

Asymptotic Theory for Time Series with Changing Mean and Variance

Violetta Dalla¹, Liudas Giraitis² and Peter M. Robinson^{3*}

¹*National and Kapodistrian University of Athens,*

²*Queen Mary, University of London,* ³*London School of Economics*

January 28, 2019

Abstract

The paper develops point estimation and asymptotic theory with respect to a semiparametric model for time series with moving mean and unconditional heteroscedasticity. These two features are modelled nonparametrically, whereas autocorrelations are described by a short memory stationary parametric time series model. We first study the usual least squares estimate of the coefficient of the first-order autoregressive model based on constant but unknown mean and variance. Allowing for both the latter to vary over time in a general way we establish its probability limit and a central limit theorem for a suitably normed and centred statistic, giving explicit bias and variance formulae. As expected mean variation is the main source of inconsistency and heteroscedasticity the main source of inefficiency, though we discuss circumstances in which the estimate is consistent for, and asymptotically normal about, the autoregressive coefficient, albeit inefficient. We then consider standard implicitly-defined Whittle estimates of a more general class of short memory parametric time series model, under otherwise more restrictive conditions. When the mean is correctly assumed to be constant, estimates that ignore the heteroscedasticity are again found to be asymptotically normal but inefficient. Allowing a slowly time-varying mean we resort to trimming out of low frequencies to achieve the same outcome. Returning to finite order autoregression, nonparametric estimates of the varying mean and variance are given asymptotic justification, and forecasting formulae developed. Finite sample properties are studied by a small Monte Carlo simulation, and an empirical example is also included.

Keywords: Semiparametric time series model, nonparametric heteroscedasticity, nonparametric moving mean, parametric autocorrelation.

JEL classifications: C14, C22.

.....
* Corresponding author: email p.m.robinson@lse.ac.uk

1 Introduction

A vast econometric and statistical literature has developed on the basis of evidence that many time series are nonstationary. There exists an unlimited number of possible forms of nonstationarity, but particular kinds have been emphasized. An early and persisting approach represents a changing mean by a deterministic function such as a polynomial in time with unknown coefficients. More recently, stochastic trends have widely featured, arising in particular from an autoregressive unit root, or a fractional model with differencing parameter exceeding 1/2. These stochastic trend models imply a variance that grows without bound, just as a polynomial trend implies the mean increases without bound.

Models have also been developed that entail mean variation and heteroscedasticity that may increase in a different kind of way, or remain bounded, or decay. The present paper models an observable scalar time series y_t , $t = 1, \dots, n$, by

$$y_t = \mu_t + h_t x_t, \quad t = 1, 2, \dots \quad (1.1)$$

The ingredients of (1.1) are described as follows. The μ_t and $h_t \geq 0$ are unknown and deterministic, whereas $\{x_t\}$ is a zero mean covariance stationary process. Thus,

$$E(y_t) = \mu_t, \quad V(y_t) = h_t^2 V(x_t), \quad t = 1, 2, \dots, \quad (1.2)$$

since x_t has constant variance. Thus in general y_t is nonstationary in both the mean and variance. However it is covariance stationary with respect to autocorrelations: for $t > s$

$$\text{corr}(y_s, y_t) = \frac{h_s h_t E(x_s x_t)}{h_s h_t \sqrt{V(x_s) V(x_t)}} = \text{corr}(x_s, x_t), \quad (1.3)$$

which depends only on $t - s$.

The prescription that μ_t and h_t are deterministic rules out numerous models, where y_t has stochastic conditional mean and/or conditional heteroscedasticity, but it nevertheless still covers a huge literature. In general μ_t , h_t and x_t can each be chosen to be parametric or nonparametric, so there are $2^3 = 8$ types of specification in this respect, going from the purely parametric, through semiparametric ones to the purely nonparametric. In early days all might well have been parametric, but, generally speaking, misspecification of μ_t is liable to lead to inconsistent estimates of not only its parameters but also those of h_t and x_t while misspecification of the latter two is liable to lead not only to inconsistent estimates of their parameters but also invalid inferences on the parameters of μ_t . Nowadays, partly reflecting the availability of longer time series data sets, in particular in finance, one would more likely expect to see a purely nonparametric model, allowing an asymptotic theory which applies more generally, though results would be sensitive to the choices of 3 different bandwidths. Semiparametric models are generally motivated by data sets that are not so large, and/or there is reasonable confidence in a partial parametrization. A common goal in semiparametric estimation is the achievement of consistency (and asymptotic normality) with

parametric rate of parameter estimates. A more ambitious goal, which has been achieved in a number of circumstances in (1.1), is asymptotic efficiency of the parameter estimates, despite the nonparametric components. An early example of the latter achievement in a model that fits our specification of (1.1) is Harvey and Robinson (1988), who took μ_t to be parametric, possibly trending, took x_t to be parametric, in particular p th order autoregressive (AR(p)), and h_t nonparametric. They showed that generalized least squares estimates (that account for both error autocorrelation and error heteroscedasticity) of the parameters describing μ_t are adaptive in the sense that they are not only asymptotically normal but achieve the Gauss-Markov asymptotic efficiency bound.

For this purpose Harvey and Robinson (1988) showed that the estimates of the AR coefficients describing x_t are consistent but they were treated there as nuisance parameters. In the present paper, for an eventually more general class of parametric models for x_t , we focus on properties of their estimates when h_t is nonparametric as in Harvey and Robinson (1988) (and sometimes treated in a similar way) but also, for the most part, μ_t varies nonparametrically.

Initially we allow for general variation in μ_t and h_t , but subsequently assume that, predominantly, they change only slowly with respect to t . In particular, we often assume that for a function $m(s)$, $s \in (0, 1)$, that satisfies some smoothness conditions,

$$\mu_t = m(t/n), \quad t = 1, \dots, n, \quad (1.4)$$

with conditions on h_t that also restrict their variation. Under (1.4), the homoscedastic special case of (1.1), with $h_t \equiv h$,

$$y_t = \mu_t + hx_t, \quad t = 1, 2, \dots, \quad (1.5)$$

has been extensively studied in the fixed-design nonparametric regression literature (beginning with, e.g., Priestley and Chao (1972), Gasser and Mueller (1979)). Initially and commonly x_t was assumed to be white noise, but subsequently x_t was allowed to be dependent, Robinson (1997) covering short and long memory and antipersistence, with Guo and Koul (2007) giving a partial extension to nonparametric heteroscedasticity. The major focus has been the asymptotic, as $n \rightarrow \infty$, statistical properties of estimates (usually kernel ones) of $m(s)$ at fixed points s or at $s = t/n$ under various conditions; under short memory, the rate of convergence is the same as under white noise but (unlike in stochastic-design nonparametric regression) asymptotic variance is affected, whereas long memory slows convergence. Note that under (1.1) and (1.4) the $y_t = y_{tn}$ form a triangular array but our notation suppresses reference to this fact. The introduction of nonparametric heteroscedasticity, as in (1.1), has also been considered in a number of references in addition to Harvey and Robinson (1988), with analogous nonparametric estimation. An alternative, possibly more appealing, heteroscedastic scenario to (1.5) models the innovations of a time series model to be heteroscedastic. Further references to this and versions of (1.1) (see also below) include Beran and Feng (2002), Phillips and Xu (2005), Xu and Phillips (2008), Zhao, Zhang and Li (2014), Zhu (2016) and Cavaliere, Nielsen and Taylor (2017).

In this paper we assume x_t is a linear process with parametric spectral density

$$f_x(u) = (2\pi)^{-1}k(u; \theta), \quad u \in \Pi, \quad (1.6)$$

for $\Pi = (-\pi, \pi]$, k a known function of frequency u , and the $p \times 1$ unknown parameter vector θ . (Since any scale parameter can be absorbed in h in (1.5), or in h_t in (1.1), we do not include an unknown scalar factor in f_x , with the effect that the best one-step-ahead linear predictor of x_t has variance 1.) Thus the model (1.1) is semiparametric, differing from the large literature on models with nonparametrically time-varying dependence parameters. We cover here only short memory processes x_t , with the effect that f_x is bounded and bounded away from zero. Our principle concern is establishing desirable asymptotic properties of estimates of θ . Thus the μ_t and h_t are mainly regarded as nuisance functions.

In the constant mean, homoscedastic special case of (1.1) and (1.5)

$$y_t = \mu + hx_t, \quad t = 1, 2, \dots, \quad (1.7)$$

y_t is covariance stationary, with spectral density $h^2 f_x(u; \theta)$. In this setting a variety of estimates of θ have been developed and their asymptotic properties established, notably ones which optimize an exact or approximate Gaussian likelihood, with consistency and asymptotic normality depending on milder assumptions than Gaussianity; these are sometimes collectively called Gaussian or Whittle estimates (though some authors reserve the latter term for estimates defined in the frequency domain), the main point being that they all have the same first-order asymptotics under stationarity. For AR(p) models these estimates include least squares, Yule-Walker and frequency-domain approximations, all of which are expressed in closed form, whereas for models such as autoregressive moving averages (ARMA) they are only implicitly defined. Here we investigate such estimates when in fact (1.1) holds. Predictably bias and variance are affected, but equally predictably given (1.3) and the fact that under our specification correlation properties of x_t are described only by θ , a suitable handling of a varying mean can lead to consistent and asymptotically normal estimation of θ , albeit with asymptotic variance (and possibly rate) inflated by heteroscedasticity. Thus we establish a degree of robustness with respect to a varying mean and variance, though it is important to stress that our failure to account for heteroscedasticity in point estimation leads to inefficiency, and thus confidence intervals that are wider, and hypothesis tests that are less powerful, than ones based on efficient estimates.

The following section examines the implications of time varying mean and heteroscedasticity and their interactions when x_t is AR(1), exploiting the simple, closed form nature of parameter estimates here to obtain not only convergence in probability and asymptotic normality but explicit formulae for asymptotic bias and variance that are more or less easy to interpret. Under more restrictive conditions on μ_t and h_t , Section 3 studies estimates of θ for implicitly defined extremum estimates of a quite general class of short memory $k(\cdot; \cdot)$, albeit under stronger conditions on the innovations in x_t , first assuming the overall model is another special case of (1.1),

$$y_t = \mu + h_t x_t, \quad t = 1, 2, \dots, \quad (1.8)$$

and then, to cover (1.1), but assuming variation in μ_t causes distortion which is confined to low frequencies we resort to trimming out such frequencies from the discrete-frequency Whittle objective function. Section 4 specializes to the case x_t is AR(p), and develops forecasting formulae. In order

to implement these we need to estimate the μ_t and h_t^2 ; we employ kernel estimates and establish their consistency and asymptotic normality. Section 5 consists of a small Monte Carlo study of finite-sample performance. In Section 6 we apply the methods to unemployment data. In both we focus on AR models. Section 7 briefly discusses some possible extensions of our work. Proofs of theorems, along with statements of technical lemmas and their proofs, are left to Section 8.

Our paper follows a number of others that combine nonparametric fixed-design regression with parametric autocorrelation in (1.5), for example Truong (1991), Altman (1993), Hall and Van Keilegom (2003), Shao and Yang (2011). They employed methods of parameter estimation that produce the same asymptotic properties to ours, parametric rate and in most cases asymptotic normality, but deal with nonparametric regression in different ways from our low frequency trimming approach. However, trimming out of low frequencies to deal with a nonparametric μ_t has been used in settings similar to ours: McCloskey (2010) considered a variety of parametric x_t , McCloskey (2013) and McCloskey and Hill (2017) establishing parametric rate and asymptotic normality of parameter estimates in case x_t follows a long memory stochastic volatility or GARCH model, respectively, for (1.5) with a variety of deterministic and stochastic μ_t , these authors, like we do below, employing discrete-frequency Whittle estimation. Similar to our work also, with respect to (1.8), Dahl and Levine (2006), Figueroa-Lopez and Levine (2013) considered a parametric linear process contaminated by deterministic nonparametric time varying heteroscedasticity like ours.

2 Bias and inefficiency

The present section indicates the consequences of ignoring mean variation and heteroscedasticity in the simplest interesting setting, namely in the case x_t is a stationary AR(1) process, so

$$x_t = \theta x_{t-1} + \varepsilon_t, \quad |\theta| < 1, \quad (2.1)$$

where ε_t is an uncorrelated white noise sequence with unit variance. Then (cf (1.3)) y_t given by (1.1) still has the autocorrelation properties of an AR(1) with AR coefficient θ , but we estimate the unknown θ as if y_t were generated by (1.7) (which would imply y_t is AR(1) with unknown mean μ and unknown innovations variance h^2). An advantage of AR processes is the closed form nature of various leading parameter estimates, facilitating an examination of the effects of mean variation and heteroscedasticity under (1.1). It would be straightforward to extend our results to an AR(p) process, for $p > 1$, though they become more complicated. Indeed there are closed form estimates (albeit ones that are inefficient even under the model (1.7)) of pure moving average models (using method of moments) and of ARMA models (using instrumental variables) so in principle an explicit study of bias would be possible in these cases also.

We employ the least squares estimate of θ ,

$$\tilde{\theta} = \frac{\sum_{t=2}^n (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^n (y_{t-1} - \bar{y})^2}, \quad (2.2)$$

where $\bar{y} = n^{-1} \sum_{t=1}^n y_t$. The same first order asymptotic properties established below would result from using instead, say, the Yule-Walker estimate or the frequency-domain Whittle estimate.

Introduce the abbreviating formulae

$$\begin{aligned} M &= \sum_{t=1}^n (\mu_t - \bar{\mu})^2, & N &= \sum_{t=2}^n (\mu_t - \bar{\mu})(\mu_{t-1} - \bar{\mu}), \\ S &= \sum_{t=1}^n h_t^2, & T &= \sum_{t=2}^n (h_t - h_{t-1})h_{t-1}, \end{aligned}$$

for $\bar{\mu} = n^{-1} \sum_{t=1}^n \mu_t$, and the 'bias' quantity

$$b_n(\theta) = \frac{(1 - \theta^2)(N - \theta M) + \theta T}{(1 - \theta^2)M + S}.$$

Theorem 2.1 *Let (2.1) hold where*

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \quad E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1, \quad \text{almost surely (a.s.)} \quad (2.3)$$

with \mathcal{F}_t the σ -field of events generated by $\{\varepsilon_s, s \leq t\}$, and the ε_t^2 are uniformly integrable, and let

$$\frac{1 + \max h_t^2}{S} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (2.4)$$

Then as $n \rightarrow \infty$,

$$\tilde{\theta} - \theta - b_n(\theta) \rightarrow_p 0. \quad (2.5)$$

Now introduce

$$U = \sum_{t=2}^n (h_t - h_{t-1})h_{t-1}x_{t-1}^2, \quad V = \sum_{t=2}^n h_t^2 h_{t-1}^2, \quad W = \sum_{t=1}^{n-1} (\mu_t - \bar{\mu})^2 h_t^2,$$

and the modified 'bias' quantity

$$c_n(\theta) = \frac{(1 - \theta^2)(N - \theta M + \theta U)}{(1 - \theta^2)M + S}.$$

Theorem 2.2 *If the conditions of Theorem 2.1 hold, and also*

$$\frac{1 + \max h_{t-1}^2 h_t^2 + S^2/n^2 + W}{V} \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (2.6)$$

and

$$M = O(n) \text{ as } n \rightarrow \infty, \quad (2.7)$$

then as $n \rightarrow \infty$,

$$((1 - \theta^2) M + S) V^{-1/2} (\tilde{\theta} - \theta - c_n(\theta)) \rightarrow_D \mathcal{N}(0, 1 - \theta^2). \quad (2.8)$$

Remarks

1. The martingale difference assumption in (2.3) does not allow conditional heteroscedasticity, is weaker than assuming the ε_t are independent, and has long been standard in the literature on estimates based on quadratic functions such as Whittle estimates following Hannan (1973), while as in the latter reference (and previously in Anderson and Walker (1964)) because θ is a dependence parameter we require only the necessary condition of finite second moments of ε_t , thereby appealing to another concern of financial econometricians, with our uniform integrability condition being weaker than the identity of distribution, and thus strict stationarity, assumed in these references.

2. Condition (2.4) implies that for $h_t \sim t^\phi$ we must have $\phi \geq -1/2$, so $S \sim n^{2\phi+1}$. There is no upper bound on ϕ in this case, so we certainly do not require h_t to be bounded, but (2.4) rules out exponential increase.

3. As expected, in general $b_n(\theta)$ is non-negligible, so $\tilde{\theta}$ is inconsistent. When x_t is not white noise this can still be the case when there is heteroscedasticity but no mean-variation, so $M = N = 0$, since we do not require that $T = o(S)$. For example in (albeit extremely unlikely) rapidly oscillating cases such as $h_t = 1(t \text{ odd}) - 1(t \text{ even})$, where $1(\cdot)$ is the indicator function, we have $T \sim -2n$, $S \sim n$. But T makes no contribution under the mild smoothness condition $\sum_{t=2}^n (h_t - h_{t-1})^2 = o(S)$. Indeed assuming

$$h_t = g(t/n), \quad (2.9)$$

with g boundedly differentiable (cf Section 4) the left side of the previous equality is $O(n^{-1})$ while if g is substantially positive S increases at rate n .

4. Of course variation in μ_t is the predominate source of bias such that even if it is generated by (1.4) with m boundedly differentiable, in general M and N are of exact order n , while $S = O(n)$ for bounded h_t . This is also the case under the alternative scenario of 'asymptotic stationarity' (Parzen (1963)), to the extent that, for some γ_0 and $\gamma_1 \in (-\gamma_0, \gamma_0)$, $n^{-1}M \rightarrow \gamma_0$, $n^{-1}N \rightarrow \gamma_1$ and bias only disappears in the measure-zero event $\theta = \gamma_1/\gamma_0$. On the other hand, as expected, the contributions of μ_t to numerator and denominator of $b_n(\theta)$ diminish as x_t approaches nonstationarity. Since $N = O(M)$ and $T = O(S)$ at least $b_n(\theta)$ is always bounded, though by no means necessarily to the extent that $\theta + b_n(\theta)$ has a limit within the stationarity region.

5. Unboundness or decay in μ_t and/or h_t cannot be ruled out, and can increase or decrease $b_n(\theta)$, depending on whether the signal-to-noise ratio diverges or decays to zero, such that, for example, if $M = o(S)$ (and $T = o(S)$ as discussed above), $b_n(\theta) \rightarrow 0$ for all θ . There is a similar outcome under sparseness (e.g. when only $o(n)$ of the μ_t deviate from a constant level).

6. As always conditions for a desirably centred central limit theorem are more stringent than those for consistency. An unfortunate requirement in Theorem 2.2 is $W = o(V)$, which can hold if noise dominates signal but will not hold in many of the situations in which M and N contribute nonnegligibly to bias in our theorems. Our proofs indicate the presence of a term that is linear in x_t with zero mean and asymptotically normal after suitable norming, and, if $W = o(V)$ is not imposed, can influence the form of the limit distribution of $\tilde{\theta}$. We have chosen to keep matters relatively simple so as to obtain centring and norming sequences that lead to the same limit distribution as under a constant mean and homoscedasticity. Likewise the condition (2.7), ruling out unbounded μ_t , is imposed to avoid other complications.

7. The components $\max h_t^2/S \rightarrow 0$ in (2.4) and $V \rightarrow \infty$ in (2.6) imply the component $S \rightarrow \infty$ in (2.4), whereas under the non-oscillating condition $c \leq h_t/h_{t-1} \leq C$, where c and C throughout respectively denote arbitrarily small and arbitrarily large generic constants, $\max h_{t-1}^2 h_t^2/V \rightarrow 0$ in (2.6) implies $\max h_t^2/S \rightarrow 0$.

8. The stochastic term U in $c_n(\theta)$ can be replaced by $T/(1-\theta^2)$, whence $c_n(\theta)$ can be replaced in Theorem 2.2 by $b_n(\theta)$, under stronger conditions, including $E\varepsilon_t^4 \leq C$, but in any case U seems likely to be dominated by other terms, in the same way as T , as discussed above.

9. The presence of M in the norming factor in Theorem 2.2 could, despite (2.7), be associated with a norming factor increasing faster than $n^{1/2}$, albeit with a non-negligible $c_n(\theta)$. But when $M = o(S)$, as under the constant mean model (1.8) for example, and if $\theta = 0$ or $U = o_p(S)$, then we deduce that

$$SV^{-1/2}(\tilde{\theta} - \theta) \rightarrow_D \mathcal{N}(0, 1 - \theta^2). \quad (2.10)$$

The main purpose of establishing a central limit theorem is statistical inference and this requires an estimate \tilde{R} of $R = SV^{-1/2}$ such that $\tilde{R}/R \rightarrow_p 1$. Since the h_t^2 are unknown they must be replaced by suitable proxies \tilde{h}_t^2 , whence we can form $\tilde{R} = \tilde{S}\tilde{V}^{-1/2}$ with $\tilde{S} = \sum_{t=1}^n \tilde{h}_t^2$, $\tilde{V} = \sum_{t=2}^n \tilde{h}_t^2 \tilde{h}_{t-1}^2$, though it remains to be shown that $\tilde{R}/R \rightarrow_p 1$. Under (2.9) with smooth g , fixed-design kernel regression of the $(y_t - \bar{y})^2$ (assuming no asymptotic bias) can, from (1.2), consistently estimate $h_t^2 \text{var}(x_1)$, where the second factor can be estimated by $(1 - \tilde{\theta}^2)^{-1}$ or simply ignored for the purpose of approximating R .

