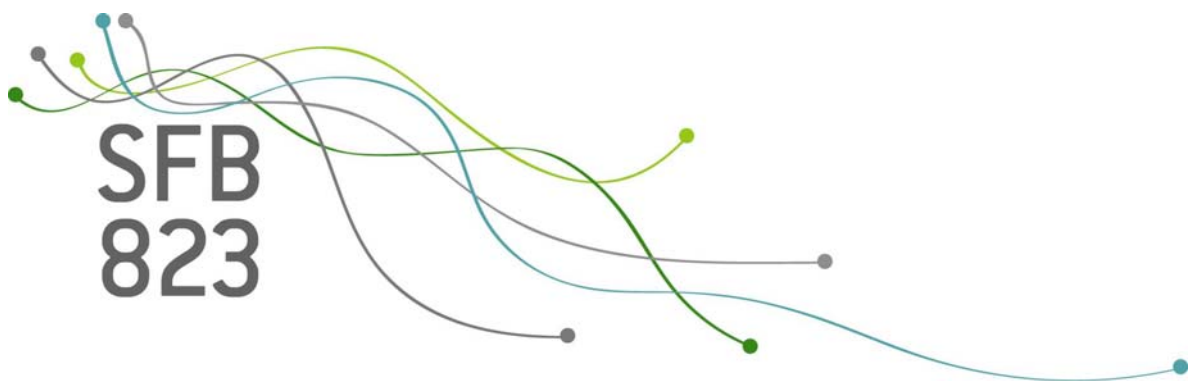


SFB  
823

# Convergence of spectral density estimators in the locally stationary framework

Rafael Kawka

Nr. 23/2019



Discussion Paper



# Convergence of Spectral Density Estimators in the Locally Stationary Framework

Rafael Kawka

August 20, 2019

Locally stationary processes are characterised by spectral densities that are functions of rescaled time. We study the asymptotic properties of spectral density estimators in the locally stationary framework. In particular, we show that for a locally stationary process with time-varying spectral density function  $f(u, \lambda)$  standard spectral density estimators consistently estimate the time-averaged spectral density  $\int_0^1 f(u, \lambda) du$ . This result is complemented by some illustrative examples and applications including HAC-inference in the multiple linear regression model and a simple visual tool for the detection of unconditional heteroskedasticity.

## 1. Introduction

For many years weakly stationary processes, i.e. processes with time-invariant second order moment structure, play a formative role in the econometrics and time series literature. Models based on stationary processes are applied in many scientific disciplines, such as finance and economics, but also in the analysis of acoustic signals or flood statistics. The assumption of stationarity often seems plausible, especially for short time series, but it is questionable in case of longer time series. For instance, the question whether stationary time series models are adequate tools to mimic the behaviour of financial series such as stock returns is extensively discussed in the literature, see Amado and Teräsvirta (2014), Mikosch and Starica (2004), Starica and Granger (2005) and van Bellegem (2012) among others. Since the assumption of stationarity is often too restrictive, various generalizations of stationary processes have been presented in the literature. For details we refer to Dahlhaus (2012) and the references therein.

A rather general family of stochastic processes adapting a time-dependent second order moment structure are *locally stationary processes* presented by Dahlhaus (1996a,b) and Dahlhaus and Polonik (2006, 2009). These processes are characterized not only by time-varying second moments, but behave approximatively stationary in a small neighborhood of each time point  $t \in \{1, \dots, T\}$ . In order to make this processes mathematically tractable, the time is rescaled to the unit interval, so that in each rescaled time point, i.e.  $t/T$ , various quantities such as the covariance, the autocovariance or the spectral density at a specific frequency  $\lambda$  can be represented by a function defined on the unit interval.

Based on the work of Dahlhaus, a broad literature on the theory of locally stationary processes has been developed in recent years. Examples are local stationary ARCH processes (Dahlhaus

and Rao, 2006) and locally stationary wavelet processes (Nason and von Sachs, 1999). But locally stationary processes also find great popularity in the applied literature. Ombao et al. (2001) investigate speech signals and EEG data whereas Adak (1998) model seismological time series with locally stationary processes. In addition, a large literature has been developed which deals with the question of whether a time series is based on a stationary or a locally stationary process. In this context, the work of Sergides and Paparoditis (2009), Dette et al. (2011) and Preuß et al. (2015) can be mentioned as examples. Dahlhaus (2012) gives an excellent overview of the recent developments in the locally stationary literature.

Interestingly, the question of how estimators for quantities of stationary processes behave in the locally stationary framework has so far received little attention in the literature. It is reasonable to assume that global estimators, i.e. estimators based on the complete time series, consistently estimate a time-averaged value of the true time-varying quantities. This has been confirmed by Dahlhaus (1997, Section 5) and Dahlhaus (2012, Example 8) for two parametric estimators (Maximum Likelihood estimator and Yule-Walker estimator of AR(p) processes). However, the question arises whether nonparametric estimators provide analogous results. In the present paper we verify this claim for kernel estimators for the spectral density of a stationary process (cf. Anderson, 1971, Chapter 9). In particular, we show that if  $f(u, \lambda)$  is the time-varying spectral density of a locally stationary process then, for some arbitrary frequency  $\lambda \in (-\pi, \pi]$ , the spectral density estimator  $\hat{f}(\lambda)$  converges to  $\int_0^1 f(u, \lambda) du$  in probability (Theorem 3.4).

The remainder of the paper is organized as follows. In Section 2 we present the definition of locally stationary processes and give some illustrative examples. In addition, we show that for every locally stationary process there exists, in a sense, a best global approximation by a stationary process. This stationary approximating process has a spectral density that corresponds to the time-averaged spectral density of the original process. In Section 3 we derive our main result, i.e. the convergence of spectral density estimators to the time-averaged spectral density. Additionally, we demonstrate the estimation performance by means of a small simulation study. Section 4 contains two applications. First, we apply the theoretical results to the multiple linear regression model by showing that HAC inference known from textbooks also works asymptotically in the locally stationary framework. Second, we introduce the *spectrum profile*, a generalization of the variance profile defined by Cavaliere and Taylor (2007a,b), which can be used as a simple tool for detecting changes in the unconditional second order moment structure. As an illustration we study the logarithmic stock returns series of four major American undertakings. Section 5 concludes. Auxiliary results as well as several proofs can be found without further reference in the appendix.

## 2. Locally Stationary Processes

Consider the following stochastic process defined via a triangular sequence  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$ , where every  $\eta_{t,T}$  has a time-varying moving average representation of the form

$$\eta_{t,T} = \sum_{j=0}^{\infty} \psi_{j,t,T} \varepsilon_{t-j}. \quad (1)$$

The process  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is called *locally stationary process* under appropriate assumptions on the sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  and the time dependent and therefore triple indexed coefficient matrices  $\psi_{j,t,T}$ . We impose the following assumption on the sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ .

**Assumption 2.1.** The sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  in the moving average representation (1) consists of  $n$ -dimensional i.i.d. random vectors with  $\mathbb{E}(\varepsilon_t) = 0$ ,  $\mathbb{E}(\varepsilon_t \varepsilon_t') = I_n$  and  $\mathbb{E}\|\varepsilon_t\|^4 < \infty$ .

To state precise assumptions on the matrices  $\psi_{j,t,T}$ , we need the following definition. Let  $\Pi$  denote the collection of all partitions of the interval  $[a, b]$ , i.e. sequences  $\{x_i\}_{i=0}^N$ , such that  $a \leq x_0 < x_1 < \dots < x_N \leq b$  with  $N \in \mathbb{N}$ . Then,

$$V_f[a, b] := \sup \left\{ \sum_{i=1}^N \|f(x_i) - f(x_{i-1})\| : \{x_i\}_{i=0}^N \in \Pi \right\}$$

is the the Variation of  $f$  over  $[a, b]$ . Thereby,  $\|\cdot\|$  is the Frobenius norm when considering a matrix and the usual Euclidean norm when considering a vector.

**Assumption 2.2.** The moving average coefficient matrices  $\psi_{j,t,T} \in \mathbb{R}^{n \times n}$  in (1) satisfy

$$\sup_{t,T} \|\psi_{j,t,T}\| \leq \frac{K}{l(j)},$$

where  $K > 0$  is independent of  $T$  and

$$l(j) := \begin{cases} 1, & j = 0, 1, \\ j^2 \log(j)^{\delta+1}, & j > 1, \end{cases}$$

with some arbitrary  $\delta > 0$ . Furthermore, there exist càdlàg functions  $\psi_j : [0, 1] \rightarrow \mathbb{R}^{n \times n}$  with

$$\begin{aligned} \|\psi_j\|_\infty &\leq \frac{K}{l(j)} \\ V_{\psi_j}[0, 1] &\leq \frac{K}{l(j)} \\ \sum_{t=1}^T \left\| \psi_{j,t,T} - \psi_j\left(\frac{t}{T}\right) \right\| &\leq \frac{K}{l(j)}. \end{aligned}$$

**Notation 2.3.** The set of all locally stationary processes of the form (1) that satisfy Assumption 2.1 and 2.2 is denoted by  $X$  and the shorthand notation  $\eta \in X$  indicates that a process  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is locally stationary.

Assumption 2.2 guarantees that the coefficient matrices  $\psi_{j,t,T}$  can be approximated by well behaving functions  $\psi_j$  and, when rescaling the time domain to the unit interval, the process  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  behaves approximately like the (stationary) process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$  with moving average representation

$$\tilde{\eta}_t(u) := \sum_{j=0}^{\infty} \psi_j(u) \varepsilon_{t-j}, \quad \text{with fixed } u \in [0, 1], \quad (2)$$

in a close neighbourhood of  $u := t/T$ .

**Remark 2.4.** The definition of locally stationary processes might look complicated at first glance because it involves the coefficient matrices  $\psi_{j,t,T}$  as well as the coefficient functions  $\psi_j$ . A rather intuitive definition with  $\psi_{j,t,T} := \psi_j(t/T)$  is, however, heavily restricting the class of processes. In particular, Künsch (1995) and Dahlhaus and Polonik (2009) show that such a definition rules out time-varying autoregressive processes (see Example 2.8 for a definition of time-varying ARMA processes).

**Remark 2.5.** The assumptions stated above are relatively mild compared to the assumptions stated in several papers dealing with locally stationary processes. Assumption 2.1 is weaker than the assumption on the sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  imposed by Dahlhaus and Polonik (2009), Dahlhaus (2009), Dette et al. (2011) and Paparoditis and Preuß (2014), who require Gaussianity or at least the existence of all moments. Assumption 2.2 is also rather mild, as, for instance, Dette et al. (2011) impose some differentiability conditions on the coefficient functions.

The function  $f : [0, 1] \times [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$  defined by<sup>1</sup>

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

with

$$A(u, \lambda) := \sum_{h=0}^{\infty} e^{-i\lambda h} \psi_h(u)$$

is called *time-varying spectral density* of  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$ . For  $h \in \mathbb{N}_0$  the function  $\Gamma_h : [0, 1] \rightarrow \mathbb{R}^{n \times n}$  with

$$\Gamma_h(u) = \int_{-\pi}^{\pi} f(u, \lambda) e^{i\lambda h} d\lambda = \sum_{j=0}^{\infty} \psi_{j+h}(u) \psi_j(u)'$$

is called *time-varying autocovariance function* of  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$ . Both functions are well defined due to Assumption 2.2 and it holds that

$$f(u, \lambda) = \frac{1}{2\pi} (\Sigma(u) + \Lambda(u, \lambda) + \Lambda(u, \lambda)^*),$$

where

$$\Lambda(u, \lambda) = \sum_{h=1}^{\infty} e^{-i\lambda h} \Gamma_h(u)$$

and  $\Sigma(u) = \Gamma_0(u)$ , which is sometimes labeled *time-varying covariance matrix*.

