Weighted Sums and Residual Empirical Processes for Time-varying Processes

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

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

March 6, 2011

Abstract

Function indexed weighted sums and sequential residual empirical processes based on time-varying AR-processes are studied. It is shown that under appropriate assumptions non-parametric estimation of the parameter functions does not influence the asymptotic distribution of the residual empirical process. An exponential inequality for weighted sums of time-varying processes provides the basis for a weak convergence result of weighted sum processes. A motivation for studying these two types of processes is provided by the fact that, as shown in this paper, large sample results for both processes can be utilized to derive the asymptotic distribution of a test for modality of the variance function that is studied by the authors in an accompanying paper.

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

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

where \( \theta_k \) are the autoregressive parameter functions, \( p \) is the order of the model, \( \sigma \) 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 \( Y_t = Y_{t,T} \) satisfying (1) in fact forms a triangular array.

The consideration of non-stationary time series models goes back to Priestly (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 Stárică (2002), Fryzlewicz et al. (2006), and Orbe et al. (2005), 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 (9 - 11) below for the case of a time-varying AR-process). We would like to point out, however, that in this paper local stationarity is only used to calculate the asymptotic covariance function in Theorem 2. All the other results hold under weaker assumptions.

The interest of this paper is the investigation of the large sample behavior of two types of processes, residual sequential empirical process and weighted sums processes, based on observations from the non-stationary process satisfying (1). One motivation for studying the behavior of both of these processes is given by the fact, that the large sample behavior of both of these processes enter the derivation of the limit distribution of a test statistic for modality of the variance function in model (1). This testing procedure is discussed in an accompanying paper Chandler and Polonik (2010); see section 4.

Residual sequential empirical process are defined as follows. Let \( Y_t = (Y_{t-1}, \ldots, Y_{t-p})' \). Given an estimator \( \hat{\theta} \) of \( \theta = (\theta_1, \ldots, \theta_p)' \) and corresponding residuals \( \hat{\eta}_t = Y_t - \hat{\theta} \left( \frac{t}{T} \right)' Y_t \) we estimate
the distribution function of the innovations \( \eta_t = \sigma \left( \frac{t}{T} \right) \epsilon_t \) by the empirical distribution function of the residuals

\[
H_T(z) = \frac{1}{T} \sum_{t=1}^{T} 1 \{ \hat{\eta}_t \leq z \}, \ z \in \mathbb{R}.
\]

Observe that we can write \( H_T(z) = \frac{1}{T} \sum_{t=1}^{T} 1 \{ \sigma \left( \frac{t}{T} \right) \epsilon_t \leq (\theta - \hat{\theta})' \left( \frac{t}{T} \right) Y_t + z \} \). This motivates to consider the following type of process. Let \([a, b] \subset [0, 1]\), and let \( Y_s, s = 1, \ldots, n = \lfloor bT \rfloor - \lceil aT \rceil + 1 \) denote the observations \( Y_t \) with \( t \in [a, b] \).

Define

\[
\nu_n(\alpha, g, z) = \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor \alpha n \rfloor} \left[ 1 \{ \sigma(a + \frac{s}{n}) \epsilon_t \leq g'(a + \frac{s}{n}) Y_s + z \} - F_{\theta + \hat{\theta}} \left( g'(a + \frac{s}{n}) Y_s + z \right) \right],
\]

where \( \alpha \in [0, 1], z \in \mathbb{R}, g : [a, b] \to \mathbb{R}^p, g \in G^p = \{ g = (g_1, \ldots, g_p)' \}, g_i \in \mathcal{G} \) with \( \mathcal{G} \) an appropriate function class such that \( \theta - \hat{\theta} \in \mathcal{G}^p \) (see below), and \( F_u(z) \) denotes the distribution function of \( \sigma(u) \epsilon_t, u \in [0, 1] \). The reason for considering a subinterval \([a, b] \subset [0, 1]\) is again motivated by our application to the results to testing for modality of the variance function. Observe that with \( 0 = (0, \ldots, 0)' \in \mathcal{G}^p \) denoting the \( p \)-vector of null-functions, \( \nu_n(z, \alpha, 0) \) equals the empirical process based on the innovations rather than the residuals. The basic form of \( \nu_n \) is standard, but in contrast to most of the related literature, we are considering a non-parametric index class \( \mathcal{G} \), and of course \( Y_t \) is non-stationary here.

Letting \( F_n(z) \) denote the empirical distribution function of the innovations \( \epsilon_s \), our results will in particular imply that under appropriate assumptions we have \( \sup_{z \in \mathbb{R}} |H_n(z) - F_n(z)| = o_P(T^{-1/2}) \) even though non-parametric estimation of the parameter functions \( \theta_k \) is involved.

The second type considered here are weighted sum processes of the form

\[
Z_n(h) = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} h\left( \frac{z}{n} \right) Y_s,
\]

\( \alpha \in [0, 1] \) and \( h \in \mathcal{H} \) where \( \mathcal{H} \) is an appropriate function class. Such processes can be considered as generalizations of partial sum processes. Large sample behavior, including exponential inequalities and weak convergence of \( Z_n \) is considered below.

The obtained results will be applied to derive the large sample distribution of a test for modality of the variance function \( \sigma^2(\cdot) \) that is proposed in the accompanying paper Chandler and Polonik (2010). In fact, the corresponding test statistic essentially is a functional of \( \alpha \to \nu_n(\alpha, \hat{\theta} - \theta, \hat{q}_\gamma) \) plus a term that can be approximated by a weighted sum, where \( \hat{q}_\gamma \) is
an estimate of an appropriate $\gamma$-quantile. It turns out that the test statistic is asymptotically
distribution free, again despite the non-parametric estimation of the parameter functions.
(For more details see below.)

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). Section 4 applies some of the results from the preceding
sections to derive a crucial approximation result for a quantity based on the residual empirical
process. This approximation provides a crucial ingredient in Chandler and Polonik (2010)
to derive the asymptotic distribution free limit of a test statistics for testing modality of the
variance function in model (1). Proofs are deferred to section 5.

Remark on measurability. Suprema of function indexed processes will enter the theo-
retical results below. We assume throughout the paper that such suprema are measurable.
Otherwise statements “in probability” have to be replaced by “outer” or “inner” probability,
respectively.

2 Residual empirical processes under time varying AR-models

There 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), Laib et al. (2008), Müller et al. (2009),
and references therein. There is not much work available on residual empirical processes for
models involving infinite dimensional parameter spaces. Except Müller et al. (2009), who
consider a the estimation of the innovation distribution in a nonparametric autoregression
model, we are only aware of Akritas and van Keilegom (2001) and Cheng (2005). Aktiras
and van Keilegom consider the estimation of the error distribution in a heteroscedastic
non-parametric regression model, and Cheng (2005) is estimating distribution and density
function of the errors in a nonparametric (homoscedastic) regression model utilizing a sample
splitting technique. Here we are considering function indexed residual empirical processes
based on non-stationary time series.

In order to formulate one of our main results for the residual empirical process $\nu_n(\alpha, g, z)$
defined in (2) above, we first introduce some notation. For a function $h : [a, b] \rightarrow \mathbb{R}$ we denote
∥h∥∞ := \sup_{u \in [a,b]} |h(u)| \quad \text{and} \quad ∥h∥^2_n := \frac{1}{n} \sum_{t \in [aT,bT]} h^2\left(\frac{t}{T}\right).

Let \mathcal{H} denote a class of functions defined on [a, b]. For a given \delta > 0, let \( N(\delta, \mathcal{H}) \) denote the minimal number \( N \) of \( \| \cdot \|_n \)-balls of radius less than or equal to \( \delta \) that are needed to cover \( \mathcal{H} \), i.e., there exists functions \( g_k, k = 1, \ldots, N \) such that the balls \( A_k = \{ h \in \mathcal{H} : \| g_k - h \|_n \leq \delta \} \) where \( \mathcal{H} \subset \bigcup_{k=1}^{N} A_k \). Then \( \log N(\delta, \mathcal{H}) \) is called the metric entropy of \( \mathcal{H} \) with respect to \( \| \cdot \|_n \).

If the balls \( A_k \) are replaced by brackets \( B_k = \{ h \in \mathcal{H} : g_k \leq h \leq \bar{g}_k \} \) for pairs of functions \( g_k \leq \bar{g}_k, k = 1, \ldots, N \) with \( \| \bar{g}_k - g_k \|_n \leq \delta \), then the minimal number \( N = N_B(\delta, \mathcal{H}) \) of such brackets with \( \mathcal{H} \subset \bigcup_{k=1}^{N} B_k \) is called a bracketing covering number, and \( \log N_B(\delta, \mathcal{H}) \) is called the metric entropy with bracketing of \( \mathcal{H} \) with respect to \( \| \cdot \|_n \).

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

\[ Y_{t,T} = \sum_{j=0}^{\infty} a_{t,T}(j) \epsilon_{t-j} \]

where \( \epsilon_t \sim \text{i.i.d.} \ (0,1) \). The distribution \( F \) of \( \epsilon_t \) 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

\[ \sup_{1 \leq t \leq T} |a_{t,T}(j)| \leq \frac{K}{\ell(j)} \]

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 \( j = 0, 1 \).

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

\[ \theta_k(\cdot) - \hat{\theta}_k(\cdot) \in \mathcal{G}, \quad k = 1, \ldots, p, \]

such that \( \sup_{g \in \mathcal{G}} \| g \|_\infty < \infty \) and for some \( \gamma \in [0,1) \) and \( C, c > 0 \) and for all \( \delta > \frac{\epsilon}{n} \)

\[ \log N_B(\delta, \mathcal{G}) \leq \begin{cases} C \delta^{-\gamma}, & \text{for } 0 < \gamma < 1 \\ C \log \left(\delta^{-1}\right), & \text{for } \gamma = 0. \end{cases} \]
For $k = 1, \ldots, p$ we have $\frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{T} \sum_{t=1}^{T} \left( \hat{\theta}_k(t) - \theta_k(t) \right) \right)^2 = O_P(m_n^{-1})$ with $m_n \to \infty$ as $T \to \infty$.

