Weighted Sums and Residual Empirical Processes for Time-varying Processes

[1]Gabriel Chandler* and [2]Wolfgang Polonik
[1]Department of Mathematics, Pomona College, Claremont, CA 91711
[2]Department of Statistics, University of California, Davis, CA 95616

E-mail: gabriel.chandler@pomona.edu and polonik@wald.ucdavis.edu

December 15, 2012

Abstract

In the context of a time-varying AR-process we study both function indexed weighted sums, and sequential residual empirical processes. As for the latter it turns out that somewhat surprisingly, under appropriate assumptions, the non-parametric estimation of the parameter functions has a negligible asymptotic effect on the estimation of the error distribution. Function indexed weighted sum processes are generalized partial sums. An exponential inequality for such weighted sums of time-varying processes provides the basis for a weak convergence result of weighted sum processes. Properties of weighted sum processes are needed to treat the residual empirical processes. As an application of our results we discuss testing for the shape of the variance function.

Keywords: Cumulants, empirical process theory, exponential inequality, locally stationary processes, nonstationary processes, investigating unimodality.

1 Introduction

Consider the following time-varying AR-model of the form

$$Y_t - \sum_{k=1}^{p} \theta_k\Big(\frac{t}{n}\Big)\, Y_{t-k} = \sigma\Big(\frac{t}{n}\Big)\, \epsilon_t, \quad t = -p+1, \ldots, 0, 1, \ldots, n, \tag{1}$$

where θ_k are the autoregressive parameter functions, p is the order of the model, σ is a function controlling the volatility, and $\epsilon_t \sim (0,1)$ i.i.d. 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, i.e. 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), Rjan and Rayner (1996), Girault et al. (1998), Eom (1999), Drees and Stărică (2002), Orbe et al. (2005), Fryzlewicz et al. (2006), and Chandler and Polonik (2006).

Dahlhaus (1997) advanced the formal analysis of time-varying processes by introducing the notion of a locally stationary process. This is a time-varying processes with time being rescaled to $[0,1]$ that satisfies certain regularity assumptions - see (14 - 16) below. We would like to point out, however, that in this paper local stationarity is only used to calculate the asymptotic covariance function in Theorem 3. All the other results hold under weaker assumptions.

Let $\mathbf{Y}_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})'$. Given an estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ and corresponding residuals $\widehat{\eta}_t = Y_t - \widehat{\boldsymbol{\theta}}\big(\frac{t}{n}\big)'\mathbf{Y}_{t-1}$, we estimate the distribution of the innovations $\eta_t = \sigma\big(\frac{t}{n}\big)\epsilon_t$ by the distribution function of the residuals. Define the sequential empirical distribution function of the residuals as

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

Let $F_n(\alpha, z) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\eta_t \le z\}$, denote the sequential empirical distribution function of the innovations η_t. A somewhat surprising result that will be shown here is that under appropriate conditions we have (see Theorem 2)

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

even though non-parametric estimation of the parameter functions θ_k is involved. See also (12) where the $\widehat{\eta}_t$ are replaced by $\widehat{\epsilon}_t$.

One can see that 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. The same applies to the estimation of the distribution function of the errors ϵ_t (cf. (12) in section 2.1). This effect is similar to what is known in the literature from the parametric case, i.e. for a standard AR(p)-process (see Koul 2002). In our case, this phenomenon is caused by the particular structure underlying the time varying processes considered. Other empirical processes based on residuals studied in the literature that also involve nonparametric estimation do not allow for such a behavior. See, for instance, Akritas and van Keilegom (2001), Cheng (2005), Müller, Schick and Wefelmayer (2007, 2009a, 2009b, 2012) and van Keilegom et al. (2008).

The fact that in this work we emphasize processes based on the $\widehat{\eta}_t$ rather than on $\widehat{\epsilon}_t$ is triggered by the applications of our results we have in mind. However, the incorporation of the estimation of σ is more or less straightforward. See section 2.1 for more details.

The proof of (2) (and also the one of (12)) involves the behavior of two types of stochastic processes, the large sample behavior of which based on observations from a non-stationary process satisfying (1) is investigated in this paper. One of these processes is the residual sequential empirical process, closely related to the sequential empirical process based on the residuals, and the other is a generalized partial sum processes or weighted sums processes. Both of them are of independent interest.

Residual sequential empirical process are defined as follows. First observe that we can write $\widehat{F}_n(\alpha,z) = \frac{1}{n}\sum_{t=1}^{\lfloor\alpha n\rfloor} \mathbf{1}\{\sigma(\frac{t}{n})\,\epsilon_t \le (\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})'(\frac{t}{n})\,\mathbf{Y}_{t-1}+z\}$. This motivates the consideration of the process

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

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 $\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} \in \mathcal{G}^p$ (see below), and $F(z)$ denotes the distribution function of the errors ϵ_t. Following Koul (2002), we call the process (3) a *residual sequential empirical process* even though residuals are not directly involved anymore. Observe that with $g = \mathbf{0} = (0, \ldots, 0)'$ denoting the p-vector of null-functions, $\nu_n(\alpha, z, \mathbf{0})$ equals the sequential empirical process based on the innovations η_t. 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 above. Our contribution here is to study these processes based on a non-stationary Y_t of the form (1).

The second type of key processes in this paper are weighted sums of the form

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

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 below.

Motivation. Properties of both processes, $\nu_n(\alpha, \mathbf{g}, z)$ and $Z_n(h)$, are used to derive conditions for (2) to hold. Moreover, properties of both processes are also used to derive the large sample behavior of a test statistic for the shape of the variance function in model (1) that is proposed in an accompanying paper, Chandler and Polonik (2012). This test statistic is based on the sequential empirical process based on the *squared* residuals - see definition of $\widehat{G}_{n,\gamma}(\alpha)$ given in (21) below. For more details, see Chapter 4.

The outline of the paper is as follows. In sections 2 and 3 we analyze the large sample behavior of the function indexed residual empirical process and weighted sums, respectively, under the time varying model (1), and apply them to show (2). Section 4 applies some of the results from the preceding sections to derive an approximation result for $\widehat{G}_{n,\gamma}(\alpha)$. This approximation provides a crucial ingredient to Chandler and Polonik (2012) where the asymptotic distribution free limit of a test statistics for testing modality of the variance function in model (1) is derived. Proofs are deferred to section 5.

Remark on measurability. Suprema of function indexed processes will enter the theoretical results below. We assume throughout the paper that such suprema are measurable.

2 Residual empirical processes under time varying AR-models

Work on empirical processes based on residuals in models involving non-parametric components has already been mentioned above. There of course exists a substantial body of work on residual empirical processes indexed by a finite dimensional parameter. For the time series setting see, for instance, Horváth et al. (2001), Stute (2001), Koul (2002), Koul and Ling (2006), Laïb et al. (2008), Müller et al. (2009a), and references therein. Here we are considering function indexed residual empirical processes based on triangular arrays of non-stationary time series of the form (1).

In order to formulate one of our main results for the residual empirical process $\nu_n(\alpha, z, \mathbf{g})$ defined in (3) above, we first introduce some notation 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 less than or equal to δ that are needed to cover $\mathcal{H}$, i.e., there exists functions g_k, $k = 1, \ldots, N$ such that the balls $\mathcal{A}_k = \{h \in \mathcal{H} : d(g_k, h) \leq \delta\}$ with $\mathcal{H} \subset \bigcup_{k=1}^N \mathcal{A}_k$. 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 $\mathcal{B}_k = \{h \in \mathcal{H} : \underline{g}_k \leq h \leq \overline{g}_k\}$ for pairs of functions $\underline{g}_k \leq \overline{g}_k$, $k = 1, \ldots, N$ with $d(\overline{g}_k, \underline{g}_k) \leq \delta$, 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

$$\|h\|_\infty := \sup_{u \in [0,1]} |h(u)| \quad \text{and} \quad \|h\|_n^2 := \frac{1}{n} \sum_{t=1}^n h^2\big(\tfrac{t}{n}\big).$$

We further denote by $N_n(\delta, \mathcal{H})$ and $N_{n,B}(\delta, \mathcal{H})$ the metric entropy and the metric entropy with bracketing, respectively, with respect to the metric defined by $\|\cdot\|_n$.

Assumptions. (i) The process $Y_t = Y_{t,T}$ has an MA-type representation

$$Y_{t,T} = \sum_{j=0}^\infty a_{t,T}(j) \epsilon_{t-j}, \tag{4}$$

where $\epsilon_t \sim_{i.i.d.} (0,1)$. The distribution 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 $a_{t,T}(\cdot)$ of the MA-type representation of $Y_{t,T}$ given in (i) satisfy

$$\sup_{1 \le t \le T} |a_{t,T}(j)| \le \frac{K}{\ell(j)}$$

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

Assumptions (i) and (ii) have been used in the literature on locally stationary processes before (e.g. Dahlhaus and Polonik, 2006, 2009). It is shown in Dahlhaus and Polonik (2005) by using a proof similar to Künsch (1995) that (i) holds for time-varying AR-processes (1) if the zeros of the corresponding AR-polynomials are bounded away from the unit circle (uniformly in the rescaled time u) and the parameter functions are of bounded variation.

Theorem 1 *Suppose that assumptions (i) and (ii) hold, and $\mathcal{G}$ denotes a function class satisfying for some $c > 0$ that*

$$\int_{c/n}^{1} \sqrt{\log N_{n,B}(u^2, \mathcal{G})}\, du < \infty \qquad \forall\, n. \tag{5}$$

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

$$\sup_{\substack{\alpha \in [0,1], z \in [-L,L],\, \mathbf{g} \in \mathcal{G}^p \\ \|g\|_n \le \delta_n}} |\nu_n(\alpha, z, \mathbf{g}) - \nu_n(\alpha, z, \mathbf{0})| = o_P(1). \tag{6}$$

This theorem is of typical nature for work on residual empirical processes (e.g. (8.2.32) in Koul 2002), although here we are dealing with time-varying AR-processes, and are considering a non-parametric index class of functions. Also keep in mind that we are considering triangular arrays (recall that $Y_t = Y_{t,T}$).

Notice that the above result is considering the difference of two processes where the 'right' centering 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. It will turn out that somewhat surprisingly, the difference of these two functions is $o_P(1/\sqrt{n})$, so that the non-parametric estimation of the parameter functions has an negligible asymptotic effect on the estimation of the (average) innovation distribution. More assumptions are needed.

Assumptions.