**Remark 2.6.** Note that for fixed  $u \in [0, 1]$  the function  $A(u, \cdot)$  is the transfer function of the linear filter

$$A(u, L) = \sum_{h=0}^{\infty} \psi_h(u) L^h.$$

Hence, the auxiliary process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$  can be represented as  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}} = A(u, L)\{\varepsilon_t\}_{t \in \mathbb{Z}}$  with autocovariance matrix at lag  $h$  given by  $\Gamma_h(u)$ . From the filter theorem (see, for instance Deistler and Scherrer, 2018, Theorem 4.1) we deduce that the spectral density of  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$  at an arbitrary frequency  $\lambda \in (-\pi, \pi]$  is given by  $f(u, \lambda)$ .

<sup>1</sup> The asterisk denotes the complex conjugate.

## 2.1. Examples of Locally Stationary Processes

We present some examples of locally stationary processes. The first two examples are well known from Dahlhaus (2000, 2012) and are considered as prime examples for locally stationary processes. The third example is a so-called heteroskedastic model that is considered by Cavaliere and Taylor (2007a,b).

**Example 2.7** (Amplitude Modulation). Let  $\{\xi_t\}_{t \in \mathbb{Z}}$  be a stationary process with moving-average representation

$$\xi_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where  $\sum_j j \|\psi_j\| < \infty$  and  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  fulfills Assumption 2.1. For some matrix valued function of bounded variation,  $\phi$  say, the amplitude modulated process  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is defined by

$$\eta_{t,T} = \phi\left(\frac{t}{T}\right) \xi_t, \quad t = 1, \dots, T.$$

Clearly,  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is locally stationary with moving average coefficients  $\psi_{j,t,T} = \phi(t/T)\psi_j$ , coefficient functions  $\psi_j(u) = \phi(u)\psi_j$  and time-varying spectral density

$$f(u, \lambda) = \phi(u)g(\lambda)\phi(u)',$$

where  $g(\lambda)$  denotes the spectral density at frequency  $\lambda$  of the stationary process  $\{\xi_t\}_{t \in \mathbb{Z}}$ .

**Example 2.8** (tvARMA). Consider the system of equations

$$\sum_{j=0}^p a_j \left(\frac{t}{T}\right) \eta_{t-j,T} = \sum_{k=0}^q b_k \left(\frac{t}{T}\right) \varepsilon_{t-k}, \quad t = 1, \dots, T, \quad (3)$$

where the sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  satisfies Assumption 2.1 and the matrix valued coefficient functions  $a_j$  and  $b_j$  with  $a_p, b_q \neq 0$  are of bounded variation. Define  $a, b : [0, 1] \times \mathbb{C} \rightarrow \mathbb{C}^{n \times n}$  via  $a(u, z) = \sum_{j=0}^p a_j(u)z^j$  and  $b(u, z) = \sum_{k=0}^q b_k(u)z^k$ . If  $\det a(u, z) \neq 0$  for all  $u \in [0, 1]$  and for all  $z$  with  $0 < |z| \leq 1 + \delta$  for some  $\delta > 0$  then, by Dahlhaus (2000, Example 2.3) and Dahlhaus and Polonik (2009, Proposition 2.4), there exists a locally stationary solution of the system (3), i.e., a locally stationary process  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  with time-varying moving average representation (1), where the coefficient matrices satisfy Assumption 2.2. The time-varying spectral density of  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is given by

$$f(u, \lambda) = \frac{1}{2\pi} a(u, e^{i\lambda})^{-1} b(u, e^{i\lambda}) b(u, e^{-i\lambda})' a(u, e^{-i\lambda})'^{-1}.$$

**Example 2.9** (Unconditional Heteroskedastic Model). Cavaliere and Taylor (2007a,b) consider univariate processes  $\{\nu_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  of the form

$$\nu_{t,T} = \sum_{j=0}^{\infty} C_j \sigma_{t-j} \varepsilon_{t-j},$$

where  $\sum_j j |C_j| < \infty$  and  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is a sequence of independent and identically distributed random variables with  $\mathbb{E}\varepsilon_t = 0$ ,  $\mathbb{E}\varepsilon_t^2 = 1$  and  $\mathbb{E}\varepsilon_t^r < K < \infty$  for some  $r \geq 4$ . The volatility term

$\sigma_t$  satisfies  $\sigma_{[uT]} = \omega(u)$  for all  $u \in [0, 1]$ , where  $\omega$  denotes a strictly positive and bounded càdlàg function. If we additionally impose that  $\omega$  is of bounded variation then  $\{\nu_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is a locally stationary process with moving average coefficients  $\psi_{j,t,T} = C_j \omega((t-j)/T)$ , coefficient functions  $\psi_j(u) = C_j \omega(u)$  and time-varying spectral density

$$f(u, \lambda) = \frac{1}{2\pi} \left| \sum_{j=0}^{\infty} C_j \omega(u) e^{-i\lambda j} \right|^2 = \frac{\omega^2(u)}{2\pi} \left| \sum_{j=0}^{\infty} C_j e^{-i\lambda j} \right|^2 = \omega^2(u) g(\lambda),$$

where  $g$  is the spectral density function of the stationary process with moving average representation  $\sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$ .

**Remark 2.10.** From Examples 2.7 and 2.9 it becomes apparent that processes with different coefficients  $\psi_{j,t,T}$  in the time-varying moving average representations (1) may have identical time-varying spectral densities.

## 2.2. Global Approximation of Locally Stationary Processes

Assumption 2.2 guarantees that a process  $\eta \in X$  can be *locally* approximated by the stationary process  $\{\eta_t(u)\}_{t \in \mathbb{Z}}$ , defined in (2), in every rescaled time point  $u = t/T$ . In this section, we switch the perspective and discuss a *global* approximation of locally stationary processes. In particular, we search for a second order stationary process that approximates a given  $\eta \in X$  optimally over the whole sample. To this end, we define the subset  $X_0 \subset X$  by

$$X_0 := \{\eta \in X : f_\eta(u, \lambda) = g(\lambda)\}.$$

The set  $X_0$  contains all locally stationary processes with time homogenous spectral density, i.e. weakly stationary processes.

**Remark 2.11.** Note that a locally stationary process cannot be identified uniquely by its time-varying spectral density due to multiple reasons. First, we do not impose any distributional assumptions on the innovation sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ . Furthermore, we have already observed in Remark 2.10 that processes with different time-varying moving average representations may share the same time-varying spectral densities. Since we are, however, only interested in a (time-varying) second order moment structure, we simply identify two processes that share the same spectral density with each other.<sup>2</sup>

The absolute summability assumption on the coefficient functions implies that every component of the time-varying spectral density is in  $L^2([0, 1] \times [-\pi, \pi])$ . We use this property to equip  $X$  with a Hilbert space structure. Consider two locally stationary processes  $\eta, \nu \in X$  with time-varying spectral densities  $f_\eta$  and  $f_\nu$ , respectively. The mapping  $X \times X \rightarrow \mathbb{C}$ , defined by

$$\langle \eta, \nu \rangle := \int_0^1 \int_{-\pi}^{\pi} \text{tr}(f(u, \lambda) g(u, \lambda)^*) d\lambda du$$

<sup>2</sup> We define  $\eta \sim \nu : \Leftrightarrow f_\eta(u, \lambda) = f_\nu(u, \lambda)$  almost everywhere with respect to the Lebesgue measure on  $[0, 1] \times [-\pi, \pi]$ . With this equivalence relation we define equivalence classes  $[\eta] := \{\xi \in X : \eta \sim \xi\}$  and the quotient set  $X/\sim := \{[\eta] : \eta \in X\}$ . In order not to make the notation unnecessarily complicated, we do not distinguish between  $X$  and  $X/\sim$ .



is an inner product inducing the norm  $\|\eta\|_X := (\langle \eta, \eta \rangle)^{1/2}$ . Consequently,  $(X, \langle \cdot, \cdot \rangle)$  is an inner product space and, since  $L^2([0, 1] \times [-\pi, \pi])$  is complete, it is also a Hilbert space. Furthermore,  $X_0$  is a closed subspace of  $X$ .

**Theorem 2.12.** *Let  $\eta \in X$  with time-varying spectral density  $f$ . The process  $\eta^0 \in X_0$  with spectral density  $f^0$ , defined by  $f^0(\lambda) := \int_0^1 f(u, \lambda) du$ , is the best approximation of  $\eta$  in  $X_0$ , i.e.,*

$$\eta^0 = \arg \min_{\xi \in X_0} \|\eta - \xi\|_X.$$

*The approximation is unique and it holds that  $\langle \eta - \eta^0, \xi \rangle = 0$  for all  $\xi \in X_0$ .*

**Remark 2.13.** The first part of this result is a multivariate generalization of Lemma 1 in Dette et al. (2011). For some  $\eta \in X$  with time-varying spectral density  $f$  they consider the distance

$$D^2 = \min_{\xi \in X_0} \|\eta - \xi\|_X^2 = \min_g \int_0^1 \int_{-\pi}^{\pi} (f(u, \lambda) - g(\lambda))^2 d\lambda du$$

as a measure of the deviation from stationarity of a locally stationary process.

Theorem 2.12 motivates the main result of this article which is presented in the subsequent section. In particular, Theorem 3.4 states that for a process  $\eta \in X$  with spectral density function  $f$  the classical kernel spectral density estimator converges to the integrated spectral density, i.e. to the spectral density  $f^0$  of the best approximating stationary process  $\eta^0 \in X_0$ .

### 3. Spectral Density Estimation

Before we present the main result of this paper we derive the probability limits of autocovariance matrix estimators in the locally stationary framework. For some  $\eta \in X$  the estimator for the autocovariance matrix at lag  $h \in \mathbb{N}_0$  is defined by

$$\hat{\Gamma}_h = \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h, T} \eta'_{t, T}.$$

If the coefficient matrices  $\psi_{j, t, T}$  in the moving average representation (1) do not depend on  $t$  and  $T$ , i.e. if  $\eta \in X_0$ , the estimator  $\hat{\Gamma}_h$  consistently estimates the autocovariance matrix of  $\eta$ . For arbitrary  $\eta \in X$  we obtain the following convergence result.

**Theorem 3.1.** *Let  $\eta \in X$  with time-varying autocovariance function  $\Gamma_h$ . Then, as  $T \rightarrow \infty$ ,*

$$\hat{\Gamma}_h \xrightarrow{P} \int_0^1 \Gamma_h(u) du$$

*for all  $h \in \mathbb{N}_0$ .*

By setting  $h = 0$  we immediately obtain the probability limit of the covariance matrix estimator  $\hat{\Sigma} = \hat{\Gamma}_0$  as a corollary.

**Corollary 3.2.** *Let  $\eta \in X$  with time-varying covariance function  $\Sigma$ . Then, as  $T \rightarrow \infty$ ,*

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \eta_{t,T} \xrightarrow{P} \int_0^1 \Sigma(u) du.$$

With this results in place we can now derive the main result of this paper, i.e. the probability limit of the spectral density estimator. For some frequency  $\lambda \in (-\pi, \pi]$  this estimator is defined by

$$\hat{f}(\lambda) = \frac{1}{2\pi} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e^{-i\lambda(t-s)} k\left(\frac{|t-s|}{M_T}\right) \eta_{t,T} \eta'_{s,T}, \quad (4)$$

where  $M_T$  and  $k$  denote the bandwidth and the kernel function. It holds that

$$\hat{f}(\lambda) = \frac{1}{2\pi} \left( \hat{\Sigma} + \hat{\Lambda}(\lambda) + \hat{\Lambda}(\lambda)^* \right),$$

where

$$\begin{aligned} \hat{\Lambda}(\lambda) &= \frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} e^{-i\lambda(t-s)} k\left(\frac{t-s}{M_T}\right) \eta_{t,T} \eta'_{t,T} \\ &= \sum_{h=1}^{T-1} k\left(\frac{h}{M_T}\right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T} \eta'_{t,T} \\ &= \sum_{h=1}^{T-1} k\left(\frac{h}{M_T}\right) e^{-i\lambda h} \hat{\Gamma}_h. \end{aligned}$$

It is well known that under appropriate assumptions on  $k$  and  $M_T$  the estimator  $\hat{f}(\lambda)$  consistently estimates the spectral density of a stationary process at frequency  $\lambda$ . Jansson (2002) provides the following assumptions on the bandwidth and the kernel function.<sup>3</sup>

**Assumption 3.3.** The kernel function  $k : [0, \infty) \rightarrow \mathbb{R}$  is continuous at zero with  $k(0) = 1$ ,  $\sup_{x \geq 0} |k(x)| < \infty$  and

$$\int_0^\infty \sup_{y \geq x} |k(y)| dx < \infty.$$

The bandwidth sequence  $\{M_T\}_{T \in \mathbb{N}}$  is strictly positive and satisfies

$$\lim_{T \rightarrow \infty} \left( \frac{1}{M_T} + \frac{M_T}{\sqrt{T}} \right) = 0.$$

**Theorem 3.4.** *Let  $\eta \in X$  with time-varying spectral density function  $f$ . If the estimator  $\hat{f}(\lambda)$  fulfills Assumption 3.3 then, as  $T \rightarrow \infty$ , it holds that*

$$\hat{f}(\lambda) \xrightarrow{P} \int_0^1 f(u, \lambda) du$$

for all  $\lambda \in (-\pi, \pi]$ .