**Theorem 1** Suppose that assumptions (i) and (ii) hold, and that for some $c > 0$ and $n$ sufficiently large $G$ satisfies

$$
\int_{\frac{1}{n}}^{1} \sqrt{\log N_B(u^2, \mathcal{G})} \, du < \infty.
$$

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

$$
\sup_{\alpha \in [0,1], \alpha \in [-L, L], \|g\| \leq \delta_n} |\nu_n(\alpha, z, g) - \nu_n(\alpha, z, 0)| = o_P(1). \tag{3}
$$

The above type of result is typical for work on residual empirical processes (e.g. (8.2.32) in Koul 2002). However, as mentioned above, compared to the existing work we are dealing with non-stationary observations, and are considering a non-parametric class of functions. Also keep in mind that we are in fact considering triangular arrays (recall that $Y_t = Y_{t,T}$).

**Discussion of the assumptions.** Assumptions (i) and (ii) have been used in the literature on locally stationary processes (e.g. Dahlhaus and Polonik, 2006, 2009). It is shown in Dahlhaus and Polonik (2006) that (i) and (ii) (and also the additional assumptions (9) - (11) needed for Theorem 3 below) hold for time-varying AR-processes (1) and more general, for time-varying ARMA-models as long as 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.

Assumption (iii) controls the complexity of the class $\mathcal{G}$. In fact this assumption implies that for some $c > 0$ we have

$$
\int_{\frac{1}{n}}^{1} \sqrt{\log N_B(u^2, \mathcal{G})} \, du < \infty \quad \text{and} \quad \int_{\frac{1}{n}}^{1} \log N(u, \mathcal{G}) \, du < \infty.
$$

Both of these assumptions are used below. Notice that the standard condition on the covering integral is $\int_{0}^{1} \sqrt{\log N_B(u, \mathcal{G})} \, du < \infty$ (or similarly without bracketing). In contrast to that, our first condition is using $N_B(u^2, \mathcal{G})$ (rather than $N_B(u, \mathcal{G})$) in the integrant, and the second does not have a square root. This makes both our the 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).

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 $(\log n)^{s/(2s+1)}$ 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 possess 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 of order $O(n^{-1})$. Consequently, in this case we have $m_n^{-1} = (\log n)^{s/(2s+1)}$. In order to verify the condition on the bracketing covering numbers from (iii), we are using 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 $\mathcal{H} = [0,1] \times [-L,L] \times \mathcal{G}^p$ denote the index space of the process $\nu_n$ where $L > 0$ is a constant. Define a metric on $\mathcal{H}$ as

$$d(h_1, h_2) = d((\alpha_1, z_1, g_1), (\alpha_2, z_2, g_2)) = |\alpha_1 - \alpha_2| + |z_1 - z_2| + \sum_{k=1}^p \|g_{1,k} - g_{2,k}\|_n. \quad (4)$$

Let $D_B(\epsilon, \mathcal{H})$ denote the bracketing covering number of $\mathcal{H}$ with respect to the metric $d$.

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

\[
\eta \geq \frac{2^6 K^*}{\sqrt{n}} \quad (5)
\]

\[
\eta \leq \frac{1}{2} K^* \sqrt{n} (\tau^2 \wedge \tau) \quad (6)
\]

\[
\eta \geq C_1 \left( \int_{\eta/2^8 K^* \sqrt{n}}^{\tau} \sqrt{\log D_B(u^2, \mathcal{H})} \, du \vee \tau \right) . \quad (7)
\]

Then, for $C_1 \geq 2^6 \sqrt{10 K^*}$ we have

\[
P\left[ \sup_{d(h_1, h_2) \leq \tau^2} |\nu_n(h_1) - \nu_n(h_2)| \geq \eta, \mathbf{F}_n \right] \leq \left( \frac{2^6(2^6 + 1) K^*}{C_1^2} + 2 \right) \exp \left( -\frac{\eta^2}{2^6(2^6 + 1) K^* \tau^2} \right),
\]

where the supremum is extended over $h_1, h_2 \in \mathcal{H}$.

3 Weighted sums under local stationarity

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

\[
Z_n(h) = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} h(a + \frac{s}{n}) Y_s, \quad h \in \mathcal{H}. \quad (8)
\]

In the iid 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). This means that we assume the existence of functions $a(\cdot, j) : (0, 1] \to \mathbb{R}$ with

\[
\sup_u |a(u, j)| \leq \frac{K}{\ell(j)}, \quad (9)
\]
\[
\sup_j \sum_{t=1}^n |a_{t,T}(j) - a(t/T, j)| \leq K, \tag{10}
\]

\[
TV(a(\cdot, j)) \leq \frac{K}{\ell(j)}, \tag{11}
\]

where for a function \( g : [0, 1] \to \mathbb{R} \) we denote by \( TV(g) \) the total variation of \( g \) on \([0, 1]\).

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

\[
c(u, k) := \int_{-\pi}^\pi f(u, \lambda) \exp(i\lambda k) d\lambda = \sum_{j=-\infty}^\infty a(u, k + j) a(u, j) \tag{12}
\]

is the time varying covariance of lag \( k \) at rescaled time \( u \in [0, 1] \).

**Theorem 2** Let \( \mathcal{H} \) denote a class of uniformly bounded, real valued functions of bounded variation defined on \([a, b]\). Assume further that for some \( c > 0 \),

\[
\int_{\frac{1}{c}}^1 \log N(u, \mathcal{H}) \, du < \infty.
\]

Then we have under assumptions (i), (ii) and (9) - (11) that as \( n \to \infty \) the process \( Z_n(h), h \in \mathcal{H} \) converges weakly to a tight, mean zero Gaussian process \( \{G(h), h \in \mathcal{H}\} \) with variance-covariance function \( C(h_1, h_2) = \frac{1}{b-a} \int_a^b h_1(u) h_2(u) S(u) \, du \), 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) Partial weighted sums of the form

\[
Z_n(\alpha, h) = \frac{1}{\sqrt{n}} \sum_{s=1}^{[\alpha n]} h(a + \frac{s}{T}) Y_s
\]
are in fact a special case of processes considered in the theorem. This can be seen by writing
\[ Z_n(\alpha, h) = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \tilde{h}_\alpha(a + \frac{s}{n})Y_s \] with \( \tilde{h}_\alpha(u) = 1(a \leq u \leq a + \alpha(b - a)) h(u). \) In other words, the partial weighted sums are weighted sums indexed a slightly modified class of sets \( \tilde{H} = \{ h_\alpha = 1(a \leq u \leq a + \alpha(b - a)) h(u), \alpha \in [0, 1], h \in H \}. \) If the class \( H \) satisfies the assumptions on the covering integral from the above theorem, then so does \( \tilde{H}. \) The limit covariance can then be written as
\[ \text{cov}(\tilde{h}_1, \alpha), \tilde{h}_2, \beta) = \frac{1}{b-a} \int_{a}^{(\alpha \wedge \beta b)} h_1(u) h_2(u) S(u) du. \]
c) Assumptions (9) - (11) are only used for calculating the covariance function of the limit process.

The main ingredients to the proof of this theorem are presented in the following results, the proofs of which are given in the appendix. These results are of independent interest.

**Theorem 3** Let \( \{ Y_t, t = 1, \ldots, T \} \) satisfy assumptions (i) and (ii) and let \( H = \{ h : [a, b] \to \mathbb{R} \} \) be totally bounded with respect to \( \| \cdot \|_n. \) Further assume that there exists a constant \( C > 0 \) such that for all \( k = 1, 2, \ldots \) we have \( E|\epsilon_s|^k \leq C^k; \) and let \( F_n = \{ \frac{1}{n} \sum_{s=1}^{n} Y_s^2 \leq M^2 \} \) where \( M > 0. \) There exists constants \( c_0, c_1, c_2 > 0 \) such that for all \( \eta > 0 \) satisfying
\begin{align*}
\eta &< 16 M \sqrt{n} \tau \\
\eta &> c_0 \left( \int_{\frac{1}{\sqrt{M^2 + \tau}}} \log N(u, H) \, du \vee \tau \right)
\end{align*}
we have
\[ P\left[ \sup_{h \in H, \|h\|_n \leq \tau} |Z_n(h)| > \eta, F_n \right] \leq c_1 \exp \left\{ - \frac{\eta}{c_2 \tau} \right\}. \]

The second result of importance for the proof of Theorem 2 deals with cumulants. For random variables \( X_1, \ldots, X_k \) we denote by \( \text{cum}(X_1, \ldots, X_k) \) their joint cumulant, and if \( X_i = X \) for all \( i = 1, \ldots, k, \) then \( \text{cum}(X_1, \ldots, X_k) = \text{cum}(X, \ldots, X) = \text{cum}_k(X), \) the \( k \)-th order cumulant of \( X. \)