(iii) There exists a class $\mathcal{G}$ with

$$\theta_k(\cdot) - \widehat{\theta}_k(\cdot) \in \mathcal{G}, \quad k = 1, \ldots, p, \quad \text{with probability tending to 1 as } n \to \infty.$$

such that $\sup_{g \in \mathcal{G}} \|g\|_\infty < \infty$ and for some $C, c > 0$

$$\int_{c/n}^{1} \log N_n(u, \mathcal{G})\, du < C < \infty \quad \forall\, n.$$

(iv) For $k = 1, \ldots, p$ we have $\frac{1}{n} \sum_{t=1}^{n} \big(\widehat{\theta}_k(\frac{t}{n}) - \theta_k(\frac{t}{n})\big)^2 = O_P(m_n^{-1})$ with $m_n \to \infty$ as $n \to \infty$.

Assumptions (iii) and (iv) are discussed below.

Theorem **2** *Suppose that the assumptions of Theorem 1 hold. If, in addition, assumptions (iii) and (iv) are satisfied, where m_n satisfies $\frac{n^{1/2}}{m_n^2 \log n} \to \infty$ as $n \to \infty$, then for any $L > 0$*

$$\sup_{\alpha \in [0,1], z \in [-L, L]} |\widehat{F}_n(\alpha, z) - F_n(\alpha, z)| = o_P(1/\sqrt{n}). \tag{7}$$

Discussion of assumptions (iii) and (iv). The assumption on the entropy integral (see assumptions (iii) and (5)) control the complexity of the class $\mathcal{G}$. Many classes $\mathcal{G}$ are know to satisfy these assumptions - see below for one example (for more examples we refer to the literature on empirical process theory). Notice that a more standard condition on the covering integral is $\int_{c/n}^{1} \sqrt{\log N_n(u, \mathcal{G})}\, du < \infty$ (or similarly with bracketing). In contrast to that, the entropy integral in assumption (iii) does not have a square root. This is similar to condition (5) where the integral is over $\sqrt{\log N_{n,B}(u^2, \mathcal{G})}$ (notice the u^2). This makes both our entropy conditions stronger than the standard assumption. The reason for this is that the exponential inequality that is underlying the derivations of our results is not of sub-Gaussian type (see Lemma 3), which in turn is caused by the dependence structure of our underlying process.

A class of non-parametric estimators satisfying conditions (iii) and (iv) 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 $\big(\frac{\log n}{n}\big)^{s/(2s+1)}$ in the L_2-norm. For $s \geq 1$ the functions in $B^s_{p,q}(C)$ have uniformly bounded total variation. Assuming that the model parameter functions also posses this property, the rate of convergence in the $\|\cdot\|_n$-norm is the same as the one of the L_2-norm, 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 $m_n^{-1} = \big(\frac{\log n}{n}\big)^{s/(2s+1)}$. In order to verify the condition on the bracketing covering numbers from (iii), we use Nickl and Pötscher (2007). Their Corollary 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.)

The proof of Theorem 1 rests on the following lemma which is of independent interest. It is modeled after similar results for empirical processes (see van de Geer, 2000, Theorem 5.11). Let $\mathcal{H}_L = [0,1] \times [-L, L] \times \mathcal{G}^p$ denote the index space of the process ν_n where $L > 0$ is some constant. Define a metric on $\mathcal{H}_L$ as

$$d_n(h_1, h_2) = d_n\big((\alpha_1, z_1, \mathbf{g}_1),\, (\alpha_2, z_2, \mathbf{g}_2)\big) = |\alpha_1 - \alpha_2| + |z_1 - z_2| + \sum_{k=1}^{p} \|g_{1,k} - g_{2,k}\|_n. \quad (8)$$

Lemma 1 *For $C_0 > 0$ let $\mathbf{A}_n = \big\{ \frac{1}{n} \sum_{s=-p+1}^{n} Y_s^2 \leq C_0^2 \big\}$ and define $K^* = 1 + (1 + p\,C_0)\frac{\|f\|_\infty}{m_*}$. Suppose that $\mathcal{H}_L$ is totally bounded with respect to d_n, that assumptions (i) and (ii) hold, and that for $C_1 > 0$,*

$$\eta \geq \frac{2^6\,K^*}{\sqrt{n}}, \quad (9)$$

$$\eta \;\leq\; \frac{1}{2}\,K^* \sqrt{n}\,(\tau^2 \wedge \tau), \quad (10)$$

$$\eta \;\geq\; C_1 \left(\int_{\eta/2^8\,K^* \sqrt{n}}^{\tau} \sqrt{\log N_B(u^2, \mathcal{H}_L, d_n)}\; du \;\vee\; \tau \right). \quad (11)$$

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

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

2.1 Estimating the distribution function of the errors ϵ_t

As has been indicated in the introduction, the fact that we present our results for the case of $\sigma(\cdot)$ not being estimated is related to the applications we have in mind (see Section 4). However, the case of also estimating the variance function, and thus estimating the distribution of ϵ_t rather than the average distribution of $\eta_t = \sigma(\frac{t}{n})\epsilon_t$, can be treated by using very similar techniques. This is discussed here briefly.

Assume we have available an estimator $\widehat{\sigma}(\cdot)$ of the variance function. Then the residuals become $\widehat{\epsilon}_t = \frac{Y_t - \widehat{\boldsymbol{\theta}}(\frac{t}{n})'\mathbf{Y}_{t-1}}{\widehat{\sigma}(\frac{t}{n})}$, and the sequential empirical distribution function of $\widehat{\epsilon}_t$ can be written as

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

Accordingly, we define the modified residual sequential empirical process as

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

where the function s lies in a class $\mathcal{S}$ such that the difference $\widehat{\sigma} - \sigma \in \mathcal{S}$ (with probability tending to 1 as $n \to \infty$). An adapted version of Theorem 1 holds for $\nu_n^*(\alpha, z, \mathbf{g}, s) - \nu_n^*(\alpha, z, \mathbf{0}, 0)$ where the adaptation consists in replacing $\mathcal{G}$ by $\mathcal{G} \cup \mathcal{S}$ in the entropy integral, and the supremum is also taken over all $s \in \mathcal{S}$ with $\|s\|_n \le \delta_n \to 0$ as $n \to \infty$.

Assuming further that conditions on $\mathcal{S}$ and the difference $\widehat{\sigma} - \sigma \in \mathcal{S}$ follow exactly the ones given in assumptions (iii) and (iv) above, then under the same assumption on the rates of convergence of our estimators, the non-parametric estimation of the AR-parameter functions and the variance function has a negligible asymptotic effect on the estimation of the distribution function of the errors ϵ_t, i.e.

$$\sup_{\alpha \in [0,1], z \in [-L, L]} \Big| \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\widehat{\epsilon}_t \le z\} - \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\{\epsilon_t \le z\}\Big| = o_P(1/\sqrt{n}). \tag{12}$$

The proofs of these results require only some straightforward adaptations, beginning with a modification of d_n (given in (8)) to a metric on $\mathcal{H}_L^* = [0,1] \times [-L, L] \times \mathcal{G}^p \times \mathcal{S}$ by adding

$\|s_1 - s_2\|_n$. With this modification Lemma 1 still holds for $\mathcal{H}_L^*$. This then is the main ingredient to the proof of Theorems 1 and 2 for processes based on the $\widehat{\epsilon}_t$ rather than on the $\widehat{\epsilon}_t$.

Just as for the estimators of the AR-parameters, the wavelet based estimator of the variance function of Dahlhaus et al. (1999) satisfies the necessary assumptions on $\widehat{\sigma}$.

3 Weighted sums under local stationarity

As discussed above, the second type of processes of importance in our context are weighted partial sums of locally stationary processes given by

$$Z_n(h) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h\big(\tfrac{t}{T}\big)\, Y_s, \quad h \in \mathcal{H}. \tag{13}$$

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), van de Geer (2000), and references therein.

We will show below 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), this means that we assume the existence of functions $a(\cdot, j) : (0,1] \to \mathbb{R}$ with

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

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

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

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 circle (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 k-th order cumulant of a random variable X.

Theorem **3** *Let $\mathcal{H}$ denote a class of uniformly bounded, real valued functions of bounded variation defined on $[0,1]$. Assume further that for some $C, c > 0$,*

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

and that for some constant $C_1 > 0$ we have $\mathrm{cum}_k(\epsilon_t) \leq k!C_1^k$ for $k = 1, 2, \ldots$ Then we have under assumptions (i) and (ii) 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-Jorgensen - see van der Vaart and Wellner (1996) for more details.

b) It is well-known that for normally distributed errors we have $\mathrm{cum}_k(\epsilon_t) = 0$ for $k = 3, 4, \ldots$, and thus they satisfy the assumptions of the theorem.

c) Weighted partial sums of the form

$$Z_n(\alpha,h) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor \alpha n \rfloor} g\big(\tfrac{t}{n}\big)Y_s$$

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 above 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 $C(h_{g_1,\alpha_1},\, h_{g_2,\alpha_2}) = \int_0^{\alpha_1 \wedge \alpha_2} g_1(u)\, g_2(u)\, S(u)\, du$.

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

The main ingredients to the proof of Theorem 3 are presented in the following results. These results are of independent interest.

Theorem **4** *Let $\{Y_t,\, t = 1, \dots, n\}$ satisfy assumptions (i) and (ii) and let $\mathcal{H} = \{h : [0,1] \to \mathbb{R}\}$ be totally bounded with respect to $\|\cdot\|_n$. Further assume that there exists a constant $C > 0$ such that for all $k = 1, 2, \dots,$ we have $\mathrm{cum}_k \le k! C^k$, and let $\mathbf{A}_n = \left\{ \frac{1}{n} \sum_{t=1}^n Y_t^2 \le M^2 \right\}$ where $M > 0$. There exists constants $c_0, c_1, c_2 > 0$ such that for all $\eta > 0$ satisfying*

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

and

$$\eta > c_0 \Big(\int_{\frac{\eta}{8\, M\, \sqrt{n}}}^{\tau} \log N_n(u, \mathcal{H})\, d\, u \,\vee\, \tau \Big) \tag{19}$$

we have

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

The next result of importance for the proof of Theorem 3 deals with cumulants. For random variables $X_1, \dots, X_k$ we denote by $\mathrm{cum}(X_1, \dots, X_k)$ their joint cumulant, and if $X_i = X$ for all $i = 1, \dots, k$, then $\mathrm{cum}(X_1, \dots, X_k) = \mathrm{cum}(X, \dots, X) = \mathrm{cum}_k(X)$, the k-th order cumulant of X.