<sup>3</sup> In his paper Jansson (2002) only considers the estimation of the spectral density at the long run frequency, i.e.  $f(\lambda)$  for  $\lambda = 0$ . However, his results can be easily generalized to the arbitrary frequency case.

Note that for a process  $\eta \in X_0$  the estimator  $\hat{f}(\lambda)$  consistently estimates the spectral density which does not vary over time, i.e. the consistency of the spectral density estimator for stationary processes is immediately implied. For arbitrary  $\eta \in X$ , however, the spectral density is time-varying and, hence,  $\hat{f}(\lambda)$  estimates the spectral density of the best approximating stationary process.

### 3.1. Finite Sample Performance

We assess the estimation performance of the spectral density estimator (4) at the frequencies  $\lambda = 0$ ,  $\lambda = \pi/4$  and  $\lambda = \pi/2$ . To this end, we consider the following univariate locally stationary processes:

$$\begin{aligned}\eta_{t,T}^{(1)} &= \cos\left(\frac{2\pi t}{T}\right)\varepsilon_t + \left(\frac{t}{T}\right)^2\varepsilon_{t-1}, \\ \eta_{t,T}^{(2)} &= \phi\left(\frac{t}{T}\right)\xi_t, \quad \phi^2(u) = 1 + \frac{1}{1 + e^{-20(u-1/2)}}, \quad \xi_t = \frac{1}{2}\xi_{t-1} + \varepsilon_t, \quad \xi_0 = 0, \\ \eta_{t,T}^{(3)} &= \left\{ \frac{1}{2}\mathbb{1}_{(0,1/2]} \left(\frac{t}{T}\right) - \frac{1}{2}\mathbb{1}_{(1/2,1]} \left(\frac{t}{T}\right) \right\} \eta_{t-1,T}^{(3)} + \varepsilon_t, \quad \eta_{0,T} = 0, \\ \eta_{t,T}^{(4)} &= \frac{t}{\sqrt{2T}}\eta_{t-1,T}^{(4)} + \varepsilon_t + \frac{t}{\sqrt{2T}}\varepsilon_{t-1}, \quad \eta_{0,T} = 0,\end{aligned}$$

where in every case the sequence  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is i.i.d. standard normal.

The first process is a time-varying MA(1) process and has already been considered in Dette et al. (2011). Its time-varying spectral density is given by

$$f_{\eta^{(1)}}(u, \lambda) = \frac{1}{2\pi} \left\{ \cos^2(2\pi u) - 2u^2 \cos(\lambda) \cos(2\pi u) + u^4 \right\}$$

and the spectral density of the best approximating stationary process is

$$f_{\eta^{(1)}}^0(\lambda) = \int_0^1 f_{\eta^{(1)}}(u, \lambda) du = \frac{7}{20\pi} - \frac{\cos(\lambda)}{2\pi^3}.$$

Consequently, it holds that  $f_{\eta^{(1)}}^0(0) = 7/(20\pi) - 1/(2\pi^3)$ ,  $f_{\eta^{(1)}}^0(\pi/4) = 7/(20\pi) - 1/(2\sqrt{2}\pi^3)$  and  $f_{\eta^{(1)}}^0(\pi/2) = 7/(20\pi)$ .

The second process is an amplitude modulated AR(1) process with autoregressive parameter 1/2. The spectral density of the stationary AR(1)-process is given by

$$f_{\xi}(\lambda) = \frac{1}{2\pi} \left( 1 - \cos(\lambda) + \frac{1}{4} \right)^{-1}$$

and the amplitude modulating function is the square root of a logistic function. Hence, from Example 2.7 we deduce that the time-varying spectral density of the second process is equal to

$$f_{\eta^{(2)}}(u, \lambda) = \phi^2(u) f_{\xi}(\lambda).$$

We can easily calculate the spectral density of the best approximating stationary process via

$$f_{\eta^{(2)}}^0(\lambda) = \int_0^1 f_{\eta^{(2)}}(u, \lambda) du = \int_0^1 \phi(x) dx f_{\xi}(\lambda) = \frac{3}{4\pi} \left(1 - \cos(\lambda) + \frac{1}{4}\right)^{-1},$$

from which we obtain  $f_{\eta^{(2)}}^0(0) = 3/\pi$ ,  $f_{\eta^{(2)}}^0(\pi/4) = 3/((5 - 2\sqrt{2})\pi)$  and  $f_{\eta^{(2)}}^0(\pi/2) = 3/(5\pi)$ .

The third process is an AR(1) process, where a structural break occurs in the middle of the sample. Hence, its time-varying spectral density is given by

$$f_{\eta^{(3)}}(u, \lambda) = \begin{cases} \frac{1}{2\pi} \left(1 - \cos(\lambda) + \frac{1}{4}\right)^{-1}, & 0 < u \leq \frac{1}{2} \\ \frac{1}{2\pi} \left(1 + \cos(\lambda) + \frac{1}{4}\right)^{-1}, & \frac{1}{2} < u \leq 1 \end{cases}$$

and, consequently,

$$f_{\eta^{(3)}}^0(\lambda) = \int_0^1 f_{\eta^{(3)}}(u, \lambda) du = \frac{1}{\pi} \frac{10}{25 - 16 \cos^2(\lambda)}.$$

Evaluating this function at the required frequencies yields  $f_{\eta^{(3)}}^0(0) = 10/(9\pi)$ ,  $f_{\eta^{(3)}}^0(\pi/4) = 10/(17\pi)$  and  $f_{\eta^{(3)}}^0(\pi/2) = 10/(25\pi)$ .

The last process is a tvARMA(1,1)-process with time-varying spectral density

$$f_{\eta^{(4)}}(u, \lambda) = \frac{1}{2\pi} \frac{1 + \sqrt{2} \cos(\lambda)u + \frac{1}{2}u^2}{1 - \sqrt{2} \cos(\lambda)u + \frac{1}{2}u^2}.$$

In this case, however, a closed form expression of the integrated spectral density is rather cumbersome and we only calculate  $f_{\eta^{(4)}}^0(\lambda)$  at the required frequencies. For  $\lambda = \pi/2$  it holds that both  $f_{\eta^{(4)}}(u, \pi/2)$  and  $f_{\eta^{(4)}}^0(\pi/2)$  are equal to  $1/(2\pi)$ . For  $\lambda = \pi/4$  we obtain, using some calculus,

$$f_{\eta^{(4)}}^0\left(\frac{\pi}{4}\right) = \int_0^1 f_{\eta^{(4)}}\left(u, \frac{\pi}{4}\right) du = \frac{1}{2\pi} \int_0^1 \frac{1 + u + \frac{1}{2}u^2}{1 - u + \frac{1}{2}u^2} du = \frac{1 + \pi - \log(4)}{2\pi}.$$

Similarly, for  $\lambda = 0$  it holds that

$$f_{\eta^{(4)}}^0(0) = \frac{1}{2\pi} \int_0^1 \frac{1 + \sqrt{2}u + \frac{1}{2}u^2}{1 - \sqrt{2}u + \frac{1}{2}u^2} du = \frac{1}{2\pi} \left\{ 9 + 4\sqrt{2} \left[ 1 + \log\left(1 - \frac{1}{\sqrt{2}}\right) \right] \right\}.$$

We simulate 10 000 time-series of each process for the sample sizes  $T \in \{50, 100, 200, 500, 1\,000\}$ . For the calculation of  $\hat{f}(\lambda)$  we use the following kernel functions:

$$\begin{aligned} \text{Bartlett:} \quad k_{BA}(x) &= \begin{cases} 1 - |x| & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases} \\ \text{Parzen:} \quad k_{PA}(x) &= \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| < 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases} \\ \text{Quadratic Spectral:} \quad k_{QS}(x) &= \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right). \end{aligned}$$

**Table 1:** Empirical RMSE of  $\hat{f}(\lambda)$  for  $\lambda \in \{0, \pi/4, \pi/2\}$ . Estimation performed using the Bartlett, Parzen and Quadratic Spectral kernels in conjunction with a sample size dependent bandwidth.

$T$	Bartlett			Parzen			Quadratic Spectral		
	$\lambda = 0$	$\lambda = \frac{\pi}{4}$	$\lambda = \frac{\pi}{2}$	$\lambda = 0$	$\lambda = \frac{\pi}{4}$	$\lambda = \frac{\pi}{2}$	$\lambda = 0$	$\lambda = \frac{\pi}{4}$	$\lambda = \frac{\pi}{2}$
Process $\eta^{(1)}$									
50	.0581	.0447	.0311	.0503	.0421	.0300	.0690	.0495	.0337
100	.0522	.0367	.0241	.0469	.0360	.0217	.0616	.0407	.0272
200	.0423	.0298	.0171	.0388	.0292	.0153	.0492	.0330	.0194
500	.0367	.0253	.0117	.0351	.0250	.0104	.0417	.0281	.0137
1 000	.0343	.0238	.0089	.0336	.0236	.0080	.0381	.0261	.0105
Process $\eta^{(2)}$									
50	.4342	.1525	.0997	.4828	.1391	.1403	.3920	.1781	.0810
100	.3574	.1114	.0659	.3989	.1169	.1005	.3138	.1262	.0528
200	.3302	.0816	.0575	.3833	.0864	.0941	.2644	.0972	.0409
500	.2680	.0552	.0445	.3133	.0737	.0638	.1967	.0681	.0263
1 000	.2239	.0409	.0377	.2598	.0661	.0443	.1515	.0507	.0194
Process $\eta^{(3)}$									
50	.1544	.0697	.0629	.1639	.0658	.0882	.1603	.0764	.0511
100	.1305	.0525	.0422	.1414	.0539	.0643	.1315	.0537	.0337
200	.1117	.0415	.0373	.1288	.0425	.0613	.1044	.0412	.0264
500	.0880	.0286	.0292	.1053	.0331	.0420	.0768	.0283	.0171
1 000	.0728	.0223	.0249	.0871	.0278	.0293	.0593	.0214	.0125
Process $\eta^{(4)}$									
50	.6651	.2191	.1145	.7244	.1970	.1738	.6116	.2571	.0832
100	.5670	.1635	.0697	.6185	.1761	.1201	.5113	.1777	.0495
200	.5238	.1231	.0618	.5933	.1340	.1125	.4374	.1399	.0393
500	.4382	.0823	.0498	.5025	.1173	.0730	.3399	.0924	.0246
1 000	.3735	.0596	.0439	.4297	.1066	.0485	.2720	.0657	.0173

We apply the sample size dependent rule  $M_T = [4(T/100)^{2/9}]$  to calculate the bandwidth parameter.