**Lemma 2** Let \( \{ Y_t, t = 1, \ldots, T \} \) have a MA-type representation given in assumption (i) with coefficients satisfying assumption (ii). For \( j = 1, 2, \ldots \) let \( h_j \) be functions defined on \([a, b]\) with \( \|h_j\|_n < \infty. \) Then there exists a constant \( 1 \leq K_0 < \infty \) such that for all \( k \geq 1, \)
\[ |\text{cum}(Z_n(h_1), \ldots, Z_n(h_k))| \leq K_0^{k-1} |\text{cum}_k(\epsilon_1)| \prod_{j=1}^{k} \|h_j\|_n. \]
If, in addition, \( \|h_j\|_\infty \le M < \infty, j = 1, \ldots, k \), then for \( k \ge 3 \),

\[
|\text{cum}(Z_n(h_1), \ldots, Z_n(h_k))| \le (K_0 M)^{k-2} n^{-\frac{k-2}{2}}.
\]

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

**Lemma 3** Suppose the assumptions of Lemma 2 hold. Let \( h \) be a function with \( \|h\|_n < \infty \).
Assume that there exists a constant \( C > 0 \) such that for all \( m = 1, 2, \ldots \) we have \( E|\epsilon_s|^m \le C^m \).
Then there exists constants \( c_1, c_2 > 0 \) such that for any \( \eta > 0 \) we have

\[
P\left[ |Z_n(h)| > \eta \right] \le c_1 \exp \left\{ - \frac{\eta}{c_2 \|h\|_n} \right\}.
\]

(15)

4 An Application: Asymptotic Distribution of a Test for Modality of the Variance Function

In an accompanying paper, Chandler and Polonik (2010), we propose a test for modality of the variance function model (1). The test statistic is essentially given by

\[
T_n = \sup_{\alpha \in [0, 1]} \left( \hat{G}_{n,\gamma}(\alpha) - \alpha \gamma \right)
\]

where \( \hat{G}_{n,\gamma}(\alpha) \) counts the number of large residuals within the first \((100 \times \alpha)\%\) of the interval \([a, b]\). If the variance function is constant, then, since one has a total of \( |n\gamma| \) large residuals, the expected value of \( n\hat{G}_{n,\gamma}(\alpha) \) approximately equals \( \alpha n \gamma \). This motivates the form of \( T_n \). We assume the interval \([a, b]\) to be fixed. For conducting the test for modality, using an appropriate interval is crucial, and Chandler and Polonik (2010) provide an estimator for the interval of interest and show that this estimation does not influence the asymptotic limit of the test statistic.

Notice that \( \hat{G}_{n,\gamma}(\alpha) \) is closely related to the sequential residual empirical process, and as can be seen from the proof of Theorem 4, weighted empirical processes enter the analysis.
of $\hat{G}_{n,\gamma}(\alpha)$ through handling the estimation of $q_{\gamma}$. Theorem 4 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 (2010). 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. $f_u(z) = \frac{1}{\sigma(u)} f(\frac{z}{\sigma(u)})$, and

$$G_{n,\gamma}(\alpha) = \frac{1}{n} \sum_{t=[aT]}^{[aN]-1} 1(\epsilon_t^2 \sigma_t^2(t) \geq q_{\gamma}^2),$$

where $q_{\gamma}$ is defined via

$$\Psi_{a,b}(z) = \int_0^1 F(\frac{z}{\sigma(a+\beta(b-a))}) d\beta - \int_0^1 F(\frac{-z}{\sigma(a+\beta(b-a))}) d\beta, \quad z \geq 0,$$

as the solution to the equation

$$\Psi_{a,b}(q_{\gamma}) = 1 - \gamma. \tag{16}$$

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 $q_{\gamma}^2$ equals the upper $\gamma$-quantile of the squared innovations $\eta_t^2 = \sigma_0^2 \epsilon_t^2$. The following approximation result is not assuming that the variance is constant, however.

**Theorem 4** Let $\gamma \in [0,1]$ and suppose that $0 \leq a < b \leq 1$ are non-random. Assume further that $E|\epsilon|^k < C^k$ for some $C > 0$ and all $k = 1, 2, \ldots$ Then, under assumptions $(i)$ - $(iv)$, with $\frac{n^{1/2}}{m^n \log n} = o(1)$ we have

$$\sqrt{n} \sup_{\alpha \in [0,1]} \left| \hat{G}_{n,\gamma}(\alpha) - G_{n,\gamma}(\alpha) + c(\alpha)(G_{n,\gamma}(1) - EG_{n,\gamma}(1)) \right| = o_p(1), \tag{17}$$

where

$$c(\alpha) = \frac{\int_a^b \left[ f_a(q_{\gamma}) + f_a(-q_{\gamma}) \right] du}{\int_a^b \left[ f_a(q_{\gamma}) + f_a(-q_{\gamma}) \right] du}.$$
Under the null-hypothesis $\sigma(u) \equiv \sigma_0 > 0$ for $u \in [a, b]$ 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 $E\epsilon_t^2 < \infty$.

Under the worst case null-hypothesis, $\sigma^2(u) = \sigma_0^2$ for all $u \in [0, 1]$, the innovations are iid, we have $c(\alpha) = \alpha$, and $E G_{n, \gamma}(\alpha) = \alpha \gamma$. Therefore the above result implies that $(\gamma (1 - \gamma))^{-1/2} \sqrt{n} (\tilde{G}_{n, \gamma}(\alpha) - \alpha \gamma)$ converges weakly to a standard Brownian Bridge. For more details we refer to Chandler and Polonik (2010).

5 Proofs

5.1 Proof of Lemma 1

We only present a brief outline. In order to simplify the notation we again assume w.l.o.g. that $a = 0$. First notice that $\nu_n(\alpha, z, g) = \sum (\sigma(a + \frac{s}{T}) \epsilon_s \leq g(a + \frac{s}{T}) Y_s - 1 + z)$ we have

$$\nu_n(\alpha, z, g) = \frac{1}{\sqrt{n}} \sum_{s=1}^n \tilde{\xi}_s \eta_{s} 1(\frac{s}{T} \leq \alpha b)$$

where $\tilde{\xi}_s = \xi_s - E(\xi_s | \mathcal{F}_{s-1})$ and $\mathcal{F}_s = \sigma(\epsilon_s, \epsilon_{s-1}, \ldots)$ denotes the $\sigma$-algebra generated by $\{\epsilon_s, \epsilon_{s-1}, \ldots\}$, and obviously also $\nu_n(\alpha_1, z_1, g_1) - \nu_n(\alpha_2, z_2, g_2)$ 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 $Z_1, \ldots, Z_T$ denote martingale differences with respect to a filtration $\{\mathcal{F}_t, t = 0, \ldots, T - 1\}$ with $|Z_t| \leq C$ for all $t = 1, \ldots, T$. Let further $S_n = \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t$ and $V_n = V_n(S_n) = \frac{1}{T} \sum_{t=1}^T E(Z_t^2 | \mathcal{F}_{t-1})$. Then we have for all $\epsilon, \tau^2 > 0$ that

$$P(S_n \geq \epsilon, V_n \leq \tau^2) \leq \exp \left( -\frac{\epsilon^2}{2\tau^2 + \frac{2C^2}{\sqrt{n}}} \right). \quad (18)$$

The form of (18) motivates that we first need to control the quadratic variation $V_n$. Let

$$\eta_{s}^{\alpha, z, g} = \tilde{\xi}_s \eta_{s} 1(\frac{s}{T} \leq \alpha b).$$
We have for \( h_1 = (\alpha_1, z_1, g_1), h_2 = (\alpha_2, z_2, g_2) \in H \) with \( d(h_1, h_2) \leq \epsilon \) that

\[
V_n = V_n(\nu_n(h_1) - \nu_n(h_2)) = \frac{1}{n} \sum_{s=1}^{n} E(\|\eta^{\alpha_1,z_1,g_1}_s - \eta^{\alpha_2,z_2,g_2}_s\| | F_{s-1})
\]

\[
\leq \frac{1}{n} \sum_{s=1}^{n} \left| 1(\frac{1}{1} \leq \alpha_1) - 1(\frac{1}{\alpha_2} \leq 2) \right| E(|\xi^{z_1,g_1}_s|) + \frac{1}{n} \sum_{s=1}^{n} E(|\xi^{z_1,g_1}_s - \xi^{z_2,g_2}_s| | F_{s-1})
\]

\[
\leq \frac{1}{n} |n (\alpha_2 - \alpha_1) + 1| + \frac{1}{n} \sum_{s=1}^{n} E(|\xi^{z_1,g_1}_s - \xi^{z_2,g_2}_s| | F_{s-1})
\]

\[
\leq |\alpha_1 - \alpha_2| + \frac{1}{n} \sum_{s=1}^{n} E(|\xi^{z_1,g_1}_s - \xi^{z_2,g_2}_s| | F_{s-1}).
\]

On \( F_n \) we have for the last sum that

\[
\frac{1}{n} \sum_{s=1}^{n} E(|\xi^{z_1,g_1}_s - \xi^{z_2,g_2}_s| | F_{s-1})
\]

\[
\leq \frac{1}{n} \sum_{s=1}^{n} \left| F_{\alpha_1} (\xi^{z_1}_s + (g_1(\frac{s}{T})) Y_{s-1}) - F_{\alpha_2} (\xi^{z_2}_s + (g_2(\frac{s}{T})) Y_{s-1}) \right|
\]

\[
+ \frac{1}{n} \sum_{s=1}^{n} \left| F_{\alpha_1} (\xi^{z_2}_s + (g_1(\frac{s}{T})) Y_{s-1}) - F_{\alpha_2} (\xi^{z_2}_s + (g_2(\frac{s}{T})) Y_{s-1}) \right|
\]

\[
\leq \sup_{u,x} \left| z_1 - z_2 \right| + \sup_{u,x} \left| \frac{1}{n} \sum_{s=1}^{n} \left| ((g_1 - g_2)(\frac{s}{T})) Y_{s-1} \right| \right|
\]

\[
\leq \frac{\|f\|_{\infty}}{m_{*}} |z_1 - z_2| + \frac{\|f\|_{\infty}}{m_{*}} \sum_{k=1}^{p} \|g_{1k} - g_{2k}\| n \sqrt{\frac{1}{n} \sum_{s=-p}^{n} Y_{s}^2} \leq (1 + C_0) \frac{\|f\|_{\infty}}{m_{*}} \epsilon. \quad (19)
\]