10. The Cauchy inequality confirms that the norming factor $SV^{-1/2}$ in (2.10) is no greater than the optimal parametric rate $n^{1/2}$, and of course it equals $n^{1/2}$ under homoscedasticity. The rate could be slower, e.g. it is $(n/\log n)^{1/2}$ if $h_t \sim t^{-1/4}$. But of far greater practical importance is the fact that even if $SV^{-1/2} \sim an^{1/2}$, the Cauchy inequality generally implies $a < 1$, so that $\tilde{\theta}$ is asymptotically inefficient under heteroscedasticity.

11. Given the large literature on adaptive estimation of semiparametric models we might attempt to obtain parameter estimates (not only in (2.1) but in the more general time series model of the rest of the paper) that are as asymptotically as efficient as if the h_t (and μ_t) are known or parametrically modelled. The fixed design nonparametric regression literature, and (1.2), indicates that under (1.4) with smooth m and (2.9) with smooth g , kernel estimates of the μ_t and $h_t^2 \text{var}(x_1)$ are consistent in the presence of quite general, nonparametric, dependence in x_t , so they can be robustly estimated at the outset by $\tilde{\mu}_t$ and $h_t^2 \text{var}(x_1)$ without specifying a parametric time series

model for x_t . Then one can form $\tilde{x}_t = (y_t - \tilde{\mu}_t) / (h_t^2 \widetilde{\text{var}}(x_1))^{1/2}$ which approximate the x_t up to scale. There is no doubt that one could then consistently estimate a nonparametric spectral density of x_t , with nonparametric rate. It remains to be shown whether estimating θ in (2.1) by least squares using the \tilde{x}_t , or proceeding analogously with respect to the more general time series model of the following section, produces root- n -consistent and asymptotically efficient estimates, but there are plenty of semiparametric models in which this outcome has been shown to be possible, for example in the related models of Harvey and Robinson (1988) and Xu and Phillips (2008). Instead, however, in the rest of the paper we pursue a more modest goal, correcting for mean variation by trimming out low frequencies in ordinary Whittle estimation based on the y_t and obtaining rate-optimal, but inefficient, parameter estimates.

3 Parameter estimation and asymptotic theory for a more general time series model

In this section we consider a more general specification of (1.6), where, since we will now consider implicitly-defined extremum estimates we denote the temporal dependence parameter vector θ_0 , $\theta \in \mathbf{R}^p$ now denoting any admissible value. We consider first the basic discrete-frequency Whittle estimate studied by Hannan (1973). As implied above we know the functional form of $k(u; \theta)$ in (1.6). We consider only $k(u; \theta)$ satisfying

$$\int_{\Pi} \log k(u; \theta) du = 0, \quad (3.1)$$

for all θ , implying that the best one-step-ahead linear predictor of x_t has variance 1.

Approximating the Gaussian pseudo likelihood of y_t under the stationary model (1.7), after eliminating h we consider the discrete-frequency Whittle estimate of θ_0 ,

$$\hat{\theta} = \arg \min_{\theta \in \Theta} Q_n(\theta),$$

where

$$Q_n(\theta) = \frac{2\pi}{[n/2]} \sum_{j=1}^{[n/2]} \frac{I_y(u_j)}{k(u_j; \theta)}, \quad (3.2)$$

in which we introduce the periodogram

$$I_y(u) = (2\pi n)^{-1} \left| \sum_{t=1}^n y_t e^{itu} \right|^2,$$

the Fourier frequencies $u_j = 2\pi j/n$, the compact set $\Theta \subset \mathbf{R}^p$, and $[\cdot]$ denotes integer part.

Note that (3.2) differs from the corresponding objective function of Hannan (1973). His summation was over $[-n/2] < j \leq [n/2]$, whereas for comparability after using evenness of summands, our summation is over $-[n/2] \leq j \leq -1, 1 \leq j \leq [n/2]$. Our omission of the $j = 0$ term is for the

familiar purpose of sample mean correction, Hannan (1973) having assumed a known mean. When n is even we have an extra summand such that there are two (identical) terms corresponding to $j = -n/2$ and $n/2$. The latter distinction has no effect on limit distribution theory.

Denote by ∇_a the partial derivative operator with respect to a vector a and by $'$ the transposition operator, and let C and c be respectively arbitrarily large and small generic positive constants. We introduce the following assumptions on the modelling of x_t .

Assumption A (i) For an independent and identically distributed (i.i.d.) sequence $\{\varepsilon_t\}$ such that $E(\varepsilon_1) = 0$, $V(\varepsilon_1) = 1$, $E(\varepsilon_1^4) < \infty$,

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j|b_j| < \infty, \quad b_0 = 1. \quad (3.3)$$

(ii) x_t has spectral density given by $f_x(u)$ in (1.6) such that $k(u; \theta)$ is a known function of u , θ ; (3.1) holds for all $\theta \in \Theta$;

$$c \leq k(u; \theta) \leq C, \quad \theta \in \Theta, \quad u \in \Pi; \quad (3.4)$$

the partial derivatives $\nabla_{\theta} k(u; \theta)$, $\nabla_u k(u; \theta)$, $\nabla_u \nabla_{\theta} k(u; \theta)$, $\nabla_{\theta} \nabla'_{\theta} k(u; \theta)$, $\nabla_u \nabla_{\theta} \nabla'_{\theta} k(u; \theta)$ exist and are bounded and continuous functions of $u \in \Pi$ and $\theta \in \Theta$; the matrix

$$\Omega = (4\pi)^{-1} \int_{\Pi} \nabla_{\theta} \log k(u; \theta_0) \nabla'_{\theta} \log k(u; \theta_0) du$$

is positive definite; and for all $u \in \Pi$ and $\theta \in \Theta$ but $\theta \neq \theta_0$, $k(u; \theta) \neq k(u; \theta_0)$.

(iii) θ_0 is an interior point of the compact set $\Theta \subset R^p$.

These conditions are largely taken from Hannan (1973) but are in some respects stronger, indeed stronger than corresponding conditions in the previous section. He did not require ε_t to have finite 4th moment, only finite variance, and relaxed our i.i.d. assumption such that ε_t and $\varepsilon_t^2 - 1$ are martingale difference sequences, as in (2.3). The inequality in (3.3) is more restrictive than the corresponding one, $\sum_{j=0}^{\infty} j b_j^2 < \infty$, of Hannan (1973). Note that the condition $b_0 = 1$ follows from (3.1); it would be possible to consider models (such as arise in a signal-plus-noise setting, for example) that do not satisfy (3.1) if a 'log' term (which is asymptotically negligible under (3.1)) were explicitly included in the Whittle objective function. However our conditions cover standard parameterizations of, for example, the stationary and invertible ARMA (though given a number of papers, one expects that asymptotics for this could also be covered as a special case of ones for fractional ARIMAs whether the true, zero, value of the differencing parameter is known or unknown).

Recall notation $S = \sum_{t=1}^n h_t^2$. We also introduce the following conditions on h_t .

Assumption B

Denoting $V^* := \sum_{t=1}^n h_t^4$ and $\Delta_{h,n} := \sum_{t=1}^{n-1} |h_t - h_{t+1}| + |h_n|$, as $n \rightarrow \infty$

$$(a) S \rightarrow \infty, \quad (b) \log n \Delta_{h,n}^2 = o(S^{1/2}), \quad (c) \max |h_t| \leq C, \quad (d) S = O(V^*). \quad (3.5)$$

Note that if h_t is bounded away from zero, as might often be expected in practice, S increases at the rate n which obtains under the homoscedastic model (1.7). This is also the case for $h_t \sim (t/n)^\phi$, $\phi > 0$, while another common power law model for heteroscedasticity, $h_t \sim t^\phi$, $\phi \neq 0$, (cf remark 2 in Section 2) does not satisfy Assumption B. If (2.9) holds with g a boundedly differentiable function on $(0, 1)$, as assumed in the following section, Δ_n is bounded but (b) allows, for example, for a step function h_t with non-negligible but finite increments $h_t - h_{t+1}$ at $o(S^{1/4}/(\log n)^{1/2})$ points t . On the other hand for $h_t = w(t/nd)$, with w bounded, non-negative, square integrable and Lipschitz continuous and $d = d_n$ satisfying $d + (n^{-1/2} \log n)/d^{5/2} \rightarrow 0$, Assumption B holds with $S \sim nd$ though it is implied that S increases faster than $n^{4/5}$. However $S \sim n$ seems by far the most plausible outcome and is implied in the following section and in the h_t employed in our Monte Carlo simulations in Section 5.

Theorem 3.1 *Let (1.8) and Assumptions A and B hold. Then, as $n \rightarrow \infty$,*

$$Q_n(\hat{\theta}) = \frac{S}{n}(1 + O_P(S^{-1/2})), \quad \frac{S}{V^{*1/2}}(\hat{\theta} - \theta_0) \rightarrow_D \mathcal{N}(0, \Omega^{-1}). \quad (3.6)$$

For a stationary time series $y_t = \mu + hx_t$, $Q_n(\hat{\theta})$ estimates the variance h^2 of the forecast error of the best one-step-ahead linear predictor, while for $y_t = \mu + h_t x_t$ under heteroscedasticity $Q_n(\hat{\theta}) \sim n^{-1}S$ estimates the average variance of the one-step-ahead forecast error over time $t = 1, \dots, n$. As in the previous section the norming factor indicates efficiency loss due to heteroscedasticity: under Assumption B,

$$|V - V^*| \leq \left| \sum_{t=1}^{n-1} h_t (h_t - h_{t+1}) \right| + h_n^2 \leq C(\Delta_{h,n} + 1) = o(S),$$

so the norming sequences in (2.10) and (3.6) are equivalent. Without a consistent estimate of the norming factor in (3.6) the theorem is useless in statistical inference, as discussed in Section 2, where some suggestions are made. And as there, the estimates of the current section are inefficient in the presence of heteroscedasticity.

We now turn to the general model (1.1), with its changing mean μ_t . The periodogram was originally motivated in stationary settings, and it is known that smooth changes in the mean, while not affecting its properties much at high frequencies, may significantly distort them at low frequencies, where its value may be boosted. Consequently Whittle estimates may lose the properties established in Theorem 3.1. To overcome this problem we trim out low frequencies u_j from the objective function $Q_n(\theta)$, a device introduced in the long memory literature by Künsch (1986) prompted by the unboundness of the spectral density at zero frequency, and used by a number of subsequent authors, with Hurvich, Lang and Soulier (2005), Iacone (2010) and McCloskey and Perron (2013) using it to cope with nonparametric μ_t in (1.4) for nonparametric long memory x_t , along with Mc-

Closkey (2013) and McCloskey and Hill (2017) for parametric x_t . Following in particular McCloskey (2013) and McCloskey and Hill (2017), for a small integer $\nu \geq 0$ introduce

$$Q_{n,\nu}^{(1)}(\theta) = \frac{2\pi}{[n/2]} \sum_{j=\nu+1}^{[n/2]} \frac{I_y(u_j)}{k(u_j; \theta)}, \quad (3.7)$$

the trimmed objective function. Introduce the 'mild trimming'

Assumption C

$$0 \leq \nu = \nu_n = o(n/S^{1/2}). \quad (3.8)$$

In Theorem 3.2 below we will show that under Assumption C the estimate

$$\hat{\theta}^{(1)} = \operatorname{argmin}_{\theta \in \Theta} Q_{n,\nu}^{(1)}(\theta)$$

preserves the properties of $\hat{\theta}$ established in Theorem 3.1. There is always ambiguity in the choice of trimming parameter, there is no optimal choice, and none of our conditions provide a clue for choosing ν in a given data set. We stress also that trimming out of low frequencies can only handle specialised forms of μ_t (see Assumption D below) and not many of the more general ones envisaged in Section 2.

Moreover trimming entails loss of information and may worsen finite sample performance. To try to mitigate this we introduce also an estimate which incorporates a correction for the trimming. Notice that under constant mean $\mu_t \equiv \mu$,

$$E \left[\sum_{j=1}^{\nu} \frac{I_y(u_j)}{k(u_j; \theta)} \right] \sim (S/n) \sum_{j=1}^{\nu} \frac{k(u_j; \theta_0)}{2\pi k(u_j; \theta)}, \quad (3.9)$$

see Lemma 8.4(i-ii) in Section 8. We use this structure to build a correction term, since periodograms at higher frequencies allow us to estimate (albeit less precisely) (S/n) by $Q_{n,\nu}^{(1)}(\theta)$ and θ_0 by $\hat{\theta}^{(1)}$, see (3.9). Thus we introduce

$$Q_{n,\nu}^{(2)}(\theta) = \frac{2\pi}{[n/2]} \left[\sum_{j=\nu+1}^{[n/2]} \frac{I_y(u_j)}{k(u_j; \theta)} + Q_{n,\nu}(\hat{\theta}^{(1)}) \sum_{j=1}^{\nu} \frac{1}{2\pi} \frac{k(u_j; \hat{\theta}^{(1)})}{k(u_j; \theta)} \right], \quad (3.10)$$

and set

$$\hat{\theta}^{(2)} = \operatorname{argmin}_{\theta \in \Theta} Q_{n,\nu}^{(2)}(\theta).$$

Intuitively one wants to choose ν as small as possible, but this is subject to the smoothness of μ_t under (1.4) in (3.11). We describe the degree of smoothness by

Assumption D μ_t is given by (1.4) where $m(u)$ can be written as a Fourier series

$$m(u) = \sum_{j=-\infty}^{\infty} e^{-i2\pi ju} \rho_j, \quad u \in (0, 1); \quad \rho_j = \rho_{-j}, \text{ all } j \neq 0; \quad \sum_{j=-\infty}^{\infty} |\rho_j| < \infty. \quad (3.11)$$

Note that the inequality in (3.11) is implied if m satisfies a Lipschitz condition of degree greater than $1/2$.

We now assume lower bounds that trade off relative to the actual rate of decay of the ρ_j and ν .

Assumption E One of the following conditions holds:

$$|\rho_k| \leq C\phi^{|k|}, \quad k = \pm 1, \dots, (\exists 0 < \phi < 1), \text{ and } \log(n/S^{1/2})/\nu \rightarrow 0, \quad (3.12)$$

$$|\rho_k| \leq C|k|^{-\gamma}, \quad k = \pm 1, \dots, (\exists \gamma > 1), \text{ and } (n/S^{1/2})^{1/(2\gamma-1)}/\nu \rightarrow 0. \quad (3.13)$$

Recall that for $h_t = g(t/n)$ where g is a continuous bounded function one has $S = \sum_{t=1}^n h_t^2 \sim n \int_0^1 h^2(u) du$, and generally for our leading case $S \sim n$ the limits in (3.12) and (3.13) can be replaced respectively by $(\log n)/\nu \rightarrow 0$ and $n^{1/(2\gamma-1)}/\nu \rightarrow 0$ respectively. Note also that in this case the upper bound rate for ν in Assumption C is $o(n^{1/2})$. McCloskey (2013) theoretically motivated an exact trimming rate under certain conditions. However even this would be insufficient to justify a particular numerical choice of ν for given n .

Theorem 3.2 (i) Let (1.8) and Assumptions A-C hold. Then as $n \rightarrow \infty$, for $k = 1, 2$,

$$Q_{n,\nu}(\hat{\theta}^{(k)}) = \frac{S}{n}(1 + O_P(S^{-1/2})), \quad \frac{S}{V^{*1/2}}(\hat{\theta}^{(k)} - \theta_0) \rightarrow_D \mathcal{N}(0, \Omega^{-1}). \quad (3.14)$$

(ii) Let (1.1) and Assumptions A-E hold. Then (3.14) remains valid.

4 Estimation and forecasting in the AR(p) case

In this section we consider the time-varying, heteroscedastic model (1.1), specializing to the AR(p) special case of $k(u; \theta)$ such that x_t in (1.1) is generated by

$$x_t = \theta_{10}x_{t-1} \dots + \theta_{p0}x_{t-p} + \varepsilon_t, \quad (4.1)$$

under the stationarity condition that all zeroes of $1 - \theta_{10}z \dots - \theta_{p0}z^p$ lie outside the unit disk, and with $\theta_0 = (\theta_{10}, \dots, \theta_{p0})'$. Note that for AR(p) x_t the objective functions of the previous section are quadratic in θ , so estimates can be defined in closed form, like those in Section 2, and asymptotic properties can be established much more easily than in our proofs for the implicitly-defined estimates covered in Section 3.

Forecasting of y_t requires estimation of the μ_t , and deriving useful properties of μ_t and h_t , requires strengthening Assumptions B, D and E.

Assumption F μ_t is given by (1.4) and h_t is given by (2.9), where $m(\cdot)$ and the positive valued function $g(\cdot)$ have bounded derivatives.

Assumption F is standard in the literature for versions of (1.1) when μ_t and/or h_t are non-parametric, especially for asymptotic normality properties like those we establish in Theorem 4.1 below. As implied previously the assumption implies that

$$|\mu_t - \mu_{t-1}| \leq Cn^{-1}, \quad |h_t - h_{t-1}| \leq Cn^{-1}, \quad t = 1, \dots, n, \quad n \geq 2. \quad (4.2)$$

We denote in this section by $\hat{\theta}$ the estimate $\hat{\theta}^{(2)}$. By Theorem 3.2,

$$\hat{\theta} - \theta_0 = O_p(n^{-1/2}). \quad (4.3)$$

Of course (4.3) indicates that replacing θ_0 by $\hat{\theta}$ will have negligible effect on estimates of μ_t and h_t , which will have slower, nonparametric, rate. The estimation of μ_t and h_t is relatively straightforward. Denote

$$z_t := y_t - \sum_{j=1}^p \hat{\theta}_j y_{t-j} = \mu_t \left(1 - \sum_{j=1}^p \hat{\theta}_j\right) + h_t \varepsilon_t + \hat{\eta}_t, \quad t = p+1, \dots, n \quad (4.4)$$

and $z_t = 0$ for $t = 1, \dots, p$, where $\hat{\eta}_t = \sum_{j=1}^p \{\hat{\theta}_j(\mu_t - \mu_{t-j}) + (h_t \theta_j - h_{t-j} \hat{\theta}_j) x_{t-j}\}$ is asymptotically negligible. Hence, we arrive at the 'regression model'

$$z_t = \mu_t^* + h_t \varepsilon_t + \hat{\eta}_t, \quad \mu_t^* := \mu_t \left(1 - \sum_{j=1}^p \hat{\theta}_j\right). \quad (4.5)$$

Therefore μ_t^* and h_t^2 can be standardly estimated using kernels,

$$\hat{\mu}_t^* := K_t^{-1} \sum_{j=1}^n k_{tj} z_j, \quad \hat{h}_t^2 := K_t^{-1} \sum_{j=1}^n k_{tj} (z_j - \hat{\mu}_j^*)^2, \quad t = 1, \dots, n, \quad (4.6)$$

with $K_t := \sum_{j=1}^n k_{tj}$ based on the weights $k_{tj} := K(|t-j|/H)$ for a user-chosen kernel function $K(x)$ and bandwidth H . In turn, we estimate μ_t by

$$\hat{\mu}_t = \frac{\hat{\mu}_t^*}{1 - \sum_{j=1}^p \hat{\theta}_j}, \quad t = 1, 2, \dots, n. \quad (4.7)$$

We impose the following standard assumptions on K and H .

Assumption G. $K(x)$ is nonnegative and differentiable, and

$$K(x), \quad |(d/dx)K(x)| \leq C(1+|x|)^{-\beta}, \quad \text{some } \beta > 2. \quad (4.8)$$

Assumption H. As $n \rightarrow \infty$, $H/n + 1/H \rightarrow 0$.

Theorem 4.1 Let (1.1) and Assumptions A-H hold. Then for any $1 \leq t = t_n \leq n$, as $n \rightarrow \infty$,

(i)

$$\begin{aligned}\widehat{\mu}_t - \mu_t &= O_p(H^{-1/2} + Hn^{-1}); \\ \widehat{h}_t^2 - h_t^2 &= O_p(H^{-1/2} + Hn^{-1}).\end{aligned}\tag{4.9}$$

(ii) If also $H = o(n^{2/3})$ and $\min h_t \geq c > 0$,

$$\frac{K_t}{K_{t,2}^{1/2}} \frac{(1 - \sum_{j=1}^p \theta_j)}{h_t} (\widehat{\mu}_t - \mu_t) \rightarrow_D \mathcal{N}(0, 1),\tag{4.10}$$

$$\frac{K_t}{K_{t,2}^{1/2}} \frac{1}{h_t^2 \text{var}(\varepsilon_1^2)^{1/2}} (\widehat{h}_t^2 - h_t^2) \rightarrow_D \mathcal{N}(0, 1),\tag{4.11}$$

where $K_t = \sum_{j=1}^n k_{tj}$ and $K_{t,2} = \sum_{j=1}^n k_{tj}^2$.

The normings have the usual rate $(nH)^{1/2}$, computation may be saved by replacing $K_{t,2}$ by $nH \int K^2(x)dx$ and except for very small or large t .