The simulation results shown in Table 1 reflect the theoretical result of Theorem 3.3, i.e. the convergence of the estimator to the spectral density of the best approximating stationary process. In all cases, the root mean square error (RMSE) decreases as the sample size increases, with the results being best when the Quadratic Spectral kernel is used. The considered frequency does not play a major role. However, it is noticeable that when the time-averaged spectral density of the second, third and fourth process is estimated, the RMSE at the long-run frequency (i.e.  $\lambda = 0$ ) is substantially higher than at the other frequencies. This is easily attributed to the autoregressive component of the underlying processes. Such effects have already been observed in the estimation of spectral densities of stationary processes and are therefore not surprising.

## 4. Applications

We present two applications of the results presented in Section 3. At first, we show that textbook HAC-inference remains valid asymptotically in a locally stationary multiple linear regression model. As a second application, we introduce the spectrum profile for locally stationary

processes which is a straightforward generalization of the variance profile for heteroskedastic processes that has been introduced by Cavaliere and Taylor (2007a).

#### 4.1. HAC Standard Errors in the Linear Regression Model

Consider the multiple linear regression model in triangular form

$$y_{t,T} = x'_{t,T}\beta + u_{t,T}, \quad t = 1, \dots, T, \quad (5)$$

where we assume that  $\mathbb{E}(u_{t,T}|x_{t,T}) = 0$  for all  $t = 1, \dots, T$  ( $T \in \mathbb{N}$ ) and that there exists a nonsingular matrix  $M$  such that

$$\frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T} \xrightarrow{P} M. \quad (6)$$

Define  $\eta_{t,T} = x_{t,T}u_{t,T}$  and assume that  $\{\eta_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  is locally stationary. By the central limit theorem for locally stationary processes (see Kawka, 2019, Theorem 2.1) it holds that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \eta_{t,T} \xrightarrow{d} \mathcal{N}\left(0, 2\pi \int_0^1 f(u, 0) ds\right).$$

With this result in place we can easily obtain the following asymptotic distribution of the centered OLS estimator for the parameter vector  $\beta$ . It holds that

$$\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T}\right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T x_{t,T}u_{t,T}\right) \xrightarrow{d} \mathcal{N}(0, V),$$

where the covariance matrix is given by

$$V = 2\pi \int_0^1 M^{-1} f(u, 0) M^{-1} du.$$

In order to do inference on  $\beta$  an estimator for  $V$  is required. Clearly, a consistent estimator is given by

$$\hat{V} = \left(\frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T}\right)^{-1} 2\pi \hat{f}(0) \left(\frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T}\right)^{-1},$$

where  $\hat{f}(0)$  is defined as in (4) with  $\eta_{t,T} = x_{t,T}u_{t,T}$ . This estimator is, however, not feasible since we do not observe the error terms  $u_{t,T}$  directly. Therefore, we can replace  $\eta_{t,T}$  by  $\hat{\eta}_{t,T} = x_{t,T}\hat{u}_{t,T}$ , where  $\hat{u}_{t,T}$  denote the OLS residuals from (5). The following theorem justifies this approach.

**Theorem 4.1.** *Assume that  $\eta \in X$  with time-varying spectral density  $f$  and let  $\{X_{t,T}\}_{1 \leq t \leq T; T \in \mathbb{N}}$  be a triangular sequence satisfying  $\sup_{t \leq T} \|D_T X_{t,T}\| = \mathcal{O}_p(1)$  for nonsingular matrices  $\{D_T\}_{T \in \mathbb{N}}$ . Assume further that  $\eta_{t,T} = \eta_{t,T}(\theta_0)$  for some unknown parameter  $\theta_0$  and let*

$$\eta_{t,T}(\theta) = \eta_{t,T}(\theta_0) - (\theta - \theta_0)X_{t,T}.$$

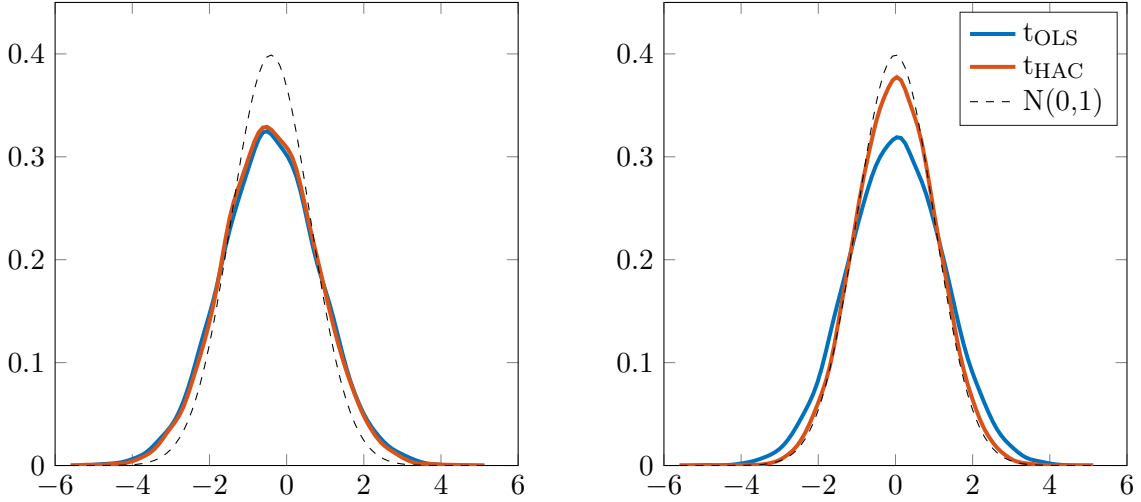
*If  $\hat{\theta}$  is an estimator for  $\theta_0$  such that  $\sqrt{T}(\hat{\theta} - \theta_0)D_T^{-1} = \mathcal{O}_p(1)$  then, as  $T \rightarrow \infty$ ,*

$$\hat{f}(\lambda)(\hat{\theta}) = \frac{1}{2\pi} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e^{-i\lambda(t-s)} k\left(\frac{|t-s|}{M_T}\right) \eta_{t,T}(\hat{\theta}) \eta_{s,T}(\hat{\theta})' \xrightarrow{P} \int_0^1 f(u, \lambda) du. \quad (7)$$

Clearly, it holds that  $\hat{\eta}_{t,T} = \eta_{t,T} - x_{t,T}x'_{t,T}(\hat{\beta} - \beta)$  and (6) implies  $\sup_{t \leq T} \|x_{t,T}x'_{t,T}\| = \mathcal{O}_P(1)$ . Consequently, a feasible estimator for the covariance matrix  $V$  is given by

$$\hat{V}(\hat{\beta}) = \left( \frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T} \right)^{-1} 2\pi \hat{f}(0)(\hat{\beta}) \left( \frac{1}{T} \sum_{t=1}^T x_{t,T}x'_{t,T} \right)^{-1},$$

where  $\hat{f}(0)(\hat{\beta})$  is defined in (7). Hence, we can compute HAC-type standard errors in exactly the same way as it is well known from standard econometrics textbooks (e.g. Verbeek, 2017, Chapter 4).



**Figure 1:** Estimated densities corresponding to the  $t$ -statistics for  $H_0 : \beta_1 + \beta_2 + \beta_3 = 3$  in (8). Left panel:  $T = 100$ . Right panel:  $T = 1000$ .

**Example 4.2** ( $t$ -test). Consider the multiple linear regression model

$$y_{t,T} = \beta_0 + \beta_1 \eta_{t,T}^{(1)} + \beta_2 \eta_{t,T}^{(2)} + \beta_3 \eta_{t,T}^{(3)} + \eta_{t,T}^{(4)}, \quad (8)$$

where the processes  $\eta^{(1)}, \dots, \eta^{(4)}$  are defined in Section 3.1 and are assumed to be independent of each other. The parameters are set to  $\beta_j = 1$  for  $j = 0, \dots, 3$ . We test the null hypothesis  $H_0 : \beta_1 + \beta_2 + \beta_3 = 3$  against the two sided alternative using a simple  $t$ -test and simulate the model 10 000 times. The empirical distributions of the test statistics, based on OLS- and HAC- standard errors respectively, are displayed in Figure 1. For small samples ( $T = 100$ ) the densities are nearly indistinguishable and differ substantially from the Gaussian bell curve. In case of larger sample sizes ( $T = 1000$ ), however, the density curve of the HAC-statistic and the Gaussian density curve are very close to each other, whereas the density curve of the standard OLS-statistic remains almost unchanged compared to the small sample size case.

## 4.2. The Spectrum Profile

As mentioned in Example 2.9, Cavaliere and Taylor (2007a,b) consider unconditionally heteroskedastic processes of the form

$$\nu_{t,T} = \sum_{j=0}^{\infty} C_j \omega \left( \frac{t-j}{T} \right) \varepsilon_{t-j}. \quad (9)$$

For fixed  $u \in [0, 1]$  the auxiliary stationary process  $\{\tilde{\nu}(u)\}_{t \in \mathbb{Z}}$  is given by

$$\tilde{\nu}_t(u) = \sum_{j=0}^{\infty} C_j \omega(u) \varepsilon_{t-j} = \omega(u) \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} = \omega(u) \xi_t,$$

with  $\{\xi_t\}_{t \in \mathbb{Z}}$  being stationary with variance  $\sigma_\xi^2$  and spectral density  $f_\xi(\lambda)$ . Hence, the time-varying variance function and the time-varying spectral density function of  $\{\nu\}_{1 \leq t \leq T; T \in \mathbb{N}}$  are given by  $\Sigma(u) = \omega^2(u) \sigma_\xi^2$  and  $f(u, \lambda) = \omega^2(u) f_\xi(\lambda)$ .

For these type of processes Cavaliere and Taylor (2007a) introduce the *variance profile*, a function that basically includes all information about the degree of unconditional heteroskedasticity of the underlying process. The variance profile is defined as the ratio of the aggregated local (short-run) variance function and the integrated or time-averaged variance, i.e.

$$V(r) = \frac{\int_0^r \Sigma(u) du}{\int_0^1 \Sigma(u) du} = \frac{\int_0^r \omega^2(u) \sigma_\xi^2 du}{\int_0^1 \omega^2(u) \sigma_\xi^2 du} = \frac{\int_0^r \omega^2(u) du}{\int_0^1 \omega^2(u) du}.$$

If  $\omega$  is constant, i.e. if the process is weakly stationary, the variance profile is simply given by the identity function and strong deviations from it indicate an unconditionally heteroskedastic behavior of the process. However, a time-varying second order moment structure does not necessarily imply a deviation of the variance profile from the identity function. For example, consider the following time-varying MA(1) process  $\eta \in X$ , defined by

$$\eta_{t,T} = \cos \left( \frac{\pi t}{T} \right) \varepsilon_t + \sin \left( \frac{\pi t}{T} \right) \varepsilon_{t-1}. \quad (10)$$

By simple calculation we obtain  $\text{Var}(\eta_{t,T}) = 1$  for all  $t = 1, \dots, T$ . Hence, the variance profile is given by  $V(r) = r$ . On the other hand, it holds that

$$\text{Cov}(\eta_{t+1,T}, \eta_{t,T}) = \cos \left( \frac{\pi t}{T} \right) \sin \left( \frac{\pi t}{T} \right) = \frac{1}{2} \sin \left( \frac{2\pi t}{T} \right),$$

which is clearly time-varying. To overcome this deficit we propose to generalize the variance profile by considering the *spectrum profile*, defined as the ratio of the aggregated spectral density and the time-averaged spectral density at a specific frequency, i.e.

$$\Xi(r, \lambda) = \frac{\int_0^r f(u, \lambda) du}{\int_0^1 f(u, \lambda) du}.$$



**Remark 4.3.** For processes of the form (9) as well as for amplitude modulated processes (cf. Example 2.7) the spectrum profile is equal to the variance profile since

$$\Xi(r, \lambda) = \frac{\int_0^r f(u, \lambda) du}{\int_0^1 f(u, \lambda) du} = \frac{\int_0^r \omega^2(u) f_\xi(\lambda) du}{\int_0^1 \omega^2(u) f_\xi(\lambda) du} = \frac{\int_0^r \omega^2(u) du}{\int_0^1 \omega^2(u) du} = V(r).$$

Consequently, for two frequencies  $\lambda_1, \lambda_2 \in (-\pi, \pi]$  it holds that  $\Xi(r, \lambda_1) = \Xi(r, \lambda_2)$ .