Thus, on \( F_n \) we have for \( h_1 = (\alpha_1, z_1, g_1), h_2 = (\alpha_2, z_2, g_2) \in H \) with \( d(h_1, h_2) \leq \epsilon \) and \( \epsilon \geq \frac{1}{n} \) that

\[
V_n(\nu_n(h_1) - \nu_n(h_2)) \leq \frac{1}{n} \sum_{s=1}^{n} E(\|\eta^{h_1}_s - \eta^{h_2}_s\| | F_{s-1})^2 \leq K^* \epsilon. \quad (20)
\]

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

$$P\left( \sup_{\alpha \in [0,1], z \in [-L,L], g \in G^p} |\nu_n(\alpha, z, g) - \nu_n(\alpha, z, 0)| \geq \eta \right)$$

$$\leq P\left( (F_n)^c \right) + P\left( \sup_{d(h_1, h_2) \leq C_n m_n^{-1}} |\nu_n(h_1) - \nu_n(h_2)| \geq \eta, F_n \right),$$

where $F_n = \{ \frac{1}{n} \sum_{s=1-p}^n Y_s^2 \leq C_0 \}$ for some $C_0 > 0$. We will see below that $\frac{1}{n} \sum_{s=1-p}^n Y_s^2 = O_P(1)$ as $n \to \infty$. Therefore, for any given $\epsilon > 0$ we can choose $C_0$ such that $P\left( (F_n)^c \right) \leq \epsilon$ for $n$ large enough. An application of Lemma 1 now gives the assertion. Similar arguments as those leading to (19) show that $\int_1^1 \sqrt{\log N_B(u^2, G)} du < \infty$ implies $\int_1^1 \sqrt{\log D_B(u^2, H)} du < \infty$ (see also (40)). It remains to show that in fact $\frac{1}{n} \sum_{s=1-p}^n Y_s^2 = O_P(1)$. To this end we show that

$$E\left[ \frac{1}{n} \sum_{s=1}^n Y_s^2 \right] < \infty, \quad (21)$$

$$\text{Var}\left[ \frac{1}{n} \sum_{s=1}^n Y_s^2 \right] = o(1). \quad (22)$$

These two facts can be seen by direct calculations as demonstrated now. First we consider (21). We have

$$EY_s^2 = E\left[ \sum_{j=-\infty}^s a_{s,T}(s-j) \epsilon_j \right]^2 = E \sum_{j=-\infty}^s \sum_{k=-\infty}^s a_{s,T}(s-j)a_{s,T}(s-k) \epsilon_j \epsilon_k$$

$$= \sum_{j=-\infty}^s a_s^2(s-j) \leq \sum_{j=0}^{\infty} \left( \frac{K}{\ell(j)} \right)^2 < C < \infty, \quad (23)$$
for some $C > 0$, where we were using assumption (ii). This implies (21). Next we indicate (22). Straightforward calculations show that

\[
E(Y_t^2 Y_s^2) - E(Y_t^2) E(Y_s^2)
= \sum_{i=-\infty}^{t} \sum_{j=-\infty}^{t} \sum_{s=-\infty}^{s} \sum_{k=-\infty}^{s} a_{t,T}(t-i)a_{t,T}(t-j)a_{s,T}(s-k)a_{s,T}(s-\ell)E(\epsilon_i \epsilon_j \epsilon_k \epsilon_\ell)
\]

\[
- \sum_{i=-\infty}^{t} a_{t,T}^2(t-i) \sum_{k=-\infty}^{s} a_{s,T}^2(s-k)
\]

\[
= (E\epsilon_t^4 - 2) A(t, s) + 2B(t, s),
\]

where

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

\[
B(t, s) = \left( \sum_{i=0}^{\infty} a_{t,T}(i)a_{s,T}(i + |t - s|) \right)^2.
\]

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

\[
A(t, s) \leq K^4 \sum_{i=0}^{\infty} \frac{1}{\ell^2(i)} \frac{1}{\ell^2(i + |t - s|)} \leq K^4 \sum_{i=0}^{\infty} \frac{1}{\ell(i)} \frac{1}{\ell(i + |t - s|)}
\]

\[
\leq K^4 \frac{1}{\ell(|t - s|)} \sum_{i=0}^{\infty} \frac{1}{\ell(i)} < C^* \frac{1}{\ell(|t - s|)},
\]

and similarly

\[
|B(t, s)| \leq \left[ K^2 \sum_{i=0}^{\infty} \frac{1}{\ell(i)} \frac{1}{\ell(i + |t - s|)} \right]^2 \leq C^* \frac{1}{\ell(|t - s|)}.
\]

Now we have

\[
\left| \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} A(t, s) \right| \leq \frac{C^*}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \frac{1}{\ell(|t - s|)} \leq \frac{C^*}{n} \sum_{k=0}^{n-1} \frac{2(n-k)}{n} \frac{1}{\ell(k)}
\]

\[
\leq \frac{2C^*}{n} \sum_{k=0}^{n-1} \frac{1}{\ell(k)} = O\left( \frac{1}{n} \right) \text{ as } n \to \infty.
\]

This implies that $\text{Var}\left[ \frac{1}{n} \sum_{s=1}^{n} Y_s^2 \right] = O(1/n)$ which is (22).
5.3 Proof of Theorem 2

Assume w.l.o.g. that $a = 0$. 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 3.

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 $k \geq 3$ 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 convergence of the covariances $\text{cov}(Z_n(h_1), Z_n(h_2))$. 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 $u \leq 0 \text{ and } u > b$, $i = 1, 2$.

We have

$$\text{cov}(Z_n(h_1), Z_n(h_2)) = \frac{1}{n} \sum_{s_1=1}^{n} \sum_{s_2=1}^{n} h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_2}{T} \right) \text{cov}(Y_{s_1}, Y_{s_2})$$

$$= \frac{1}{n} \sum_{s_1=1}^{n} \sum_{k=s_1-n}^{s_1-1} h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_1-k}{T} \right) \text{cov}(Y_{s_1}, Y_{s_1-k})$$

$$= \frac{1}{n} \sum_{s_1=1}^{n} \sum_{|k| \leq \sqrt{n}} h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_1-k}{T} \right) \text{cov}(Y_{s_1}, Y_{s_1-k}) + R_{1n} \tag{25}$$

where for $n$ sufficiently large

$$|R_{1n}| \leq \frac{1}{n} \sum_{s_1=1}^{n} \sum_{|k| > \sqrt{n}} \left| h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_1-k}{T} \right) \right| \left| \text{cov}(Y_{s_1}, Y_{s_1-k}) \right|.$$

From Proposition 5.4 of Dahlhaus and Polonik (2009) we obtain that $\sup_s |\text{cov}(Y_s, Y_{s-k})| \leq \frac{K}{\sqrt{|k|}}$ for some constant $K$. Since both $h_1$ and $h_2$ are bounded and $\sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{|k|}} < \infty$, we can conclude that $R_{1n} = o(1)$. The main term in (25) can be approximated as

$$\frac{1}{n} \sum_{s_1=1}^{n} \sum_{|k| \leq \sqrt{n}} h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_1-k}{T} \right) c \left( \frac{s_1}{T}, k \right) + R_{2n} \tag{26}$$

where

$$|R_{2n}| \leq \frac{1}{n} \sum_{s_1=1}^{n} \sum_{|k| \leq \sqrt{n}} \left| h_1 \left( \frac{s_1}{T} \right) \cdot h_2 \left( \frac{s_1-k}{T} \right) \right| \left| \text{cov}(Y_{s_1}, Y_{s_1-k}) - c \left( \frac{s_1}{T}, k \right) \right|.$$
Proposition 5.4 of Dahlhaus and Polonik (2009) also says that for $|k| \leq \sqrt{n}$ we have
\[
\sum_{s_1=0}^{n} \left| \text{cov}(Y_{s_1}, Y_{s_1-k}) - c\left(\frac{s_1}{n}, k\right) \right| \leq K(1 + \frac{|k|}{n})
\]
for some $K > 0$. Applying this result we obtain that
\[
|R_{2,n}| \leq \frac{1}{n} \sum_{s_1=1}^{n} \sum_{k=-\sqrt{n}}^{\sqrt{n}} \left| h_1\left(\frac{s_1}{n}\right) \cdot h_2\left(\frac{s_1-k}{T}\right) \right| \left| \text{cov}(Y_{s_1}, Y_{s_1-k}) - c\left(\frac{s_1}{n}, k\right) \right|
\leq K_1 \frac{1}{n} \sum_{k=-\sqrt{n}}^{\sqrt{n}} \sum_{s_1=1}^{n} \left| \text{cov}(Y_{s_1}, Y_{s_1-k}) - c\left(\frac{s_1}{n}, k\right) \right| \leq K_1 \frac{1}{n} \sum_{k=-\sqrt{n}}^{\sqrt{n}} (1 + \frac{|k|}{\ell(k)}) = o(1)
\]
as $n \to \infty$. Next we replace $h_2\left(\frac{s_1-k}{T}\right)$ in the main term of (26) by $h_2\left(\frac{s_1}{T}\right)$. The approximation error can be bounded by
\[
\frac{1}{n} \sum_{s_1=1}^{n} \sum_{|k| \leq \sqrt{n}} \left| h_1\left(\frac{s_1}{n}\right) \right| \left| h_2\left(\frac{s_1-k}{T}\right) - h_2\left(\frac{s_1}{T}\right) \right| \frac{K}{\ell(k)} = o(1).
\]
Here we are using the fact that $\sup_u |c(u, k)| \leq \frac{K}{\ell(k)}$ (see 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
\[
\text{cov}(Z_n(h_1), Z_n(h_2)) = \frac{1}{n} \sum_{s_1=1}^{[\alpha_1 n]} h_1\left(\frac{s_1}{n}\right) \cdot h_2\left(\frac{s_1}{T}\right) \sum_{k \leq \sqrt{n}} c\left(\frac{s_1}{n}, k\right) + o(1).
\]
Since $S(u) = \sum_{k=-\infty}^{\infty} c\left(\frac{s_1}{T}, k\right) < \infty$ we also have
\[
\text{cov}(Z_n(h_1), Z_n(h_2)) = \sum_{k=-\infty}^{\infty} \frac{1}{n} \sum_{s_1=1}^{n} h_1\left(\frac{s_1}{n}\right) \cdot h_2\left(\frac{s_1}{T}\right) c\left(\frac{s_1}{n}, k\right) + o(1).
\]
Finally, we utilize the fact that $TV(c(\cdot, k)) \leq \frac{K}{\ell(k)}$ 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 $s_1$ by the integral. Recalling that
\[
n = \lfloor bT \rfloor - \lceil aT \rceil + 1 = (b-a)T + O\left(\frac{1}{T}\right)
\]
gives the assertion.
5.4 Proof of Lemma 2