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 $k \ge 1$,*

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

If, in addition, $\|h_j\|_\infty \leq M < \infty$, $j = 1, \ldots, k$, *then for* $k \geq 3$,

$$\big|\,\mathrm{cum}(Z_n(h_1), \ldots, Z_n(h_k))\,\big| \leq (K_0)^{k-2}\, M^k \,\big|\,\mathrm{cum}_k(\epsilon_1)\,\big|\, n^{-\frac{k-2}{2}}.$$

The behavior of the cumulants given in the above lemma is needed for the following crucial exponential inequality.

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

$$P\Big[\,|Z_n(h)| > \eta\,\Big] \leq c_1\, \exp\Big\{-\frac{\eta}{c_2\,\|h\|_n}\Big\}. \tag{20}$$

The assumptions on the cumulants in particular hold for errors with $\mathrm{E}|\epsilon_t|^k \leq \left(\frac{C}{2}\right)^k$ *for all* $k = 1, 2, \ldots$

4 Applications

We present two applications that motivate the above study of the two types of processes. The first is a test for the constancy of the variance function, i.e. homoskedasticity. The second, closely related though more involved than the first, is a test for determining the modality of the variance function.

4.1 A test of homoskedasticity

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 weak stationarity.) To this end, consider the process

$$\widehat{G}_{n,\gamma}(\alpha) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}(\widehat{\eta}_t^2 \geq \widehat{q}_\gamma^2), \quad \alpha \in [0,1], \tag{21}$$

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

$Y_1, \ldots, Y_n$. If the variance function is constant, then, since one has 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, $T_n = \sup_{\alpha\in[0,1]} \left|\widehat{G}_{n,\gamma}(\alpha) - \alpha\gamma\right|$. The behavior of this test statistic follows the discussion of our second application.

4.2 Asymptotic distribution of a test statistic for modality of the variance function

We argue briefly that the test for constancy of variance from subsection 4.1 can be modified to a test for modality of the variance function. For more details, please refer to Chandler and Polonik (2012). A key realization is that for a function to be monotonic (which, say, a unimodal function is on either side of the mode), the variance function must be 'at least' constant. In other words, a constant variance function provides the boundary case under the null for a test of monotonicity. We conduct this test on appropriate intervals $[a, b] \subset [0, 1]$, where this crucial choice depends on the underlying variance function. Under the alternative hypothesis the interval needs to be located around a deviation from the null-hypothesis, meaning around an antimode. Chandler and Polonik (2012) provide an estimator for the interval of interest and show that this estimation does not influence the asymptotic limit of the test statistic. Via this estimation, and supposing we are testing for a monotonically increasing function (locally on $[a, b]$), it is appropriate to modify the test statistic to $T_n = \sup_{\alpha\in[a,b]} \left(\widehat{G}_{n,\gamma}(\alpha) - \alpha\gamma\right)$.

To simplify the notation we assume in the following that $[a, b] = [0, 1]$. 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 5, weighted empirical processes enter the analysis of $\widehat{G}_{n,\gamma}(\alpha)$ through handling the estimation of q_γ. Theorem 5 below 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). In particular, it implies that the large sample behavior 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. This is somewhat a surprise, and it is connected to the particular structure of our model.

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

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

where q_γ is defined by means of

$$\Psi(z) = \int_0^{\mathbf{1}} F\big(\tfrac{z}{\sigma(u)}\big)\, du - \int_0^1 F\big(\tfrac{-z}{\sigma(u)}\big)\, du, \quad z \geq 0, \tag{22}$$

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 **5** *Let* $\gamma \in [0, 1]$ *and suppose that* $0 \leq a < b \leq 1$ *are non-random. Assume further that* $\mathrm{cum}_k|\epsilon_t| < k!C^k$ *for some* $C > 0$ *and all* $k = 1, 2, \ldots$ *Then, under assumptions (i) - (iv), with* $\frac{n^{1/2}}{m_n^2 \log n} = o(1)$ *we have*

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

where

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

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) are 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 (worst case) null-hypothesis, $\sigma^2(u) = \sigma_0^2$ for all $u \in [0, 1]$, the innovations are i.i.d., we have $c(\alpha) = \alpha$, and $\mathrm{E}G_{n,\gamma}(\alpha) = \alpha\gamma$. Therefore the above result implies that $(\gamma(1-\gamma))^{-1/2}\sqrt{n}(\widehat{G}_{n,\gamma}(\alpha) - \alpha\gamma)$ converges weakly to a standard Brownian Bridge. Further details can be found in Chandler and Polonik (2012).

5 Proofs

5.1 Proof of Lemma 1

We only present a brief outline. First notice that $\nu_n(\alpha, z, \mathbf{g})$ is a sum of bounded martingale differences, because with $\xi_t^{z,\mathbf{g}} = \mathbf{1}\big(\sigma(\frac{t}{n})\,\epsilon_t \le \mathbf{g}'(\frac{t}{n})\mathbf{Y}_{t-1} + z\big)$ we have

$$\nu_n(\alpha, z, \mathbf{g}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{\xi}_t^{z,\mathbf{g}} \mathbf{1}\big(\tfrac{t}{n} \le \alpha\big)$$

where $\widetilde{\xi}_t^{z,\mathbf{g}} = \xi_t^{z,\mathbf{g}} - \mathrm{E}(\xi_t^{z,\mathbf{g}}|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ denotes the σ-algebra generated by $\{\epsilon_t, \epsilon_{t-1}, \ldots\}$, and obviously also $\nu_n(\alpha_1, z_1, \mathbf{g}_1) - \nu_n(\alpha_2, z_2, \mathbf{g}_2)$ are sums of martingale differences. The proof of this 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 $|Z_t| \le C$ for all $t = 1, \ldots, n$. Let further $S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t$ and $V_n = V_n(S_n) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}(Z_t^2|\, \mathcal{F}_{t-1})$. Then we have for all $\epsilon, \tau^2 > 0$ that*

$$P\big(\, S_n \ge \epsilon,\, V_n \le \tau^2\,\big) \le \exp\left(-\frac{\epsilon^2}{2\tau^2 + \frac{2\epsilon\, C}{\sqrt{n}}}\right). \tag{25}$$

The form of (25) motivates that we first need to control the quadratic variation V_n. Let

$$\eta_t^{\alpha,z,\mathbf{g}} = \widetilde{\xi}_t^{z,\mathbf{g}}\, \mathbf{1}\big(\tfrac{t}{n} \le \alpha\big).$$

We have for $h_1 = (\alpha_1, z_1, \mathbf{g}_1), h_2 = (\alpha_2, z_2, \mathbf{g}_2) \in \mathcal{H}$ with $d\big(h_1, h_2\big) \le \epsilon$ that

$$\begin{aligned} V_n =& V_n(\nu_n(h_1) - \nu_n(h_2)) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\big(|\eta_t^{\alpha_1,z_1,\mathbf{g}_1} - \eta_t^{\alpha_2,z_2,\mathbf{g}_2}|\big|\, \mathcal{F}_{t-1}\big) \\ \le& \frac{1}{n} \sum_{t=1}^{n} \big|\, 1\big(\tfrac{t}{n} \le \alpha_1\,\big) - 1\big(\tfrac{t}{n} \le \alpha_2\,\big)\,\big|\, \mathrm{E}(|\xi_t^{z_1,\mathbf{g}_1}|\big|\mathcal{F}_{t-1}) + \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}(|\xi_t^{z_1,\mathbf{g}_1} - \xi_t^{z_2,\mathbf{g}_2}|\big|\mathcal{F}_{t-1}) \\ \le& \frac{1}{n}\big|n\,(\alpha_2 - \alpha_1) + 1\big| + \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}(|\xi_t^{z_1,\mathbf{g}_1} - \xi_t^{z_2,\mathbf{g}_2}|\big|\mathcal{F}_{t-1}) \end{aligned}$$

$$\leq |\alpha_1 - \alpha_2| + \frac{1}{n} + \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}(|\xi_t^{z_1,\mathbf{g}_1} - \xi_t^{z_2,\mathbf{g}_2}|\big|\mathcal{F}_{t-1}).$$

On $\mathbf{A}_n$ we have for the last sum that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}(|\xi_t^{z_1,\mathbf{g}_1} - \xi_t^{z_2,\mathbf{g}_2}|\big|\mathcal{F}_{t-1}) \\
\leq\ & \frac{1}{n}\sum_{t=1}^{n} \big|F_{a+\frac{t}{n}}\big(z_1 + (\mathbf{g}_1(\tfrac{t}{n}))'\mathbf{Y}_{t-1}\big) - F_{\frac{t}{n}}\big(z_2 + (\mathbf{g}_1(\tfrac{t}{n}))'\mathbf{Y}_{t-1}\big)\big| \\
&\qquad + \frac{1}{n}\sum_{t=1}^{n} \big|F_{\frac{s}{T}}\big(z_2 + (\mathbf{g}_1(\tfrac{t}{n}))'\mathbf{Y}_{t-1}\big) - F_{\frac{t}{n}}\big(z_2 + (\mathbf{g}_2(\tfrac{t}{n}))'\mathbf{Y}_{t-1}\big) \\
\leq\ & \sup_{u,x} f_u(x)\ |z_1 - z_2| + \sup_{u,x} f_u(x)\ \frac{1}{n}\sum_{t=1}^{n} \big|\ \big((\mathbf{g}_1 - \mathbf{g}_2)(\tfrac{t}{n})\big)'\mathbf{Y}_{t-1}| \\
\leq\ & \frac{\|f\|_\infty}{m_*}\,|z_1 - z_2| + \frac{\|f\|_\infty}{m_*}\ \sum_{k=1}^{p} \|g_{1k} - g_{2k}\|_n \sqrt{\frac{1}{n}\sum_{t=-p}^{n} Y_t^2}\ \leq\ (1 + C_0)\,\frac{\|f\|_\infty}{m_*}\,\epsilon. \qquad (26)
\end{aligned}$$

Thus, on $\mathbf{A}_n$ we have for $h_1 = (\alpha_1, z_1, \mathbf{g}_1), h_2 = (\alpha_2, z_2, \mathbf{g}_2) \in \mathcal{H}_L$ with $d_n\big(h_1, h_2\big) \leq \epsilon$ and $\epsilon \geq \frac{1}{n}$ that

$$V_n(\nu_n(h_1) - \nu_n(h_2)) \leq \frac{1}{n}\sum_{s=1}^{n} \mathrm{E}\big(\big|\overline{\eta}_s^{h_1} - \underline{\eta}_s^{h_2}\,\big|\,\mathcal{F}_{s-1}\big)^2 \leq K^*\,\epsilon. \qquad (27)$$