The following generalization of Theorem 3.4 yields a consistent estimator of the spectrum profile.

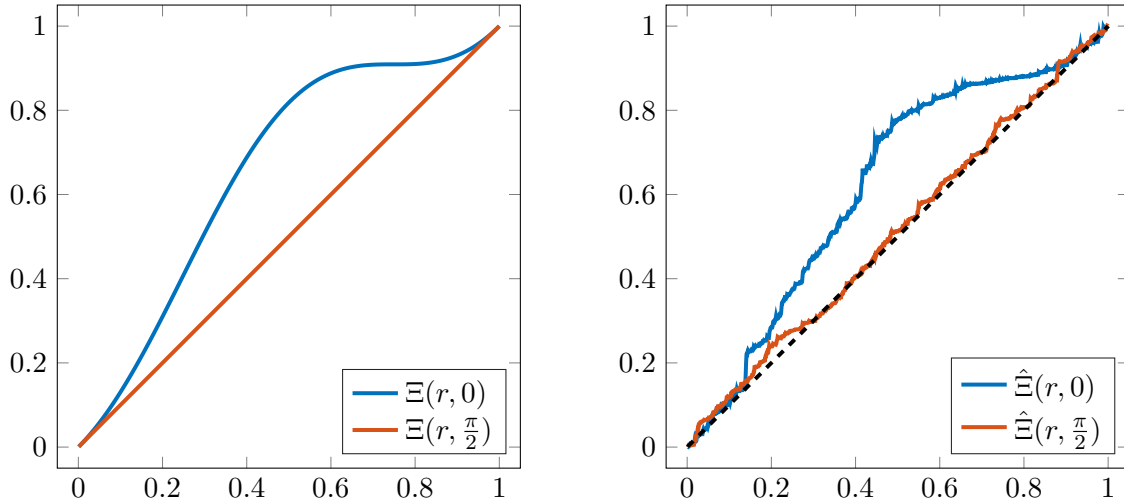
**Corollary 4.4.** Let  $\eta \in X$  with time-varying spectral density function  $f$  and let  $r \in (0, 1]$ . Denote by  $\hat{f}_r(\lambda)$  the spectral density estimator based on the first  $[rT]$  observations only, i.e.,

$$\hat{f}_r(\lambda) := \frac{1}{2\pi} \frac{1}{[rT]} \sum_{t=1}^{[rT]} \sum_{s=1}^{[rT]} e^{-i\lambda(t-s)} k\left(\frac{|t-s|}{M_T}\right) \eta_{t,T} \eta'_{s,T}.$$

Then, under Assumption 3.3, it holds that

$$\hat{f}_r(\lambda) \xrightarrow{P} \frac{1}{r} \int_0^r f(u, \lambda) du,$$

for all  $\lambda \in (-\pi, \pi]$ .



**Figure 2:** Spectrum profile of a time series satisfying the difference equation (10), generated with i.i.d. standard normal  $\varepsilon_t$  for  $T = 1000$ . Left panel: Spectrum profile at frequencies  $\omega = 0$  and  $\omega = \frac{\pi}{2}$ . Right panel: The corresponding estimates.

With this result in place we can consistently estimate the spectrum profile via

$$\hat{\Xi}(r, \lambda) = \frac{[rT] \hat{f}_r(\lambda)}{T \hat{f}(\lambda)}.$$

Note that the estimator function  $\hat{\Xi}(\cdot, \lambda)$  is a step function for every  $T \in \mathbb{N}$ . We can easily define a continuous version of the spectrum profile estimator by linear interpolation. Consistency is in this case – of course – not affected.

**Example 4.5.** We illustrate the estimation of the spectrum profile for a time series generated as a realization of a process that satisfies equation (10). To calculate the theoretical spectrum profile note that the time-varying spectral density function is given by

$$f(r, \lambda) = \left| \cos(\pi r) + \sin(\pi r) e^{i\lambda} \right|^2 = 1 + \sin(2\pi r) \cos(\lambda).$$

Hence, elementary calculus yields

$$\Xi(r, \lambda) = \frac{\int_0^r 1 + \cos(\lambda) \sin(2\pi x) dx}{\int_0^1 1 + \cos(\lambda) \sin(2\pi x) dx} = r + \frac{\cos(\lambda)}{2\pi} (1 - \cos(2\pi r)).$$

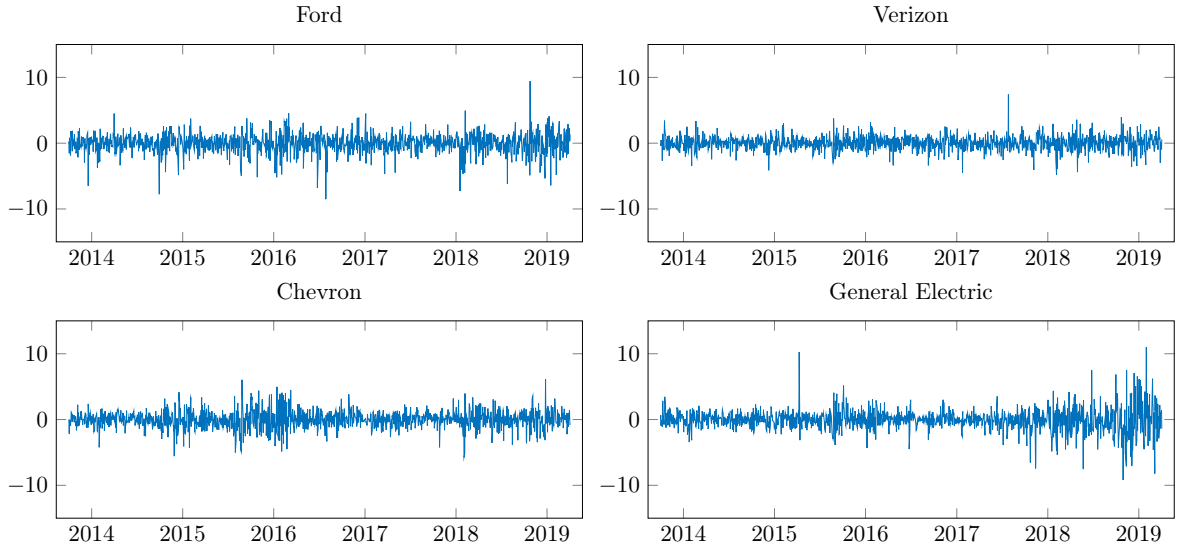
Note that for  $\lambda = \pi/2$  the spectrum profile and the variance profile coincide and are given by the identity function, i.e.  $\Xi(r, \pi/2) = r$ . Figure 2 shows the estimated spectrum profile at frequencies  $\lambda = 0$  and  $\lambda = \pi/2$  for a simulated time series.

**Remark 4.6.** To visualize the spectrum profile for all frequencies simultaneously it is beneficial to modify  $\Xi(u, \lambda)$  by subtracting the identity function, i.e. to consider

$$\Xi_M(r, \lambda) = \Xi(r, \lambda) - r = \frac{\int_0^r f(u, \lambda) du - r \int_0^1 f(u, \lambda) du}{\int_0^1 f(u, \lambda) du}.$$

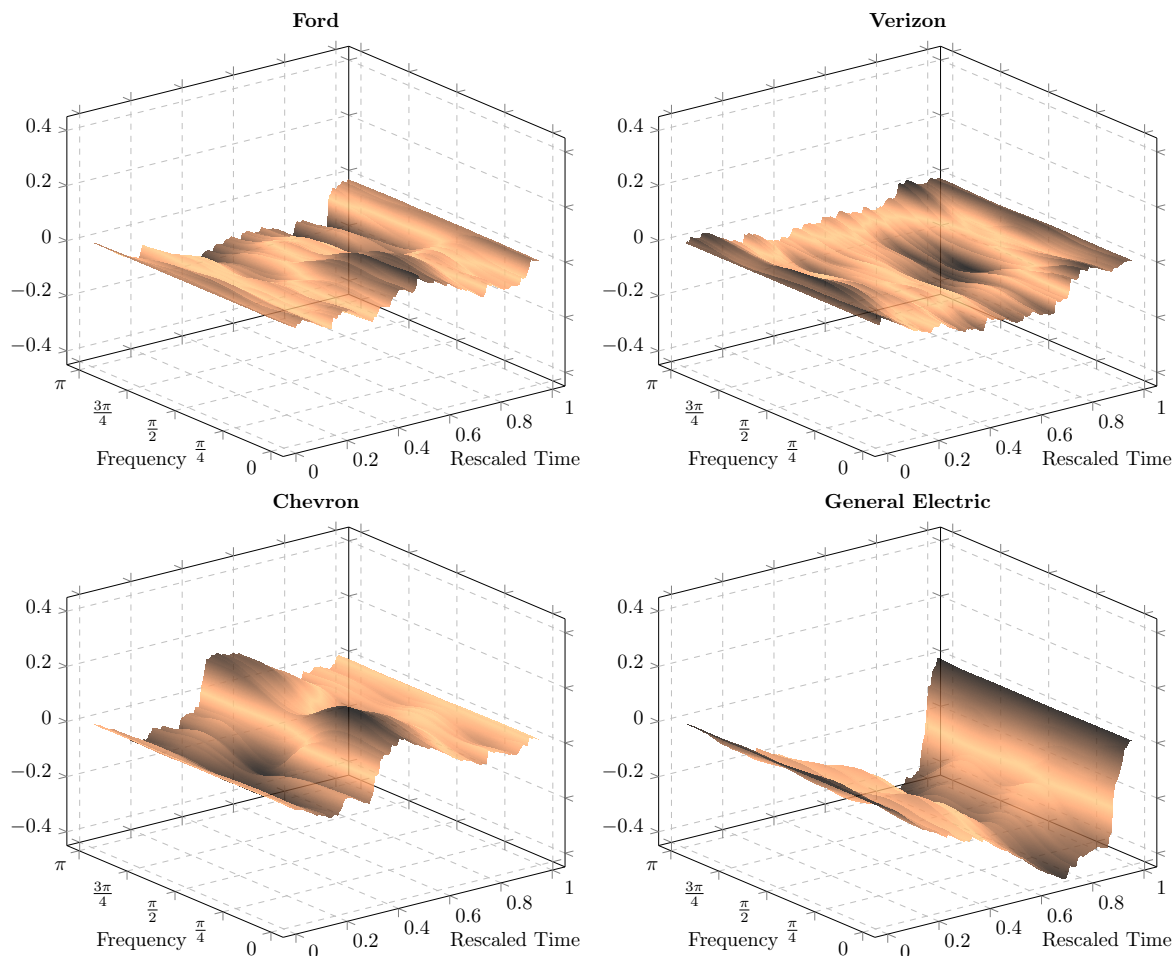
The *modified spectrum profile* can now be simply estimated by

$$\hat{\Xi}_M(r, \lambda) = \frac{r \hat{f}_r(\lambda) - r \hat{f}(\lambda)}{\hat{f}(\lambda)}.$$



**Figure 3:** Daily logarithmic returns (in %) of closing stock prices of Ford (top left), Verizon (top right), Chevron (bottom left) and General Electric (bottom right).