For ease of notation we assume w.l.o.g. that \( a = 0 \). By utilizing multilinearity of cumulants we obtain that

\[
\text{cum}(Z_n(h_1), \ldots, Z_n(h_k)) = n^{-\frac{k}{2}} \sum_{s_1, s_2, \ldots, s_k=1}^{n} h_1(\frac{s_1}{T}) \cdots h_k(\frac{s_k}{T}) \text{cum}(Y_{s_1}, \ldots, Y_{s_k}).
\]

In order to estimate \( \text{cum}(Y_{s_1}, \ldots, Y_{s_k}) \) we utilize the special structure \( Y_{s_i} \). Since the \( \epsilon_j \) are independent we again obtain by using multilinearity of the cumulants together with the fact that \( \text{cum}(\epsilon_{j_1}, \ldots, \epsilon_{j_k}) = 0 \) unless all the \( j_\ell, \ell = 1, \ldots, k \) are equal, that

\[
\text{cum}(Y_{s_1}, \ldots, Y_{s_k}) = \sum_{j=0}^{\min\{s_1, \ldots, s_k\}} a_{s_1,T}(s_1 - j) \cdots a_{s_k,T}(s_k - j) \text{cum}(\epsilon_j, \ldots, \epsilon_j)
\]

\[
= \text{cum}_k(\epsilon_1) \sum_{j=0}^{\min\{s_1, \ldots, s_k\}} a_{s_1,T}(s_1 - j) \cdots a_{s_k,T}(s_k - j).
\] (27)

Thus \( |\text{cum}(Y_{s_1}, \ldots, Y_{s_k})| \leq |\text{cum}_k(\epsilon_1)| \sum_{j=0}^{\infty} \prod_{i=1}^{k} K_{\ell(s_i - j)} \), and consequently,

\[
|\text{cum}(Z_n(h_1), \ldots, Z_n(h_k))| \leq n^{-\frac{k}{2}} |\text{cum}_k(\epsilon_1)| \sum_{j=-\infty}^{\infty} \prod_{i=1}^{k} \left[ \sum_{s_i=0}^{n} |h_i(\frac{s_i}{T})| \frac{K}{\ell(s_i - j)} \right]
\]

\[
= n^{-\frac{k}{2}} |\text{cum}_k(\epsilon_1)| \sum_{j=-\infty}^{\infty} \prod_{i=1}^{k} \left[ \sum_{s_i=0}^{n} |h_i(\frac{s_i}{T})| \frac{K}{\ell(s_i - j)} \right]
\]

\[
\times \prod_{i=3}^{k} \left[ \sum_{s_i=0}^{n} |h_i(\frac{s_i}{T})| \frac{K}{\ell(s_i - j)} \right].
\]

Utilizing Cauchy-Schwarz inequality we have for the last product

\[
\prod_{i=3}^{k} \left[ \sum_{s_i=0}^{n} |h_i(\frac{s_i}{T})| \frac{1}{\ell(s_i - j)} \right] \leq \prod_{i=3}^{k} \sqrt{\sum_{s_i=0}^{n} |h_i(\frac{s_i}{T})|^2} \sqrt{\sum_{s_i=0}^{n} \left( \frac{K}{\ell(s_i - j)} \right)^2}
\]

19
\[
\leq n^{k-1} \prod_{i=3}^{k} \|h_i\|_n \sqrt{\sum_{s=-\infty}^{\infty} \left( \frac{K}{\ell(s)} \right)^2} \\
\leq K_0^{k-2} n^{k-1} \prod_{i=3}^{k} \|h_i\|_n. 
\] (28)

where we used the fact that
\[
\sqrt{\sum_{s=-\infty}^{\infty} \left( \frac{K}{\ell(s)} \right)^2} \leq \sum_{s=-\infty}^{\infty} \frac{K}{\ell(s)} \leq K_0 
\]
for some \( K_0 < \infty \). Notice that the bound (28) does not depend on the index \( j \) anymore, so that
\[
|\text{cum}(Z_n(h_1), \ldots, Z_n(h_k))| \leq K_0^{k-2} n^{-1} \prod_{i=3}^{k} \|h_i\|_n |\text{cum}_k(\epsilon_1)| \sum_{j=-\infty}^{\infty} \prod_{i=1}^{2} \left[ \sum_{s_i=0}^{n} |h_i(s_i)| \right] \frac{K}{\ell(s_i - j)}.
\]

As for the last sum, we have by again using the fact that
\[
\sum_{j=-\infty}^{\infty} \frac{K}{\ell(s_1 - j) \ell(s_2 - j)} \leq \frac{K^*}{\ell(s_1 - s_2)} 
\]
(cf. (24)) for some \( K^* > 0 \), and the Cauchy-Schwarz inequality that
\[
\leq K^* \sqrt{\sum_{s_1=0}^{n} \sum_{s_2=0}^{n} h_1(s_1) h_2(s_2)} \frac{1}{\ell(s_1 - s_2)} \sqrt{\sum_{s_1=0}^{n} \sum_{s_2=0}^{n} h_1(s_1) h_2(s_2)} \frac{1}{\ell(s_1 - s_2)}.
\]

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 \( i = 1, \ldots, k \), then, instead of the estimate (28), we have
\[
\prod_{i=3}^{k} \left[ \sum_{s_i=0}^{n} |h_i(\frac{s_i}{\ell})| \frac{1}{\ell(s_i-j)} \right] \leq M^{k-2} \prod_{i=3}^{k} \sum_{s_i=0}^{n} \frac{1}{\ell(s_i-j)} \leq (MK_{0})^{k-2}
\]

with \( K_{0} = \sum_{s=-\infty}^{\infty} \frac{1}{\ell(s)} \).

5.5 Proof of Lemma 3

First observe that since \(|\text{cum}_k(\epsilon_s)| \leq E(|\epsilon_s|^k) + \sum_{j=1}^{k-1} (k-1) E|\epsilon_s|^j |\text{cum}_j(\epsilon_s)|\) it is straightforward to see that assumption \( E|\epsilon_s|^k \leq C^k \) implies that \(|\text{cum}_k(\epsilon_s)| \leq 3^k C^k\). It follows by utilizing Lemma 2 that

\[
\Psi_{Z_n(h)}(t) = \log E e^{tZ_n(h)} = \frac{1}{K_{0}} \sum_{k=1}^{\infty} \frac{t^k}{k!} \text{cum}_k(Z_n(h)) \\
\leq \frac{1}{K_{0}} \sum_{k=1}^{\infty} \frac{t^k}{k!} (3CK_{0} ||h_j||_n)^k = K_{0}^{-1} \left(e^{3tK_{0} ||h||_n} - 1 \right).
\]

We obtain for any \( t > 0 \)

\[
P[|Z_n(h)| > \eta] \leq 2 e^{-t\eta} E(e^{Z_n(h)}) = 2 \exp\{-t\eta\} \exp\{\Psi_{Z_n(h)}(t)\} \\
\leq 2 \exp\{-t\eta + K_{0}^{-1} (e^{3tC||h||_n} - 1)\}.
\]

Choosing \( t = \frac{1}{3C||h||_n} \) gives the assertion:

\[
P[|Z_n(h)| > \eta] \leq 2 \exp \left\{ - \frac{\eta}{3C||h||_n} \right\} e^{K_{0}^{-1}(e-1)}.
\]

5.6 Proof of Theorem 3

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 \( c_1 \exp\{-c_2 \left( \frac{\eta}{||h||_n} \right)^2\} \). 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 (13) just makes sure that the upper limit in the integral from (14) is larger than the lower limit.
5.7 Proof of Theorem 4

First we indicate how the two processes studied in this paper enter the picture, and simul-
taneously we provide an outline of the proof.