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

5.2 Proof of Theorem 1

We will utilize Lemma 1. For $\eta > 0$ we have

$$\begin{aligned}
&P\big(\sup_{\alpha\in[0,1],z\in[-L,L],\mathbf{g}\in\mathcal{G}^p} |\nu_n(\alpha, z, \mathbf{g}) - \nu_n(\alpha, z, \mathbf{0})| \geq \eta\big) \\
&\quad \leq P\big(\big(\mathbf{A}_n\big)^c\big) + P\Big(\sup_{\substack{d_n(h_1,h_2)\leq C\,m_n^{-1} \\ h_1,h_2\in\mathcal{H}_L}} |\nu_n(h_1) - \nu_n(h_2)| \geq \eta, \mathbf{A}_n\Big),
\end{aligned}$$

where again $\mathbf{A}_n = \{\frac{1}{n}\sum_{t=1-p}^{n} Y_t^2 \leq C\}$ for some $C > 0$. We will see below that $\frac{1}{n}\sum_{t=-p+1}^{n} Y_t^2 = O_P(1)$ as $n \to \infty$. Therefore, for any given $\epsilon > 0$ we can choose C such that $P\big(\big(\mathbf{A}_n\big)^c\big) \leq \epsilon$ for n large enough. An application of Lemma 1 now gives the assertion. Similar arguments as those leading to (19) show that $\int_{\frac{c}{n}}^1 \sqrt{\log N_{n,B}(u^2,\mathcal{G})}\,du < \infty$ implies $\int_{\frac{c}{n}}^1 \sqrt{\log N_B(u^2,\mathcal{H}_L,d_n)}\,du < \infty$ (see also (46)). It remains to show that in fact $\frac{1}{n}\sum_{t=-p+1}^{n} Y_t^2 = O_P(1)$. To this end we show that

$$\mathrm{E}\Big[\frac{1}{n}\sum_{t=1}^{n} Y_t^2\Big] < \infty, \tag{28}$$

$$\mathrm{Var}\Big[\frac{1}{n}\sum_{t=1}^{n} Y_t^2\Big] = o(1). \tag{29}$$

These two facts can be seen by direct calculations as demonstrated now. First we consider (28). We have

$$\begin{aligned}\mathrm{E}Y_t^2 = \mathrm{E}\Big[\sum_{j=-\infty}^{t} a_{t,n}(t-j)\epsilon_j\Big]^2 &= \mathrm{E}\sum_{j=-\infty}^{t}\sum_{k=-\infty}^{t} a_{t,n}(t-j)a_{t,n}(t-k)\epsilon_j\epsilon_k \\ &= \sum_{j=-\infty}^{t} a_{t,T}^2(t-j) \leq \sum_{j=0}^{\infty}\Big(\frac{K}{\ell(j)}\Big)^2 < C < \infty,\end{aligned} \tag{30}$$

for some $C > 0$, where we were using assumption (ii). This implies (28). Next we indicate (29). Straightforward calculations show that

$$\begin{aligned}&\mathrm{E}(Y_t^2\,Y_s^2) - \mathrm{E}(Y_t^2)\,\mathrm{E}(Y_s^2)\\ =\;& \sum_{i=-\infty}^{t}\sum_{j=-\infty}^{t}\sum_{k=-\infty}^{s}\sum_{\ell=-\infty}^{s} a_{t,T}(t-i)a_{t,T}(t-j)a_{s,T}(s-k)a_{s,T}(s-\ell)\mathrm{E}(\epsilon_i\,\epsilon_j\,\epsilon_k\,\epsilon_\ell)\\ &\qquad - \sum_{i=-\infty}^{t} a_{t,T}^2(t-i)\sum_{k=-\infty}^{s} a_{s,T}^2(s-k)\\ =\;& (\mathrm{E}\epsilon_t^4 - 2)\;A(t,s) + 2B(t,s),\end{aligned}$$

where

$$A(t,s) = \sum_{i=0}^{\infty} a_{t,T}^2(i)a_{s,T}^2(i+|t-s|),$$

$$B(t,s) = \Big(\sum_{i=0}^{\infty} a_{t,T}(i) a_{s,T}(i+|t-s|)\Big)^2.$$

We obtain that for some $C^* < \infty$,

$$\begin{aligned} A(t,s) &\le K^4 \sum_{i=0}^{\infty} \frac{1}{\ell^2(i)} \frac{1}{\ell^2(i+|t-s|)} \le K^4 \sum_{i=0}^{\infty} \frac{1}{\ell(i)} \frac{1}{\ell(i+|t-s|)} \\ &\le K^4 \frac{1}{\ell(|t-s|)} \sum_{i=0}^{\infty} \frac{1}{\ell(i)} < C^* \frac{1}{\ell(|t-s|)}, \end{aligned}$$

and similarly

$$|B(t,s)| \le \Big[K^2 \sum_{i=0}^{\infty} \frac{1}{\ell(i)} \frac{1}{\ell(i+|t-s|)} \Big]^2 \le C^* \frac{1}{\ell(|t-s|)}. \tag{31}$$

Now we have

$$\begin{aligned} \Big| \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} A(t,s) \Big| &\le \frac{C^*}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{1}{\ell(|t-s|)} \le \frac{C^*}{n} \sum_{k=0}^{n-1} \frac{2(n-k)}{n} \frac{1}{\ell(k)} \\ &\le \frac{2\,C^*}{n} \sum_{k=0}^{n-1} \frac{1}{\ell(k)} = O\Big(\frac{1}{n}\Big) \quad \text{as } n \to \infty. \end{aligned}$$

This implis that $\mathrm{Var}\big[\frac{1}{n}\sum_{s=1}^{n} Y_s^2\big] = O(1/n)$ which is (29).

5.3 Proof of Theorem 3

Showing weak convergence of the process $Z_n(h)$ means proving asymptotic tightness and convergence of the finite dimensional distribution (e.g. see van der Vaart and Wellner, 1996). Tightness follows from Theorem 4.

It remains to show 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 $k \ge 3$ 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 convergence of the covariances $\mathrm{cov}(Z_n(h_1), Z_n(h_2))$. The range of the summation indices below 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 $u \le 0$ and $u > 1$, $i = 1, 2$. We have

$$\begin{aligned}
\operatorname{cov}(Z_n(h_1), Z_n(h_2)) &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} h_1\big(\tfrac{t}{n}\big)\cdot h_2\big(\tfrac{s}{n}\big)\operatorname{cov}(Y_s, Y_t)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{k=t-n}^{t-1} h_1\big(\tfrac{t}{n}\big)\cdot h_2\big(\tfrac{t-k}{n}\big)\operatorname{cov}(Y_t, Y_{t-k})\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\le\sqrt{n}} h_1\big(\tfrac{t}{n}\big)\cdot h_2\big(\tfrac{t-k}{n}\big)\operatorname{cov}(Y_t, Y_{t-k}) \quad + \quad R_{1n} \qquad (32)
\end{aligned}$$

where for n sufficiently large

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

From Proposition 5.4 of Dahlhaus and Polonik (2009) we obtain that $\sup_t |\operatorname{cov}(Y_t,\, Y_{t-k})| \le \frac{K}{\ell(k)}$ 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 (32) can be approximated as

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\le\sqrt{n}} h_1\big(\tfrac{t}{T}\big)\cdot h_2\big(\tfrac{t-k}{n}\big)\; c\big(\tfrac{t}{n},\, k\big) + R_{2n} \qquad (33)$$

where

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

Proposition 5.4 of Dahlhaus and Polonik (2009) also says that for $|k| \le \sqrt{n}$ we have $\sum_{t=0}^{n}\big|\operatorname{cov}(Y_t, Y_{t-k}) - c\big(\tfrac{t}{n},\, k\big)\big| \le K\big(1 + \frac{|k|}{n}\big)$ for some $K > 0$. Applying this result we obtain that

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

as $n \to \infty$. Next we replace $h_2(\frac{t-k}{n})$ in the main term of (33) by $h_2(\frac{t}{n})$. The approximation error can be bounded by

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{|k|\le\sqrt{n}}\Big|\,h_1\big(\tfrac{t}{n}\big)\Big|\cdot\Big|\,h_2\big(\tfrac{t-k}{n}\big)-h_2\big(\tfrac{t}{n}\big)\,\Big|\,\tfrac{K}{\ell(|k|)}=o(1).$$

Here we are using the fact that $\sup_u|c(u,k)|\le\frac{K}{\ell(|k|)}$ (see Proposition 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

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

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

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

Finally, we utilize the fact that $TV(c(\cdot,k))\le\frac{K}{\ell(|k|)}$ which is another result from Proposition 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.

5.4 Proof of Lemma 2

By utilizing multilinearity of cumulants we obtain that

$$\operatorname{cum}(Z_n(h_1),\ldots,Z_n(h_k))=n^{-\frac{k}{2}}\sum_{t_1,t_2,\ldots,t_k=1}^{n}h_1\big(\tfrac{t_1}{n}\big)\cdots h_k\big(\tfrac{t_k}{n}\big)\operatorname{cum}(Y_{t_1},\ldots,Y_{t_k}).$$

In order to estimate $\operatorname{cum}(Y_{t_1},\ldots,Y_{t_k})$ we utilize the special structure of the Y_t-variables. Since the ϵ_j are independent we again obtain by again using multilinearity of the cumulants together with the fact that $\operatorname{cum}(\epsilon_{j_1},\ldots,\epsilon_{j_k})=0$ unless all the $j_\ell,\;\ell=1,\ldots,k$ are equal, that

$$\begin{aligned}\operatorname{cum}(Y_{t_1},\ldots,Y_{t_k})&=\sum_{j=0}^{\min\{t_1,\ldots,t_k\}}a_{t_1,n}(t_1-j)\cdots a_{t_k,n}(t_k-j)\operatorname{cum}(\epsilon_j,\ldots,\epsilon_j)\\&=\operatorname{cum}_k(\epsilon_1)\sum_{j=0}^{\min\{t_1,\ldots,t_k\}}a_{t_1,n}(t_1-j)\cdots a_{t_k,n}(t_k-j).\end{aligned}\tag{34}$$