Theorem 4.1 implies that the estimate $\widehat{h}_t := (\widehat{h}_t^2)^{1/2}$ of h_t has the property

$$\frac{K_t}{K_{t,2}^{1/2}} \frac{2}{h_t \text{var}(\varepsilon_1^2)^{1/2}} (\widehat{h}_t - h_t) \rightarrow_D \mathcal{N}(0, 1)\tag{4.12}$$

and allows us to establish the following corollary for residuals

$$\widehat{\varepsilon}_t = \widehat{h}_t^{-1} (z_t - \widehat{\mu}_t^*), \quad t = 1, 2, \dots, n.$$

Note that the estimates $\widehat{\mu}_t, \widehat{h}_t^2$ may not be robust to misspecification of the time series model for x_t , unlike the simpler ones mentioned in Section 2. The following proposition shows that residuals $\widehat{\varepsilon}_t$ are consistent estimates of ε_t in (4.1).

Corollary 4.1 Let the assumptions of Theorem 4.1 hold. Then, for any $1 \leq t = t_n \leq n$, as $n \rightarrow \infty$,

$$\widehat{\varepsilon}_t - \varepsilon_t = O_P(H^{-1/2} + Hn^{-1}).\tag{4.13}$$

Smoothed nonparametric estimates are always sensitive to choice of bandwidth. In the numerical work of Sections 5 and 6 below we use leave-one-out cross-validation (CV), which is described as minimizing with respect to H

$$\begin{aligned}\sum_{t=1}^n (z_t - \widehat{\mu}_{t,-t}^*)^2 \text{ for } \widehat{\mu}_t^* &\text{ with } \widehat{\mu}_{t,-t}^* = (\sum_{j=1, j \neq t}^n k_{tj})^{-1} \sum_{j=1, j \neq t}^n k_{tj} z_j; \\ \sum_{t=1}^n ((z_t - \widehat{\mu}_t^*)^2 - \widehat{h}_{t,-t}^2)^2 \text{ for } \widehat{h}_t^2 &\text{ with } \widehat{h}_{t,-t}^2 = (\sum_{j=1, j \neq t}^n k_{tj})^{-1} \sum_{j=1, j \neq t}^n k_{tj} (z_j - \widehat{\mu}_j^*)^2.\end{aligned}$$

However, we do not imply CV is a necessary way of choosing H . In some settings CV has been shown to be optimal in the sense of asymptotically minimizing integrated mean squared error, but we establish no such property here. Moreover it is computationally relatively intensive; there is no reason to assume it will produce the best forecasts; being data-dependent it does not satisfy Assumption H; even its rate may not satisfy Assumption H; while other evidence suggests it will not minimize departure from the normal approximations justified in Theorem 4.1. Other, simpler, ways of choosing H are available, and one wise alternative to an automatic method entails employing a few arbitrary but sensible-looking H , thereby also gauging sensitivity.

Given y_1, \dots, y_{t-1} , by (1.1) and (4.1), the theoretical best one-step-ahead forecast $\hat{y}_{t|t-1}^{(opt)} = E[y_t | \varepsilon_s, s \leq t-1]$ of y_t is given by

$$\hat{y}_{t|t-1}^{(opt)} = \mu_t + h_t \sum_{j=1}^p \theta_j x_{t-j}. \quad (4.14)$$

It yields the forecast error $y_t - \hat{y}_{t|t-1}^{(opt)} = h_t \varepsilon_t$ with variance $V(h_t \varepsilon_t) = h_t^2$. Such a forecast depends on unknown parameters μ_t , h_t and θ_j and x_t .

A simple one-step-ahead forecast of y_t based on estimation results for the model (1.1) is

$$\hat{y}_{t|t-1} = \widehat{\mu_{t-1,t-1}^*} + \sum_{j=1}^p \hat{\theta}_{j|t-1} y_{t-j}, \quad (4.15)$$

where $\hat{\theta}_{1|t-1}, \dots, \hat{\theta}_{p|t-1}$ are estimated using sample y_1, \dots, y_{t-1} and

$$\widehat{\mu_{t-1,t-1}^*} = \left(\sum_{j=1}^{t-1} k_{tj} \right)^{-1} \left(\sum_{j=1}^{t-1} k_{tj} z_j \right)$$

denotes the estimate of μ_{t-1}^* computed using $z_j = y_j - \sum_{k=1}^p \hat{\theta}_{k|t-1} y_{j-k}$, $j = p+1, \dots, t-1$ and the bandwidth H that minimizes the variance of the in-sample forecast error

$$(t - t_0)^{-1} \sum_{s=t_0}^{t-1} (z_s - \widehat{\mu_{s-1,s-1}^*})^2, \quad (4.16)$$

where t_0 is a starting point. We attempt no study of asymptotic properties of these statistics. Indeed, Theorem 4.1 does not directly cover the one-sided estimates $\widehat{\mu_{t-1,t-1}^*}$ in (4.15), but they will have only nonparametric rate, as thus will the estimation error in (4.15), thus losing some advantage in the parametric rate for estimating θ_0 and converging slower than is possible with a correctly specified parametric μ_t .

An alternative, simpler, estimate of μ_t mentioned in Section 2, is the well known $\tilde{\mu}_t = \sum_{j=1}^n y_j k_{tj} / K_t$. Its bias treatment is similar to $\hat{\mu}_t$'s, while the stochastic component of $(K_t / h_t K_{t,2}^{1/2})(\tilde{\mu}_t - \mu_t)$ can be shown to be asymptotically normal with zero mean and variance $2\pi f_x(0)$ under mild conditions, where $f_x(u)$ is the spectral density of x_t . This would follow from an extension to heteroscedastic

weighting (under (4.2)) of (the short memory case of) Theorem 1 of Robinson (1997), under mild conditions including $f_x(u)$ continuous and positive at zero frequency. (Indeed the same theorem can be extended to provide a CLT under nonparametric long memory and antipersistence, with studentization discussed in the same reference.) Thus no parametric assumptions on x_t are imposed, unlike in the derivation of $\hat{\mu}_t$, with its parametric whitening. Of course $f_x(0)$ is unknown. If, though, x_t is assumed to be AR(p) (4.1), then $2\pi f_x(0) = k(0; \theta_0) = (1 - \sum_{j=1}^p \theta_{j0})^{-2}$ to correspond with (4.10). If the AR(p) assumption is incorrect then the normal approximation fails, like that in (4.10). But we have the option of replacing $f_x(0)$ by a consistent nonparametric estimate (e.g. one based on AR(p) fitting as above but with p regarded as diverging slowly relative to n), whence we can justify equally efficient inference to that provided by $\hat{\mu}_t$ but with a simpler point estimate and without imposing parametric assumptions on x_t .

5 Monte-Carlo simulations

First we report finite-sample properties of estimates of the parameter θ_0 in the model

$$y_t = \mu_t + h_t x_t, \quad x_t = \theta_0 x_{t-1} + \varepsilon_t, \quad \theta_0 = 0.5, \quad \varepsilon_t \sim IIDN(0, 1) \quad (5.1)$$

for variety of choices of μ_t and h_t but only for a single sample size $n = 512$. Even given the inefficiency of our estimates when h_t is not constant, $n = 512$ might seem a large sample for a one-parameter model, especially when autocorrelation is modest. Thus we also carried out some simulations with $n = 256$, as well as with $\theta_0 = 0.8$, using the same functions μ_t and h_t , number of replications and trimming parameters. As expected there is some deterioration overall with respect to such changes, in particular the results tend to be worse for larger trimming, the estimates of θ_0 tend to be less precise, and the normal approximation is less accurate. Further reducing n , and choosing θ_0 closer to the unit root, would surely lead to further deterioration (simulation results are available upon request).

We estimate $\theta_0 = 0.5$ by $\hat{\theta}_{ad} = \hat{\theta}^{(2)}$ and $\hat{\theta}_{wh} = \hat{\theta}^{(1)}$, for arbitrary choices $\nu = 0, 1, \dots, 9$ of the trimming parameter introduced in Section 3. By Theorem 3.2, both estimates, as n and ν increase, satisfy the normal approximation of (3.14),

$$\frac{S}{V^{*1/2}}(\hat{\theta}^{(k)} - \theta_0) \rightarrow_D \mathcal{N}(0, 1 - \theta_0^2), \quad k = 1, 2. \quad (5.2)$$

The true values of h_t were used in (5.2) for evaluation of the 95% confidence interval for θ_0 . Of course this makes no logical sense because the paper assumes the h_t are unknown, on which basis a practitioner cannot compute such intervals based on the true h_t . On the other hand if the practitioner knew the h_t she would be foolish to use the parameter estimates justified in the paper, but instead should use ones based on the y_t/h_t , thereby also obtaining more efficient estimates. Thus we also obtained some Monte-Carlo results with estimated S , V^* , namely $\hat{S} = \sum_{t=1}^n \hat{h}_t^2$, $\hat{V}^* = \sum_{t=1}^n \hat{h}_t^4$, with \hat{h}_t as defined in Section 4; equally one could use instead $\tilde{V} = \sum_{t=2}^n \hat{h}_t^2 \hat{h}_{t-1}^2$. The paper falls short of establishing (5.2) after such replacements; we do not verify the property

$\tilde{R}/R \rightarrow_p 1$ referred to in remark 9 of Section 2. However we computed coverage probabilities as if the normal approximation holds for the studentized quantity $(\hat{S}/\hat{V}^{*1/2})(\hat{\theta} - \theta_0)$, and compared them with those using the true value of h_t . We found almost identical coverage probabilities. In Tables 1-4 we report coverage using the estimate \hat{h}_t with kernel function $K(u) = \exp(-|u|)$ and bandwidth H selected from $1, \dots, [2n^{2/3}]$ with cross-validation.

Note that the unconstrained minimizers of $Q_{n,\nu}^{(1)}(\theta)$ and $Q_{n,\nu}^{(2)}(\theta)$ are respectively described in closed form as

$$\begin{aligned}\tilde{\theta}^{(1)} &= \frac{\sum_{j=\nu+1}^{[n/2]} I_y(u_j) \cos u_j}{\sum_{j=\nu+1}^{[n/2]} I_y(u_j)}, \\ \tilde{\theta}^{(2)} &= \left(\frac{2\pi}{[n/2]} \left[\sum_{j=\nu+1}^{[n/2]} I_y(u_j) + \frac{Q_{n,\nu}(\hat{\theta}^{(1)})}{2\pi} \sum_{j=1}^{\nu} |1 - \hat{\theta}^{(1)} e^{iu_j}|^{-2} \right]^{-1} \right. \\ &\quad \left. \times \left(\frac{2\pi}{[n/2]} \left[\sum_{j=\nu+1}^{[n/2]} I_y(u_j) + \frac{Q_{n,\nu}(\hat{\theta}^{(1)})}{2\pi} \sum_{j=1}^{\nu} |1 - \hat{\theta}^{(1)} e^{iu_j}|^{-2} \cos u_j \right] \right),\end{aligned}$$

(where we might replace $\hat{\theta}^{(1)}$ by $\tilde{\theta}^{(1)}$ in the latter). These need not necessarily fall within the stationary region $(-1, 1)$ (though they are very likely to if θ_0 is not near the boundary and n is reasonably large) but they differ from the constrained minimizers (over $\Theta = [\eta - 1, 1 - \eta]$, for $\eta \in (0, 1)$) $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$ only when the former pair exceed $1 - \eta$ or are less than $\eta - 1$. Thus

$$\hat{\theta}^{(k)} = \tilde{\theta}^{(k)} 1\left(\tilde{\theta}^{(k)} \in \Theta\right) + (1 - \eta) 1\left(\tilde{\theta}^{(k)} > 1 - \eta\right) + (\eta - 1) 1\left(\tilde{\theta}^{(k)} < \eta - 1\right), \quad k = 1, 2,$$

so $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$ can be obtained simply by trimming the closed form estimates $\tilde{\theta}^{(1)}$, $\tilde{\theta}^{(2)}$, numerical optimization, and the asymptotic arguments (uniform convergence for consistency, mean value theorem for asymptotic normality) tailored for implicitly-defined extremum estimates, see Section 3, can be replaced by simpler ones. In case of μ_t constant and $\nu = 0$, by expanding and using the property $\sum_{j=1}^n e^{itu_j} = n1(t = 0, \text{ mod } n)$, when n is odd

$$\tilde{\theta}^{(k)} = \tilde{\theta} \left(\frac{\sum_{t=2}^n (y_{t-1} - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \right) + \frac{(y_1 - \bar{y})(y_n - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} = \tilde{\theta} + O_p(n^{-1}), \quad k = 1, 2, \quad (5.3)$$

with $\tilde{\theta}$ given by (2.2). For n even $\tilde{\theta}^{(k)}$ is more complicated (due to our extra summand in the definition of $Q_n(\theta)$) but the final outcome in (5.3) is the same. It follows that $\tilde{\theta}^{(k)}$ has the same limit distribution as in (2.10). Moreover, for small enough $\delta > 0$, under Assumption A(iii),

$$1\left(\left|\tilde{\theta}^{(k)}\right| > 1 - \eta\right) \leq 1\left(\left|\tilde{\theta}^{(k)} - \theta_0\right| > \delta\right) \leq \left(\tilde{\theta}^{(k)} - \theta_0\right)^2 / \delta^2 = O_p(n^{-1}), \quad k = 1, 2,$$

from (2.10), so not only is the limit distribution of $\hat{\theta}^{(k)}$ also (2.10), but this actually follows from the proof of Theorem 2.2 and minor side calculations. Nevertheless, in our Monte Carlo study all θ_0 estimates were obtained by numerical optimization with $\eta = 0.1$.

Tables 1-4 report the Monte Carlo average value $E\hat{\theta}$, root mean squared error (RMSE) and the 95% coverage probabilities (CP) for the estimates $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ computed for the values $\nu = 0, 1, \dots, 9$ of trimming parameter based on 1,000 replications. We consider constant $h_t = 1$, the sinusoidal $h_t = 0.5 + |\sin(2\pi t/n)|$ and the break $h_t = 1 + 1(t/n > 0.5)$. Each of these h_t satisfies Assumption B, but the third h_t doesn't satisfy Assumption F. Notice that for $\nu = 0$, $\hat{\theta}_{ad} = \hat{\theta}_{wh} = \hat{\theta}$, so that the estimates $\hat{\theta}_{ad}$, $\hat{\theta}_{wh}$ obtained with $\nu = 0$ correspond to $\hat{\theta}$ of 3.2. The first two rows of these tables show the distortions of the estimator $\hat{\theta}$ arising from the changing mean μ_t .

Table 1 centers on the case $\mu_t = 0$. If μ_t is constant, then obviously no trimming is needed. The main difference between $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ lies in that the trimming with higher ν leads to increase of the negative bias of $\hat{\theta}_{wh}$ while the bias of $\hat{\theta}_{ad}$ remains close to zero. This distorts substantially the 95% CP based on $\hat{\theta}_{wh}$ even for moderate values of the trimming parameter ν , e.g. for $\nu = 4$ the CP drops below 90% for all three cases of h_t . For $\hat{\theta}_{ad}$, the CPs are not affected by the trimming and remain close to, but always below, nominal.

Tables 2-3 show that under changing mean μ_t , the bias in $\hat{\theta}_{wh}$ is sometimes substantially positive, leading to distorted CPs, but it decreases monotonically with increasing ν , while $E\hat{\theta}_{ad}$ monotonically drops to θ_0 , and then stays flat (for $\nu > 3$) with the CPs close to the nominal probability 0.95. Without trimming both estimates $\hat{\theta}_{ad}$, $\hat{\theta}_{wh}$ are biased. Overall, the performance of $\hat{\theta}_{ad}$ is found to be preferable in being less sensitive to ν , though in a number of cases CPs are worse.

Table 4 reports estimation results for μ_t with a strong break which does not satisfy Assumption D of Section 3. Nevertheless, we observe a somewhat similar behavior of $E\hat{\theta}$ and RMSE as in the case of the smooth μ_t in Tables 2 and 3. The bias and RMSE of $\hat{\theta}_{ad}$ are more stable for moderate to stronger trimming as opposed to the estimate $\hat{\theta}_{wh}$. Although the CPs remain distorted, the ν for $\hat{\theta}_{ad}$ selected by the above rule confirms consistency of the estimator with satisfactory MSE ¹.

Figures 1 and 2 depict the average value and one realization of the estimates $\hat{\theta}_{ad}$, $\hat{\theta}_{wh}$ for $\nu = 0, 1, \dots, 9$ for the model (5.1). To select ν , we suggest plotting $\hat{\theta}_{ad}$ for $\nu = 0, 1, \dots, 9$ and choosing the smallest ν from the region of ν s where the estimates $\hat{\theta}_{ad}$ show some degree of stability, e.g. $\nu = 3$ in Figure 1, and $\nu = 2$ in Figure 2, or using a priory pre-selected values $\nu = 4, \dots, 7$. (Performance of any given ν will depend on the form of μ_t and the signal-to-noise ratio.)

Last, we explore through only one replicate estimation of the μ_t and h_t , and forecasting of y_t , as described in Section 4. Of course there is minimal scientific value in results based on a single replicate. After estimating θ_0 by $\hat{\theta}_{ad}$ with $\nu = 4$, according to Section 4, estimation of μ_t and h_t was reduced to a nonparametric regression model (4.5) with an independent noise. The latter allows estimation of μ_t and h_t by a standard kernel method with bandwidth selected by leave-one-out cross-validation. Figure 3 reports estimation results on μ_t and h_t based on a sample of size $n = 512$ for y_t generated by (5.1) with $\mu_t = \sin(2\pi t/n)$, $h_t = 0.5 + |\sin(\pi t/n)|$. Figure 4 reports results with $\mu_t = \sin(5\pi t/n)$, $h_t = 0.5 + |\sin(2\pi t/n)|$ and $x_t \sim AR(2)$, $\theta_1 = 0.2$, $\theta_2 = 0.4$. The kernel

¹In the case of possible strong breaks in μ_t , we suggest pre-filtering the data by some step function. The number of potential breaks and their location for such step function can be found by minimizing residual sums of squares using the LWZ criterion and trimming parameter 0.2 (see Bai and Perron 1998, 2003). The latter should be subtracted from the data.

function $K(u) = \exp(-|u|)$ is used, and H is selected from integers $1, \dots, [2n^{2/3}]$ using leave-one-out cross-validation. To construct a 95% confidence band (CB) for the μ_t and h_t (which consists of approximate 95% confidence intervals based on (4.10) and (4.12) for each t), we estimate θ by $\hat{\theta}_{ad}$ in (4.10), h_t by \hat{h}_t in (4.10) and (4.12), while in (4.12) we replace $\text{var}(\varepsilon_1^2)$ by the sample variance of $\hat{\varepsilon}_t^2$ with $\hat{\varepsilon}_t$ given in (4.13).

Figures 3 and 4 show that the estimates $\hat{\mu}_t$ and \hat{h}_t track well the true values of μ_t and h_t , which fall within the 95% CB, except for the edges of the sample. The standardized residuals $\hat{\varepsilon}_t$ computed based on (4.13) show very weak autocorrelation and their sample variance is close to 1.

We also examine the one-step-ahead forecast $\hat{y}_{t|t-1}$ of y_t given in (4.15) and based on the subsample y_1, \dots, y_{t-1} . In Figures 5 and 6 we plot the forecasts $\hat{y}_{t|t-1}$ of y_t generated by (5.1) and $\mu_t = \sin(5\pi t/n) + 2t/n$, $h_t = 0.5 + |\sin(2\pi t/n)|$ and $\mu_t = 1 + 3 \times 1(t/n > 7/8)$, $h_t = 1$, respectively. For $n = 512$ we start forecasting at $t = 401$. To compute $\hat{y}_{t|t-1}$, we re-estimate θ_0 by $\hat{\theta}_{ad}$ with $\nu = 4$ and μ_s and h_s for $s = 1, \dots, t-1$ and choose H from integers $1, \dots, [(t-1)^{3/4}]$ by cross-validation setting $t_0 = t - 50$ in (4.16). We compare the results with the forecast $\hat{y}_{t|t-1,ar}$ produced by a simple AR(1) model estimated by OLS.

Table 5 reports the averages over 100 replications of the mean squared error (MSE)

$$\text{MSE} = \frac{1}{512-401} \sum_{t=401}^{512} (y_t - \hat{y}_{t|t-1})^2$$

of the forecast $\hat{y}_{t|t-1}$ in (4.15), MSE of the forecast $\hat{y}_{t|t-1,ar}$ and the theoretical minimum

$$\text{MSE}(\text{opt}) = \frac{1}{512-401} \sum_{t=401}^{512} (y_t - \hat{y}_{t|t-1}^{(\text{opt})})^2 = \frac{1}{512-401} \sum_{t=401}^{512} (h_t \varepsilon_t)^2$$

for a variety of functions μ_t and h_t under (5.1), with also $\theta_0 = 0, 0.8, -0.5$. The last equality follows from (5.1) and (4.14). It is evident that when μ_t is constant, $\hat{y}_{t|t-1}$ and $\hat{y}_{t|t-1,ar}$ are very similar, but when μ_t varies $\hat{y}_{t|t-1}$ is superior in all cases and especially so when there is zero or negative autocorrelation in x_t .

