}) = \sqrt{n}\,(F_n - \overline{F}_n)(\alpha,z)$ is a sequential empirical process, or a Kiefer–Müller process, based on independent, but not necessarily identically distributed, random variables. This process is asymptotically stochastically equicontinuous, uniformly in α with respect to $\rho_n(v,w) = \left|\overline{F}_n(1,v) - \overline{F}_{1,n}(1,w)\right|$. That is, for every $\eta > 0$, there exists an $\epsilon > 0$ with

$$\lim_{n\to\infty}\sup P\left[\sup_{\alpha\in[0,1],\,\rho_n(z_1,z_2)\le\epsilon} |\nu_n(\alpha,z_1,\mathbf{0}) - \nu_n(\alpha,z_2,\mathbf{0})| > \eta\right] = 0. \tag{63}$$

In fact, with $\overline{\rho}_n\,(\,(\alpha_1,z_1),\,(\alpha_2,z_2)\,) = |\alpha_1 - \alpha_2| + \rho_n(z_1,z_2)$, we have

$$\begin{aligned}\sup_{\alpha\in[0,1]}\ \sup_{z_1,z_2,\in\mathbb{R},\,\rho_n(z_1,z_2)\le\epsilon} & |\nu_n(\alpha,z_1,\mathbf{0}) - \nu_n(\alpha,z_2,\mathbf{0})|\\ & \le \sup_{\substack{\alpha_1,\alpha_2\in[0,1],\,z_1,z_2\in\mathbb{R}\\ \overline{\rho}_n(\,(\alpha,z_1),\,(\alpha,z_2)\,)\le\epsilon}} |\nu_n(\alpha_1,z_1,\mathbf{0}) - \nu_n(\alpha_2,z_2,\mathbf{0})|.\end{aligned}$$

Thus, (63) follows from asymptotic stochastic $\overline{d}_n$-equicontinuity of $\nu_n(\alpha,z,\mathbf{0})$. This in turn follows from a proof similar to, but simpler than, the proof of Lemma 1. In fact, it can be seen from (35) that, for $\mathbf{g}_1 = \mathbf{g}_2 = \mathbf{0}$, we simply can use the metric $\overline{\rho}\,(\,(\alpha_1,z_1),\,(\alpha_2,z_2)\,) = |\alpha_1 - \alpha_2| + \rho_n(z_1,z_2)$ in the estimation of the quadratic variation, which in the simple case of $\mathbf{g}_1 = \mathbf{g}_2 = \mathbf{0}$ amounts to the estimation of the variance, because the randomness only comes in through the ϵ_t. With this modification, the proof of the $\overline{\rho}_n$-equicontinuity of $\nu_n(\alpha,z_1,\mathbf{0})$ follows the proof of Lemma 1.

Thus, if $\widehat{q}_\gamma$ is consistent for q_γ with respect to $\overline{\rho}_n$, then it follows that both (60) and (61) are $o_P(1)$. The proof of the consistency is omitted here. (It is available from the authors.) This completes the proof. □