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

$$\big|\mathrm{cum}(Z_n(h_1),\ldots,Z_n(h_k))\big| \le n^{-\frac{k}{2}}\ \big|\mathrm{cum}_k(\epsilon_1)\big| \sum_{j=-\infty}^{\infty}\prod_{i=1}^{k}\Big[\sum_{t_i=0}^{n}\big|h_i\big(\tfrac{t_i}{n}\big)\big|\frac{K}{\ell(|t_i-j|)}\Big]$$

$$= n^{-\frac{k}{2}}\ \big|\mathrm{cum}_k(\epsilon_1)\big| \sum_{j=-\infty}^{\infty}\prod_{i=1}^{2}\Big[\sum_{t_i=0}^{n}\big|h_i\big(\tfrac{t_i}{n}\big)\big|\frac{K}{\ell(|t_i-j|)}\Big]$$

$$\times\prod_{i=3}^{k}\Big[\sum_{t_i=0}^{n}\big|h_i\big(\tfrac{t_i}{n}\big)\big|\frac{K}{\ell(|t_i-j|)}\Big].$$

Utilizing Cauchy-Schwarz inequality we have for the last product

$$\prod_{i=3}^{k}\Big[\sum_{t_i=0}^{n}\big|h_i\big(\tfrac{t_i}{n}\big)\big|\frac{1}{\ell(|t_i-j|)}\Big] \le \prod_{i=3}^{k}\sqrt{\sum_{t_i=0}^{n}h_i\big(\tfrac{t_i}{n}\big)^2}\ \sqrt{\sum_{t_i=0}^{n}\Big(\frac{K}{\ell(|t_i-j|)}\Big)^2}$$

$$\le n^{\frac{k}{2}-1}\prod_{i=3}^{k}\|h_i\|_n\sqrt{\sum_{t=-\infty}^{\infty}\Big(\frac{K}{\ell(|t|)}\Big)^2}$$

$$\le K_0^{k-2}n^{\frac{k}{2}-1}\prod_{i=3}^{k}\|h_i\|_n. \tag{35}$$

where we used the fact that $\sqrt{\sum_{t=-\infty}^{\infty}\Big(\frac{K}{\ell(|t|)}\Big)^2} \le \sum_{t=-\infty}^{\infty}\frac{K}{\ell(|t|)} \le K_0$ for some $K_0<\infty$. Notice that the bound (35) does not depend on the index j anymore, so that

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

As for the last sum, we have by again using the fact that $\sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)}\frac{K}{\ell(|t_2-j|)} \le \frac{K^*}{\ell(|t_1-t_2|)}$ for some $K^*>0$ (cf. (31)) together with the Cauchy-Schwarz inequality that

$$\sum_{j=-\infty}^{\infty}\prod_{i=1}^{2}\Big[\sum_{t_i=0}^{n}\big|h_i\big(\tfrac{t_i}{n}\big)\big|\frac{K}{\ell(|t_i-j|)}\Big]$$

$$= \sum_{t_1=0}^{n}\sum_{t_2=0}^{n}h_1\big(\tfrac{t_1}{n}\big)h_2\big(\tfrac{t_2}{n}\big)\sum_{j=-\infty}^{\infty}\frac{K}{\ell(|t_1-j|)}\frac{K}{\ell(|t_2-j|)}$$

$$
\begin{aligned}
&\le \sum_{t_1=0}^{n}\sum_{t_2=0}^{n} h_1\big(\tfrac{t_1}{n}\big)h_2\big(\tfrac{t_2}{n}\big)\frac{K^*}{\ell(|t_1-t_2|)}\\
&\le K^* \sqrt{\sum_{t_1=0}^{n}\sum_{t_2=0}^{n} h_1\big(\tfrac{t_1}{n}\big)^2 \,\frac{1}{\ell(|t_1-t_2|)}}\;\sqrt{\sum_{t_1=0}^{n}\sum_{t_2=0}^{n} h_1\big(\tfrac{t_2}{n}\big)^2 \,\frac{1}{\ell(|t_1-t_2|)}}\\
&= K^* \sqrt{\sum_{t_1=0}^{n} h_1\big(\tfrac{t_1}{n}\big)^2 \sum_{t_2=0}^{n}\frac{1}{\ell(|t_1-t_2|)}}\;\sqrt{\sum_{t_1=0}^{n} h_1\big(\tfrac{t_2}{n}\big)^2 \sum_{t_2=0}^{n}\frac{1}{\ell(|t_1-t_2|)}}\\
&\le K_0\, n\,\|h_1\|_n\,\|h_2\|_n.
\end{aligned}
$$

This completes the proof of the first part of the lemma. The second part follows similar to the above by observing that if $\|h_i\|_\infty < M$ for all $i = 1,\dots,k$, then, instead of the estimate (35), we have

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

with $K_0 = \sum_{t=-\infty}^{\infty}\frac{1}{\ell(|t|)}$.

5.5 Proof of Lemma 3

Using Lemma 2 the assumptions on the cumulants imply that

$$
\begin{aligned}
\Psi_{Z_n(h)}(t) = \log \mathrm{E}e^{t\,Z_n(h)} &= \sum_{k=0}^{\infty}\frac{t^k}{k!}\,\mathrm{cum}_k(Z_n(h))\\
&\le \frac{1}{K_0}\sum_{k=1}^{\infty}(tCK_0\,\|h\|_n)^k,
\end{aligned}
$$

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

$$
\begin{aligned}
P\big[|Z_n(h)| > \eta\big] \le 2\,e^{-t\eta}\,\mathrm{E}\big(\,e^{Z_n(h)}\,\big) &= 2\,\exp\{-t\eta\}\,\exp\{\Psi_{Z_n(h)}(t)\}\\
&\le 2\,\exp\{-t\eta + K_0^{-1}\sum_{k=1}^{\infty}(tCK_0\,\|h\|_n)^k\}.
\end{aligned}
$$

Choosing $t = \frac{1}{2CK_0\|h\|_n}$ gives the assertion:

$$P\big[|Z_n(h)| > \eta\big] \le 2\, e^{1/K_0} \exp\Big\{ -\frac{\eta}{2CK_0\|h\|_n} \Big\}.$$

The fact that

$$\big|\mathrm{cum}_j(\epsilon_t)\big| \le j!\, C^j, \quad j = 1, 2, \ldots \tag{36}$$

holds if $\mathrm{E}|\epsilon_t|^k \le \big(\frac{C}{2}\big)^k$, $k = 1, 2, \ldots$ can be seen by induction as follows. First observe that since $|\mathrm{cum}_1(\epsilon_t)| = \big|E(\epsilon_t)\big| = 0$, (36) holds for $j = 1$. Now assume that (36) holds for all $j = 1, \ldots, k-1$. Using $|\mathrm{cum}_k(\epsilon_t)| \le \mathrm{E}(|\epsilon_t|^k) + \sum_{i=1}^{k-1} \binom{k-1}{i-1} \mathrm{E}|\epsilon_t|^{k-i} |\mathrm{cum}_i(\epsilon_t)|$ we get

$$\begin{aligned}
|\mathrm{cum}_k(\epsilon_t)| &\le \Big(\frac{C}{2}\Big)^k + \sum_{j=1}^{k-1} \binom{k-1}{j-1} \mathrm{E}|\epsilon_t|^{k-j} |\mathrm{cum}_j(\epsilon_t)| \\
&\le \Big(\frac{C}{2}\Big)^k + \sum_{j=1}^{k-1} \binom{k-1}{j-1} \Big(\frac{C}{2}\Big)^{k-j} j!\, C^j \\
&\le \Big(\frac{C}{2}\Big)^k + (k-1)!\, C^k \sum_{j=1}^{k-1} \frac{j}{(k-j)!\, 2^{k-j}} \\
&\le \Big(\frac{C}{2}\Big)^k + k!\, C^k \sum_{\ell=1}^{k-1} \frac{2^{-\ell}}{\ell!} \\
&\le \Big(\frac{C}{2}\Big)^k + k!\, C^k (\sqrt{e} - 1) \\
&\le k!\, C^k \qquad\qquad (\text{for } k \ge 2).
\end{aligned}$$

5.6 Proof of Theorem 4

Using the exponential inequality from Lemma 3 we can mimic the proof of Lemma 3.2 from van de Geer (2000). As compared to van de Geer, our exponential bound is of the form $c_1 \exp\{-c_2 \frac{\eta}{\|h\|_n}\}$ rather than $c_1 \exp\{-c_2 \big(\frac{\eta}{\|h\|_n}\big)^2\}$. 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 Theorem 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. Condition (18) just makes sure that the upper limit in the integral from (19) is larger than the lower limit.

5.7 Proof of Theorem 2 and Theorem 5

First we indicate how the two processes studied in this paper enter the picture, and simultaneously we provide an outline of the proof. Recall that

$$\widehat{F}_n(\alpha, z) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}(\widehat{\eta}_t \leq z) \qquad \text{and} \qquad F_n(\alpha, z) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}(\sigma(\tfrac{t}{n})\, \epsilon_t \leq z).$$

With this notation we can write

$$\begin{aligned} \sqrt{n}\, (\, \widehat{G}_{n,\gamma}(\alpha) - G_{n,\gamma}(\alpha)\,) & \\ = \Big[\sqrt{n}\, \big(\, F_n(\alpha, \widehat{q}_\gamma) - H_n(\alpha, \widehat{q}_\gamma)\,\big) &- \sqrt{n}\, \big(\, F_n(\alpha, -\widehat{q}_\gamma) - H_n(\alpha, -\widehat{q}_\gamma)\,\big)\,\Big] \\ - \Big[\sqrt{n}\, \big(\, F_n(\alpha, \widehat{q}_\gamma) - F_n(\alpha, q_\gamma)\,\big) &- \sqrt{n}\, \big(\, F_n(\alpha, -\widehat{q}_\gamma) - F_n(\alpha, -q_\gamma)\,\big)\,\Big] \\ =: \qquad I_n(\alpha) \qquad &- \qquad II_n(\alpha) \end{aligned} \tag{37}$$

with $I_n(\alpha)$ and $II_n(\alpha)$, respectively, denoting the two quantities inside the two $[\cdot]$-brackets. The assertion of Theorem 5 follows from

$$\sup_{\alpha\in[0,1]} |I_n(\alpha)| = o_P(1) \qquad \text{and} \tag{38}$$

$$\sup_{\alpha\in[0,1]} |II_n(\alpha) - c(\alpha)\, \big(G_{n,\gamma}(1) - \mathrm{E}G_{n,\gamma}(1)\big)| = o_P(1). \tag{39}$$

Property (39) can be shown by using empirical process theory based on independent, but not identically distributed random variables. This will be done below. First we consider (38). We will see that this involves properties of both residual empirical processes and weighted sum processes.

Proof of $\sup_{\alpha\in[0,1]} |I_n(\alpha)| = o_P(1)$ and proof of Theorem 2.