We compare these forecasts with the direct forecast

$$\hat{y}_{t|t-1}^{(\text{dir})} = \hat{\mu}_{t-1,t-1}(H_{t-1}^*) + \hat{e}_{t|t-1}, \quad \hat{e}_{t|t-1} = b_0 + b_1 e_{t-1}$$

of y_t , where

$$\hat{\mu}_{s-1,s-1}(H) = \left(\sum_{j=1}^{s-1} k_{sj} \right)^{-1} \left(\sum_{j=1}^{s-1} k_{sj} y_j \right), \quad s = 51, \dots, t,$$

is a one-sided kernel estimate of μ_s obtained using data y_1, \dots, y_{s-1} ,

$$H_{s-1}^* = \underset{j=s-50}{\text{argmin}} \sum_{j=s-50}^{s-1} (y_j - \hat{\mu}_{j-1,j-1}(H))^2$$

and $\hat{e}_{t|t-1}$ is an AR(1) OLS forecast of e_t based on $e_s = y_s - \hat{\mu}_{s-1,s-1}(H_{s-1}^*)$, $s = 51, \dots, t-1$, with b_0 and b_1 denoting the OLS estimates of AR(1) parameters. The MSE of the direct forecast $\hat{y}_{t|t-1}^{(\text{dir})}$

is comparable or larger than that of the proposed forecast $\hat{y}_{t|t-1}$.

Table 1: Monte-Carlo $E\hat{\theta}$, RMSE and 95%CP for the estimators $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ for (5.1) with $\mu_t = 0$.

		$h_t = 1$			$h_t = 0.5 + \sin(2\pi t/n) $			$h_t = 1 + 1(t/n > 0.5)$		
ν		$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP
0	$\hat{\theta}_{ad}$	0.49	0.04	93.1	0.49	0.05	93.3	0.49	0.05	93.1
	$\hat{\theta}_{wh}$	0.49	0.04	93.1	0.49	0.05	93.3	0.49	0.05	93.1
1	$\hat{\theta}_{ad}$	0.49	0.04	92.7	0.49	0.05	92.5	0.49	0.05	93.0
	$\hat{\theta}_{wh}$	0.49	0.04	92.9	0.49	0.05	92.8	0.49	0.05	93.0
2	$\hat{\theta}_{ad}$	0.49	0.04	92.6	0.49	0.05	92.6	0.49	0.05	92.6
	$\hat{\theta}_{wh}$	0.48	0.05	91.9	0.48	0.05	91.7	0.48	0.05	91.4
3	$\hat{\theta}_{ad}$	0.49	0.04	92.5	0.49	0.05	91.7	0.49	0.05	92.1
	$\hat{\theta}_{wh}$	0.47	0.05	89.3	0.47	0.05	90.5	0.47	0.05	90.1
4	$\hat{\theta}_{ad}$	0.49	0.04	92.4	0.49	0.05	91.5	0.49	0.05	92.7
	$\hat{\theta}_{wh}$	0.47	0.05	86.5	0.47	0.06	88.2	0.47	0.06	88.2
5	$\hat{\theta}_{ad}$	0.49	0.04	92.6	0.49	0.05	91.6	0.49	0.05	92.5
	$\hat{\theta}_{wh}$	0.46	0.06	82.7	0.46	0.06	84.7	0.46	0.06	85.9
6	$\hat{\theta}_{ad}$	0.49	0.04	92.3	0.49	0.05	91.4	0.49	0.05	91.8
	$\hat{\theta}_{wh}$	0.45	0.06	78.8	0.45	0.07	80.4	0.46	0.07	82.7
7	$\hat{\theta}_{ad}$	0.49	0.04	91.5	0.48	0.05	91.2	0.49	0.05	92.3
	$\hat{\theta}_{wh}$	0.45	0.07	74.4	0.45	0.07	75.6	0.45	0.07	79.4
8	$\hat{\theta}_{ad}$	0.48	0.05	91.3	0.48	0.05	90.8	0.48	0.05	91.5
	$\hat{\theta}_{wh}$	0.44	0.07	70.2	0.44	0.08	71.4	0.44	0.08	74.9
9	$\hat{\theta}_{ad}$	0.48	0.05	90.6	0.48	0.05	90.0	0.48	0.05	90.5
	$\hat{\theta}_{wh}$	0.43	0.08	63.9	0.43	0.08	66.7	0.44	0.08	70.6

Table 2: Monte-Carlo $E\hat{\theta}$, RMSE and 95%CP for the estimators $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ for (5.1) with $\mu_t = \sin(5\pi t/n)$.

		$h_t = 1$			$h_t = 0.5 + \sin(2\pi t/n) $			$h_t = 1 + 1(t/n > 0.5)$		
ν		$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP
0	$\hat{\theta}_{ad}$	0.63	0.13	7.2	0.60	0.10	30.3	0.56	0.07	70.1
	$\hat{\theta}_{wh}$	0.63	0.13	7.2	0.60	0.10	30.3	0.56	0.07	70.1
1	$\hat{\theta}_{ad}$	0.62	0.13	10.2	0.59	0.10	34.9	0.55	0.07	72.5
	$\hat{\theta}_{wh}$	0.61	0.12	13.0	0.58	0.09	40.7	0.55	0.06	78.1
2	$\hat{\theta}_{ad}$	0.56	0.07	62.5	0.54	0.06	79.3	0.52	0.05	91.0
	$\hat{\theta}_{wh}$	0.55	0.06	74.0	0.53	0.05	86.9	0.51	0.05	93.0
3	$\hat{\theta}_{ad}$	0.50	0.04	92.7	0.50	0.05	92.5	0.50	0.05	92.2
	$\hat{\theta}_{wh}$	0.48	0.04	91.6	0.48	0.05	91.9	0.48	0.05	90.7
4	$\hat{\theta}_{ad}$	0.49	0.04	93.0	0.49	0.05	91.9	0.49	0.05	92.9
	$\hat{\theta}_{wh}$	0.47	0.05	87.5	0.47	0.05	88.6	0.47	0.06	88.2
5	$\hat{\theta}_{ad}$	0.49	0.04	92.6	0.49	0.05	91.2	0.49	0.05	92.8
	$\hat{\theta}_{wh}$	0.46	0.06	83.7	0.46	0.06	85.2	0.46	0.06	86.1
6	$\hat{\theta}_{ad}$	0.49	0.04	92.1	0.49	0.05	91.4	0.49	0.05	92.0
	$\hat{\theta}_{wh}$	0.46	0.06	78.5	0.45	0.06	80.7	0.46	0.07	83.5
7	$\hat{\theta}_{ad}$	0.49	0.04	91.6	0.49	0.05	91.3	0.49	0.05	92.3
	$\hat{\theta}_{wh}$	0.45	0.07	74.6	0.45	0.07	75.3	0.45	0.07	79.6
8	$\hat{\theta}_{ad}$	0.48	0.05	91.3	0.48	0.05	90.6	0.48	0.05	91.5
	$\hat{\theta}_{wh}$	0.44	0.07	70.6	0.44	0.08	71.5	0.44	0.08	75.0
9	$\hat{\theta}_{ad}$	0.48	0.05	91.0	0.48	0.05	90.1	0.48	0.05	90.3
	$\hat{\theta}_{wh}$	0.43	0.08	64.3	0.43	0.08	66.7	0.44	0.08	70.9

Table 3: Monte-Carlo $E\hat{\theta}$, RMSE and 95%CP for the estimators $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ for (5.1) with $\mu_t = 2t/n$.

		$h_t = 1$			$h_t = 0.5 + \sin(2\pi t/n) $			$h_t = 1 + 1(t/n > 0.5)$		
ν		$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP
0	$\hat{\theta}_{ad}$	0.59	0.10	28.7	0.57	0.08	57.4	0.54	0.06	83.5
	$\hat{\theta}_{wh}$	0.59	0.10	28.7	0.57	0.08	57.4	0.54	0.06	83.5
1	$\hat{\theta}_{ad}$	0.53	0.05	81.0	0.52	0.05	89.2	0.51	0.05	92.6
	$\hat{\theta}_{wh}$	0.53	0.05	86.0	0.52	0.05	91.0	0.50	0.05	93.3
2	$\hat{\theta}_{ad}$	0.52	0.04	88.9	0.51	0.05	91.6	0.50	0.05	92.7
	$\hat{\theta}_{wh}$	0.51	0.04	92.8	0.50	0.04	93.3	0.49	0.05	93.2
3	$\hat{\theta}_{ad}$	0.51	0.04	91.8	0.51	0.04	92.4	0.50	0.05	92.8
	$\hat{\theta}_{wh}$	0.49	0.04	93.3	0.49	0.05	92.9	0.48	0.05	91.6
4	$\hat{\theta}_{ad}$	0.50	0.04	93.1	0.50	0.05	92.4	0.50	0.05	92.5
	$\hat{\theta}_{wh}$	0.48	0.04	92.1	0.48	0.05	91.3	0.47	0.05	89.2
5	$\hat{\theta}_{ad}$	0.50	0.04	93.4	0.50	0.05	92.3	0.49	0.05	93.4
	$\hat{\theta}_{wh}$	0.47	0.05	88.4	0.47	0.05	88.8	0.47	0.06	87.1
6	$\hat{\theta}_{ad}$	0.50	0.04	93.5	0.49	0.05	92.3	0.49	0.05	91.7
	$\hat{\theta}_{wh}$	0.46	0.05	83.6	0.46	0.06	86.3	0.46	0.06	84.0
7	$\hat{\theta}_{ad}$	0.49	0.04	92.6	0.49	0.05	91.6	0.49	0.05	91.9
	$\hat{\theta}_{wh}$	0.46	0.06	78.0	0.45	0.07	80.3	0.45	0.07	80.4
8	$\hat{\theta}_{ad}$	0.49	0.04	92.7	0.49	0.05	91.5	0.49	0.05	92.0
	$\hat{\theta}_{wh}$	0.45	0.07	75.1	0.45	0.07	75.2	0.45	0.07	76.8
9	$\hat{\theta}_{ad}$	0.49	0.04	92.2	0.49	0.05	91.6	0.48	0.05	90.7
	$\hat{\theta}_{wh}$	0.44	0.07	69.2	0.44	0.08	69.9	0.44	0.08	72.8

Table 4: Monte-Carlo $E\hat{\theta}$, RMSE and 95%CP for the estimators $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ for (5.1) with $\mu_t = 1 + 3 \times 1(t/n > 0.5)$.

		$h_t = 1$			$h_t = 0.5 + \sin(2\pi t/n) $			$h_t = 1 + 1(t/n > 0.5)$		
ν		$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP	$E\hat{\theta}$	RMSE	CP
0	$\hat{\theta}_{ad}$	0.81	0.31	0.0	0.77	0.27	0.0	0.69	0.20	0.0
	$\hat{\theta}_{wh}$	0.81	0.31	0.0	0.77	0.27	0.0	0.69	0.20	0.0
1	$\hat{\theta}_{ad}$	0.61	0.11	17.0	0.58	0.09	43.9	0.55	0.06	78.4
	$\hat{\theta}_{wh}$	0.60	0.11	21.0	0.58	0.08	48.5	0.54	0.06	81.9
2	$\hat{\theta}_{ad}$	0.61	0.11	15.8	0.58	0.09	42.5	0.55	0.06	78.2
	$\hat{\theta}_{wh}$	0.60	0.10	23.5	0.57	0.08	52.7	0.53	0.06	84.5
3	$\hat{\theta}_{ad}$	0.56	0.07	60.6	0.54	0.06	79.3	0.52	0.05	89.3
	$\hat{\theta}_{wh}$	0.54	0.06	78.8	0.52	0.05	89.8	0.50	0.05	93.5
4	$\hat{\theta}_{ad}$	0.56	0.07	60.9	0.54	0.06	79.2	0.52	0.05	90.0
	$\hat{\theta}_{wh}$	0.53	0.05	82.3	0.52	0.04	91.9	0.50	0.05	93.3
5	$\hat{\theta}_{ad}$	0.53	0.05	80.6	0.52	0.05	88.5	0.51	0.05	92.7
	$\hat{\theta}_{wh}$	0.51	0.04	94.0	0.50	0.04	95.0	0.48	0.05	91.9
6	$\hat{\theta}_{ad}$	0.53	0.05	80.2	0.52	0.05	88.9	0.51	0.05	92.3
	$\hat{\theta}_{wh}$	0.50	0.04	94.9	0.49	0.04	94.0	0.47	0.05	89.9
7	$\hat{\theta}_{ad}$	0.52	0.05	87.8	0.51	0.05	91.1	0.50	0.05	92.5
	$\hat{\theta}_{wh}$	0.48	0.05	91.8	0.47	0.05	90.0	0.46	0.06	85.4
8	$\hat{\theta}_{ad}$	0.52	0.05	88.0	0.51	0.05	91.1	0.50	0.05	92.2
	$\hat{\theta}_{wh}$	0.48	0.05	89.7	0.47	0.05	87.2	0.46	0.06	83.1
9	$\hat{\theta}_{ad}$	0.51	0.04	90.0	0.50	0.04	91.4	0.49	0.05	92.2
	$\hat{\theta}_{wh}$	0.46	0.06	82.0	0.45	0.06	80.5	0.45	0.07	77.3

Figure 1: Average values $\hat{\theta}_{ad}$, $\hat{\theta}_{wh}$ (left) and one simulation of $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ (right) for trimming parameter $\nu = 0, 1, \dots, 9$ and (5.1) with $\mu_t = \sin(5\pi t/n)$, $h_t = 1$.

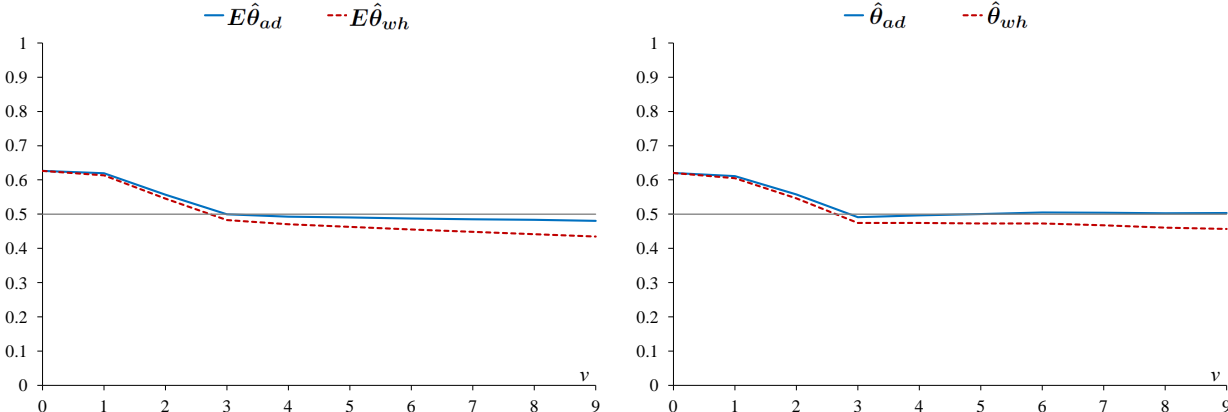


Figure 2: One simulation of $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ for trimming parameter $\nu = 0, 1, \dots, 29$ and (5.1) with $\mu_t = 0$, $h_t = 1$.

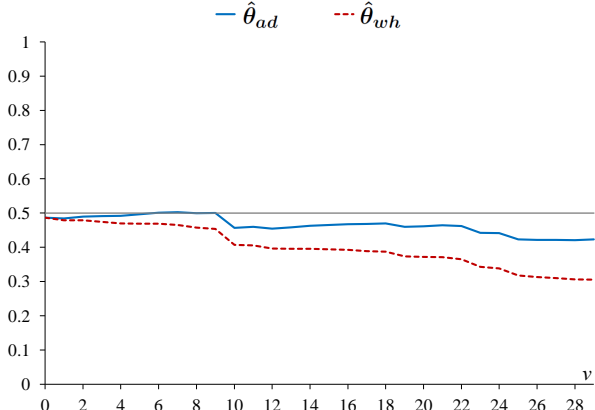


Figure 3: Estimation results for (5.1) with $\mu_t = \sin(2\pi t/n)$, $h_t = 0.5 + |\sin(\pi t/n)|$. (Top left) y_t , $\hat{\mu}_t$, μ_t and 95% confidence band for μ_t ; (Top right) h_t , \hat{h}_t and 95% confidence band for h_t ; (Bottom left) Sample ACF for residuals $\hat{\varepsilon}_t$ at lags 1,2,...,10 and 95% confidence band for zero correlation; (Bottom right) Estimates $\hat{\theta}_{ad}$ and $\hat{\theta}_{wh}$ of θ_0 for trimming parameter $\nu = 0, 1, \dots, 9$.

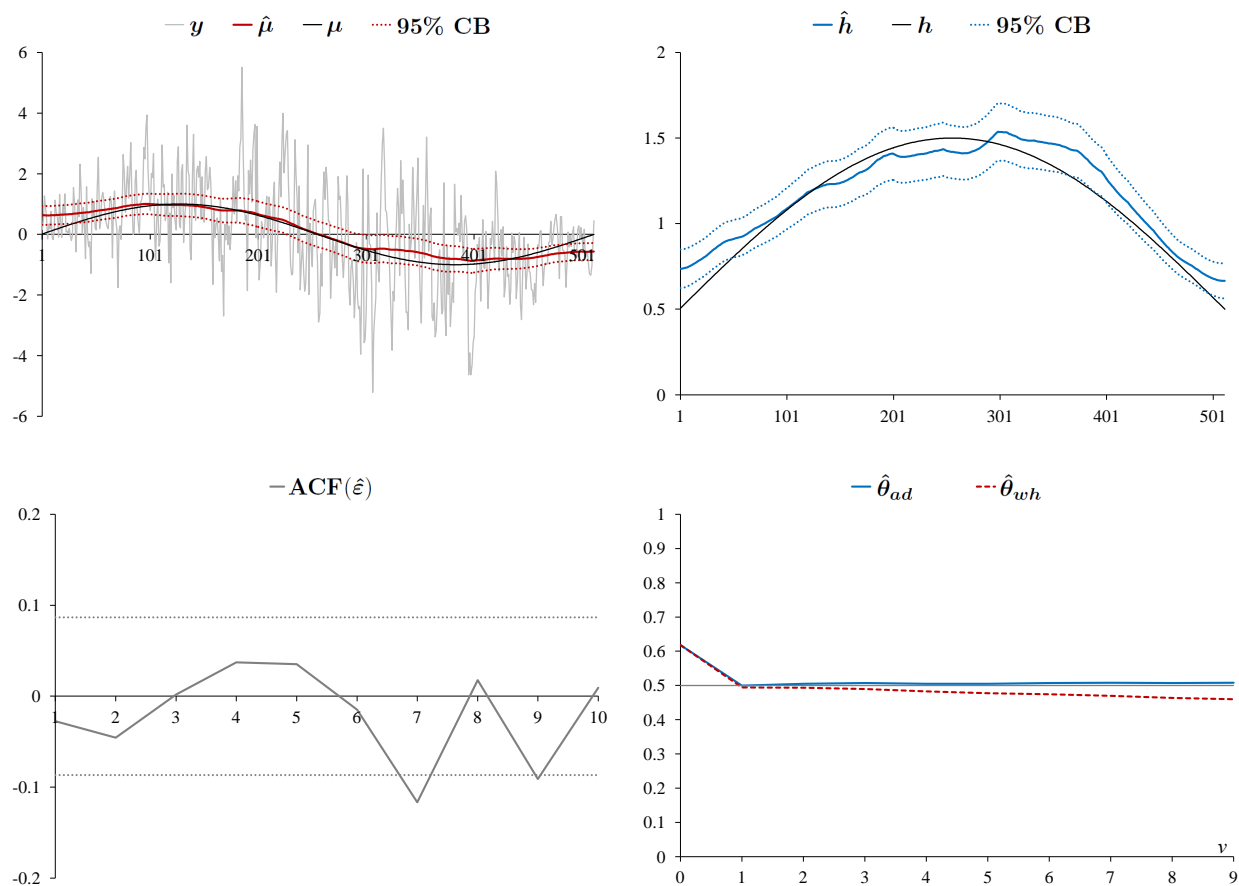


Figure 4: Estimation results for $y_t = \mu_t + h_t x_t$ with $\mu_t = \sin(5\pi t/n)$, $h_t = 0.5 + |\sin(2\pi t/n)|$, $x_t \sim AR(2)$, $\theta_{10} = 0.2$, $\theta_{20} = 0.4$. (Top left) y_t , $\hat{\mu}_t$, μ_t and 95% confidence band for μ_t ; (Top right) h_t , \hat{h}_t and 95% confidence band for h_t ; (Bottom left) Sample ACF for residuals $\hat{\varepsilon}_t$ at lags 1, 2, ..., 10 and 95% confidence band for zero correlation; (Bottom right) Estimates $\hat{\theta}_{1,ad}$, $\hat{\theta}_{2,ad}$ and $\hat{\theta}_{1,wh}$, $\hat{\theta}_{2,wh}$ of θ_1 , θ_2 for trimming parameter $\nu = 0, 1, \dots, 9$.