Recall that we have relabeled the observations $Y_t$ inside the rescaled interval $[a, b] \subset [0, 1]$, i.e. $aT \leq t \leq bT$ to $Y_s$, $s = 1, \ldots, n$. Also, $Y_{s-k}$ with $k \geq s$ denotes $Y_{[aT]-(k-s)}$. Let

$$
\hat{H}_n(\alpha, z) = \frac{1}{n} \sum_{s=1}^{\lfloor \alpha n \rfloor} 1(\tilde{\eta}_s \leq z) \quad \text{and} \quad F_n(\alpha, z) = \frac{1}{n} \sum_{s=1}^{\lfloor \alpha n \rfloor} 1(\sigma(a + \frac{s}{T}) \epsilon_s \leq z).
$$

With this notation

$$
\sqrt{n} \left( \hat{G}_{n,\gamma}(\alpha) - G_{n,\gamma}(\alpha) \right) = \left[ \sqrt{n} \left( F_n(\alpha, \tilde{q}_\gamma) - H_n(\alpha, \tilde{\eta}_\gamma) \right) - \sqrt{n} \left( F_n(\alpha, -\tilde{q}_\gamma) - H_n(\alpha, -\tilde{\eta}_\gamma) \right) \right] \\
- \left[ \sqrt{n} \left( F_n(\alpha, \tilde{q}_\gamma) - F_n(\alpha, q_\gamma) \right) - \sqrt{n} \left( F_n(\alpha, -\tilde{q}_\gamma) - F_n(\alpha, -q_\gamma) \right) \right] \\
=: \quad I_n(\alpha) - II_n(\alpha) \quad (29)
$$

with $I_n(\alpha)$ and $II_n(\alpha)$, respectively, denoting the two quantities inside the two $[\cdot]$-brackets. The assertion of Theorem 4 follows from

$$
\sup_{\alpha \in [0,1]} |I_n(\alpha)| = o_P(1) \quad \text{and} \quad \sup_{\alpha \in [0,1]} |II_n(\alpha) - c(\alpha) (G_{n,\gamma}(1) - EG_{n,\gamma}(1))| = o_P(1). \quad (30)
$$

Property (31) can be shown by using empirical process theory based on independent, but not identically distributed random variables. Verification of (30) involves both residual empirical processes and weighted sum processes. To see the latter, observe that for each $z \in \mathbb{R}$

$$
\hat{H}_n(\alpha, z) = \frac{1}{n} \sum_{s=1}^{\lfloor \alpha n \rfloor} 1(\tilde{\eta}_s \leq z) = \frac{1}{n} \sum_{s=1}^{\lfloor \alpha n \rfloor} 1(\sigma(a + \frac{s}{T}) \epsilon_s \leq \sigma(a + \frac{s}{T}) \epsilon_s - \tilde{\eta}_s + z) \\
= \frac{1}{n} \sum_{s=1}^{\lfloor \alpha n \rfloor} 1(\sigma(a + \frac{s}{T}) \epsilon_s \leq \left((\theta - \hat{\theta})(a + \frac{s}{T})\right)'Y_{s-1} + z).
$$

Recall that by assumption $\hat{\theta}_n(\cdot) - \theta(\cdot) \in \mathcal{G}^p$, where for $g = (g_1, \ldots, g_p)' : [a, b]^p \to \mathbb{R}$ we
write \( \{ g \in \mathcal{G}^p \} \) for \( \{ g_i \in \mathcal{G}, i = 1, \ldots, p \} \). We also write
\[
F_n(\alpha, z, g) = \frac{1}{n} \sum_{s=1}^{\lfloor an \rfloor} 1 \{ \sigma(\frac{t}{T}) \epsilon_t \leq g'(\frac{t}{T}) Y_t + z \}.
\]

By taking into account that by assumption \( \hat{\theta} - \theta = O_P(m^{-1}) \) the above implies that with probability tending to one we have for any \( C > 0 \) that
\[
\sup_{\alpha \in [0,1], z \in [-L, L]} \sqrt{n} \left| F_n(\alpha, z) - \hat{H}_n(\alpha, z) \right| \leq \sup_{\alpha \in [0,1], z \in [-L, L], g \in \mathcal{G}^p, \| g \|_\infty \leq C m^{-1}} \sqrt{n} \left| F_n(\alpha, z) - F_n(\alpha, z, g) \right|. \tag{32}
\]

Thus, if \( L \) is such that \( q \gamma \in (-L, L) \) and \( P(\hat{q} \gamma \in [-L, L]) \to 1 \), as \( n \to \infty \), then the right hand side in (32) being \( o_P(1) \) implies that \( \sup_{\alpha \in [0,1], z \in [-L, L], g \in \mathcal{G}^p, \| g \|_\infty \leq C m^{-1}} \sqrt{n} \left| F_n(\alpha, z) - F_n(\alpha, z, g) \right| \).

In order to control the right hand side in (32), an appropriate centering is needed. Let \( F_u(z) = F(\frac{z}{\sigma(u)}) \) denote the cdf corresponding to the pdf \( f_u(z) \), defined above Theorem 4. Let
\[
E_n(\alpha, z, g) := \frac{1}{n} \sum_{s=1}^{\lfloor an \rfloor} \mathbb{E} \left[ 1 \{ \sigma(\frac{t}{T}) \epsilon_t \leq g'(\frac{t}{T}) Y_t + z \} \mid \epsilon_{s-1}, \epsilon_{s-2}, \ldots \right]
\]
\[
= \frac{1}{n} \sum_{s=1}^{\lfloor an \rfloor} F_{a+\frac{s}{T}} \left( (g(a + \frac{s}{T}))' Y_{s-1} + z \right) \tag{33}
\]
and let \( 0 = (0, \ldots, 0)' \in \mathcal{G}^p \) denote the \( p \)-vector of null-functions. Write
\[
F_n(\alpha, z) - F_n(\alpha, z, g) = \left[ (F_n(\alpha, z, 0) - E_n(\alpha, z, 0)) - (F_n(\alpha, z, g) - E_n(\alpha, z, g)) \right] + \left[ E_n(\alpha, z, 0) - E_n(\alpha, z, g) \right]
\]
\[
= T_{n1}(\alpha, z, g) + T_{n2}(\alpha, z, g) \tag{34}
\]
with \( T_{n1} \) and \( T_{n2} \) denoting the two terms inside the two \([\ ]\) brackets. This centering makes \( T_{n1} \) a sum of martingale differences (cf. proof of Lemma 1 given above). The quantity \( \nu_n(\alpha, z, g) = \sqrt{n} \left( F_n(\alpha, z, g) - E_n(\alpha, z, g) \right) \) is the residual empirical process discussed in section 2. As for \( T_{n2}(\alpha, z, g) \) notice that we have
\[
\sqrt{n} T_{n2}(\alpha, z, g) = \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor an \rfloor} \left[ F_{a+\frac{s}{T}} \left( z + (g(a + \frac{s}{T}))' Y_{s-1} \right) - F_{a+\frac{s}{T}}(z) \right]
\]
\[
\frac{1}{\sqrt{n}} \sum_{s=1}^{\lceil n\alpha \rceil} f_{a + \frac{z}{T}}(z) (g(a + \frac{z}{T}))' Y_{s-1} + \frac{1}{\sqrt{n}} \sum_{s=1}^{\lceil n\alpha \rceil} (f_{a + \frac{z}{T}}(\zeta_s) - f_{a + \frac{z}{T}}(z)) (g(a + \frac{z}{T}))' Y_{s-1} \tag{35}
\]

with \(\zeta_s\) between \(z\) and \((g(a + \frac{z}{T}))' Y_{s-1}\). The second term in (35) is a remainder term, while the first term is a weighted sum of the \(Y_s\)'s.

The above motivates that (30) can be verified by appropriate control of both the residual empirical process and a weighted sum process.

After this outline we now present the missing details of the proof. Without loss of generality we assume that \(aT > p\) so that the residuals are all well defined, and we assume \(a = 0\). Choose \(L\) large enough such that \(q_\gamma \in (-L, L)\). As outlined above, it suffices to show that both

\[
\sup_{\alpha \in [0, 1]} |I_n(\alpha)| = o_P(1) \quad \text{and} \quad \sup_{\alpha \in [0, 1]} |II_n(\alpha) - c(\alpha) (G_{n, \gamma}(1) - EG_{n, \gamma}(1))| = o_P(1).
\]

**Proof of sup\(\alpha \in [0, 1]\) |\(I_n(\alpha)\)| = \(o_P(1)\).** It follows from (32) and (34) that together

\[
\sup_{\alpha \in [0, 1], z \in [-L, L], g \in G_p} \left| \sqrt{n} T_{n1}(\alpha, z, g) \right| = o_P(1) \quad \text{and} \quad \sup_{\alpha \in [0, 1], z \in [-L, L], g \in G_p} \left| \sqrt{n} T_{n2}(\alpha, z, g) \right| = o_P(1) \tag{36}
\]

imply the desired result, provided we can show that \(P(\hat{q}_\gamma \notin [-L, L]) = o(1)\) as \(n \to \infty\).

Since \(q_\gamma \in (-L, L)\), the desired property follows from consistency of \(\hat{q}_\gamma\) as an estimator for \(q_\gamma\). This will be shown at the end of this proof.