Observe that

$$\begin{aligned} \widehat{F}_n(\alpha, z\,) = \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\big(\, \widehat{\eta}_t \leq z\,\big) &= \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\big(\, \sigma(\tfrac{t}{n})\, \epsilon_t \leq \sigma(\tfrac{t}{n})\, \epsilon_t - \widehat{\eta}_t + z\,\big) \\ &= \frac{1}{n} \sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\big(\, \sigma(\tfrac{t}{n})\, \epsilon_t \leq \big((\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})(\tfrac{t}{n})\big)' \mathbf{Y}_{t-1} + z\,\big). \end{aligned}$$

Recall that by assumption $P\big(\widehat{\boldsymbol{\theta}}_n(\cdot) - \boldsymbol{\theta}(\cdot) \in \mathcal{G}^p\big) \to 1$ as $n \to \infty$, where for $\mathbf{g} = (g_1, \ldots, g_p)' : [0,1]^p \to \mathbb{R}$ we write $\{\mathbf{g} \in \mathcal{G}^p\}$ for $\{g_i \in \mathcal{G}, i = 1, \ldots, p\}$. We also write

$$F_n(\alpha, z, \mathbf{g}) = \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathbf{1}\big\{\sigma(\tfrac{t}{n})\,\epsilon_t \le \mathbf{g}'(\tfrac{t}{n})\,\mathbf{Y}_{t-1} + z\big\}.$$

By taking into account that by assumption $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_P(m_n^{-1})$ the above implies that for any $\epsilon > 0$ we can find a $C > 0$ such that with probability at least $1 - \epsilon$ we have for large enough n that

$$\sup_{\alpha \in [0,1],\, z \in [-L,L]} \sqrt{n}\,\Big|\, F_n(\alpha, z) - \widehat{F}_n(\alpha, z)\,\Big| \le \sup_{\substack{\alpha \in [0,1],\, z \in [-L,L], \\ \mathbf{g} \in \mathcal{G}^p,\, \|\mathbf{g}\|_n \le C m_n^{-1}}} \sqrt{n}\,\Big|\, F_n(\alpha, z) - F_n(\alpha, z, \mathbf{g})\,\Big|. \quad (40)$$

Thus, if L is such that $q_\gamma \in (-L, L)$ and $P(\widehat{q}_\gamma \in [-L, L]) \to 1$, as $n \to \infty$, then the right hand side in (40) being $o_P(1)$ for any $C > 0$ implies that $\sup_\alpha |I_n(\alpha)| = o_P(1)$, and this then also proves Theorem 2. In order to control the right hand side in (40), we first introduce an appropriate centering. Let

$$\begin{aligned} E_n(\alpha, z, \mathbf{g}) :=& \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} \mathrm{E}\big[\mathbf{1}\big\{\sigma(\tfrac{t}{n})\,\epsilon_t \le \mathbf{g}'(\tfrac{t}{n})\,\mathbf{Y}_{t-1} + z\big\}\,\big|\,\epsilon_{t-1}, \epsilon_{t-2}, \ldots\big] \\ =& \frac{1}{n}\sum_{t=1}^{\lfloor \alpha n \rfloor} F\left(\frac{\mathbf{g}(\tfrac{t}{n}))'\mathbf{Y}_{t-1} + z}{\sigma(\tfrac{t}{n})}\right) \end{aligned} \quad (41)$$

and let $\mathbf{0} = (0, \ldots, 0)' \in \mathcal{G}^p$ denote the p-vector of null-functions. Write

$$\begin{aligned} F_n(\alpha, z) - F_n(\alpha, z, \mathbf{g}) =& \Big[\big(F_n(\alpha, z, \mathbf{0}) - E_n(\alpha, z, \mathbf{0})\big) - \big(F_n(\alpha, z, \mathbf{g}) - E_n(\alpha, z, \mathbf{g})\big)\Big] \\ &+ \big[E_n(\alpha, z, \mathbf{0}) - E_n(\alpha, z, \mathbf{g})\big] \\ =:& \;\; T_{n1}(\alpha, z, \mathbf{g}) + T_{n2}(\alpha, z, \mathbf{g}) \end{aligned} \quad (42)$$

with T_{n1} and T_{n2} denoting the two terms inside the two [ ] brackets. This centering makes T_{1n} a sum of martingale differences (cf. proof of Lemma 1 given above). The quantity $\nu_n(\alpha, z, \mathbf{g})) = \sqrt{n}\big(F_n(\alpha, z, \mathbf{g}) - E_n(\alpha, z, \mathbf{g})\big)$ in fact is the residual empirical process discussed in section 2.

It follows from (40) and (42) that together

$$\sup_{\substack{\alpha\in[0,1],\, z\in[-L,L],\\ \mathbf{g}\in\mathcal{G}^p,\, \|\mathbf{g}\|_n\le Cm_n^{-1}}} |\sqrt{n}\,T_{n1}(\alpha,z,\mathbf{g})| = o_P(1) \qquad \text{and} \tag{43}$$

$$\sup_{\substack{\alpha\in[0,1],\, z\in[-L,L],\\ \mathbf{g}\in\mathcal{G}^p,\, \|\mathbf{g}\|_n\le Cm_n^{-1}}} \Big|\sqrt{n}\,T_{n2}(\alpha,z,\mathbf{g})\Big| = o_P(1) \tag{44}$$

imply the desired result, provided we can show that $P(\widehat{q}_\gamma \notin [-L,L]) = o(1)$ as $n\to\infty$. Since $q_\gamma \in (-L,L)$, the desired property follows from consistency of $\widehat{q}_\gamma$ as an estimator for q_γ. This will be shown at the end of this proof.

Verification of (43). We have with $\mathcal{H}_L$ as defined immediately above Lemma 1 in Section 2 that

$$\sup_{\substack{\alpha\in[0,1],\, z\in[-L,L],\\ \mathbf{g}\in\mathcal{G}^p,\, \|\mathbf{g}\|_n\le Cm_n^{-1}}} |\nu_n(\alpha,z,\mathbf{g})-\nu_n(\alpha,z,\mathbf{0})| \le \sup_{\substack{d_n(h_1,h_2)\le C\,m_n^{-1}\\ h_1,h_2\in\mathcal{H}_L}} |\nu_n(h_1)-\nu_n(h_2)|.$$

As above let $\mathbf{A}_n = \{\frac{1}{n}\sum_{t=1-p}^n Y_t^2 \le C_0\}$ for some $C_0>0$. It is shown in the proof of Theorem 1 that $\frac{1}{n}\sum_{s=-p+1}^n Y_s^2 = O_P(1)$, and thus $P(\mathbf{A}_n) = o(1)$. It remains to show that and for $\eta>0$ we have

$$P\Big(\sup_{\substack{d_n(h_1,h_2)\le C_\epsilon\, m_n^{-1}\\ h_1,h_2\in\mathcal{H}_L}} |\nu_n(h_1)-\nu_n(h_2)| \ge \eta, \mathbf{A}_n\Big) = o(1). \tag{45}$$

This, however, is an immediate application of Theorem 1, and (43) is verified. Notice here that the finiteness of the covering integral of $\mathcal{H}_L$ with respect to $\|\cdot\|_n$ follows by standard arguments. In fact, it is not difficult to see that for some $C_0>0$

$$\log N_n(C_1\delta,\mathcal{H}_L) \le -C_0\log\epsilon + \log N_n(\epsilon,\mathcal{G}). \tag{46}$$

Verification of (44). We have

$$\begin{aligned}\sqrt{n}\,T_{n2}(\alpha,z,\mathbf{g}) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}\Big[F\Big(\frac{\mathbf{g}\big(\frac{t}{n}\big)'\mathbf{Y}_{t-1}+z}{\sigma\big(\frac{t}{n}\big)}\Big) - F\Big(\frac{z}{\sigma\big(\frac{t}{n}\big)}\Big)\Big]\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil} f_{\frac{t}{n}}(z)\,\big(\mathbf{g}\big(\tfrac{t}{n}\big)\big)'\mathbf{Y}_{t-1} + \frac{1}{\sqrt{n}}\sum_{t=1}^{\lceil n\alpha\rceil}(f_{\frac{t}{n}}(\zeta_t)-f_{\frac{t}{n}}(z))\big(\mathbf{g}\big(\tfrac{t}{n}\big)\big)'\mathbf{Y}_{t-1}\end{aligned} \tag{47}$$

with ζ_t between z and $z+\big(\mathbf{g}\big(\frac{t}{n}\big)\big)'\mathbf{Y}_{t-1}$. The second term in (47) is a remainder term that will be treated below. The first term is a weighted sum of the Y_s's and we will now use Theorem 4 to control this term. We can write the first term as $\sum_{k=1}^{p} Z_{k,n}(h)$ where

$$Z_{k,n}(h) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(\tfrac{t}{n}) Y_{t-k}, \quad k = 1, \ldots, p, \tag{48}$$

for $h \in \mathcal{H}$ with $\mathcal{H} = \{h_{\alpha,z,g}(u) = \mathbf{1}_\alpha(u)\, f_u(z) g(u),\ \alpha \in [0,1],\ z \in \mathbb{R},\ g \in \mathcal{G}\}$ where we use the shorthand notation $\mathbf{1}_\alpha(u) = \mathbf{1}\big(u \le \alpha\big)$. We will apply Theorem 4 to show that each $Z_{k,n}(h)$ tends to zero uniformly in $h \in \mathcal{H}$. Since by our assumptions the functions $\{f_u(z), z \in [-L, L]\}$ are uniformly bounded, we have $\sup_{\alpha,z,g} \|h_{\alpha,z,g}\|_n^2 \le \sup_{u,z} |f_u(z)|^2\, m_n^{-1} = o(1)$. Since $\mathrm{E} Z_n(h_{\alpha,z,g}) = 0$, an application of Theorem 4 now shows that the first term of (47) converges to zero in probability uniformly in $(\alpha, z, \mathbf{g})$.

Now we treat the second term in (47). Recall that our assumptions imply that the functions $\{f_u,\ u \in [0,1]\}$ are uniformly Lipschitz continuous with Lipschitz constant c, say. Therefore we can estimate the last term in (47) by

$$\begin{aligned} \frac{c}{\sqrt{n}} \sum_{t=1}^{n} \big|\, \big(\mathbf{g}\big(\tfrac{t}{n}\big)\big)' \mathbf{Y}_{t-1}\, \big|^2 &\le \frac{c}{\sqrt{n}} \sum_{t=1}^{n} \Big(\sum_{k=1}^{p} g_k^2\big(\tfrac{t}{n}\big) \sum_{j=0}^{p} Y_{t-j}^2 \Big) \\ &\le c\, p \sup_{-p \le t \le T} Y_t^2\ \sqrt{n} \sum_{k=1}^{p} \|g_k\|_n^2 = c\, p\, \frac{\sqrt{n}}{m_n^2}\, O_P(\log n) = o_P(1), \end{aligned}$$

where the last inequality uses the fact that $\sup_{-p \le t \le T} Y_t^2 = O_P(\log n)$, which follows as in the proof of Lemma 5.9 of Dahlhaus and Polonik (2009).