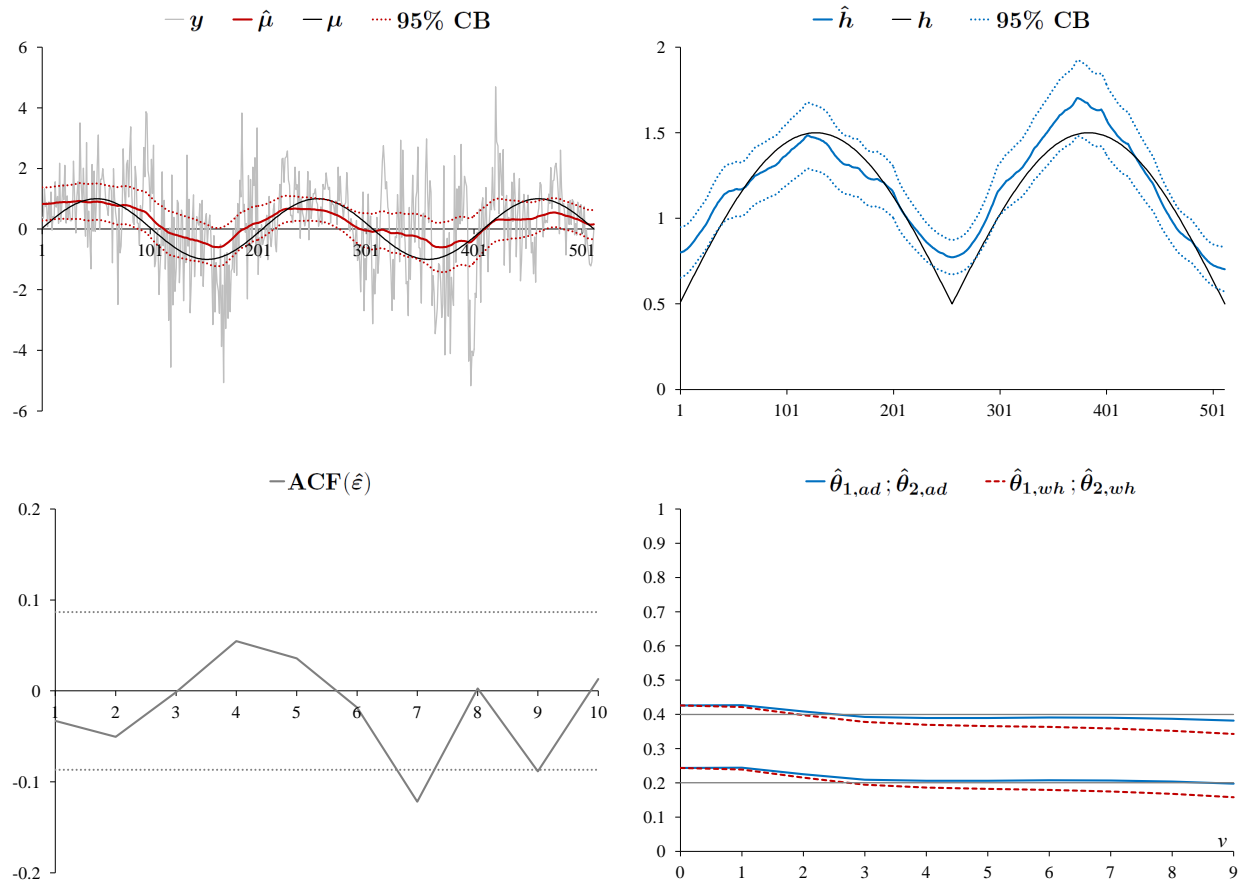


Figure 5: One-step-ahead forecasts $\hat{y}_{t|t-1}$ and AR(1) forecast $\hat{y}_{t|t-1,ar}$ for (5.1) with $\mu_t = \sin(5\pi t/n) + 2t/n$, $h_t = 0.5 + |\sin(2\pi t/n)|$. Plot of y_t and μ_t .

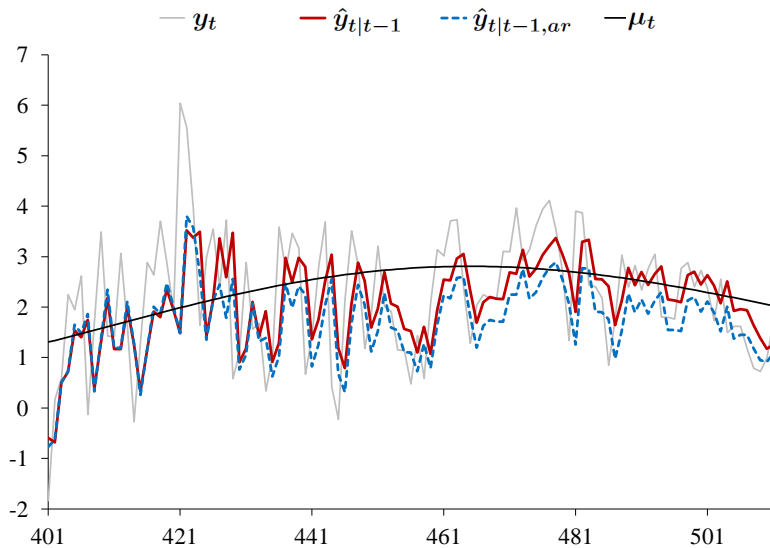


Figure 6: One-step-ahead forecasts $\hat{y}_{t|t-1}$ and AR(1) forecast $\hat{y}_{t|t-1,ar}$ for (5.1) with $\mu_t = 1 + 3 \times 1(t/n > 7/8)$, $h_t = 1$. Plot of y_t and μ_t .

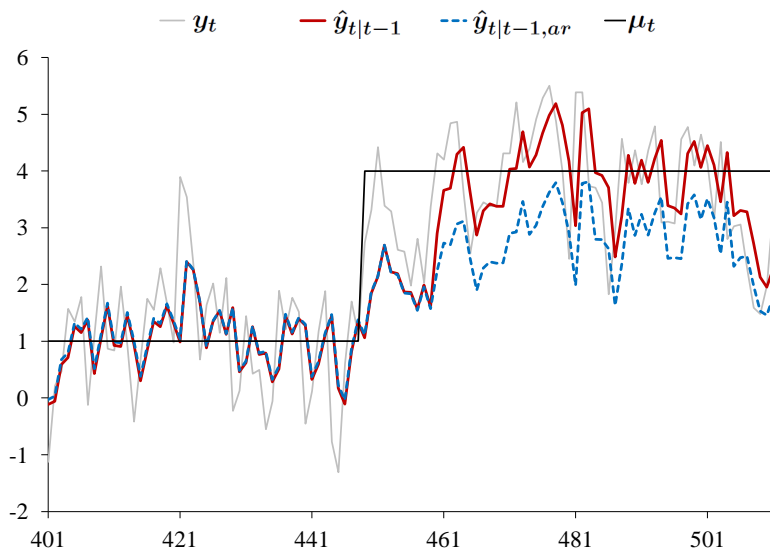


Table 5: Average MSE of the forecasts $\hat{y}_{t|t-1}$, $\hat{y}_{t|t-1,ar}$, $\hat{y}_{t|t-1}^{(dir)}$, and $\text{MSE}(\text{opt})$ for $\hat{\mu}_{t-1,t-1}$; $\mu_{1t} = \sin(5\pi t/n) + 2t/n$, $\mu_{2t} = 1 + 3 \times 1(t/n > 7/8)$, $\mu_{3t} = 1 + 1(17/20 \leq t/n < 19/20) + 2 \times 1(t/n \geq 19/20)$, $h_{1t} = 0.5 + |\sin(2\pi t/n)|$, $h_{2t} = 0.5 + |\sin(5\pi t/n)|$.

μ_t	h_t	θ_0	$\text{MSE}(\hat{y}_{t t-1})$	$\text{MSE}(\hat{y}_{t t-1,ar})$	$\text{MSE}(\hat{y}_{t t-1}^{(dir)})$	$\text{MSE}(\text{opt})$
0	1	0	1.026	1.001	1.028	0.998
		0.5	1.026	1.001	1.204	0.998
		0.8	1.029	1.002	1.098	0.998
		-0.5	1.025	1.001	1.012	0.998
0	h_{1t}	0	1.297	1.264	1.302	1.259
		0.5	1.297	1.264	1.521	1.259
		0.8	1.300	1.265	1.400	1.259
		-0.5	1.297	1.265	1.280	1.259
μ_{1t}	h_{1t}	0	1.449	2.386	1.451	1.259
		0.5	1.377	1.593	1.527	1.259
		0.8	1.325	1.336	1.398	1.259
		-0.5	1.492	4.118	1.664	1.259
μ_{2t}	1	0	1.404	3.359	1.393	0.998
		0.5	1.230	1.739	1.312	0.998
		0.8	1.153	1.219	1.176	0.998
		-0.5	1.610	6.888	1.789	0.998
μ_{3t}	h_{2t}	0	1.519	2.327	1.519	1.303
		0.5	1.438	1.596	1.605	1.303
		0.8	1.388	1.379	1.456	1.303
		-0.5	1.608	3.650	1.691	1.303

6 Empirical application

In this section, we illustrate how the proposed estimation and forecasting methodology in Sections 3-4 can be applied to real data. We examine the monthly seasonally adjusted US civilian unemployment rate during the period 1970-2016². A variety of univariate parametric models has been used to model the US unemployment rate, see for example Montgomery et al. (1998), van Dijk, Franses and Paap (2002) and Deschamps (2008). Since the dynamics of the unemployment rate are mainly driven by the business cycle, various nonlinear univariate models have been considered besides ARIMA models. Montgomery et al. (1998) study the quarterly unemployment US rate for the period 1948-1993. Among other things, they compare ARIMA models for unemployment with threshold autoregression (TAR) and Markov switching autoregression (MSA) models for changes in unemployment. These authors find that for one-step-ahead forecasting unemployment, TAR and MSA models outperform ARIMA models in terms of mean squared error (MSE) during periods of economic contractions, but underperform during period of economic expansions, so that for the whole period the ARIMA approach gives the same forecasting performance as TAR and slightly better than MSA. Van Dijk, Franses and Paap (2002) examine the monthly US unemployment rate for the period 1970-1999 using the fractionally integrated threshold autoregression (FI-TAR), the fractionally integrated smooth transition autoregression (FI-STAR), the fractionally integrated autoregression (ARFI) as well as the STAR model. The authors find that the introduction of long memory does not lead to an improved fit, but no comparison is done in terms of out-of-sample forecasting. Deschamps (2008) compares the performance of the logistic smooth transition autoregression (LSTAR), the Markov switching autoregression (MSAR) and AR models for forecasting the yearly change of the logistic transformation of deseasonalized monthly US unemployment rate during the period 1960-2004. He finds that LSTAR and MSAR models have better mean absolute forecasting error than AR ones, but the difference is small leading the author to conclude that all three models can provide a fairly good forecasting of unemployment, although no comparison was done in periods of economic contractions and expansions like in Montgomery et al. (1998). From a nonparametric perspective, Golan and Perloff (2004) use simplex methods to forecast the quarterly US unemployment rate using the same data as in Montgomery et al. (1998) and compare their results without though discriminating between economic contractions and expansions. For one-step-ahead forecasting their method gives slightly worse MSE than the simple ARIMA(1,1,0) for unemployment. Overall, it seems that the simple ARIMA(1,1,0) for one-step-ahead forecasting of US unemployment is the best from the univariate parametric and nonparametric methods that have been used, although this is not the case during periods of economic contractions³. Moreover, in Montgomery et al. (1998) and Golan and Perloff (2004), the univariate approach is compared with forecasting from multivariate parametric models and with forecasts from surveys; the authors find that survey forecasts have the smallest MSE overall.

We denote by U_t the unemployment rate at month t and evaluate its first differences $y_t = \Delta U_t$ resulting to a sample of $n = 567$ observations. We fit model (1.1) to changes in unemployment

²The source of the data is Federal Reserve Bank of St. Louis.

³We should note that for multi-steps-ahead forecasting, the performance of the ARIMA(1,1,0) deteriorates and becomes worse than the above parametric and nonparametric methods.

y_t assuming that x_t follows an AR(1) model. The estimation results are presented in Figure 7. We first estimate the AR(1) parameter θ using the adjusted Whittle estimator $\hat{\theta}_{ad}$ with trimming parameters $\nu = 0, 1, \dots, 9$. The estimates $\hat{\theta}_{ad}$ drop initially with ν , become flat for $\nu = 3, 4$ and then start dropping again. We pick trimming parameter $\nu = 4$ for which $\hat{\theta}_{ad} = 0.16$. We then estimate the functions μ_t and h_t with kernel function $K(u) = \exp(-|u|)$, and bandwidth selected from $15, \dots, [2n^{2/3}]$ in cross-validation. For the evaluation of the standard error in the 95% CB for μ_t we employ the estimates \hat{h}_t of h_t and $\hat{\theta}_{ad}$ for θ in (4.10), while to evaluate SD in (4.12) for h_t we use the estimate \hat{h}_t and replace $\text{var}(\varepsilon_1^2)$ by the sample variance of ε_t^2 . We observe that the estimated mean $\hat{\mu}_t$ of changes in unemployment fluctuates around 0, with a tendency of positive changes observed during periods of recessions and negative changes during periods of growth⁴. It is interesting to notice that $\hat{\mu}_t$ starts becoming positive shortly before the beginning of recessions and negative shortly after the end of the recessions. The estimated \hat{h}_t seems to be also non-constant with a tendency of higher values during periods of lengthy recessions and lower values during periods of growth or short recessions. The standardized residuals $\hat{\varepsilon}_t$ show some first order negative autocorrelation⁵, while their Jarque-Bera statistic does not reject normality.

We also explore one-step-ahead forecasting of changes in unemployment rate y_t . We start forecasting at $t = 401$ in every step re-estimating the AR parameter and the functions μ_t and h_t . We use $\nu = 4$, $K(u) = \exp(-|u|)$, select H from $1, \dots, [(t-1)^{3/4}]$ and set $t_0 = t - 50$ in cross-validation (4.16). The results of one-step-ahead forecasting, $\hat{y}_{t|t-1}$, and those from a simple AR(1) model estimated by OLS, $\hat{y}_{t|t-1,ar}$, are compared in Figure 8 with the data. It shows that the forecast $\hat{y}_{t|t-1}$ is superior to the AR(1) forecast $\hat{y}_{t|t-1,ar}$ during and around the period of recession, while in other times the two forecasts look similar. The MSE of all the forecasts $\hat{y}_{t|t-1}$, $t = 401, \dots, 567$, is 0.026, while that of $\hat{y}_{t|t-1,ar}$ is 0.028, so that we have an improvement of around 7%. Concentrating on the recession period, we find MSE of 0.040 for $\hat{y}_{t|t-1}$ and 0.077 for $\hat{y}_{t|t-1,ar}$, an improvement of 48%. We tried also the direct forecasting method $\hat{y}_{t|t-1}^{(dir)}$ which was nearly identical to $\hat{y}_{t|t-1}$ because of weak autocorrelation in the data.

⁴We use the NBER chronology for the dating of recessions.

⁵We also estimated model (1.1) assuming x_t follows an AR(2) and an ARMA(1,1) model, but the first order autocorrelation in the standardized residuals was not removed.

Figure 7: Model fitting to the first differences of unemployment rate y_t , $t = 1, \dots, 567$. Model: $y_t = \mu_t + h_t x_t$, $x_t \sim AR(1)$. Shaded areas correspond to recessions. (Top left) Estimated mean $\hat{\mu}_t$ and 95% confidence band for μ_t ; (Top right) Estimated \hat{h}_t and 95% confidence band for h_t ; (Bottom left) Sample ACF for standardized residuals $\hat{\varepsilon}_t$ at lags 1,2,...,10 and 95% confidence band for zero autocorrelation; (Bottom right) Adjusted Whittle estimate ($\hat{\theta}_{ad}$) and 95% confidence band for θ_0 with trimming parameter $\nu = 0, 1, \dots, 9$.

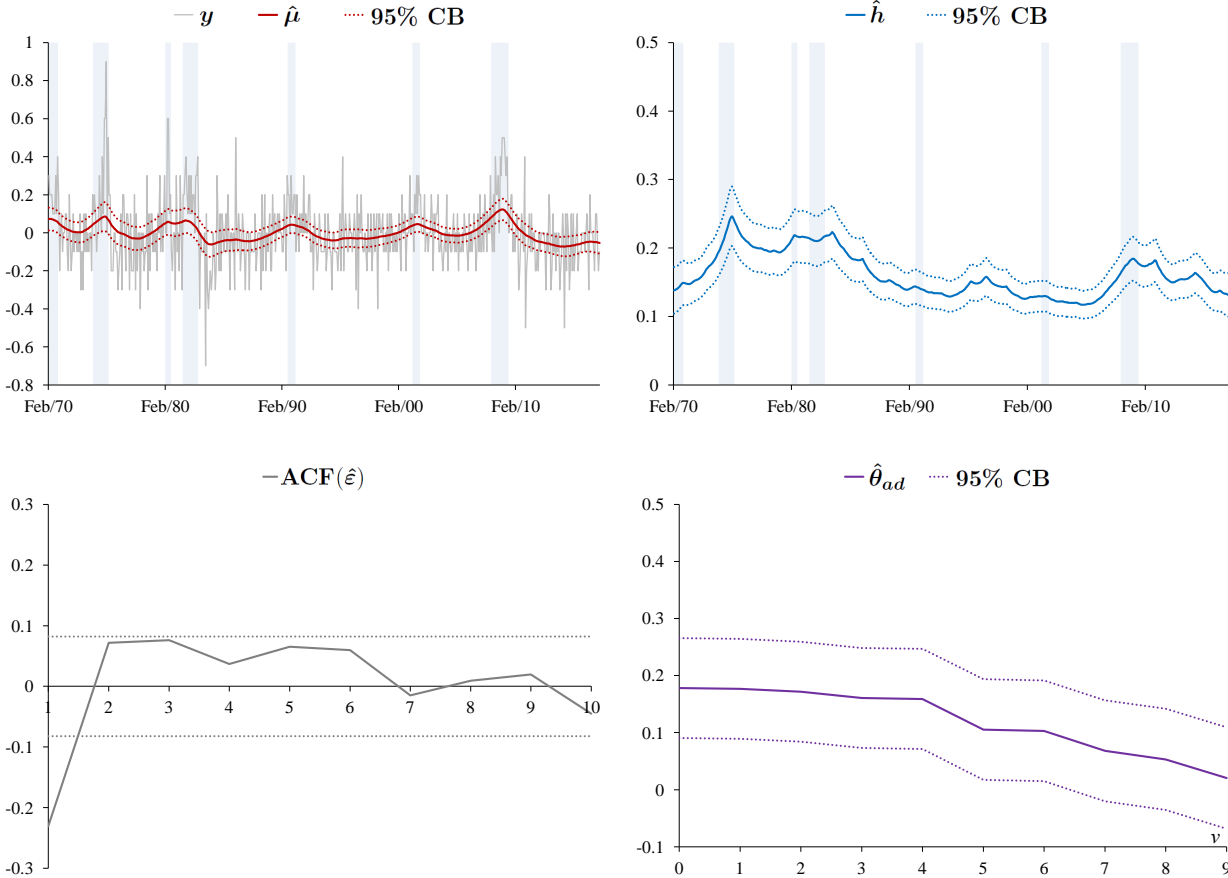
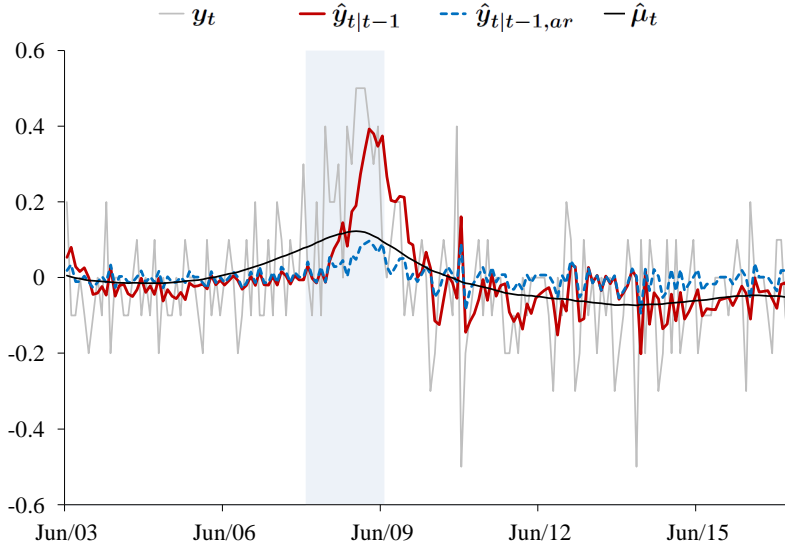


Figure 8: One-step-ahead forecasts $\hat{y}_{t|t-1}$ of the first differences of unemployment rate y_t , forecast based on AR(1) model, $\hat{y}_{t|t-1,ar}$, and estimated mean $\hat{\mu}_t$. Shaded area correspond to recession.



7 Final comments

We have developed estimation and inference for a nonstationary semiparametric time series model in which mean and variance can nonparametrically change with time and stationary autocorrelations are described parametrically. A number of directions for further research come to mind.

1. The practitioner may well lack reliable prior information on the parametric specification of $k(u; \theta)$, so there is risk of inconsistent estimation, biased forecasts and invalid inferences. In case μ_t is constant, (1.3) suggests that sample autocorrelations of y_t will consistently estimate population autocorrelations of x_t , giving hope that model choice procedures, such as AIC, or modifications thereof, will help to choose $k(u; \theta)$ from the y_t data. When μ_t is not constant this will not work, and it appears one should begin by estimating μ_t , which can be done consistently by simple nonparametric regression of y_t (rather than the $k(u; \theta)$ – dependent approach of Section 4).