Notice that by assumption (iv), for any given \(\epsilon > 0\) and \(n\) large enough there exists a constant \(C_\epsilon > 0\) such that with

\[
A_n(\epsilon) := \{ \| \hat{\theta}_{\epsilon,k}(u) - \theta_k(u) \|_n \geq C_\epsilon m_n^{-1} \text{ for all } k = 1, \ldots, p \}
\]

we have

\[
P(A_n(\epsilon)) \leq \epsilon.
\]

Thus, on \(A_n^C(\epsilon)\) we can assume that \(\|g\|_n \leq C_\epsilon m_n^{-1}\), which means that on \(A_n^C(\epsilon)\) we can
assume $d((\alpha, z, g), (\alpha, z, 0)) \leq C\epsilon m_n^{-1}$. Thus, on $A^\epsilon_n(\epsilon)$

$$\sup_{\alpha \in [0, 1], z \in [-L, L], g \in \mathcal{G}} |\nu_n(\alpha, z, g) - \nu_n(\alpha, z, 0)| \leq \sup_{d(h_1, h_2) \leq C\epsilon m_n^{-1}} |\nu_n(h_1) - \nu_n(h_2)|.$$  

It is shown in the proof of Theorem 1 that $\frac{1}{n} \sum_{s=n-p}^n Y_s^2 = O_P(1)$, and thus $P(F_n) = o(1)$. By definition of $T_{n2}(\alpha, z, g)$ it remains to show that and for $\eta > 0$ we have

$$P\left( \sup_{d(h_1, h_2) \leq C\epsilon m_n^{-1}} |\nu_n(h_1) - \nu_n(h_2)| \geq \eta, F_n \right) = o(1). \quad (38)$$

This, however, is an immediate application of Theorem 1, and (36) is verified. Now we consider (37). Again we assume that we are on $A^\epsilon_n(\epsilon)$. We have already seen in (35) that

$$\sqrt{n} T_{n2}(\alpha, z, g) = \frac{1}{\sqrt{n}} \sum_{s=1}^{[n\alpha]} f_{a+\frac{s}{n}}(z) \left( g\left(\frac{s}{n}\right) \right)' Y_{s-1} + \frac{1}{\sqrt{n}} \sum_{s=1}^{[n(\alpha+1)]} \left( f_{a+\frac{s}{n}}(z) - f_{a+\frac{s}{n}}(z) \right) \left( g\left(\frac{s}{n}\right) \right)' Y_{s-1} \quad (39)$$

with $\zeta_s$ between $z$ and $z + \left( g\left(\frac{s}{n}\right) \right)' Y_{s-1}$. The second term in (39) will be treated below. The first sum on the right hand side can be written as $\sum_{k=1}^p Z_{k,n}(h)$ where

$$Z_{k,n}(h) = \frac{1}{\sqrt{n}} \sum_{s=1}^n h\left(\frac{s}{n}\right) Y_{s-k}, \quad k = 1, \ldots, p. \quad (40)$$

for $h \in \mathcal{H}$ with $\mathcal{H} = \{h_{\alpha,z,g}(u) = 1_{\alpha}(u) f_u(z) g(u), \alpha \in [0, 1], z \in \mathbb{R}, g \in \mathcal{G}\}$ where we use the shorthand notation $1_{\alpha}(u) = 1(u \leq \alpha b)$. We will apply Theorem 3 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 $\sup_{\alpha,z,g} \|h_{\alpha,z,g}\|_n^2 \leq \sup_{\alpha,z} |f_u(z)| m_n^{-1} = o(1)$. Since $E Z_n(h_{\alpha,z,g}) = 0$, Theorem 3 implies the result if we have shown that $\mathcal{H}$ has a finite covering integral.

The finiteness of the covering integral of $\mathcal{H}$ with respect to $\| \cdot \|_n$ follows by standard arguments. In fact, it is not difficult to see that for some $C_0 > 0$

$$\log N(C_1 \delta, \mathcal{H}) \leq -C_0 \log \epsilon + \log N(\epsilon, \mathcal{G}). \quad (41)$$

Our assumptions now allow an application of Theorem 3, showing that the first term of (39) converges to zero in probability uniformly in $(\alpha, z, g)$.
Now we treat the second term in (39). 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 (39) by

\[
\frac{c}{\sqrt{n}} \sum_{s=1}^{n} \left| (g(s))' Y_{s-1} \right|^2 \leq \frac{c}{\sqrt{n}} \sum_{s=1}^{n} \left( \sum_{k=1}^{p} g_k(s) \right)^2 \sum_{j=0}^{p} Y_{s-j}^2 \\
\leq cp \sup_{-p \leq t \leq T} Y_t^2 \sqrt{n} \sum_{k=1}^{p} \|g_k\|^2_n = cp \frac{\sqrt{n}}{m_n^2} O_P(\log n) = o_P(1).
\]

where the last inequality uses the fact that \(\sup_{-p \leq t \leq T} Y_t^2 = O_P(\log n)\), 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) (G_{n,\gamma}(1) - EG_{n,\gamma}(1)) | = o_P(1).** Define

\[
\tilde{F}_n(\alpha, z) := EF_n(\alpha, z) = \frac{1}{n} \sum_{s=1}^{[\alpha n]} F_{n+\frac{s}{T}}(z).
\]

(Recall that \(F_n(z) = F(\frac{z}{\sigma(u)})\).) We can write

\[
II_n(\alpha) = \sqrt{n} \left( (F_n - \tilde{F}_n)(\alpha, q_{\gamma}) - (F_n - \tilde{F}_n)(\alpha, q_{\gamma}) \right) \\
+ \sqrt{n} \left( (F_n - \tilde{F}_n)(\alpha, -q_{\gamma}) - (F_n - \tilde{F}_n)(\alpha, -q_{\gamma}) \right) \\
+ \sqrt{n} \left( (\tilde{F}_n(\alpha, q_{\gamma}) - \tilde{F}_n(\alpha, q_{\gamma})) - \sqrt{n} (\tilde{F}_n(\alpha, -q_{\gamma}) - \tilde{F}_n(\alpha, -q_{\gamma})) \right)
\]

(42)

The process \(\nu_n(\alpha, z, 0) = \sqrt{n} (F_n - \tilde{F}_n)(\alpha, z)\) 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 \(\alpha\) with respect to \(d_n(v, w) = |\tilde{F}_n(1, v) - \tilde{F}_n(1, w)|\), i.e. for every \(\eta > 0\) there exists an \(\epsilon > 0\) with

\[
\limsup_{n \to \infty} P \left[ \sup_{\alpha \in [0,1], \tilde{d}_n(z_1, z_2) \leq \epsilon} |\nu_n(\alpha, z_1, 0) - \nu_n(\alpha, z_2, 0)| > \eta \right] = 0.
\]

(45)

In fact, with \(\tilde{d}_n(\alpha_1, z_1), (\alpha_2, z_2) = |\alpha_1 - \alpha_2| + d_n(z_1, z_2)\) we have

\[
\sup_{\alpha \in [0,1]} \sup_{z_1, z_2, e \in \mathbb{R}, d_n(z_1, z_2) \leq \epsilon} |\nu_n(\alpha, z_1, 0) - \nu_n(\alpha, z_2, 0)| \\
\leq \sup_{\alpha_1, \alpha_2 \in [0,1], z_1, z_2, e \in \mathbb{R}, \tilde{d}_n((\alpha_1, z_1), (\alpha_2, z_2)) \leq \epsilon} |\nu_n(\alpha_1, z_1, 0) - \nu_n(\alpha_2, z_2, 0)|.
\]

26
Thus, (45) follows from asymptotic stochastic \( \overline{d}_n \)-equicontinuity of \( \nu_n(\alpha, z, 0) \). This in turn follows from a proof similar to, but simpler than, the proof of Lemma 1. In fact, it can be seen from (19) that for \( g_1 = g_2 = 0 \) we simply can use the metric \( \overline{d}( (\alpha_1, z_1), (\alpha_2, z_2) ) = |\alpha_1 - \alpha_2| + d_n(z_1, z_2) \) in the estimation of the quadratic variation, which in the simple case of \( g_1 = g_2 = 0 \) amounts to the estimation of the variance, because the randomness only comes in through the \( \epsilon_s \). With this modification the proof of the \( \overline{d}_n \)-equicontinuity of \( \nu_n(\alpha, z_1, 0) \) follows the proof of Lemma 1.

Thus, if \( \hat{q}_\gamma \) is consistent for \( q_\gamma \) with respect to \( \overline{d}_n \), then it follows that both (42) and (43) are \( o_P(1) \). We now prove this consistency of \( \hat{q}_\gamma \).

First observe that \( \frac{1}{b} \int_0^b F(\frac{\sigma_\gamma}{\sigma(u)}) \, dv = \int_0^1 F(\frac{\sigma_\gamma}{\sigma(bu)}) \, du \) is close to \( F_n(1, q_\gamma) \). In fact, since by assumption \( u \to F(\frac{\sigma_\gamma}{\sigma(u)}) \) is of bounded variation, the difference is of the order \( O(1/n) \). Consequently we have

\[
| (1 - \gamma) - ( F_n(1, q_\gamma) - F_n(1, -q_\gamma) ) | = \left| \Psi_{0,b}(1, q_\gamma) - ( F_n(1, q_\gamma) - F_n(1, -q_\gamma) ) \right| \leq c n^{-1} \tag{46}
\]

for some \( c \geq 0 \). In fact (46) holds uniformly in \( \gamma \). This follows from the fact that the functions \( u \to F_u(\frac{z}{\sigma(u)}) \) are Lipschitz continuous uniformly in \( z \). (Notice that in the case where \( \sigma(\cdot) \equiv \sigma_0 \) is constant on \([a, b]\) then \( \Psi_{0,b}(z) = \int_0^1 F(\frac{\sigma_0}{\sigma(bu)}) \, du - \int_0^1 F(\frac{\sigma_0}{\sigma(bu)}) \, du = F(\frac{\sigma_0}{\sigma_0}) - F(\frac{\sigma_0}{\sigma_0}) = F_n(1,1) - F_n(1,-1) \). 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

\[
\overline{d}_n(\hat{q}_\gamma, q_\gamma) = | F_n(1, \hat{q}_\gamma) - F_n(1, q_\gamma) | = o_P(1). \tag{47}
\]

By assumption \( \Psi_{0,b}(z) \) is a strictly monotonic function in \( z \). Together with (46) this implies that (47) is equivalent to

\[
\left| ( F_n(1, \hat{q}_\gamma) - F_n(1, -\hat{q}_\gamma) ) - ( F_n(1, q_\gamma) - F_n(1, -q_\gamma) ) \right| = o_P(1),
\]

which (by using (46)) follows from

\[
\left| ( F_n(1, \hat{q}_\gamma) - F_n(1, -\hat{q}_\gamma) ) - (1 - \gamma) \right| = o_P(1), \tag{48}
\]

Since by definition of \( \hat{q}_\gamma \) we have \( \hat{H}_n(1, \hat{q}_\gamma) - \hat{H}_n(1, -\hat{q}_\gamma) = 1 - \gamma \), (48) follows from

\[
\sup_z | \hat{H}_n(1, z) - F_n(1, z) | = o_P(1). \tag{49}
\]
To see (49) notice that \( \sup_z |\hat{H}_n(1, z) - F_n(1, z)| = o_P(1) \). This follows from (32), (34), (36) and (37). 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 (47), and as outlined above, this implies that both (42) and (43) are \( o_P(1) \), uniformly in \( z \).