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

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

We can write

$$II_n(\alpha) = \sqrt{n}\, \big((F_n - \overline{F}_n)(\alpha, \widehat{q}_\gamma) - (F_n - \overline{F}_n)(\alpha, q_\gamma)\, \big) \tag{49}$$

$$+ \sqrt{n} \big((F_n - \overline{F}_n)(\alpha, -\widehat{q}_\gamma)\ - (F_n - \overline{F}_n)(\alpha, -q_\gamma) \big) \tag{50}$$

$$+ \sqrt{n}\, (\overline{F}_n(\alpha, \widehat{q}_\gamma) - \overline{F}_n(\alpha, q_\gamma))\ -\ \sqrt{n}\, (\overline{F}_n(\alpha, -\widehat{q}_\gamma) - \overline{F}_n(\alpha, -q_\gamma)) \tag{51}$$

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 $\rho_n(v,w) = \big|\overline{F}_n(1,v) - \overline{F}_{1,n}(1,w)\big|$, i.e. for every $\eta > 0$ there exists an $\epsilon > 0$ with

$$\limsup_{n\to\infty} P\Big[\sup_{\alpha\in[0,1],\, \rho_n(z_1,z_2)\le\epsilon} |\nu_n(\alpha, z_1, \mathbf{0}) - \nu_n(\alpha, z_2, \mathbf{0})| > \eta \Big] = 0. \tag{52}$$

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

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

Thus, (52) 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 (26) that for $\mathbf{g}_1 = \mathbf{g}_2 = \mathbf{0}$ we simply can use the metric $\overline{\rho}\big(\,(\alpha_1, z_1),\, (\alpha_2, z_2)\,\big) = |\alpha_1 - \alpha_2| + \rho_n(z_1, z_2)$ 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 (49) and (50) are $o_P(1)$. We now prove this consistency of $\widehat{q}_\gamma$.

Recall that $\Psi(u) = \int_0^1 F(\frac{z}{\sigma(u)})\,du - \int_0^1 F(\frac{-z}{\sigma(u)})\,du$. Observe that $\int_0^1 F(\frac{q_\gamma}{\sigma(u)})\,du$ is close to $\overline{F}_n(1, q_\gamma)$. In fact, since by assumption $u \to F(\frac{q_\gamma}{\sigma(u)})$ is of bounded variation, their difference is of the order $O(1/n)$. Consequently we have

$$\begin{aligned} &\big|\,(1-\gamma) - \big(\,\overline{F}_n(1, q_\gamma) - \overline{F}_n(1, -q_\gamma)\,\big)\,\big| \\ &= \Big|\,\Psi(1, q_\gamma) - \big(\,\overline{F}_n(1, q_\gamma) - \overline{F}_n(1, -q_\gamma)\,\big)\,\Big| \le c\,n^{-1} \end{aligned} \tag{53}$$

for some $c \ge 0$. In fact (53) holds uniformly in γ. This follows from the fact that the functions $u \to F\big(\frac{z}{\sigma(u)}\big)$ are Lipschitz continuous uniformly in z. (Notice that in the case where $\sigma(\cdot) \equiv \sigma_0$ is constant on $[0,1]$ then $\Psi(z) = F(\frac{z}{\sigma_0}) - F(\frac{-z}{\sigma_0}) = \overline{F}_n(1,z) - \overline{F}_n(1,-z)$. In other words, in this case we can choose $c = 0$.) We now show that under our assumptions we have for any fixed $0 < \gamma < 1$ that

$$\overline{\rho}_n(\widehat{q}_\gamma,\, q_\gamma) = \big|\overline{F}_n(1,\widehat{q}_\gamma) - \overline{F}_n(1,q_\gamma)\big| = o_P(1). \tag{54}$$

By assumption $\Psi(z)$ is a strictly monotonic function in z. Together with (53) this implies that (54) is equivalent to $\Big|\,\big(\,\overline{F}_n(1,\widehat{q}_\gamma) - \overline{F}_n(1,-\widehat{q}_\gamma)\,\big) - \big(\,\overline{F}_n(1,q_\gamma) - \overline{F}_n(1,-q_\gamma)\,\big)\,\Big| = o_P(1)$, which (by using (53)) follows from

$$\Big|\,\big(\,\overline{F}_n(1,\widehat{q}_\gamma) - \overline{F}_n(1,-\widehat{q}_\gamma)\,\big) - \big(\,1-\gamma\,\big)\,\Big| = o_P(1), \tag{55}$$

Since by definition of $\widehat{q}_\gamma$ we have $\widehat{F}_n(1,\widehat{q}_\gamma) - \widehat{F}_n(1,-\widehat{q}_\gamma) = 1-\gamma$, (55) follows from

$$\sup_z |\widehat{F}_n(1,z) - \overline{F}_n(1,z)| = o_P(1). \tag{56}$$

To see (56) notice that $\sup_z |\widehat{F}_n(1,z) - F_n(1,z)| = o_P(1)$. This follows from (40), (42), (43) and (44). Utilizing triangular inequality it remains to show that $\sup_z |F_n(1,z) - \overline{F}_n(1,z)| = o_P(1)$. This uniform law of large numbers result follows from the arguments given below. It can also be seen directly by observing that for every fixed z we have $|F_n(1,z) - \overline{F}_n(1,z)| = o_P(1)$ (which easily follows by observing that the variance of this quantity tends to zero as $n \to \infty$) together with a standard argument as in the proof of the classical Glivenko-Cantelli theorem for continuous random variables, utilizing monotonicity of $\overline{F}_n(1,z)$ and $F_n(1,z)$. This completes the proof of (54), and as outlined above, this implies that both (49) and (50) are $o_P(1)$, uniformly in α.

It remains to consider the quantity in (51). First, we derive an upper bound for $\widehat{q}_\gamma$. Let $B_n(\epsilon) = \{|\overline{F}_n(\widehat{q}_\gamma) - \overline{F}_n(q_\gamma)| < \epsilon;\ \sup_z |(\widehat{F}_n - \overline{F}_n)(1,z)| < \delta_\epsilon/6\}$ where $\delta_\epsilon > 0$ is such that $|\overline{F}_n(q_{\gamma\pm\delta_\epsilon}) - \overline{F}_n(q_\gamma)| < \epsilon$. We already know that $P(B_n(\epsilon)) \to 1$ as $n \to \infty$ for any $\epsilon > 0$. Now choose $\epsilon > 0$ small enough (such that $\gamma > \delta_\epsilon/6$ in order for the below to be well defined). On $B_n(\epsilon)$ we have

$$\begin{aligned}
\widehat{q}_\gamma &= \inf\{z \geq 0 : \widehat{F}_n(1,z) - \widehat{F}_n(1,-z) \geq 1-\gamma;\ |\overline{F}_n(z) - \overline{F}_n(q_\gamma)| < \epsilon\} \\
&\leq \inf\{z \geq 0 : \overline{F}_n(1,z) - \overline{F}_n(1,-z) \geq 1-\gamma - \big[\,(\widehat{F}_n - \overline{F}_n)(1,q_\gamma) - (\widehat{F}_n - \overline{F}_n)(1,-q_\gamma)\,\big] \\
&\qquad\qquad + 2 \sup_{|\overline{F}_n(v) - \overline{F}_n(w)| \leq \epsilon} |(\widehat{F}_n - \overline{F}_n)(1,v) - |(\widehat{F}_n - \overline{F}_n)(1,w)|\} \\
&\leq \inf\{z \geq 0 : \Psi(z) \geq 1-\gamma - \big[\,(\widehat{F}_n - \overline{F}_n)(1,q_\gamma) - (\widehat{F}_n - \overline{F}_n)(1,-q_\gamma)\,\big]
\end{aligned}$$

$$
\begin{aligned}
&\qquad + 2 \sup_{|\overline{F}_n(v)-\overline{F}_n(w)|\le\epsilon} |(\widehat{F}_n - \overline{F}_n)(1,v) - |(\widehat{F}_n - \overline{F}_n)(1,w)| + c\,n^{-1}\} \\
&= \Psi^{-1}(1-\gamma-Q_n+r_n),
\end{aligned}
$$

where for short $r_n = 2\sup_{|\overline{F}_n(v)-\overline{F}_n(w)|\le\epsilon} |(\widehat{F}_n - \overline{F}_n)(1,v) - |(\widehat{F}_n - \overline{F}_n)(1,w)| + cn^{-1}$ with c from (53), and $Q_n = \big[\,(\widehat{F}_n - \overline{F}_n)(1,q_\gamma) - (\widehat{F}_n - \overline{F}_n)(1,-q_\gamma)\,\big]$. Now let

$$
\Psi(\alpha,z) = \int_0^\alpha F\big(\tfrac{z}{\sigma(u)}\big)\,du - \int_0^\alpha F\big(\tfrac{-z}{\sigma(u))}\big)\,du, \quad z \ge 0,\ \alpha\in[0,1]. \tag{57}
$$