2. Given that Whittle estimates for parametric stationary long memory models have analogous asymptotic properties to those for short memory ones (following Fox and Taqqu (1986)), it would seem possible to extend our results to allow x_t to have stationary long memory.

3. First-differencing of macroeconomic series is widely practised, as in Section 6, with a view to removing a unit root. If in (1.1) actually $z_t = x_t - x_{t-1}$ is stationary with short memory

$$y_t - y_{t-1} = \mu_t - \mu_{t-1} + h_t z_t + (h_t - h_{t-1}) x_{t-1} = h_t z_t + O_p\left(t^{1/2}/n\right),$$

under (1.4) and (2.9) and bounded differentiability; properties of estimates of parameters describing z_t remain to be studied. More generally nonstationary fractional models for x_t with unknown

memory parameter can be considered, where a time varying mean may or may not be dominated by a stochastic trend. It has been found that the Whittle estimates of Section 3 are inconsistent in the presence of nonstationarity, unlike suitably tapered versions (Velasco and Robinson (2000)) or conditional-sum-of-squares estimates (Hualde and Robinson (2011)).

4. Observable explanatory variables may seem relevant to explaining y_t . With μ_t a known function of such explanatory variables and nonparametrically smoothly time-varying coefficients, as in Robinson (1989), where the latter can be estimated using kernels, similar results to ours should be available; the same kind of approach was used by Dahlhaus (1997) to model time-varying dependence parameters. We could allow h_t to similarly depend on explanatory variables.

8 Proofs

8.1 Proofs of Section 2.

PROOF OF THEOREM 2.1. We have

$$\tilde{\theta} - \theta = \sum_{t=2}^n (y_t - \bar{y} - \theta (y_{t-1} - \bar{y})) (y_{t-1} - \bar{y}) / \sum_{t=1}^{n-1} (y_t - \bar{y})^2. \quad (8.1)$$

Defining, $\bar{hx} = n^{-1} \sum_{t=1}^n h_t x_t$, the denominator of (8.1) is

$$\sum_{t=1}^{n-1} (\mu_t - \bar{\mu} + h_t x_t - \bar{hx})^2 = M + 2 \sum_{t=1}^{n-1} (\mu_t - \bar{\mu}) (h_t x_t - \bar{hx}) + \sum_{t=1}^{n-1} (h_t x_t - \bar{hx})^2.$$

Since x_t has mean zero and bounded spectral density it is easily seen that \bar{hx} has mean zero and variance $O(S/n^2)$, and $\sum_{t=1}^{n-1} (\mu_t - \bar{\mu}) h_t x_t$ has mean zero and variance W , where, for example, $W = O(\max h_t^2 M) = o(MS)$ under (2.4). Since also $\sum_{t=1}^{n-1} (\mu_t - \bar{\mu}) = \bar{\mu} - \mu_n = O(M^{1/2})$ we conclude that the denominator of (8.1) is

$$\begin{aligned} M + \sum_{t=2}^n h_{t-1}^2 x_{t-1}^2 + o_p(M^{1/2} S^{1/2}) + O_p(M^{1/2} S^{1/2}/n) + O_p(S/n) \\ = M + \sum_{t=2}^n h_{t-1}^2 x_{t-1}^2 + o_p(M + S). \end{aligned}$$

Now

$$\sum_{t=1}^{n-1} h_t^2 (x_t^2 - E x_t^2) = \sum_{j=0}^{\infty} \theta^{2j} \sum_{t=1}^{n-1} h_t^2 \eta_{t-j} + \sum_{j,k=0, j \neq k}^{\infty} \theta^j \theta^k \sum_{t=1}^{n-1} h_t^2 \varepsilon_{t-j} \varepsilon_{t-k}, \quad (8.2)$$

where $\eta_t = \varepsilon_t^2 - 1$. With $1(\cdot)$ denoting the indicator function, for some $K \in (0, \infty)$, write $\eta'_t = \eta_t 1(|\eta_t| \leq K) - E(\eta_t 1(|\eta_t| \leq K) | \mathcal{F}_{t-1})$, so $E(\eta'_t | \mathcal{F}_{t-1}) = 0$ a.s. It follows that the η'_t are uncorrelated with zero mean and variance bounded by $4K^2$. Thus uniformly in j , $E(\sum_{t=1}^{n-1} h_t^2 \eta'_{t-j})^2 \leq 4K^2 \sum_{t=1}^{n-1} h_t^4 = O(K^2 \max h_t^2 S) = o(K^2 S^2)$. On the other hand, with $\eta''_t = \eta_t 1(|\eta_t| > K) -$

$E(\eta_t 1(|\eta_t| > K) | \mathcal{F}_{t-1})$ we have $E|\sum_{t=1}^{n-1} h_t^2 \eta_{t-j}''| \leq 2 \max E|\eta_t''| S = o(S)$ as $K \rightarrow \infty$, by uniform integrability of the ε_t^2 . Since $\eta_t' + \eta_t'' = \eta_t - E(\eta_t | \mathcal{F}_{t-1}) = \eta_t$ by (2.3) it follows that the first term on the right of (8.2) is $o_p(S)$. Omitting the easy proof that the second term on the right of (8.2) has mean zero and variance $o(S^2)$, it follows that

$$S^{-1} \sum_{t=1}^{n-1} h_t^2 (x_t^2 - E x_t^2) \rightarrow_p 0. \quad (8.3)$$

It immediately follows that the denominator of (8.1) is $M + S/(1 - \theta^2) + o_p(S + M)$.

The numerator of (8.1) is

$$\begin{aligned} & \sum_{t=2}^n (\mu_t - \bar{\mu} + h_t x_t - \bar{h}x - \theta(\mu_{t-1} - \bar{\mu} + h_{t-1} x_{t-1} - \bar{h}x)) (\mu_{t-1} - \bar{\mu} + h_{t-1} x_{t-1} - \bar{h}x) \\ &= N - \theta M + \sum_{t=2}^n (h_t x_t - \bar{h}x - \theta(h_{t-1} x_{t-1} - \bar{h}x)) (h_{t-1} x_{t-1} - \bar{h}x) \end{aligned} \quad (8.4)$$

$$+ \sum_{t=2}^n (\mu_t - \bar{\mu} - \theta(\mu_{t-1} - \bar{\mu})) (h_{t-1} x_{t-1} - \bar{h}x) \quad (8.5)$$

$$+ \sum_{t=2}^n (h_t x_t - \bar{h}x - \theta(h_{t-1} x_{t-1} - \bar{h}x)) (\mu_{t-1} - \bar{\mu}). \quad (8.6)$$

Both terms (8.5) and (8.6) have mean zero and variance $O(W) = o(MS)$, while, since $h_t x_t - \theta h_{t-1} x_{t-1} = h_t(x_t - \theta x_{t-1}) + \theta(h_t - h_{t-1})x_{t-1}$, the sum in (8.4) is

$$\sum_{t=2}^n h_t h_{t-1} \varepsilon_t x_{t-1} + \theta \sum_{t=2}^n (h_t - h_{t-1}) h_{t-1} x_{t-1}^2 + O_p(S/n). \quad (8.7)$$

For the second term in (8.7) we have $S^{-1} \sum_{t=2}^n (h_t - h_{t-1}) h_{t-1} (x_{t-1}^2 - E x_{t-1}^2) \rightarrow_p 0$, in the same way as (8.3), while the first term is $o_p(S)$ because it has mean zero and variance $O(V) = O(\max h_t^2 S) = o(S^2)$. Thus the numerator of (8.1) is $N - \theta M + \theta T/(1 - \theta^2) + o_p(S + M)$. The proof is complete. \square

PROOF OF THEOREM 2.2. From the proof of Theorem 2.1 we can write

$$\begin{aligned} & \frac{(1 - \theta^2) M + S}{V^{1/2}} \left(\tilde{\theta} - \theta - \frac{(1 - \theta^2) (N - \theta M + \theta U) + O_p(W^{1/2} + M/n + S/n)}{(1 - \theta^2) M + S + o_p(M + S)} \right) \\ &= V^{-1/2} (1 - \theta^2) \sum_{t=2}^n h_t h_{t-1} \varepsilon_t x_{t-1} / (1 + o_p(1)). \end{aligned}$$

Given (2.6) it remains to prove that

$$(1 - \theta^2)^{1/2} V^{-1/2} \sum_{t=2}^n h_t h_{t-1} \varepsilon_t x_{t-1} \rightarrow_D \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty. \quad (8.8)$$

Since $E(\varepsilon_t x_{t-1} | \mathcal{F}_{t-1}) = 0$ a.s. we employ a martingale central limit theorem (e.g. Brown (1971)), which entails requires proving that, as $n \rightarrow \infty$,

$$\sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 | \mathcal{F}_{t-1}) / \sum_{t=2}^n h_t^2 h_{t-1}^2 E \varepsilon_t^2 x_{t-1}^2 \rightarrow_p 1, \quad (8.9)$$

and, for any $\epsilon > 0$,

$$V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1(|h_t h_{t-1} \varepsilon_t x_{t-1}| \geq \epsilon V^{1/2})) \rightarrow 0. \quad (8.10)$$

Applying (2.3), (8.9) follows because

$$V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 (x_{t-1}^2 - E x_{t-1}^2) \rightarrow_p 0,$$

whose proof is identical to that of (8.3), given (2.6). To establish the Lindeberg condition (8.10) we first bound the left side by

$$V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1_{\varepsilon t}) + V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1_{xt}) \quad (8.11)$$

for $1_{\varepsilon t} = 1(\varepsilon_t^2 \geq \epsilon V^{1/2} / \max h_t h_{t-1})$, $1_{xt} = 1(x_{t-1}^2 \geq \epsilon V^{1/2} / \max h_t h_{t-1})$. Now uniform integrability of the ε_t^2 implies the same of the x_t^2 because

$$x_t^2 = \sum_{j=0}^K \theta^{2j} \varepsilon_{t-j}^2 + \sum_{j=0}^K \theta^{2j} \varepsilon_{t-j}^2 + \sum_{j,k=0, j \neq k}^{\infty} \theta^{j+k} \varepsilon_{t-j} \varepsilon_{t-k},$$

where for any finite K the first term is uniformly integrable, the second has expectation that is $O(\theta^{2K}) = o(1)$ uniformly in t as $K \rightarrow \infty$, and the final one has finite variance. After applying (2.3) the second term in (8.11) is bounded by $\max E(x_{t-1}^2 1_{xt}) \rightarrow 0$ from (2.6). The first term in (8.11) equals, for any $K \in (0, \infty)$,

$$V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1(x_{t-1}^2 \leq K) 1_{\varepsilon t}) + V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1(x_{t-1}^2 > K) 1_{\varepsilon t}).$$

The first term is bounded by $K \max E(\varepsilon_t^2 1_{\varepsilon t}) \rightarrow 0$ as $n \rightarrow \infty$, the second is bounded by

$$V^{-1} \sum_{t=2}^n h_t^2 h_{t-1}^2 E(\varepsilon_t^2 x_{t-1}^2 1(x_{t-1}^2 > K)) \leq \max E(x_{t-1}^2 1(x_{t-1}^2 > K)) \rightarrow \infty \text{ as } K \rightarrow \infty,$$

after applying (2.3) once again. This completes the proof of (8.10), thence of (8.8) and thence of the theorem. \square

8.2 Proofs of Section 3.

PROOF OF THEOREM 3.1. Though only an initial consistency proof is needed, we begin by establishing a parametric rate. We prove

$$\hat{\theta} - \theta_0 = O_P(S^{-1/2}), \quad (8.12)$$

Set $u_n(\theta) = S \|\theta - \theta_0\|^2$, where $\|\cdot\|$ denotes the Euclidean norm. It suffices to show that for any $\epsilon > 0$ there exists $K = K_\epsilon$ such that

$$P\{u_n(\hat{\theta}) \geq K\} \leq \epsilon \quad (8.13)$$

for n sufficiently large. Denote $\Psi_K = \{\theta : \theta \in \Theta, u_n(\psi) \geq K\}$. Then,

$$\begin{aligned} P\{u_n(\hat{\theta}) \geq K\} &\leq P(\inf_{\theta \in \Psi_K} \{Q_n(\theta) - Q_n(\theta_0)\} \leq 0) \\ &= P(\inf_{\theta \in \Psi_K} \{u_n(\theta)^{-1}(Q_n(\theta) - Q_n(\theta_0))\} \leq 0). \end{aligned} \quad (8.14)$$

Write

$$Q_n(\theta) - Q_n(\theta_0) = \frac{1}{[n/2]} \left\{ U_n(\theta) + \frac{S}{n} T_n(\theta) \right\}, \quad (8.15)$$

where

$$U_n(\theta) = \sum_{j=1}^{[n/2]} \left\{ \frac{k_j(\theta_0)}{k_j(\theta)} - 1 \right\} \left\{ \frac{2\pi I_{y,j}}{k_j(\theta_0)} - \frac{S}{n} \right\}, \quad T_n(\theta) = \sum_{j=1}^{[n/2]} \left\{ \frac{k_j(\theta_0)}{k_j(\theta)} - 1 \right\}, \quad (8.16)$$

and $I_{y,j} = I_y(u_j)$, $k_j(\theta) = k(u_j; \theta)$. Since $\inf_{\theta} (h_1(\theta) + h_2(\theta)) \geq \inf_{\theta} h_1(\theta) - \sup h_2(\theta)$, (8.14) is bounded by

$$P\{u_n(\hat{\theta}) \geq K\} \leq P\left(\sup_{\theta \in \Psi_K} \left| \frac{U_n(\theta)}{u_n(\theta)} \right| \geq \inf_{\theta \in \Psi_K} \left| \frac{(S/n)T_n(\theta)}{u_n(\theta)} \right| \right). \quad (8.17)$$

It is shown in Lemma 8.2 below that as $n \rightarrow \infty$, $K \rightarrow \infty$,

$$\inf_{\theta \in \Psi_K} (u_n(\theta)^{-1}(S/n)T_n(\theta)) \geq c_1, \quad E \sup_{\theta \in \Psi_K} |u_n(\theta)^{-1/2}U_n(\theta)| \leq c_2,$$

where $c_1 > 0$, $c_2 > 0$ do not depend on K, n . Thus,

$$P\{u_n(\hat{\theta}) \geq K\} \leq P\left(\sup_{\theta \in \Psi_K} \left| \frac{U_n(\theta)}{u_n(\theta)^{1/2}K^{1/2}} \right| \geq c_1 \right), \quad (8.18)$$

because $u_n(\theta) \geq K$ for $\theta \in \Psi_K$. So, as $n \rightarrow \infty$, $K \rightarrow \infty$,

$$P\{u_n(\hat{\theta}) \geq K\} \leq (c_1 K^{1/2})^{-1} E \left[\sup_{\Psi_K} |u_n(\theta)^{-1/2}U_n(\theta)| \right] \leq (c_2/c_1) K^{-1/2} \rightarrow 0.$$

This proves (8.12).

To show the first claim in (3.6), write

$$Q_n(\hat{\theta}) - (S/n) = \{Q_n(\theta_0) - (S/n)\} + \{Q_n(\hat{\theta}) - Q_n(\theta_0)\} =: v_{n,1} + v_{n,2}.$$

We shall prove that

$$v_{n,k} = (S_n/n)O_P(S^{-1/2}), \quad k = 1, 2 \quad (8.19)$$

which implies (8.12). Since $f_x(u_j) = (2\pi)^{-1}k_j(\theta_0)$,

$$v_{n,1} = Q_n(\theta_0) - (S/n) = [n/2]^{-1} \sum_{j=1}^{[n/2]} \left(\frac{I_{y,j}}{f_x(u_j)} - (S/n) \right) = [n/2]^{-1} q_{n,[n/2]},$$

where $q_{n,[n/2]}$ is as in (8.52). By (8.54), $E|q_{n,[n/2]}| \leq CS^{1/2}$. Hence, $v_{n,1} = O_P(S^{1/2}/n) = (S/n)O_P(S^{-1/2})$ which proves (8.19) for $v_{n,1}$. In addition, it implies $Q_n(\theta_0) = O_P(S/n)$.

Smoothness assumptions on $k_\theta(u)$ imply $|k_j(\hat{\theta})^{-1} - k_j(\theta_0)^{-1}| \leq C\|\hat{\theta} - \theta_0\|k_j(\theta_0)^{-1}$. Thus,

$$|v_{n,2}| = \frac{2\pi}{[n/2]} \left| \sum_{j=1}^{[n/2]} (k_j(\hat{\theta})^{-1} - k_j(\theta_0)^{-1}) I_{y,j} \right| \leq C\|\hat{\theta} - \theta_0\|Q_n(\theta_0) = O_P(S^{-1/2})(S/n). \quad (8.20)$$

Hence, $v_{n,2}$ satisfies (8.19). This completes the proof of (8.12).

Next we prove the second claim in (3.6). By the mean value theorem,

$$\nabla_\theta Q_{n,y}(\hat{\theta}) - \nabla_\theta Q_{n,y}(\theta_0) = \nabla_\theta \nabla'_\theta Q_{n,y}(\theta^*)(\hat{\theta} - \theta_0)$$

for some θ^* such that $\|\theta^* - \hat{\theta}\| \leq \|\hat{\theta} - \theta_0\|$; strictly speaking a different value is generally needed for each element. With probability tending to 1, $\nabla_\theta Q_{n,y}(\hat{\theta}) = 0$. We will show that

$$(2\pi) \frac{[n/2]}{2\pi} \nabla_\theta \nabla'_\theta Q_{n,y}(\theta^*) = S \Omega(1 + o_P(1)), \quad (8.21)$$

$$(2\pi)(V^*)^{-1/2} \frac{[n/2]}{2\pi} \nabla_\theta Q_{n,y}(\theta_0) \rightarrow_D \mathcal{N}(0, \Omega). \quad (8.22)$$

Since $S = O(V^*)$ by (3.5)(d), and $\hat{\theta} - \theta_0 = O_P(S^{-1/2})$, this implies $\Omega(S/V^{*1/2})(\hat{\theta} - \theta_0) \rightarrow_D \mathcal{N}(0, \Omega^{-1})$ which proves (3.6).

To show (8.21), it suffices to prove that

$$nS^{-1} \|\nabla_\theta \nabla'_\theta Q_n(\theta^*) - (S/[n/2])\Omega\| = o_P(1). \quad (8.23)$$

We have

$$\begin{aligned} \|\nabla_\theta \nabla'_\theta Q_n(\theta^*) - (S/[n/2])\Omega\| &\leq \|\nabla_\theta \nabla'_\theta (Q_n(\theta^*) - Q_n(\theta_0))\| \\ &+ \|\nabla_\theta \nabla'_\theta Q_n(\theta_0) - E\nabla_\theta \nabla'_\theta Q_n(\theta_0)\| + \|E\nabla_\theta \nabla'_\theta Q_n(\theta_0) - (S/[n/2])\Omega\| =: j_{n,1} + j_{n,2} + j_{n,3}. \end{aligned}$$

It remains to show that for $k = 1, 2, 3$,

$$j_{n,k} = o_P(S/n). \quad (8.24)$$

By Assumption A, $j_{n,1} \leq C\|\theta^* - \theta_0\|J_n$, $J_n = n^{-1} \sum_{j=1}^{[n/2]} I_{hx}(u_j)$. By Theorem 8.1(i), $EJ_n \leq CS/n$ which implies $J_n = O_P(S/n)$. Since $\|\theta^* - \theta_0\| = O_P(S^{-1/2})$ this proves (8.24) for $j_{n,1}$. By Assumption B and Theorem 8.1(ii), $j_{n,2} = O_P(V^{*1/2}n^{-1}) = o_P(Sn^{-1})$ which proves (8.24) for

$k = 2$.

To verify (8.24) for $j_{n,3}$, notice that by Theorem 8.1(i),

$$E\nabla_{\theta}\nabla'_{\theta}Q_n(\theta_0) = \frac{2\pi}{[n/2]} \frac{1}{2\pi} \int_0^{\pi} \nabla_{\theta}\nabla'_{\theta}k(u; \theta_0)^{-1} f_x(u) du S + o(S/n) = \frac{S}{[n/2]} \Omega + o\left(\frac{S}{n}\right),$$

noting that from (3.1), standard arguments give

$$\begin{aligned} \int_0^{\pi} (\nabla_{\theta}\nabla'_{\theta}k^{-1}(u; \theta_0))k(u; \theta_0) du &= \int_0^{\pi} k(u; \theta_0)^{-2} \nabla_{\theta}k(u; \theta_0) \nabla'_{\theta}k(u; \theta_0) du \\ &= \int_0^{\pi} \nabla_{\theta} \log k(u; \theta_0) \nabla'_{\theta} \log k(u; \theta_0) du = 2\pi \Omega. \end{aligned}$$

Thus, $j_{n,3} = o(S/n)$ which completes the proof of (8.24) and (8.21).