**Example 4.7 (Stock Market Data).** We consider the logarithmic returns of the daily closing price of four major public companies (Ford, Verizon, Chevron and General Electric) obtained



**Figure 4:** Estimated modified spectrum profiles of the Ford (top left), Verizon (top right), Chevron (bottom left) and General Electric (bottom right) series.

from Yahoo! Finance. Each series (cf. Figure 3) spans the period from October 1st, 2013 to April 3rd, 2019 and contains 1.385 observations. The Ford and the Verizon series do not display any severe changes in volatility whereas the General Electric and the Chevron series clearly change their appearance over time. Thus, modeling the two latter series by stationary time series models seems to be inappropriate. The estimated (modified) spectrum profiles, which are depicted in Figure 4, affirm this presumption. In particular, the spectrum profiles of the Ford and the Verizon series are comparatively flat and close to zero irrespective of rescaled time and frequency. The spectrum profiles of the Chevron and General Electric series, however, clearly vary over time and frequency. Hence, locally stationary models appear to be more suitable for these series.

**Remark 4.8.** Example 4.7 reveals that heteroskedastic processes of the form (9) can only model time series with time-variable second order moment structure to a limited extent. In particular, the time-variation in the spectrum profile of the General Electric series is almost independent of the frequency, indicating that heteroskedastic processes might be a good choice for modeling this series. On the contrary, the shape of the spectrum profile of the Chevron

series heavily hinges upon the frequency. Hence, heteroskedastic processes do not mimic the behaviour of this series properly.

## 5. Conclusion

In this article we have shown that for any locally stationary process, that can be considered as an element in a Hilbert space, there exists an optimal approximation by a stationary process. This approximation is given by the orthogonal projection from the space of all locally stationary processes on the closed subspace that contains all stationary processes. If a locally stationary process has a time-varying spectral density  $f(u, \lambda)$  then the spectral density of the best approximating stationary process is given by  $\int_0^1 f(u, \lambda) du$ , i.e. by the time-averaged value of  $f(u, \lambda)$ . Furthermore, spectral density estimators that are designed to estimate the (time-invariant) spectral density of a stationary process, consistently estimate this particular quantity (Theorem 3.4). Based upon this result we have presented two interesting applications. On the one hand, the consistent estimation of the time-averaged spectral density justifies the usage of HAC standard errors for linear hypothesis testing in the multiple linear regression model under unconditional heteroskedasticity. On the other hand, it allows us to estimate the spectrum profile which is a simple visual tool for the detection of unconditional heteroskedasticity in empirical time series analysis.

## Acknowledgements

This work has been supported in part by the Collaborative Research Center *Statistical modeling of nonlinear dynamic processes* (SFB 823, Teilprojekt A3/A4) of the German Research Foundation (DFG).

## References

- Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association* 93, 1488 – 1501.
- Amado, C. and T. Teräsvirta (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance* 25, 15–35.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley & Sons.
- Cavaliere, G. and A. R. Taylor (2007a). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics* 140, 919–947.
- Cavaliere, G. and A. R. Taylor (2007b). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis* 29(2), 300–330.
- Dahlhaus, R. (1996a). Maximum likelihood estimation and model selection for locally stationary processes. *Nonparametric Statistics* 6, 171 – 191.

- Dahlhaus, R. (1996b). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications* 62, 139–168.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* 25, 1–37.
- Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics* 28, 1762 – 1794.
- Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics* 151, 101–112.
- Dahlhaus, R. (2012). Locally stationary processes. In T. S. Rao, S. S. Rao, and C. Rao (Eds.), *Handbook of Statistics, Volume 30 of Time Series Analysis: Methods and Applications*, Chapter 13, pp. 351–413. Elsevier.
- Dahlhaus, R. and W. Polonik (2006). Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes. *The Annals of Statistics* 34, 2790 – 2824.
- Dahlhaus, R. and W. Polonik (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* 15, 1–39.
- Dahlhaus, R. and S. S. Rao (2006). Statistical inference for time-varying arch processes. *The Annals of Statistics* 34, 1075 – 1114.
- Deistler, M. and W. Scherrer (2018). *Modelle der Zeitreihenanalyse*. Springer Birkhäuser.
- Dette, H., P. Preuß, and M. Vetter (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* 106, 1113–1124.
- Hansen, B. E. (1992). Consistent covariance matrix estimator for dependent heterogeneous processes. *Econometrica* 60, 967–972.
- Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.
- Kawka, R. (2019). Limit theorems for locally stationary processes. *SFB 823 Discussion Paper 09/19*.
- Künsch, H. R. (1995). A note on causal solutions for locally stationary ar-processes. Preprint ETH Zürich.
- Mikosch, T. and C. Starica (2004). Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *The Review of Economics and Statistics* 86, 378–390.
- Nason, G. P. and R. von Sachs (1999). Wavelets in time-series analysis. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 357, 2511 – 2526.
- Ombao, H. C., J. A. Raz, R. von Sachs, and B. A. Malow (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association* 96, 543 – 560.

Paparoditis, E. and P. Preuß (2014). Estimation of the bispectrum for locally stationary processes. *Statistics and Probability Letters* 89, 8–16.

Preuß, P., R. Puchstein, and H. Dette (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* 110, 654 – 668.

Sergides, M. and E. Paparoditis (2009). Frequency domain tests of semiparametric hypothesis for locally stationary processes. *Scandinavian Journal of Statistics* 36, 800 – 821.

Starica, C. and C. Granger (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics* 87, 503–522.

van Bellegem, S. (2012). Locally stationary volatility modeling. In *Handbook of Volatility Models*. Wiley.

Verbeek, M. (2017). *A Guide to Modern Econometrics*. Wiley.

## A. Auxiliary Lemmata

**Lemma A.1.** *Under Assumption 3.3 it holds that*

$$\lim_{T \rightarrow \infty} \frac{1}{T^{1/2}} \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| = 0.$$

*Proof.* Follows immediately from Jansson (2002, Lemma 1). □

**Lemma A.2.** *Let  $\eta \in X$  and  $\tilde{\Lambda}(\lambda)$  be the modified kernel estimator of  $\Lambda(\lambda)$  based on the auxiliary process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$ , i.e.*

$$\tilde{\Lambda}(\lambda) := \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)'$$

*Under Assumption 3.3 it holds that:*

$$\lim_{T \rightarrow \infty} \mathbb{E} \left\| \hat{\Lambda}(\lambda) - \tilde{\Lambda}(\lambda) \right\| = 0.$$

*Proof.* It holds that

$$\begin{aligned} & \left\| \eta_{t+h,T} \eta'_{t,T} - \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\| \\ & \leq \left\| \eta_{t+h,T} \eta'_{t,T} - \eta_{t+h,T} \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\| + \left\| \eta_{t+h,T} \tilde{\eta}_t \left( \frac{t}{T} \right)' - \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\| \\ & \leq \|\eta_{t+h,T}\| \left\| \eta'_{t,T} - \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\| + \left\| \eta_{t+h,T} - \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \right\| \left\| \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\|. \end{aligned}$$

Assumption 2.2 directly implies that

$$\|\eta_{t,T}\| \leq \sum_{j=0}^{\infty} \|\psi_{j,t,T}\| \|\varepsilon_{t-j}\| \leq \sum_{j=0}^{\infty} \frac{K}{l(j)} \|\varepsilon_{t-j}\|$$

and

$$\left\| \tilde{\eta}_t \left( \frac{t}{T} \right) \right\| \leq \sum_{j=0}^{\infty} \|\psi\|_{\infty} \|\varepsilon_{t-j}\| \leq \sum_{j=0}^{\infty} \frac{K}{l(j)} \|\varepsilon_{t-j}\|.$$

Similarly, by Assumption 2.2 we get

$$\left\| \eta_{t,T} - \tilde{\eta}_t \left( \frac{t}{T} \right) \right\| \leq \sum_{j=0}^{\infty} \left\| \psi_{j,t,T} - \psi_j \left( \frac{t}{T} \right) \right\| \|\varepsilon_{t-j}\|.$$

Consequently, it follows that

$$\begin{aligned} \mathbb{E} \left\| \hat{\Lambda}(\lambda) - \check{\Lambda}(\lambda) \right\| &= \mathbb{E} \left\| \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \left[ \eta_{t+h,T} \eta'_{t,T} - \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] \right\| \\ &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \mathbb{E} \left\| \eta_{t+h,T} \eta'_{t,T} - \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right\| \\ &\leq \frac{2}{T} \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \sum_{j=0}^{\infty} \frac{K}{l(j)} \sum_{k=0}^{\infty} \sum_{t=1}^T \left\| \psi_{k,t,T} - \psi_k \left( \frac{t}{T} \right) \right\| \\ &\leq \frac{2}{T} \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \left( \sum_{j=0}^{\infty} \frac{K}{l(j)} \right)^2. \end{aligned}$$

The expression on the right hand side goes to zero by Lemma A.1.  $\square$

**Lemma A.3.** *Let  $\eta \in X$  and let  $\check{\Lambda}(\lambda)$  be defined by*

$$\check{\Lambda}(\lambda) := \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)',$$

for the auxiliary process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$ . With Assumption 3.3 in place, it holds that

$$\lim_{T \rightarrow \infty} \mathbb{E} \left\| \tilde{\Lambda}(\lambda) - \check{\Lambda}(\lambda) \right\| = 0.$$

*Proof.* It holds that

$$\begin{aligned} &\left\| \tilde{\Lambda}(\lambda) - \check{\Lambda}(\lambda) \right\| \\ &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\| \left[ \psi_j \left( \frac{t+h}{T} \right) - \psi_j \left( \frac{t}{T} \right) \right] \varepsilon_{t+h-j} \varepsilon'_{t-k} \psi_k \left( \frac{t}{T} \right)' \right\| \\ &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left\| \psi_j \left( \frac{t+h}{T} \right) - \psi_j \left( \frac{t}{T} \right) \right\| \left\| \psi_k \left( \frac{t}{T} \right)' \right\| \|\varepsilon_{t+h-j} \varepsilon'_{t-k}\| \\ &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \frac{1}{T} \sum_{t=1}^{T-h} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} V_{\psi_j} [0, 1] \|\psi_k\|_{\infty} \|\varepsilon_{t+h-j} \varepsilon'_{t-k}\|. \end{aligned}$$

Thus, we conclude:

$$\begin{aligned} \mathbb{E} \left\| \tilde{\Lambda}(\lambda) - \check{\Lambda}(\lambda) \right\| &\leq C \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \sum_{j=0}^{\infty} V_{\psi_j} [0, 1] \sum_{k=0}^{\infty} \|\psi_k\|_{\infty} \\ &\leq \frac{C}{T} \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \left( \sum_{j=0}^{\infty} \frac{K}{l(k)} \right)^2, \end{aligned}$$

with  $C$  being a finite constant due to the moment assumption on  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ . Since the two series on the right hand side are finite by Assumption 2.2, the claim follows by Lemma A.1 and Lemma A.2.  $\square$

**Lemma A.4.** *Let  $\eta \in X$  with corresponding auxiliary process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$  and let  $\delta > 0$ . Then, for  $0 \leq h \leq T-1$ , it holds that*

$$\mathbb{E} \left\| \frac{1}{T^{1/2+\delta}} \sum_{t=1}^{T-h} \left\{ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' - \mathbb{E} \left[ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] \right\} \right\| \leq \frac{C}{T^{\delta}},$$

where

$$C = n \left( \sum_{k=0}^{\infty} \|\psi_k\|_{\infty} \right)^2 (\mathbb{E} \|\varepsilon_t\|^4)^{1/2}.$$

*Proof.* First note that

$$\tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \left( \frac{t}{T} \right) \varepsilon_{t+h-j} \varepsilon'_{t-k} \psi_k \left( \frac{t}{T} \right)'$$

and, consequently,

$$\begin{aligned} \mathbb{E} \left[ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \left( \frac{t}{T} \right) \mathbb{E} [\varepsilon_{t+h-j} \varepsilon'_{t-k}] \psi_k \left( \frac{t}{T} \right)' \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \left( \frac{t}{T} \right) [I_n \mathbb{1}_{\{j=k+h\}}] \psi_k \left( \frac{t}{T} \right)' \\ &= \sum_{k=0}^{\infty} \psi_{k+h} \left( \frac{t}{T} \right) \psi_k \left( \frac{t}{T} \right)'. \end{aligned}$$