It remains to consider the quantity in (44). First, we derive an upper bound for \( \tilde{q}_\gamma \). Let \( B_n(\epsilon) = \{|F_n(\tilde{q}_\gamma) - \overline{F}_n(\tilde{q}_\gamma)| < \epsilon; \sup_z |(\hat{H}_n(1, z) - F_n(1, z)| < \delta^*/6\} \) where \( \delta^* > 0 \) is such that \( |\overline{F}_n(q_{\gamma+\delta^*}) - \overline{F}_n(q_{\gamma})| < \epsilon \). The above shows that \( P(B_n(\epsilon)) \to 1 \) as \( n \to \infty \) for any \( \epsilon > 0 \). Now choose \( \epsilon > 0 \) small enough (such that \( \gamma > \delta^*/6 \) in order for the below to be well defined). On \( B_n(\epsilon) \) we have

\[
\tilde{q}_\gamma = \inf \{ z \geq 0 : \hat{H}_n(1, z) - \hat{H}_n(1, -z) \geq 1 - \gamma; |F_n(z) - \overline{F}_n(q_{\gamma})| < \epsilon \}
\leq \inf \{ z \geq 0 : F_n(1, z) - F_n(1, -z) \geq 1 - \gamma - \left[ (\hat{H}_n - F_n)(1, q_{\gamma}) - (\hat{H}_n - F_n)(1, -q_{\gamma}) \right] + 2 \sup_{|F_n(v) - F_n(w)| \leq \epsilon} |(\hat{H}_n - F_n)(1, v) - (\hat{H}_n - F_n)(1, w)| \}
\leq \inf \{ z \geq 0 : \Psi_{0, \delta}(z) \geq 1 - \gamma - \left[ (\hat{H}_n - F_n)(1, q_{\gamma}) - (\hat{H}_n - F_n)(1, -q_{\gamma}) \right] + 2 \sup_{|F_n(v) - F_n(w)| \leq \epsilon} |(\hat{H}_n - F_n)(1, v) - (\hat{H}_n - F_n)(1, w)| + cn^{-1} \}
\leq \Psi_{0, \delta}^{-1}(1 - \gamma - Q_n + r_n),
\]

where for short \( r_n = 2 \sup_{|F_n(v) - F_n(w)| \leq \epsilon} |(\hat{H}_n - F_n)(1, v) - (\hat{H}_n - F_n)(1, w)| + cn^{-1} \) with \( c \) from (46), and \( Q_n = \left[ (\hat{H}_n - F_n)(1, q_{\gamma}) - (\hat{H}_n - F_n)(1, -q_{\gamma}) \right] \). Now let

\[
\Psi_{0, \delta}(\alpha, z) = \int_0^\alpha F(\frac{\tilde{z} - u}{\sigma(b_u)}) \, du - \int_0^\alpha F(\frac{-\tilde{z} - u}{\sigma(b_u)}) \, du, \quad z \geq 0, \quad \alpha \in [0, 1]. \tag{50}
\]

Observe that \( \Psi_{0, \delta}(z) = \Psi_{0, \delta}(1, z) \). For ease of notation we omit the subscripts on \( \Psi \) in what follows. Since by definition of \( q_{\gamma} \) we have \( q_{\gamma} = \Psi^{-1}(1 - \gamma) \), and since \( \Psi(\alpha, z) \) is strictly increasing in \( z \) for any \( \alpha \) we have on \( B_n(\epsilon) \),

\[
[F_n(\alpha, q_{\gamma}) - \overline{F}_n(\alpha, q_{\gamma})] - [F_n(\alpha, -q_{\gamma}) - \overline{F}_n(\alpha, -q_{\gamma})] \]
\[
\Psi(\alpha, \hat{q}_\gamma) - \Psi(\alpha, q_\gamma) + cn^{-1}
\]
\[
\leq \Psi\left(\alpha, \Psi^{-1}(1 - \gamma - Q_n + r_n)\right) - \Psi(\alpha, \Psi^{-1}(1 - \gamma)) + cn^{-1}
\]
\[
= - \frac{\partial_z \Psi(\alpha, \Psi^{-1}(\xi_n^+)\right)}{\Psi'(\Psi^{-1}(\xi_n^+)\right)} (Q_n - r_n) + cn^{-1}
\]
\[
= - \int_0^b \left[ f_u(\Psi^{-1}(\xi_n^+)) + f_u(-\Psi^{-1}(\xi_n^+))\right] du + \int_0^b \left[ f_u(\Psi^{-1}(\xi_n^+)) + f_u(-\Psi^{-1}(\xi_n^+))\right] du (Q_n - r_n) + cn^{-1}
\]
\[
=: - (c(\alpha) + o_P(1)) (Q_n - r_n) + cn^{-1}
\]

with \(\xi_n^+ \in [1 - \gamma, 1 - \gamma - Q_n + r_n]\), \(f_u(z) = \frac{\partial_z}{\partial_z} F_u(z) = \frac{1}{\sigma'(z)} f\left(\frac{z}{\sigma'(z)}\right)\), 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

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

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

\[
[F_n(\alpha, \hat{q}_\gamma) - F_n(\alpha, q_\gamma)] - [F_n(\alpha, -\hat{q}_\gamma) - F_n(\alpha, -q_\gamma)] \geq -(c(\alpha) + o_P(1)) (Q_n + r_n) - cn^{-1}.
\]

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

\[
\sqrt{n} \left[ F_n(\alpha, \hat{q}_\gamma) - F_n(\alpha, q_\gamma)\right] - \left[ F_n(\alpha, -\hat{q}_\gamma) - F_n(\alpha, -q_\gamma)\right] = - \frac{c(\alpha)}{\sqrt{n}} \sum_{s=1}^n \left( 1(\sigma^2(\frac{s}{n}) \epsilon_s^2 \leq q_\gamma^2) - P(\sigma^2(\frac{s}{n}) \epsilon_s^2 \leq q_\gamma^2) \right) + o_P(1)
\]

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 \(|\hat{q}_\gamma - q_\gamma| = o_P(1)\) as \(n \to \infty\) (cf. comment right after (37)). To see that, first notice that our regularity conditions imply that \(\inf_u f_u(q_\gamma) = d_\ast > 0\). Again using the fact that our assumptions imply the uniform Lipschitz continuity of the functions \(z \to f_u(z)\) we obtain the existence

29
of an $\epsilon_0 > 0$ such that $f_u(z) > \frac{d}{2}$ for all $|z - q_\gamma| \leq \epsilon_0$ and all $u \in [0, 1]$. Consequently, if $|\hat{q}_\gamma - q_\gamma| \leq \epsilon_0$ then $|F_{\hat{q}_\gamma}(\hat{q}_\gamma) - F_{\hat{q}_\gamma}(q_\gamma)| = \left| \int_{\hat{q}_\gamma}^{q_\gamma} f_u(u) \, du \right| \geq \frac{d}{2} |\hat{q}_\gamma - q_\gamma| \forall s = 0, 1, \ldots, n$. On the other hand, if $|\hat{q}_\gamma - q_\gamma| > \epsilon_0$ then, because the $f_u(z)$ are all non-negative, $|F_{\hat{q}_\gamma}(\hat{q}_\gamma) - F_{\hat{q}_\gamma}(q_\gamma)| = \left| \int_{\hat{q}_\gamma}^{q_\gamma} f_u(u) \, du \right| \geq \frac{d}{2} \epsilon_0 \forall s = 0, 1, \ldots, n$. Thus, for any $\epsilon > 0$ there exists an $\eta > 0$ with

$$\{|\hat{q}_\gamma - q_\gamma| > \epsilon\} \subset \left\{|F_{\hat{q}_\gamma}(\hat{q}_\gamma) - F_{\hat{q}_\gamma}(q_\gamma)| > \eta \quad \forall s = 0, 1, \ldots, n\right\}. \tag{51}$$

Since all the function $z \to F_{\hat{q}_\gamma}(z)$ are strictly monotone, we have

$$\overline{d}_n(q_\gamma, q_\gamma) = |\overline{F}_n(q_\gamma) - \overline{F}_n(q_\gamma)| = \frac{1}{n} \sum_{s=1}^{n} |F_{\hat{q}_\gamma}(\hat{q}_\gamma) - F_{\hat{q}_\gamma}(q_\gamma)|.$$

Consequently, (51) implies that if $|\hat{q}_\gamma - q_\gamma|$ is not $o_P(1)$, then also $\overline{d}_n(q_\gamma, q_\gamma)$ is not $o_P(1)$. This is a contradiction to (47) and this completes our proof.

6 References:


of time varying parameters under shape restrictions. *J. Econometrics* 126, 53 - 77.