Next we prove (8.22). Write

$$(2\pi)(V^*)^{-1/2} \frac{[n/2]}{2\pi} \nabla_{\theta} Q_{n,y}(\theta_0) = (s_{1,n}, \dots, s_{p,n})', \quad s_{j,n} := \frac{2\pi}{V^{*1/2}} \sum_{j=1}^{[n/2]} (d/d\theta_j) k^{-1}(u_j, \theta_0) I_y(u_j).$$

Condition (3.1) implies that $\int_0^{\pi} (d/d\theta_j) k^{-1}(u, \theta_0) f_x(u) du = 0$, $j = 1, \dots, p$. Hence, by Theorem 8.1 (ii-iii), for $j = 1, \dots, p$, $s_{j,n} \rightarrow_D \mathcal{N}(0, v_j^2)$ where $v_j^2 = (2\pi) \int_0^{\pi} ((d/d\theta_j) k^{-1}(u, \theta_0) f_x(u))^2 du = (4\pi)^{-1} \int_{-\pi}^{\pi} ((d/d\theta_j) \log k(u, \theta))^2 du$. In turn, (8.22) can be verified using the Cramer-Wold device. This completes the proof of the theorem. \square

The next proposition provides a useful criterion for the proof of Theorem 3.2.

Proposition 8.1 *Let Assumptions A and B hold. Let $\tilde{\theta} = \operatorname{argmin}_{\theta \in \Theta} \tilde{Q}_n(\theta)$ where*

$$\tilde{Q}_n(\theta) := Q_{n,hx}(\theta) + \frac{2\pi}{[n/2]} c_n(\theta)$$

and $c_n(\theta)$ is such that

$$E[\sup_{\theta \in \Theta} |c_n(\theta)|] \leq CS^{1/2}, \tag{8.25}$$

$$E[\sup_{\theta \in \Theta} |c_n(\theta) - c_n(\theta_0)|/S^{1/2} \|\theta - \theta_0\|] \leq C. \tag{8.26}$$

Then,

$$\tilde{\theta} - \theta_0 = O_P(S^{-1/2}), \quad \tilde{Q}_n(\tilde{\theta}) = \frac{S}{n} (1 + O_P(S^{-1/2})). \tag{8.27}$$

In addition, let $\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} Q_{n,hx}(\theta)$. If $E[\sup_{\theta \in \Theta} \|\nabla_{\theta} c_n(\theta)\|] = o(S^{1/2})$, then

$$\tilde{\theta} - \hat{\theta} = o_P(S^{-1/2}). \tag{8.28}$$

PROOF OF PROPOSITION 8.1. To prove the first claim in (8.27), using (8.15), write

$$\begin{aligned}\tilde{Q}_n(\theta) - \tilde{Q}_n(\theta_0) &= Q_{n,hx}(\theta) - Q_{n,hx}(\theta_0) + \frac{2\pi}{[n/2]}(c_n(\theta) - c_n(\theta_0)) \\ &= \frac{1}{[n/2]} \left\{ U_n(\theta) + 2\pi(c_n(\theta) - c_n(\theta_0)) + \frac{S}{n}T_n(\theta) \right\}.\end{aligned}$$

The same argument as in the proof of Theorem 3.1 implies that $P\{u_n(\hat{\theta}) \geq K\} \leq p_{n,1} + p_{n,2}$, where

$$p_{n,1} := P\left(\sup_{\theta \in \Psi_K} \left| \frac{U_n(\theta)}{u_n(\theta)^{1/2}K^{1/2}} \right| \geq c_1/2\right), \quad p_{n,2} := P\left(\sup_{\theta \in \Psi_K} 2\pi \left| \frac{c_n(\theta) - c_n(\theta_0)}{u_n(\theta)^{1/2}K^{1/2}} \right| \geq c_1/2\right).$$

We showed in the proof of Theorem 3.1 that $p_{n,1} \leq (2c_1/c_2)K^{-1/2}$. By (8.26),

$$p_{n,2} \leq CK^{-1/2}E[\sup_{\theta \in \Theta} |c_n(\theta) - c_n(\theta_0)|/S^{1/2}\|\theta - \theta_0\|] \rightarrow 0 \text{ as } K \rightarrow \infty.$$

This implies $P\{u_n(\hat{\theta}) \geq K\} \rightarrow 0$ as $n, K \rightarrow \infty$.

To verify the second claim in (8.27), write

$$\tilde{Q}_n(\tilde{\theta}) - (S/n) = (Q_{n,hx}(\tilde{\theta}) - (S/n)) + \frac{2\pi}{[n/2]}c_n(\tilde{\theta}).$$

Since $\|\tilde{\theta} - \theta_0\| = O_P(S^{-1/2})$, then $|Q_{n,hx}(\tilde{\theta}) - (S/n)| = (S/n)O_P(S^{-1/2})$, see the proof of the first claim in (3.6). In turn, by (8.25), $n^{-1}c_n(\tilde{\theta}) = (S/n)O_P(S^{-1/2})$, which proves the second relation in (8.27).

It remains to prove (8.28). With probability tending to 1, $\nabla_{\theta}\tilde{Q}_n(\tilde{\theta}) = 0$, $\nabla_{\theta}Q_{n,hx}(\hat{\theta}) = 0$, and

$$\begin{aligned}0 &= \nabla_{\theta}\tilde{Q}_n(\tilde{\theta}) - \nabla_{\theta}Q_{n,hx}(\hat{\theta}) + \frac{2\pi}{[n/2]}\nabla_{\theta}c_n(\theta) = Q_{n,hx}(\theta^*)(\tilde{\theta} - \hat{\theta}) + \frac{2\pi}{[n/2]}\nabla_{\theta}c_n(\theta) \\ &= \frac{S}{[n/2]}\Omega(\tilde{\theta} - \hat{\theta}) + r_n(\theta), \quad r_n(\theta) = (\nabla_{\theta}\nabla'_{\theta}Q_{n,hx}(\theta^*) - \frac{S}{[n/2]}\Omega)(\tilde{\theta} - \hat{\theta}) + \frac{2\pi}{[n/2]}\nabla_{\theta}c_n(\theta).\end{aligned}$$

By Assumption A, Ω is positive definite. Hence, by the mean value theorem, for some θ^* such that $\|\theta^* - \hat{\theta}\| \leq \|\tilde{\theta} - \hat{\theta}\|$,

$$\|\tilde{\theta} - \hat{\theta}\| \leq \frac{n}{S}\|\Omega^{-1}\| \sup_{\theta \in \Theta} \|r_n(\theta)\| \leq C \frac{n}{S} \{ \|\tilde{\theta} - \hat{\theta}\| \|\nabla_{\theta}\nabla'_{\theta}Q_n(\theta^*) - \frac{S}{[n/2]}\Omega\| + S^{-1} \sup_{\theta \in \Theta} \|\nabla_{\theta}c_n(\theta)\| \}.$$

By (8.27) and (8.12), $\|\tilde{\theta} - \hat{\theta}\| \leq \|\tilde{\theta} - \theta_0\| + \|\hat{\theta} - \theta_0\| = O_P(S^{-1/2})$, by (8.23) $(n/S)\|\nabla_{\theta}\nabla'_{\theta}Q_n(\theta^*) - \frac{S}{[n/2]}\Omega\| = o_P(1)$, and by the assumptions of the proposition, $S^{-1} \sup_{\theta \in \Theta} \|\nabla_{\theta}c_n(\theta)\| = o_P(S^{-1/2})$. Hence, $\|\tilde{\theta} - \hat{\theta}\| = o_P(S^{-1/2})$ which proves (8.28) and completes the proof of the proposition. \square

PROOF OF THEOREM 3.2 For $k = 1, 2$, we can write

$$Q_{n,\nu}^{(k)}(\theta) = Q_{n,hx}(\theta) + \frac{2\pi}{[n/2]}c_n^{(k)}(\theta), \quad c_n^{(k)}(\theta) := \frac{[n/2]}{2\pi}(Q_{n,\nu}^{(k)}(\theta) - Q_{n,hx}(\theta)). \quad (8.29)$$

By Proposition 8.1, to prove

$$\widehat{\theta}^{(k)} - \theta_0 = O_P(S^{-1/2}) \quad (8.30)$$

it suffices to verify that $c_n^{(1)}(\theta)$ in (i) and $c_n^{(2)}(\theta)$ in (ii) satisfy conditions (8.25) and (8.26).

We have

$$\begin{aligned} c_n^{(1)}(\theta) &= \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} \frac{I_y(u_j) - I_{hx}(u_j)}{k(u_j; \theta)} - \sum_{j=1}^{\nu} \frac{I_{hx}(u_j)}{k(u_j; \theta)}, \\ c_n^{(2)}(\theta) &= c_n^{(1)}(\theta) + \tilde{c}_n(\theta), \quad \tilde{c}_n(\theta) := Q_{n,\nu}^{(1)}(\widehat{\theta}^{(1)}) \sum_{j=1}^{\nu} \frac{1}{2\pi} \frac{k(u_j; \widehat{\theta}^{(1)})}{k(u_j; \theta)}. \end{aligned} \quad (8.31)$$

Since for $\theta \in \Theta$, $k^{-1}(u_j; \theta) \leq C$ and $|k^{-1}(u_j; \theta) - k^{-1}(u_j; \theta_0)| \leq C\|\theta - \theta_0\|$, we obtain

$$\begin{aligned} |c_n^{(1)}(\theta)| &\leq C(J_{n,1} + J_{n,2}), \quad J_{n,1} := \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} |I_y(u_j) - I_{hx}(u_j)|, \quad J_{n,2} := \sum_{j=1}^{\nu} |I_{hx}(u_j)|, \\ |c_n^{(1)}(\theta) - c_n^{(1)}(\theta_0)| &\leq C(J_{n,1} + J_{n,2})\|\theta - \theta_0\|. \end{aligned}$$

Similarly,

$$\begin{aligned} |\tilde{c}_n(\theta)| &\leq C\nu Q_{n,\nu}^{(1)}(\widehat{\theta}^{(1)}) \leq CJ_{n,3}, \quad J_{n,3} := \nu n^{-1} \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} I_y(u_j); \\ |\tilde{c}_n(\theta) - \tilde{c}_n(\theta_0)| &\leq CJ_{n,3}\|\theta - \theta_0\|. \end{aligned}$$

We will show that for $k = 1, 2, 3$,

$$EJ_{n,k} = o(S^{1/2}), \quad n \rightarrow \infty, \quad (8.32)$$

which proves (8.25) and (8.26).

We start with $EJ_{n,1}$.

i) Let $y_t = \mu + h_t x_t$. The equality

$$\sum_{t=1}^n e^{itu_j} = 0, \quad j = 1, \dots, n-1 \quad (8.33)$$

implies $I_y(u_j) = I_{hx}(u_j)$ and hence, $J_{n,1} = 0$.

ii) Let $y_t = \mu_t + h_t x_t$. Denote $I_\mu(u_j) = (2\pi)^{-1} \left| \sum_{t=1}^n e^{itu_j} \mu_t \right|$. Since

$$|I_y(u_j) - I_{hx}(u_j)| = |I_{\mu+hx}(u_j) - I_{hx}(u_j)| \leq 2(I_\mu(u_j)I_{hx}(u_j))^{1/2} + I_\mu(u_j),$$

then $E|I_y(u_j) - I_{hx}(u_j)| \leq C(I_\mu^{1/2}(u_j)(S/n)^{1/2} + I_\mu(u_j))$. Recall that $I_\mu(u_j) = (2\pi/n)|D_{\mu,n}(u_j)|^2$ where $D_{\mu,n}(u_j)$ is as in Lemma 8.3(iii).

Suppose that μ_t and ν satisfy (3.12). Then by Lemma 8.3(iii), $|D_{\mu,n}(u_j)| \leq Cn\phi^j$, and

$$EJ_{n,1} \leq C \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} (n^{1/2}\phi^j(S/n)^{1/2} + n\phi^{2j}) \leq C(S^{1/2}\phi^\nu + n\phi^{2\nu}) = o(S^{1/2}) \quad (8.34)$$

because by (3.12), $n\phi^{2\nu} = o(S^{1/2})$ and $\phi^\nu = o(1)$.

Let μ_t and ν satisfy (3.13). Then by Lemma 8.3(iii), $|D_{\mu,n}(u_j)| \leq Cnj^{-\gamma}$. So,

$$EJ_{n,1} \leq C \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} (n^{1/2}j^{-\gamma}(S/n)^{1/2} + nj^{-2\gamma}) \leq C(S^{1/2}\nu^{-\gamma+1} + n\nu^{-2\gamma+1}) = o(S^{1/2}) \quad (8.35)$$

because by (3.13), $n\nu^{-2\gamma+1} = o(S^{1/2})$ and $\nu \rightarrow \infty$. This proves (8.32) for $EJ_{n,1}$.

Next we prove (8.32) for $EJ_{n,2}$. By Lemma 8.4(i)-(ii) and (3.5)(b), $EI_{hx}(u_j) \leq C((S/n) + n^{-1} \log n \Delta_{h,n}^2) \leq C(S/n)$. Hence, $EJ_{n,2} \leq C\nu(S/n) = o(S^{1/2})$, by (3.8).

To prove (8.32) for $EJ_{n,3}$, notice that $J_{n,3} \leq \nu n^{-1}(J_{n,1} + \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} I_{hx}(u_j))$. We showed, that $EJ_{n,1} = o(S^{1/2})$. Thus, using $I_{hx}(u_j) \leq C(S/n)$, we obtain

$$EJ_{n,3} \leq C\nu n^{-1}(S^{1/2} + \sum_{j=\nu+1}^{\lfloor n/2 \rfloor} (S/n)) \leq C((\nu/n)S^{1/2} + \nu(S/n)) = o(S^{1/2}),$$

by (3.8). This verifies (8.32) for $EJ_{n,3}$ and completes the proof of (8.30) in the parts (i) and (ii) of the theorem.

In addition, (8.27) implies the first claim in (3.14).

To show the second claim in (3.14), write

$$\widehat{\theta}^{(k)} - \theta_0 = \widehat{\theta} - \theta_0 + (\widehat{\theta}^{(k)} - \widehat{\theta}), \quad k = 1, 2$$

where $\widehat{\theta} = \operatorname{argmin}_{\theta \in \Theta} Q_{n,hx}(\theta)$. By Theorem 3.1, $(S/V^{*1/2})(\widehat{\theta}^{(k)} - \theta_0) \rightarrow \mathcal{N}(0, \Omega^{-1})$. We will show that

$$\widehat{\theta}^{(k)} - \widehat{\theta} = o_P(S^{-1/2}) \quad (8.36)$$

which implies $(S/V^{*1/2})(\widehat{\theta}^{(k)} - \widehat{\theta}) = o_P((S/V^{*1/2})S^{-1/2}) = o_P(1)$ by (3.5)(d). The latter proves (3.14).

By Proposition 8.1, to verify (8.36) it suffices to show that $c_n^{(k)}(\theta)$, $k = 1, 2$ in (8.29) have property $E[\sup_{\theta \in \Theta} \|\nabla_{\theta} c_n^{(k)}(\theta)\|] = o(S^{1/2})$. The latter follows from (8.32), noting that under Assumption A, $\sup_{\theta \in \Theta} \|\nabla_{\theta} c_n^{(k)}(\theta)\| \leq C(J_{n,1} + J_{n,2} + J_{n,3})$. This proves (3.14) and completes the proof of the theorem. \square

8.3 Proofs of Section 4.

The following lemma is used to prove Theorem 4.1.

Lemma 8.1 (i) Under the assumptions of Theorem 4.1(i), $\hat{\mu}_t^*$ given in (4.6) satisfies

$$\hat{\mu}_t^* - \mu_t^* = O_p(H^{-1/2} + Hn^{-1}). \quad (8.37)$$

(ii) In addition, if $H = o(n^{2/3})$ and $\min_t h_t \geq c > 0$, then

$$\frac{K_t}{K_{t,2}^{1/2}} \frac{1}{h_t} (\hat{\mu}_t^* - \mu_t^*) \rightarrow_D \mathcal{N}(0, 1). \quad (8.38)$$

(iii) In addition, \widehat{h}_t^2 satisfies (4.9) in Theorem 4.1(i) and (4.11) in Theorem 4.1(ii).

PROOF OF LEMMA 8.1. (i) We start with the proof of (8.37). By (4.5), $z_j - \mu_t^* = h_j \varepsilon_j + e_{tj}$ where $e_{tj} := (\mu_j^* - \mu_t^*) + \hat{\eta}_j$. Write

$$\hat{\mu}_t^* - \mu_t^* := K_t^{-1} \sum_{j=1}^n k_{tj} (z_j - \mu_t^*) = s_{n,t} + r_{n,t} \quad (8.39)$$

where $s_{n,t} := K_t^{-1} \sum_{j=1}^n k_{tj} h_j \varepsilon_j$, $r_{n,t} = K_t^{-1} \sum_{j=1}^n k_{tj} e_{tj}$. We shall show that

$$(K_t/K_{t,2}^{1/2}) s_{n,t} \rightarrow_D \mathcal{N}(0, h_t^2), \quad (8.40)$$

$$|r_{n,t}| \leq O_P(Hn^{-1} + n^{-1/2}) \xi_{nt} \quad (8.41)$$

where $O_P(Hn^{-1} + n^{-1/2})$ does not depend on t and $E\xi_{nt}^2 \leq C$ for all t and n . Since $(K_t/K_{t,2}^{1/2}) = sH^{1/2}(1+o(1))$ where $s = \int K(u)du / (\int K^2(u)du)^{1/2} > 0$, (8.40) and (8.41) imply (8.37) and (8.38).

To prove (8.40), write $(K_t/K_{t,2}^{1/2})S_{nt} = \sum_{j=1}^n b_{nj} \varepsilon_j$ where $b_{nj} := K_{t,2}^{-1/2} k_{tj} h_j$. Notice that

$$\sum_{j=1}^n b_{nj} \rightarrow 1, \quad \sum_{j=1}^{n-1} |b_{nj} - b_{n,j+1}| + |b_{nn}| = o(1), \quad (8.42)$$

see e.g. the proof Lemma 12.2.1 in Giraitis, Koul and Surgailis (2012). By Proposition 4.3.1. in Giraitis Koul and Surgailis (2012), (8.42) implies (8.40).

To show (8.41), notice that by (4.2), $|\mu_j^* - \mu_t^*| = |\delta_n| |\mu_j - \mu_t| \leq C(|t-j|/n) = O(H/n)(|t-j|/H)$, while

$$|\hat{\eta}_j| \leq C\{n^{-1} + (n^{-1} + a_n)v_j\}, \quad a_n = \sum_{k=1}^p |\theta_{k0} - \hat{\theta}_k|, \quad v_j = \sum_{k=1}^p |x_{j-k}| \quad (8.43)$$

where by (4.3), $a_n = O_P(n^{-1/2})$ and $E v_j^2 \leq C$. Hence, $e_{tj} = O_P(Hn^{-1} + n^{-1/2}) e_{tj}^*$ with $e_{tj}^* = (|t-j|/H) + v_j + 1$. Thus,

$$|r_{n,t}| \leq O_P(Hn^{-1} + n^{-1/2}) \xi_{nt}, \quad \xi_{nt} := K_t^{-1} \sum_{j=1}^n k_{tj} e_{tj}^* \quad (8.44)$$

where $E\xi_{nt}^2 \leq C(K_t^{-1} \sum_{j=1}^n k_{tj} (|t-j|/H + 1))^2 \leq C$ for all t and n , which proves (8.41).

(iii) Recall that $E\varepsilon_t^2 = 1$. Then,

$$\widehat{h}_t^2 - h_t^2 = K_t^{-1} \sum_{j=1}^n k_{tj} \{(z_j - \hat{\mu}_j^*)^2 - h_t^2\} = s_{h^2;nt} + r'_{nt}, \quad (8.45)$$

where $s_{h^2;nt} = K_t^{-1} \sum_{j=1}^n k_{tj} h_j^2 (\varepsilon_j^2 - 1)$ and $r'_{nt} = K_t^{-1} \sum_{j=1}^n k_{tj} \{(z_j - \hat{\mu}_j^*)^2 - h_j^2 \varepsilon_j^2\}$.

The same argument as in the proof of (8.40) implies that

$$(K_t/K_{t,2}^{1/2}) s_{h^2;nt} \rightarrow_D \mathcal{N}(0, h_t^4 \text{var}(\varepsilon_1^2)). \quad (8.46)$$

We will show that

$$r'_{nt} = o_P(H^{-1/2}) + O_P(Hn^{-1}), \quad (8.47)$$

which together with (8.46) proves the part (iii) of this lemma.

To prove (8.47), write $(z_j - \hat{\mu}_j^*)^2 - h_j^2 \varepsilon_j^2 = (z_j - \hat{\mu}_j^* - h_j \varepsilon_j)^2 + 2(z_j - \hat{\mu}_j^* - h_j \varepsilon_j) h_j \varepsilon_j$. Then, by (4.5) and (8.39),

$$(z_j - \hat{\mu}_j^*)^2 - h_j^2 \varepsilon_j^2 = (-s_{n,j} - r_{n,j} + \hat{\eta}_j)^2 + 2(-s_{n,j} - r_{n,j} + \hat{\eta}_j) h_j \varepsilon_j = -2s_{n,j} h_j \varepsilon_j + \eta_j$$

where $\eta_j = (-s_{n,j} - r_{n,j} + \hat{\eta}_j)^2 + 2(-r_{n,j} + \hat{\eta}_j) h_j \varepsilon_j$. Hence,

$$r'_{nt} = r'_{nt,1} - 2r'_{nt,2}, \quad r'_{nt,1} = K_t^{-1} \sum_{j=1}^n k_{tj} \eta_j, \quad r'_{nt,2} = K_t^{-1} \sum_{j=1}^n k_{tj} s_{n,j} h_j \varepsilon_j.$$

Using (8.43) and (8.44), we can bound $\eta_j \leq C s_{n,j}^2 + O_P(Hn^{-1} + n^{-1/2}) \tilde{\xi}_{nj}$ with $E\tilde{\xi}_{nj}^2 \leq C$ and $E s_{n,j}^2 \leq C K_j^{-1} \leq C H^{-1}$ where C does not depend on j, n . Clearly, this implies $r'_{nt,1} = O_P(Hn^{-1}) + o_P(H^{-1/2})$.