This immediately implies:

$$\begin{aligned} &\left\| \frac{1}{T^{1/2+\delta}} \sum_{t=1}^{T-h} \left\{ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' - \mathbb{E} \left[ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] \right\} \right\| \\ &\leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{T^{\delta}} \left\| \frac{1}{T^{1/2}} \sum_{t=1}^{T-h} \psi_j \left( \frac{t}{T} \right) [\varepsilon_{t+h-j} \varepsilon'_{t-k} - I_n \mathbb{1}_{\{j=k+h\}}] \psi_k \left( \frac{t}{T} \right)' \right\|. \end{aligned}$$



Thus, we obtain:

$$\begin{aligned}
& \mathbb{E} \left\| \frac{1}{T^{1/2+\delta}} \sum_{t=1}^{T-h} \left\{ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' - \mathbb{E} \left[ \tilde{\eta}_{t+h} \left( \frac{t}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] \right\} \right\| \\
& \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{T^\delta} \mathbb{E} \left\| \frac{1}{T^{1/2}} \sum_{t=1}^{T-h} \psi_j \left( \frac{t}{T} \right) [\varepsilon_{t+h-j} \varepsilon'_{t-k} - I_n \mathbb{1}_{\{j=k+h\}}] \psi_k \left( \frac{t}{T} \right)' \right\| \\
& \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{T^\delta} \left( \mathbb{E} \left\| \frac{1}{T^{1/2}} \sum_{t=1}^{T-h} \psi_j \left( \frac{t}{T} \right) [\varepsilon_{t+h-j} \varepsilon'_{t-k} - I_n \mathbb{1}_{\{j=k+h\}}] \psi_k \left( \frac{t}{T} \right)' \right\|^2 \right)^{1/2} \\
& \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{T^\delta} \left( \mathbb{E} \left[ \|\psi_j\|_\infty^2 \|\psi_k\|_\infty^2 \|\varepsilon_{1+h-j} \varepsilon'_{1-k} - I_n \mathbb{1}_{\{j=k+h\}}\|^2 \right] \right)^{1/2},
\end{aligned}$$

where the second last line follows by the Cauchy-Schwarz inequality and the last by the i.i.d. assumption on  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  and the submultiplicativity of the matrix norm. Finally, we continue as follows:

$$\begin{aligned}
& \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{T^\delta} \left( \mathbb{E} \left[ \|\psi_j\|_\infty^2 \|\psi_k\|_\infty^2 \|\varepsilon_{1+h-j} \varepsilon'_{1-k} - I_n \mathbb{1}_{\{j=k+h\}}\|^2 \right] \right)^{1/2} \\
& \leq \frac{1}{T^\delta} \sum_{j=0}^{\infty} \|\psi_j\|_\infty \sum_{k=0}^{\infty} \|\psi_k\|_\infty \left( \mathbb{E} \|\varepsilon_{1+h-j}\|^2 \|\varepsilon_{1-k}\|^2 \right)^{1/2} \\
& \leq \frac{1}{T^\delta} \sum_{j=0}^{\infty} \|\psi_j\|_\infty \sum_{k=0}^{\infty} \|\psi_k\|_\infty \left( \sum_{l=1}^n \sum_{m=1}^n \sqrt{\mathbb{E}(\varepsilon_{1+h-j,l}^4)} \sqrt{\mathbb{E}(\varepsilon_{1-k,l}^4)} \right)^{1/2} \\
& \leq \frac{1}{T^\delta} \sum_{j=0}^{\infty} \|\psi_j\|_\infty \sum_{k=0}^{\infty} \|\psi_k\|_\infty \left( n^2 \cdot \max_{l=1, \dots, n} \mathbb{E}(\varepsilon_{1,l}^4) \right)^{1/2} \\
& \leq \frac{1}{T^\delta} \left( \sum_{j=0}^{\infty} \|\psi_j\|_\infty \right)^2 \cdot n \cdot (\mathbb{E} \|\varepsilon_1\|^4)^{1/2},
\end{aligned}$$

where the third line follows again from the Cauchy-Schwarz inequality and the definition of the Euclidean norm. This completes the proof.  $\square$

**Lemma A.5.** *Let  $\eta \in X$  with auxiliary process  $\{\tilde{\eta}_t(u)\}_{t \in \mathbb{Z}}$ . Under Assumption 3.3 it holds that:*

$$\lim_{T \rightarrow \infty} \mathbb{E} \check{\Lambda}(\lambda) = \int_0^1 \Lambda(u, \lambda) du,$$

with  $\check{\Lambda}(\lambda)$  defined in Lemma A.3.

*Proof.* Since the series  $\sum_k \|\psi_{k+h}\|_\infty \|\psi_k\|_\infty$  converges for all  $h \in \mathbb{N}$ , it holds that

$$\int_0^1 \Lambda(u, \lambda) du = \int_0^1 \sum_{h=1}^{\infty} e^{-i\lambda h} \sum_{k=0}^{\infty} \psi_{k+h}(u) \psi_k(u)' du = \sum_{h=1}^{\infty} e^{-i\lambda h} \sum_{k=0}^{\infty} \int_0^1 \psi_{k+h}(u) \psi_k(u)' du.$$

Therefore, we consider the following approximation of  $\int_0^1 \Lambda(u, \lambda) du$ :

$$\Lambda^a(\lambda) := \frac{1}{T} \sum_{t=1}^T \sum_{h=1}^{\infty} e^{-i\lambda h} \Gamma_h \left( \frac{t}{T} \right),$$

with

$$\Gamma_h(u) := \sum_{k=0}^{\infty} \psi_{k+h}(u) \psi_k(u)'$$

denoting the time-varying covariance matrix. Clearly, for fixed  $\lambda \in [-\pi, \pi]$ , it holds that

$$\left\| \Lambda^a(\lambda) - \int_0^1 \Lambda(u, \lambda) du \right\| \rightarrow 0,$$

as  $T \rightarrow \infty$ . Thus, it remains to prove that  $\|\check{\Lambda}(\lambda) - \Lambda^a(\lambda)\|$  converges to zero. It holds that

$$\|\mathbb{E}\check{\Lambda}(\lambda) - \Lambda^a(\lambda)\| = \left\| \sum_{h=1}^{\infty} e^{-i\lambda h} \left( \mathbb{1}_{\{h \leq T-1\}} k \left( \frac{h}{M_T} \right) \frac{1}{T} \sum_{t=1}^{T-h} \Gamma_h \left( \frac{t}{T} \right) - \frac{1}{T} \sum_{t=1}^T \Gamma_h \left( \frac{t}{T} \right) \right) \right\|$$

and after applying the triangular inequality it follows that

$$\|\mathbb{E}\check{\Lambda}(\lambda) - \Lambda^a(\lambda)\| \leq S_1 + S_2,$$

where

$$\begin{aligned} S_1 &= \sum_{h=1}^{\infty} \left| \mathbb{1}_{\{h \leq T-1\}} k \left( \frac{h}{M_T} \right) - 1 \right| \left\| \frac{1}{T} \sum_{t=1}^T \Gamma_h \left( \frac{t}{T} \right) \right\|, \\ S_2 &= \sum_{h=1}^{\infty} \left| \mathbb{1}_{\{h \leq T-1\}} k \left( \frac{h}{M_T} \right) \right| \left\| \frac{1}{T} \sum_{t=T-h+1}^T \Gamma_h \left( \frac{t}{T} \right) \right\|. \end{aligned}$$

First, we show that  $S_2$  goes to zero as  $T \rightarrow \infty$ . It holds that

$$\begin{aligned} S_2 &= \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \left\| \frac{1}{T} \sum_{t=T-h+1}^T \Gamma_h \left( \frac{t}{T} \right) \right\| \\ &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \frac{h}{T} \|\Gamma_h\|_{\infty} \\ &\leq \sup_{x \geq 0} |k(x)| \frac{1}{T} \sum_{h=1}^T h \|\Gamma_h\|_{\infty}. \end{aligned} \tag{11}$$

Since

$$\sum_{h=1}^{\infty} h \|\Gamma_h\|_{\infty} \leq \sum_{h=1}^{\infty} \sum_{k=0}^{\infty} h \|\psi_{k+h}\|_{\infty} \|\psi_k\|_{\infty} \leq \sum_{h=0}^{\infty} h \|\psi_h\|_{\infty} \sum_{j=0}^{\infty} \|\psi_j\|_{\infty} < \infty$$

it follows that  $h \|\Gamma_h\|_{\infty} \rightarrow 0$  as  $h \rightarrow \infty$ . Consequently the right hand side of (11) converging to zero as a Caesaro limit.

Finally, we show the convergence of  $S_1$ . It holds that:

$$\begin{aligned} S_1 &= \sum_{h=1}^{\infty} \left| \mathbb{1}_{\{h \leq T-1\}} k\left(\frac{h}{M_T}\right) - 1 \right| \left\| \frac{1}{T} \sum_{t=1}^T \Gamma_h\left(\frac{t}{T}\right) \right\| \\ &\leq \sum_{h=1}^q \left| \mathbb{1}_{\{h \leq T-1\}} k\left(\frac{h}{M_T}\right) - 1 \right| \|\Gamma_h\|_{\infty} + \sum_{h=q+1}^{\infty} \left| \mathbb{1}_{\{h \leq T-1\}} k\left(\frac{h}{M_T}\right) - 1 \right| \|\Gamma_h\|_{\infty}, \end{aligned}$$

for some  $Q \in \mathbb{N}$ . For all  $\varepsilon > 0$  there exists a  $Q \in \mathbb{N}$  such that

$$\left| \sum_{h=Q+1}^{\infty} \|\Gamma_h\|_{\infty} \right| < \varepsilon.$$

Furthermore, for  $T > Q$  it holds that

$$\max_{1 \leq h \leq Q} \left| k\left(\frac{h}{M_T}\right) - 1 \right| \leq \sup_{0 \leq x \leq Q/M_T} |k(x) - 1| \rightarrow 0,$$

since  $M_T^{-1} \rightarrow 0$  as  $T \rightarrow \infty$ . Combining the last two results we conclude that  $S_1$  converges to zero which at last proves the claim.  $\square$

## B. Proofs of Main Theorems

*Proof of Theorem 2.12.* The proof is a straightforward multivariate extension of the proof of Lemma 1 in Dette et al. (2011). Let  $\xi \in X_0$  with spectral density  $g(\lambda)$ . Then it holds that

$$\begin{aligned} \|\eta - \xi\|_X^2 &= \int_0^1 \int_{-\pi}^{\pi} \|f(u, \lambda) - g(\lambda)\|^2 d\lambda du \\ &= \sum_{l=1}^n \sum_{m=1}^n \int_0^1 \int_{-\pi}^{\pi} |f_{lm}(u, \lambda) - g_{lm}(\lambda)|^2 d\lambda du \\ &= \sum_{l=1}^n \sum_{m=1}^n \int_0^1 \int_{-\pi}^{\pi} |f_{lm}(u, \lambda) - g_{lm}^*(\lambda) + g_{lm}^*(\lambda) - g_{lm}(\lambda)|^2 d\lambda du \\ &= \sum_{l=1}^n \sum_{m=1}^n \int_0^1 \int_{-\pi}^{\pi} |f_{lm}(u, \lambda) - g_{lm}^*(\lambda)|^2 + |g_{lm}^*(\lambda) - g_{lm}(\lambda)|^2 d\lambda du \\ &\geq \sum_{l=1}^n \sum_{m=1}^n \int_0^1 \int_{-\pi}^{\pi} |f_{lm}(u, \lambda) - g_{lm}^*(\lambda)|^2 d\lambda du, \end{aligned}$$

where we have equality if and only if  $g_{lm}(\lambda) = f_{lm}^0(\lambda)$  for all  $l, m = 1, \dots, n$ . Thereby, the subscripts  $l$  and  $m$  denote the corresponding components of the matrix valued function. The second to last equality follows from the simple fact that for  $a, b \in \mathbb{C}$  we have  $|a + b|^2 = |a|^2 + |b|^2 + 2(\operatorname{Re}(a)\operatorname{Re}(b) + \operatorname{Im}(a)\operatorname{Im}(b))$  and

$$\int_0^1 \operatorname{Re}(f_{lm}(u, \lambda) - f_{lm}^0(\lambda)) du = \int_0^1 \operatorname{Im}(f_{lm}(u, \lambda) - f_{lm}^0(\lambda)) du = 0.$$