DOI: 10.1111/jtsa.12200

REFERENCES

Akritas MG, van Keilegom I. 2001. Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics* **28**: 549–567.
Alexander KS, Pyke R. 1986. A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Annals of Probability* **14**: 582–597.
Boldin MV. 1982. Estimation of the distribution of noise in an autoregressive scheme. *Theory of Probability and Its Applications* **27**: 866–871.
Chandler G, Polonik W. 2006. Discrimination of locally stationary time series based on the excess mass functional. *Journal of the American Statistical Association* **101**: 240–253.
Chandler G, Polonik W. 2012. Mode Identification of volatility in time-varying autoregression. *Journal of the American Statistical Association* **107**: 1217–1229.
Cheng F. 2005. Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *Journal of Statistical Planning and Inference* **128**: 327–349.
Dahlhaus R. 1997. Fitting time series models to nonstationary processes. *Annals of Statistics* **25**: 1–37.
Dahlhaus R, Neumann M, Sachs VR. 1999. Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* **5**: 873–906.
Dahlhaus R, Neumann M. 2001. Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic Processes and their Applications* **91**: 277–309.
Dahlhaus R, Polonik W. 2005. Nonparametric maximum likelihood estimation and empirical spectral processes for Gaussian locally stationary processes. *Technical Report*.
Dahlhaus R, Polonik W. 2006. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Annals of Statistics* **34**: 2790–2824.
Dahlhaus R, Polonik W. 2009. Empirical spectral processes for locally stationary time series. *Bernoulli* **15**: 1–39.
Drees H, Stărică C. 2002. A simple non-stationary model for stock returns. Preprint.
Eom KB. 1999. Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models. *Multidimensional Systems and Signal Processing* **10**: 357–378.
Freedman DA. 1975. On tail probabilities for martingales. *Annals of Probability* **3**: 100–118.
Fryzlewicz P, Sapatinas T, Subba Rao S. 2006. A Haar–Fisz technique for locally stationary volatility estimation. *Biometrika* **93**: 687–704.
Girault J-M, Ossant F, Ouahabi A, Kouamé D, Patat F. 1998. Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* **45**: 650–659.
Grenier Y. 1983. Time-dependent ARMA modeling of nonstationary signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **31**: 899–911.
Hall MG, Oppenheim AV, Willsky AS. 1983. Time varying parametric modeling of speech. *Signal Processing* **5**: 267–285.
Horváth L, Kokoszka P, Teyessiére G. 2001. Empirical process of the squared residuals of an ARCH sequence. *Annals of Statistics* **29**: 445–469.
Koul HL. 2002. *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edn., Lecture Notes in Statistics, vol. 166. New York: Springer.
Koul HL, Ling S. 2006. Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* **34**: 994–1012.
Künsch HR. 1995. A note on causal solutions for locally stationary AR processes. Preprint. ETH Zürich.
Laïb N, Lemdani M, Olud-Saïd E. 2008. On residual empirical processes of GARCH-SM models: application to conditional symmetry tests. *Journal of Time Series Analysis* **29**: 762–782.
Müller U, Schick A, Wefelmeyer W. 2007. Estimating the error distribution function in semiparametric regression. *Statistics & Decisions* **25**: 1–18.
Müller U, Schick A, Wefelmeyer W. 2009a. Estimating the innovation distribution in nonparametric autoregression. *Probability Theory and Related Fields* **144**: 53–77.
Müller U, Schick A, Wefelmeyer W. 2009b. Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statistics & Probability Letters* **79**: 957–964.
Müller U, Schick A, Wefelmeyer W. 2012. Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference* **142**: 552–566.
Neumeyer N, van Keilegom I. 2010. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis* **101**: 1067–1078.
Nickl R, Pötscher BM. 2007. Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *Journal of Theoretical Probability* **20**: 177–199.

DOI: 10.1111/jtsa.12200

Orbe S, Ferreira E, Rodrigues-Poo J. 2005. Nonparametric estimation of time varying parameters under shape restrictions. *Journal Econometrics* **126**: 53–77.
Paparoditis E. 2009. Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15**: 1190–1221.
Paparoditis E. 2010. Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association* **105**: 839–851.
Puchstein R, Preuß P. 2016. Testing for stationarity in multivariate locally stationary processes. *Journal of Time Series Analysis* **37**: 3–29.
Priestley MB. 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **27**: 204–237.
Preuß P, Vetter M, Dette H. 2013. A test for stationarity based on empirical processes. *Bernoulli* **19**: 2715–2749.
Rajan JJ, Rayner PJ. 1996. Generalized feature extraction for time-varying autoregressive models. *IEEE Transactions on Signal Processing* **44**: 2498–2507.
Schick A, Wefelmeyer W. 2002. Estimating the innovation distribution in nonlinear autoregressive models. *Annals of the Institute of Statistical Mathematics* **54**: 245–260.
Stute W. 2001. Residual analysis for ARCH(p)-time series. *Test* **1**: 393–403.
Subba Rao T. 1970. The fitting of nonstationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society. Series B (Methodological)* **32**: 312–322.
van de Geer S. 2000. *Empirical Processes in M-Estimation.* Cambridge, UK: Cambridge University Press.
van der Vaart AW, Wellner J. 1996. *Weak Convergence and Empirical Processes*. New York: Springer.
van Keilegom I, González Manteiga W, Sánchez Sellero C. 2008. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17**: 401–415.
von Sachs M, Neumann MH. 2000. A wavelet-based test for stationarity. *Journal of Time Series Analysis* **21**: 597–613.
Wefelmeyer W. 1994. An efficient estimator for the expectation of a bounded function under the residual distribution of an autoregressive process. *Annals of the Institute of Statistical Mathematics* **46**: 300–315.

DOI: 10.1111/jtsa.12200