Observe that $\Psi(z) = \Psi(1,z)$, where $\Psi(z)$ is defined in (22). Since by definition of q_γ we have $q_\gamma = \Psi^{-1}(1-\gamma)$, and since $\Psi(\alpha,z)$ is strictly increasing in z for any α we have on $B_n(\epsilon)$,

$$
\begin{aligned}
&\big[\,\overline{F}_n(\alpha,\widehat{q}_\gamma) - \overline{F}_n(\alpha,q_\gamma)\,\big] - \big[\,\overline{F}_n(\alpha,-\widehat{q}_\gamma) - \overline{F}_n(\alpha,-q_\gamma)\,\big] \\
= \;& \Psi(\alpha,\widehat{q}_\gamma) - \Psi(\alpha,q_\gamma) + c\,n^{-1} \\
\le \;& \Psi\Big(\alpha,\Psi^{-1}\big(\,1-\gamma-Q_n+r_n\,\big)\Big) - \Psi\big(\alpha,\Psi^{-1}(1-\gamma)\big) + c\,n^{-1} \\
= \;& -\frac{\frac{\partial}{\partial z}\Psi(\alpha,\Psi^{-1}(\xi_n^+))}{\Psi'(\Psi^{-1}(\xi_n^+))}\,(Q_n - r_n) + c\,n^{-1} \\
= \;& -\frac{\int_0^\alpha \big[\,f_u(\Psi^{-1}(\xi_n^+)) + f_u(-\Psi^{-1}(\xi_n^+))\,\big]\,du}{\int_0^1 \big[f_u(\Psi^{-1}(\xi_n^+)) + f_u(-\Psi^{-1}(\xi_n^+))\big]\,du}\quad (Q_n - r_n) + c\,n^{-1} \\
=: \;& -\,(c(\alpha) + o_P(1))\,(Q_n - r_n) + c\,n^{-1}
\end{aligned}
$$

with $\xi_n^+ \in [1-\gamma, 1-\gamma-\ Q_n + r_n]$, $f_u(z) = \frac{\partial}{\partial z}F\big(\frac{z}{\sigma(u)}\big) = \frac{1}{\sigma(u)}\,f\big(\frac{z}{\sigma(u)}\big)$, and the $o_P(1)$-term equals $c_n(\alpha) - c(\alpha)$ with $c_n(\alpha)$ the ratio from the second to last line in the above formula, and

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

The fact that $|c_n(\alpha) - c(\alpha)| = o_P(1)$ follows from our smoothness assumptions together with the fact that $|Q_n - r_n| = o_P(1)$. Similarly, we obtain a lower bound of the form

$$
\big[\,\overline{F}_n(\alpha,\widehat{q}_\gamma) - \overline{F}_n(\alpha,q_\gamma)\,\big] - \big[\,\overline{F}_n(\alpha,-\widehat{q}_\gamma) - \overline{F}_n(\alpha,-q_\gamma)\,\big] \ge -(c(\alpha) + o_P(1))\,(Q_n + r_n) - c\,n^{-1}.
$$

From the above we know that $\sqrt{n}\,Q_n = \sqrt{n}\,(F_n - \overline{F}_n)(1,q_\gamma) - \sqrt{n}\,(F_n - \overline{F}_n)(1,-q_\gamma) + o_P(1)$ and $\sqrt{n}\,r_n = o_P(1)$, so that

$$\begin{aligned}
&\sqrt{n}\,\big[\,F_n(\alpha,\widehat{q}_\gamma) - \overline{F}_n(\alpha,q_\gamma)\,\big] - \big[\,F_n(\alpha,-\widehat{q}_\gamma) - \overline{F}_n(\alpha,-q_\gamma)\,\big] \\
&\qquad = \; -\frac{c(\alpha)}{\sqrt{n}}\sum_{t=1}^{n}\big(\,\mathbf{1}(\sigma^2\big(\tfrac{s}{n}\big)\,\epsilon_t^2 \le q_\gamma^2) - P(\sigma^2\big(\tfrac{t}{n}\big)\,\epsilon_t^2 \le q_\gamma^2)\,\big) + o_P(1) \\
&\qquad = \; -\,c(\alpha)\,\sqrt{n}\,(G_{n,\gamma}(1) - \mathrm{E}G_{n,\gamma}(1)\,) + o_P(1).
\end{aligned}$$

Finally, observe that under H_0, where $\sigma(u) \equiv \sigma_0$, we have $c(\alpha) = \alpha$, because f_u does not depend on $u \in [a,b]$. The proof is complete once we have shown that $|\widehat{q}_\gamma - q_\gamma| = o_P(1)$ as $n \to \infty$ (cf. comment right after (44)). To see that, first notice that our regularity conditions imply that $\inf_u f_u(q_\gamma) =: d_* > 0$. Again using the fact that our assumptions imply the uniform Lipschitz continuity of the functions $z \to f_u(z)$ we obtain the existence of an $\epsilon_0 > 0$ such that $f_u(z) > \frac{d_*}{2}$ for all $|z - q_\gamma| \le \epsilon_0$ and all $u \in [0,1]$. Consequently, if $|\widehat{q}_\gamma - q_\gamma| \le \epsilon_0$ then $\big|F\big(\frac{\widehat{q}_\gamma}{\sigma(\frac{t}{n})}\big) - F\big(\frac{q_\gamma}{\sigma(\frac{t}{n})}\big)\big| = \Big|\int_{q_\gamma}^{\widehat{q}_\gamma} f_{\frac{t}{n}}(u)\,du\Big| \ge \frac{d_*}{2}\,|\widehat{q}_\gamma - q_\gamma|\ \forall\ t = 0,1,\ldots,n$. On the other hand, if $|\widehat{q}_\gamma - q_\gamma| > \epsilon_0$ then, because the $f_{u(z)}$ are all non-negative, $\big|F\big(\frac{\widehat{q}_\gamma}{\sigma(\frac{s}{n})}\big) - F\big(\frac{q_\gamma}{\sigma(\frac{s}{n})}\big)\big| = \Big|\int_{q_\gamma}^{\widehat{q}_\gamma} f_{\frac{s}{T}}(u)\,du\Big| \ge \frac{d_*}{2}\,\epsilon_0 \quad \forall\ s = 0,1,\ldots,n$. Thus, for any $\epsilon > 0$ there exists an $\eta > 0$ with

$$\{|\widehat{q}_\gamma - q_\gamma| > \epsilon\} \subset \Big\{\big|F\big(\tfrac{\widehat{q}_\gamma}{\sigma(\frac{s}{n})}\big) - F\big(\tfrac{q_\gamma}{\sigma(\frac{s}{n})}\big)\big| > \eta\ \ \forall\, t = 0,1,\ldots,n\Big\}. \tag{58}$$

Since all the function $z \to F_{\frac{s}{T}}(z)$ are strictly monotone, we have

$$\overline{d}_n(\widehat{q}_\gamma, q_\gamma) = \big|\overline{F}_n(\widehat{q}_\gamma) - \overline{F}_n(q_\gamma)\big| = \frac{1}{n}\sum_{t=1}^{n}\big|F\big(\tfrac{\widehat{q}_\gamma}{\sigma(\frac{s}{n})}\big) - F\big(\tfrac{q_\gamma}{\sigma(\frac{s}{n})}\big)\big|.$$

Consequently, (58) implies that if $|\widehat{q}_\gamma - q_\gamma|$ is not $o_P(1)$, then also $\overline{d}_n(\widehat{q}_\gamma, q_\gamma)$ is not $o_P(1)$. This is a contradiction to (54) and this completes our proof.

6 References:

AKRITAS, M.G. and VAN KEILEGOM, I. (2001): Non-paramteric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.

ALEXANDER, K.S. and PYKE, R. (1986): A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Ann. Probab.* **14**, 582–597.

CHANDLER, G. and POLONIK, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *J. Amer. Statist. Assoc.* **101**, 240–253.

CHANDLER, G. and POLONIK, W. (2012): Mode Identification of volatility in time-varying autoregression. *J. Amer. Statist. Assoc.* **107**, 1217-1229.

CHENG, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Statist. Plann. Inference* **128**, 327–349.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.

DAHLHAUS, R., NEUMANN, M. and V. SACHS, R. (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* **5**, 873–906.

DAHLHAUS, R. and POLONIK, W. (2005). Nonparametric maximum likelihood estimation and empirical spectral processes for Gaussian locally stationary processes. Technical report.

DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes. *Ann. Statist.* **34**, 2790–2824.

DAHLHAUS, R. and POLONIK, W. (2009). Empirical spectral processes for locally stationary time series *Bernoulli* **15**, 1–39.

DREES, H. and STĂRICĂ, C. (2002). A simple non-stationary model for stock returns. Preprint.

EOM, K.B. (1999). Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models, *Multidimensional Systems and Signal Processing* **10**, 357–378.

FRYZLEWICZ, P., SAPATINAS, T., and 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. and PATAT, F. (1998). Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization. *IEEE Trans. Ultrasonic, Ferroeletric, and Frequency Control* **45**, 650–659.

GRENIER, Y. (1983): Time-dependent ARMA modeling of nonstationary signals, *IEEE Trans. on Acoustics, Speech, and Signal Processing* **31** , 899–911.

HALL, M.G., OPPENHEIM, A.V. and WILLSKY, A.S. (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. *Ann. Statist.* **29**, 445–469.

KOUL, H. L. (2002). *Weighted empirical processes in dynamic nonlinear models*, 2nd edn., Lecture Notes in Statistics, 166, Springer, New York.

KOUL, H.L., LING, S. (2006): Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.* **34**, 994–1012.

KÜNSCH, H.R. (1995): A note on causal solutions for locally stationary AR processes. Preprint. ETH Zürich.

LAÏB, N., LEMDANI, M. and OLUD-SAÏD, E. (2008): On residual empirical processes of GARCH-SM models: Application to conditional symmetry tests. *J. Time Ser. Anal.* **29**, 762–782.

MÜLLER, U., SCHICK, A. and WEFELMEYER, W. (2007): Estimating the error distribution function in semiparametric regression. *Statistics & Decisions* **25**, 1–18.

MÜLLER, U., SCHICK, A. and WEFELMEYER, W. (2009a): Estimating the innovation distribution in nonparametric autoregression. *Probab. Theory Relat. Fields* **144**, 53–77

MÜLLER, U., SCHICK, A. and WEFELMEYER, W. (2009b): Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Letters* **79**, 957–964.

MÜLLER, U., SCHICK, A. and WEFELMEYER, W. (2012): Estimating the error distribution function in semiparametric additive regression models. *J. Statist. Plan. Inference* **142**, 552–566.

NICKL, R. and PÖTSCHER, B.M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Thoer. Probab.* **20**, 177–199.

ORBE, S., FERREIRA, E., and RODRIGUES-POO, J. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *J. Econometrics* **126**, 53–77.

PRIESTLEY, M.B. (1965). Evolutionary spectra and non-stationary processes. *J. Royal Statist. Soc. - Ser. B* **27**, 204–237.

RAJAN, J.J. and Rayner, P.J. (1996). Generalized feature extraction for time-varying autoregressive models, *IEEE Trans. on Signal Processing* **44**, 2498–2507.

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, *J. Royal Statist. Soc. - Ser. B* **32**, 312–322.

VAN DE GEER, S. (2000): *Empirical processes in M-estimation.* Cambridge University Press.

VAN DER VAART, A.W. and WELLNER, J. (1996). *Weak convergence and empirical processes.* Springer, New York.

VAN KEILEGOM, I., GONZÁLEZ MANTEIGA, W. and SÁNCHEZ SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution, *Test* **17**, 401–415.