The remainder of the Theorem follows from the Projection Theorem in Hilbert spaces.  $\square$

*Proof of Theorem 3.1.* It holds that

$$\left\| \hat{\Gamma}_h - \int_0^1 \Gamma_h(u) du \right\| \leq \left\| \hat{\Gamma}_h - \tilde{\Gamma}_h \right\| + \left\| \tilde{\Gamma}_h - \mathbb{E}\tilde{\Gamma}_h \right\| + \left\| \mathbb{E}\tilde{\Gamma}_h - \int_0^1 \Gamma_h(u) du \right\|,$$

where

$$\tilde{\Gamma}_h := \frac{1}{T} \sum_{t=1}^{T-h} \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)'$$

Following the lines of the proof of Lemma A.2 we obtain

$$\mathbb{E} \left\| \hat{\Gamma}_h - \tilde{\Gamma}_h \right\| \leq \frac{1}{T} \sum_{t=1}^T \sum_{j=0}^{\infty} \frac{C}{l(j)} \sum_{k=0}^{\infty} \left\| \psi_{k,t,T} - \psi_k \left( \frac{t}{T} \right) \right\| \leq \frac{1}{T} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{C}{l(j)l(k)},$$

with  $C > 0$ . Consequently,  $\mathbb{E} \left\| \hat{\Gamma}_h - \tilde{\Gamma}_h \right\|$  converges to zero as  $T \rightarrow \infty$ .

Lemma A.4 immediately implies (with  $\delta := 1/2$ ) that

$$\lim_{T \rightarrow \infty} \mathbb{E} \left\| \tilde{\Gamma}_h - \mathbb{E}\tilde{\Gamma}_h \right\| = 0.$$

Finally,

$$\lim_{T \rightarrow \infty} \left\| \mathbb{E}\tilde{\Gamma}_h - \int_0^1 \Gamma_h(u) du \right\| = 0$$

follows by standard calculus arguments.  $\square$

*Proof of Theorem 3.4.* Since  $2\pi\hat{f}(\lambda) = \hat{\Sigma} + \hat{\Lambda}(\lambda) + \hat{\Lambda}(\lambda)^*$  and  $\hat{\Sigma} \xrightarrow{P} \Sigma$  by Corollary 3.2, it suffices to show  $\hat{\Lambda}(\lambda) \xrightarrow{P} \Lambda(\lambda)$ . It holds that

$$\mathbb{E} \left\| \hat{\Lambda}(\lambda) - \int_0^1 \Lambda(u, \lambda) du \right\| \leq S_1 + S_2 + S_3 + S_4,$$

where

$$S_1 := \mathbb{E} \left\| \hat{\Lambda}(\lambda) - \tilde{\Lambda}(\lambda) \right\|, \quad S_2 := \mathbb{E} \left\| \tilde{\Lambda}(\lambda) - \check{\Lambda}(\lambda) \right\|, \quad S_3 := \mathbb{E} \left\| \check{\Lambda}(\lambda) - \mathbb{E}\check{\Lambda}(\lambda) \right\|,$$

and

$$S_4 := \left\| \mathbb{E}\check{\Lambda}(\lambda) - \int_0^1 \Lambda(u, \lambda) du \right\|.$$

The convergence of  $S_1$  and  $S_2$  and  $S_4$  to zero follows directly from Lemmas A.2, A.3 and A.5. Thus, it remains to show that  $S_3 \rightarrow 0$  as  $T \rightarrow \infty$ . It holds that

$$\begin{aligned} S_3 &\leq \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right| \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \left\{ \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' - \mathbb{E} \left[ \tilde{\eta}_{t+h} \left( \frac{t+h}{T} \right) \tilde{\eta}_t \left( \frac{t}{T} \right)' \right] \right\} \right\| \\ &\leq n \left( \sum_{k=0}^{\infty} \|\psi_k\|_{\infty} \right)^2 (\mathbb{E}\|\varepsilon_t\|^4)^{1/2} \frac{1}{T^{1/2}} \sum_{h=1}^{T-1} \left| k \left( \frac{h}{M_T} \right) \right|, \end{aligned}$$

by Lemma A.4. The convergence to zero and as a consequence the claim of the theorem follows from Lemma A.1.  $\square$

*Proof of Corollary 4.4.* The proof is essentially in line with the proof of Theorem 3.4. To this end we consider estimators  $\hat{\Lambda}_{|r}(\lambda)$  and  $\hat{\Sigma}_{|r}(\lambda)$ , which are constructed in the same way as  $\hat{f}_{|r}(\lambda)$ . It suffices to show the claim for the most difficult term  $\hat{\Lambda}_{|r}(\lambda)$ . We can easily define  $\check{\Lambda}_{|r}(\lambda)$  and  $\check{\Lambda}_{|r}(\lambda)$  as we have done in Lemma A.2 and Lemma A.3. In fact, these Lemmas, as well as Lemma A.4, are still valid when  $\hat{\Lambda}(\lambda)$  is replaced by  $\hat{\Lambda}_{|r}(\lambda)$ . We thus have to show that

$$\check{\Lambda}_{|r}(\lambda) \rightarrow \frac{1}{r} \int_0^r \Lambda(u, \lambda) du.$$

Since we have

$$\Lambda_{|r}^a(\lambda) = \frac{1}{[rT]} \sum_{t=1}^{[rT]} \sum_{h=1}^{\infty} e^{-i\lambda h} \gamma_h \left( \frac{t}{T} \right) = \frac{T}{[rT]} \frac{1}{T} \sum_{t=1}^{[rT]} \sum_{h=1}^{\infty} e^{-i\lambda h} \gamma_h \left( \frac{t}{T} \right) \rightarrow \frac{1}{r} \int_0^r \Lambda(u, \lambda) du,$$

it remains to show that  $\|\check{\Lambda}_{|r}(\lambda) - \Lambda_{|r}^a(\lambda)\|$  goes to zero. This is, however, done in exactly the same fashion as in the proof of Lemma A.5.  $\square$

*Proof of Theorem 4.1.* We adapt the proof of Hansen (1992, Theorem 3), now for triangular arrays and arbitrary frequency  $\lambda$ . Let

$$\hat{\Sigma}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \eta_{t,T}(\hat{\theta}) \eta_{t,T}(\hat{\theta})'$$

and

$$\hat{\Lambda}(\lambda)(\hat{\theta}) = \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T}(\hat{\theta}) \eta_{t,T}(\hat{\theta})'.$$

It holds that

$$\begin{aligned} & \eta_{t+h,T}(\hat{\theta}) \eta_{t,T}(\hat{\theta})' - \eta_{t+h,T} \eta_{t,T}' \\ &= -\eta_{t+h,T} X_{t,T}' (\hat{\theta} - \theta_0)' - (\hat{\theta} - \theta_0) X_{t+h,T} \eta_{t,T}' + (\hat{\theta} - \theta_0) X_{t+h,T} X_{t,T}' (\hat{\theta} - \theta_0)'. \end{aligned}$$

Hence,

$$\begin{aligned} & \frac{\sqrt{T}}{M_T} \left\| \hat{\Lambda}(\lambda)(\hat{\theta}) - \hat{\Lambda}(\lambda) \right\| \tag{12} \\ & \leq \frac{\sqrt{T}}{M_T} \left\| \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T} X_{t,T}' (\hat{\theta} - \theta_0)' \right\| \\ & + \frac{\sqrt{T}}{M_T} \left\| \sum_{h=1}^{T-1} k \left( \frac{h}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} (\hat{\theta} - \theta_0) X_{t+h,T} \eta_{t,T}' \right\| \\ & + \frac{\sqrt{T}}{M_T} \left\| \sum_{h=1}^{T-1} k \left( \frac{|h|}{M_T} \right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} (\hat{\theta} - \theta_0) X_{t+h,T} X_{t,T}' (\hat{\theta} - \theta_0)' \right\|. \end{aligned}$$

For the first term we get

$$\begin{aligned}
& \frac{\sqrt{T}}{M_T} \left\| \sum_{h=1}^{T-1} k\left(\frac{h}{M_T}\right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T} X'_{t,T} (\hat{\theta} - \theta_0)' \right\| \\
& \leq \frac{\sqrt{T}}{M_T} \sum_{h=1}^{T-1} \left| k\left(\frac{h}{M_T}\right) \right| \left\| \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T} X'_{t,T} D'_T (D'_T)^{-1} (\hat{\theta} - \theta_0)' \right\| \\
& \leq \frac{1}{M_T} \sum_{h=1}^{T-1} \left| k\left(\frac{h}{M_T}\right) \right| \left\| \sqrt{T} (\hat{\theta} - \theta_0) D_T^{-1} \right\| \left\| \frac{1}{T} \sum_{t=1}^{T-h} D_T X_{t,T} X'_{t,T} D'_T \right\|^{1/2} \left\| \frac{1}{T} \sum_{t=1}^{T-h} \eta_{t+h,T} \eta'_{t+h,T} \right\|^{1/2}.
\end{aligned}$$

For the first factor it holds that

$$\lim_{T \rightarrow \infty} \frac{1}{M_T} \sum_{h=1}^{T-1} \left| k\left(\frac{h}{M_T}\right) \right| = \int_0^\infty |k(u)| du.$$

The second and third factors are  $\mathcal{O}_P(1)$  by assumption and the last term is bounded by Corollary 3.2. Hence, the whole expression is bounded in probability. The second term in (12) is  $\mathcal{O}_P(1)$  by the same arguments. For the third term, we obtain

$$\begin{aligned}
& \frac{\sqrt{T}}{M_T} \left\| \sum_{h=1}^{T-1} k\left(\frac{h}{M_T}\right) e^{-i\lambda h} \frac{1}{T} \sum_{t=1}^{T-h} (\hat{\theta} - \theta_0) X_{t+h,T} X'_{t,T} (\hat{\theta} - \theta_0)' \right\| \\
& \leq \frac{\sqrt{T}}{M_T} \sum_{h=1}^{T-1} \left| k\left(\frac{h}{M_T}\right) \right| \left\| \frac{1}{T} \sum_{t=1}^{T-h} (\hat{\theta} - \theta_0) \delta_T^{-1} D_T X_{t+h,T} X'_{t,T} D'_T (D'_T)^{-1} (\hat{\theta} - \theta_0)' \right\| \\
& \leq \frac{1}{M_T} \sum_{h=1}^{T-1} \left| k\left(\frac{h}{M_T}\right) \right| \left\| \sqrt{T} (\hat{\theta} - \theta_0) D_T^{-1} \right\| \left\| \frac{1}{T} \sum_{t=1}^T D_T X_{t,T} X'_{t,T} D'_T \right\| \left\| (\hat{\theta} - \theta_0) D_T^{-1} \right\|.
\end{aligned}$$

The first three terms are bounded (in probability) by the same arguments as before and the last term converges to zero in probability. Hence, the whole expression is  $o_P(1)$ . Consequently, (12) is  $\mathcal{O}_P(1)$ . Similarly, we can show that  $\|\hat{\Sigma}(\hat{\theta}) - \hat{\Sigma}\| = o_P(1)$ . Finally, note that

$$\hat{f}(\lambda)(\hat{\theta}) = \frac{1}{2\pi} \left( \hat{\Sigma}(\hat{\theta}) + \hat{\Lambda}(\lambda)(\hat{\theta}) + \hat{\Lambda}(\lambda)(\hat{\theta})^* \right)$$

and the claim follows from Assumption 3.3 and Theorem 3.4.  $\square$