To evaluate $r'_{nt,2}$, write $s_{n,j} = K_j^{-1} k_{jj} \varepsilon_j + s'_{n,j}$, where $s'_{n,j} = K_j^{-1} \sum_{l=1:l \neq j}^n k_{jl} \varepsilon_l$. So,

$$\begin{aligned} E[r'_{nt,2}^2] &\leq 2E(K_t^{-1} \sum_{j=1}^n k_{tj} s'_{n,j} \varepsilon_j)^2 + 2E(K_t^{-1} \sum_{j=1}^n k_{tj} \{K_j^{-1} k_{jj} \varepsilon_j\})^2 \\ &\leq C K_t^{-2} \sum_{j=1}^n k_{tj}^2 E(s'_{n,j})^2 E\varepsilon_j^2 + C(K_t^{-2} \sum_{j=1}^n k_{tj}^2 k_{jj}^2 K_j^{-2} E\varepsilon_j^2) \leq C H^{-2} \end{aligned}$$

which implies that $r'_{nt,2} = o_P(H^{-1/2})$. This completes the proof of (8.47) and the lemma. \square

PROOF OF THEOREM 4.1. Set $\delta_n = 1 - \sum_{k=1}^p \hat{\theta}_k$, $\delta = 1 - \sum_{k=1}^p \theta_{k0}$. By (4.7),

$$\hat{\mu}_t - \mu_t = \frac{\hat{\mu}_t^*}{\delta_n} - \frac{\mu_t^*}{\delta} = \frac{\hat{\mu}_t^* - \mu_t^*}{\delta} + \frac{\hat{\mu}_t^* - \mu_t^*}{\delta} \left(\frac{\delta}{\delta_n} - 1 \right) + \frac{\mu_t^*}{\delta} \left(\frac{\delta}{\delta_n} - 1 \right).$$

By stationarity of (x_t) , $\delta \neq 0$, and by (4.3), $\delta/\delta_n - 1 = O_P(n^{-1/2})$. Hence (4.9) for $\hat{\mu}_t$ follows from (8.37), while (4.10) follows from (8.38) of Lemma 8.1, noting that $K_t/K_{t,2}^{1/2} = O(H^{1/2}) = o(n^{1/2})$. Finally, (4.9) and (4.11) for \widehat{h}_t^2 are shown in Lemma 8.1(iii). \square

PROOF OF COROLLARY 4.1. By (4.4) and (4.5),

$$\hat{\varepsilon}_t - \varepsilon_t = \frac{z_t - \hat{\mu}_t^*}{\hat{h}_n} - \varepsilon_t = \frac{h_t \varepsilon_t + \hat{\eta}_t + \mu_t^* - \hat{\mu}_t^*}{\hat{h}_n} - \varepsilon_t = \frac{(h_t - \hat{h}_t) \varepsilon_t + \hat{\eta}_t + (\mu_t^* - \hat{\mu}_t^*)}{\hat{h}_n}.$$

By (8.37), $\mu_t^* - \hat{\mu}_t^* = O_P(H^{-1/2} + Hn^{-1})$ while (8.43) implies $\hat{\eta}_t = O_P(n^{-1/2}) = o_P(H^{-1/2})$. It is easy to see that (4.9) implies $\hat{h}_t^{-1} = O_P(1)$ and $h_t - \hat{h}_t = O_P(H^{-1/2} + Hn^{-1})$. Moreover, $\varepsilon_t = O_P(1)$. This yields $\hat{\varepsilon}_t - \varepsilon_t = O_P(H^{-1/2} + Hn^{-1})$ which completes the proof of (4.13). \square

8.4 Auxiliary results

Lemma 8.2 *Let $T_n(\theta)$ and $U_n(\theta)$ be given by (8.16). Suppose that assumptions of Theorem 3.1 hold. Then, as $n \rightarrow \infty$ and $K \rightarrow \infty$,*

$$\inf_{\theta \in \Psi_K} \{(S/n)T_n(\theta)/u_n(\theta)\} \geq c_1, \quad (8.48)$$

$$E \sup_{\theta \in \Psi_K} \{U_n(\theta)/u_n(\theta)^{1/2}\} \leq c_2 \quad (8.49)$$

where constants $c_1 > 0$ and $c_2 > 0$ do not depend on n and K .

PROOF. We start with the proof of (8.48). Set $u_n^*(\theta) = n\|\theta - \theta_0\|^2$. Using similar argument as in the proof of Lemma 7.1 in Giraitis, Hidalgo and Robinson (2001), it follows that

$$\inf_{\theta \in \Theta: |u_n^*(\theta)| \geq K} \{T_n(\theta)/u_n^*(\theta)\} \geq c. \quad (8.50)$$

Observe that $(S/n)T_n(\theta)/u_n(\theta) = T_n(\theta)/u_n^*(\theta)$. Moreover, $S \leq cn$, and therefore, $\Psi_K = \{\theta \in \Theta : u_n(\theta) \geq K\} \subset \{\theta \in \Theta : u_n^*(\theta) \geq K/c\}$. Hence, (8.50) implies (8.48).

Proof of (8.49). Write $U_n(\theta)$ in (8.16) as $U_n(\theta) = \sum_{j=1}^{[n/2]} b_{j,\theta} v_{n,j}$ where

$$b_{j,\theta} = \frac{k_j(\theta_0)}{k_j(\theta)} - 1, \quad v_{n,j} = \frac{2\pi I_{y,j}}{k_j(\theta_0)} - \frac{S}{n} = \frac{I_{y,j}}{f_x(u_j)} - \frac{S}{n}. \quad (8.51)$$

By summation by parts,

$$|U_n(\theta)| \leq \sum_{l=1}^{[n/2]-1} |b_{l,\theta} - b_{l+1,\theta}| |q_{n,l}| + b_{[n/2],\theta} |q_{n,[n/2]}|, \quad q_{n,l} = \sum_{j=1}^l v_{n,j}. \quad (8.52)$$

Smoothness properties of $k(u; \theta)$ imply that $|b_{j,\theta} - b_{j+1,\theta_0}| \leq C\|\theta - \theta_0\|n^{-1}$, $|b_{j,\theta}| \leq C\|\theta - \theta_0\|$ where C does not depend on j and $\theta \in \Theta$. Hence,

$$|U_n(\theta)| \leq C\|\theta - \theta_0\| (n^{-1} \sum_{l=1}^{[n/2]-1} |q_{n,l}| + |q_{n,[n/2]}|), \quad (8.53)$$

where $q_{n,l}$ does not depend on θ . We will show that

$$E|q_{n,l}| \leq CS^{1/2}, \quad l = 1, \dots, [n/2] \quad (8.54)$$

where C does not depend on l and n . Indeed,

$$q_{n,l} = \sum_{j=1}^l f_{x,j}^{-1}(I_y(u_j) - EI_y(u_j)) + \sum_{j=1}^l f_{x,j}^{-1}(EI_y(u_j) - \frac{S}{n}f_{x,j}) =: q_{n,1;l} + q_{n,2;l}.$$

By Theorem 8.1(ii), $E|q_{n,1;l}| \leq \text{var}(q_{n,1;l})^{1/2} \leq CS^{1/2}$. On the other hand, by Lemma 8.4(i)-(ii), assumption $f_{x,j}^{-1} \leq C$ and (3.5),

$$|q_{n,2;l}| \leq C \sum_{j=1}^l n^{-1} \log n \Delta_{h,n}^2 \leq C \log n \Delta_{h,n}^2 \leq CS^{1/2},$$

which implies (8.54). Thus,

$$E\left[\sup_{\theta \in \Psi_K} |U_n(\theta)|/u_n(\theta)^{1/2}\right] \leq S^{-1/2} \left(n^{-1} \sum_{l=1}^{[n/2]-1} E|q_{n,l}| + E|q_{n,[n/2]}\right) \leq C$$

which proves (8.49). □

Recall notation $D_{h,n}(u) := \sum_{t=1}^n e^{itu} h_t$, $|u| \leq \pi$.

Lemma 8.3 *Let Assumption B holds. Then*

- (i) $\sup_{|u| \leq \pi} |D_{h,n}(u)u| \leq C\Delta_{h,n}$,
- (ii) $\int_{\Pi} |D_{h,n}(u)| du \leq C \log n \Delta_{h,n}$,
- (iii) $|D_{\mu,n}(u_j)| \leq \begin{cases} Cn\phi^j & \text{if (3.12) holds,} \\ Cnj^{-\gamma} & \text{if (3.13) holds, for } j = 1, \dots, [n/2]. \end{cases}$

Proof. (i) Using summation by parts, we can write

$$D_{h,n}(u) = \sum_{t=1}^{n-1} (h_t - h_{t+1})s_t(u) + h_n s_n(u), \quad \text{where } s_t(u) = \sum_{j=1}^t e^{iju}.$$

Moreover, $|s_t(u)| = |\sin(tu/2)/\sin(u/2)| \leq Ct/(1+t|u|) \leq C|u|^{-1}$, $|u| \leq \pi$ where $C > 0$ does not depend on $1 \leq t \leq n$. Hence,

$$|D_{h,n}(u)| \leq C \left\{ \sum_{t=1}^{n-1} |h_t - h_{t+1}| |u|^{-1} + |h_n| |u|^{-1} \right\} \leq C\Delta_{h,n} |u|^{-1}, \quad |u| \leq \pi, \quad (8.55)$$

since $t/(1+t|u|) \leq |u|^{-1}$, which proves (i).

(ii) For $1 \leq t \leq n$, $\int_{\Pi} t(1+t|u|)^{-1} du \leq 2 \int_0^{n\pi} (1+u)^{-1} du \leq C \log n$. Therefore, by (8.55),

$$\int_{\Pi} |D_{h,n}(u)| du \leq C \sum_{t=1}^{n-1} |h_t - h_{t+1}| \log n + C|h_n| \log n \leq C\Delta_{h,n} \log n.$$

(iii) By (3.11), for $j = 1, \dots, [n/2]$,

$$D_{\mu,n}(u_j) = \sum_{t=1}^n e^{itu_j} \left(\sum_{p=-\infty}^{\infty} e^{-ipu_t} \rho_p \right) = \sum_{p=-\infty}^{\infty} \rho_p \left(\sum_{t=1}^n e^{itu_{j-p}} \right) = n \sum_{k=-\infty}^{\infty} \rho_{j+kn}$$

where the last equality follows from (8.33) and equality $e^{itu_j} = e^{itu_{j+kn}}$, $k = \pm 1, \pm 2, \dots$

Let (3.12) hold. Then $|\rho_{j+kn}| \leq C\phi^{|j+kn|}$. Notice that $|j+kn| \geq (|k|-1)n+j$ for $|k| \geq 1$ and $j \leq [n/2]$. Indeed, $|j+kn| \geq |k|n-j \geq (|k|-1)n+n-j \geq (|k|-1)n+j$. Hence, $\sum_{k=-\infty}^{\infty} |\rho_{j+kn}| \leq C\phi^j \sum_{k=0}^{\infty} \phi^{kn} \leq C\phi^j/(1-\phi)$ which implies (iii).

Let (3.13) hold. Then, $|\rho_{j+kn}| \leq C|j+kn|^{-\gamma}$. Observe that $|j+kn| \geq j(2|k|-1)$ for $|k| \geq 1$. Indeed, $n/j \geq 2$, and thus, $|j+kn| \geq |kn|-j \geq j(|k|(n/j)-1) \geq j(2|k|-1)$. Hence, $\sum_{k=-\infty}^{\infty} |\rho_{j+kn}| \leq Cj^{-\gamma}(1+\sum_{k=1}^{\infty} (2k-1)^{-\gamma}) \leq Cj^{-\gamma}$. This proves (iii) and completes the proof of the lemma. \square

Set $f_{x,j} = f_x(u_j)$, $A_{x,j} = A_x(u_j)$ where $A_x(u) = \sum_{k=0}^{\infty} e^{-iuk} b_k$.

Lemma 8.4 *Let $y_t = h_t x_t$, $t = 1, \dots, n$ and Assumptions A and B hold. Then for $1 \leq j \leq k \leq [n/2]$,*

$$\begin{aligned} (i) \quad & E|w_{h\varepsilon,j}|^2 = (2\pi)^{-1}(S/n), \quad E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}] = (2\pi n)^{-1}D_{h^2,n}(u_{j-k}), \\ (ii) \quad & |E[w_{y,j}\overline{w_{y,k}}] - 2\pi f_{x,\ell} E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}]| \leq Cn^{-1} \log n \Delta_{h,n}^2, \quad \ell = j, k; \\ (iii) \quad & |E[w_{y,j}\overline{w_{h\varepsilon,k}}] - A_{x,\ell} E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}]| \leq Cn^{-1} \log n \Delta_{h,n}^2, \quad \ell = j, k. \end{aligned}$$

Proof. (i) Since $E[\varepsilon_t \varepsilon_s] = 1(t=s)$, we have

$$\begin{aligned} E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}] &= (2\pi n)^{-1} \sum_{t,s=1}^n e^{itu_j - isu_k} h_t h_s E[\varepsilon_t \varepsilon_s] = (2\pi n)^{-1} D_{h^2,n}(u_{j-k}), \quad (8.56) \\ E|w_{h\varepsilon,j}|^2 &= (2\pi n)^{-1} \sum_{k=1}^n h_k^2 = (2\pi)^{-1}(S/n). \end{aligned}$$

(ii) Recall that $E x_t = 0$ and $E[x_t x_s] = \int_{-\pi}^{\pi} e^{i(t-s)u} f_x(u) du$. Hence,

$$\begin{aligned} E[w_{y,j}\overline{w_{y,k}}] &= (2\pi n)^{-1} \sum_{t,s=1}^n e^{itu_j - isu_k} h_t h_s E[x_t x_s] \quad (8.57) \\ &= (2\pi n)^{-1} \int_{\Pi} D_{h,n}(u+u_j) D_{h,n}(-u-u_k) f_x(u) du. \end{aligned}$$

Since $\int_{\Pi} e^{itu} du = 0$ for $t = 1, 2, \dots$; $= 2\pi$ for $t = 0$, then

$$\begin{aligned} \int_{\Pi} D_{h,n}(u + u_j) D_{h,n}(-u - u_k) du &= \int_{\Pi} \sum_{t,s=1}^n e^{i(t-s)u + i(tu_j - su_k)} h_t h_s du \\ &= 2\pi \sum_{t=1}^n e^{itu_{j-k}} h_t^2 = 2\pi D_{h^2,n}(u_{j-k}). \end{aligned}$$

Combining this equality with (8.57), we obtain

$$\begin{aligned} E[w_{y,j} \overline{w_{y,k}}] &= (2\pi n)^{-1} \int_{\Pi} D_{h,n}(u + u_j) D_{h,n}(-u - u_k) (f_x(u) - f_x(u_k)) du \\ &\quad + f_x(u_k) n^{-1} D_{h^2,n}(u_{j-k}) =: q_{n;jk} + f_x(u_k) n^{-1} D_{h^2,n}(u_{j-k}). \end{aligned} \quad (8.58)$$

We shall show below that

$$|q_{n;jk}| \leq C n^{-1} \log n \Delta_{h,n}^2 \quad (8.59)$$

which together with (8.58) and the part (i) of the lemma proves (ii) for $\ell = k$. (For $\ell = j$, (8.58) follows using similar argument.) First we show that

$$|D_{h,n}(-u - u_k)(f_x(u) - f_x(u_k))| \leq C \Delta_{h,n}, \quad |u| \leq \pi. \quad (8.60)$$

Indeed, by assumption, $f_x(u) = f_x(-u)$ and $|(d/dx)f_x(u)| \leq C$.

Let $|u + u_k| \leq \pi$. Then $|f_x(u) - f_x(u_k)| = |f_x(-u) - f_x(u_k)| \leq C|u + u_k|$ and (8.60) holds by Lemma 8.3(i).

Let $|u + u_k| > \pi$. Then $u, u_k \in (0, \pi]$ which implies $|2\pi - u - u_k| \leq \pi$ and $|2\pi - u - u_k| \geq |u - u_k|$. Therefore, $|f_x(u) - f_x(u_k)| \leq C|u - u_k| \leq C|2\pi - u - u_k|$. Thus, $|D_{h,n}(-u - u_k)(f_x(u) - f_x(u_k))| \leq |D_{h,n}(2\pi - u - u_k)(2\pi - u - u_k)| \leq C \Delta_{h,n}$ by Lemma 8.3(i). This proves (8.60).

Using (8.60) in (8.58) and subsequently applying Lemma 8.3(i-ii), we obtain

$$|q_{n;jk}| \leq C n^{-1} \Delta_{h,n} \int_{\Pi} |D_{h,n}(u + u_j)| du \leq 2C n^{-1} \Delta_{h,n} \int_{\Pi} |D_{h,n}(u)| du \leq C n^{-1} \log n \Delta_{h,n}^2$$

which completes the proof of (8.59).

(iii) We have $E[x_t \varepsilon_s] = \int_{\Pi} e^{i(t-s)u} f_{x\varepsilon}(u) du$, where $f_{x\varepsilon}(u) = (2\pi)^{-1} A_x(u)$ is the cross-spectral density. Hence,

$$\begin{aligned} E[w_{y,j} \overline{w_{h\varepsilon,k}}] &= (2\pi n)^{-1} \sum_{t,s=1}^n e^{itu_j - isu_k} h_t h_s E[x_t \varepsilon_s] \\ &= (2\pi n)^{-1} \int_{\Pi} D_{h,n}(u + u_j) D_{h,n}(-u - u_k) f_{x\varepsilon}(u) du. \end{aligned}$$

Thus, (iii) follows by the same argument as in the proof of (ii), noting that under Assumption B, $|(d/du)f_{x\varepsilon}(u)| \leq C$. \square

In the following theorem we discuss asymptotic properties of the sum

$$S_{n,y} := \sum_{j=1}^{\lfloor n/2 \rfloor} v(u_j) I_y(u_j) \quad (8.61)$$

where $I_y(u_j)$ is the periodogram based on observations y_1, \dots, y_n and $v(u)$, $u \in [0, \pi]$ is a continuous bounded function with a bounded derivative. Denote $\text{cum}_4(\varepsilon_0) = (\text{var}(\varepsilon_0^2) - 2)$.

Theorem 8.1 *Let Assumptions A and B hold. Then the sum $S_{n,y}$ in (8.61) has the following properties:*

- (i) $ES_{n,y} = 2\pi \int_0^\pi v(u)f_x(u)du S + o(S^{1/2})$,
- (ii) $\text{var}(S_{n,y}) = AV^* + o(V^*)$, $A = 2\pi \int_0^\pi v^2(u)f_x^2(u)du + \text{cum}_4(\varepsilon_0) \left(2\pi \int_0^\pi v(u)f_x(u)du\right)^2$,
- (iii) $\frac{S_{n,y} - ES_{n,y}}{(\text{var}(S_{n,y}))^{1/2}} \rightarrow_D \mathcal{N}(0, 1)$.

In particular, if $\int_0^\pi v(u)f_x(u)du = 0$, then

$$(V^*)^{-1/2}S_{n,y} \rightarrow_D \mathcal{N}(0, v^2), \quad v^2 = 2\pi \int_0^\pi v^2(u)f_x^2(u)du. \quad (8.62)$$

Proof. (i) By Lemma 8.4(i)-(ii), $E[I_y(u_j)] = f_x(u_j)S/n + O(\log n \Delta_{n,h}^2 n^{-1})$ where $\log n \Delta_{n,h}^2 = o(S^{1/2})$ by assumption (3.5)(b). Hence,

$$ES_{n,y} := \sum_{j=1}^{\lfloor n/2 \rfloor} v(u_j)EI_y(u_j) = (n^{-1} \sum_{j=1}^{\lfloor n/2 \rfloor} v(u_j)f_x(u_j))S + o(S^{1/2}).$$

Since the derivative of the function $v_1(u) = v(u)f_x(u)$ is bounded, $n^{-1} \sum_{j=1}^{\lfloor n/2 \rfloor} v(u_j)f_x(u_j) = 2\pi \int_0^\pi v_1(u)du + O(n^{-1})$. Noting that $S/n = O(1) = o(S^{1/2})$ and $S = O(V^*)$, this proves the claim (i) of the lemma.

(ii) Denote $S_{n,h\varepsilon}^{(appr)} := \sum_{j=1}^{\lfloor n/2 \rfloor} v(u_j)|A_x(u_j)|^2 I_{h\varepsilon}(u_j)$. We shall show that

$$(a) \quad \text{var}(S_{n,y} - S_{n,h\varepsilon}^{(appr)}) = o(V^*); \quad (b) \quad \text{var}(S_{n,h\varepsilon}^{(appr)}) = AV^* + o(V^*) \quad (8.63)$$

which implies the claim (ii) of the lemma. Using Proposition 6.2.1 of Giraitis, Koul and Surgailis (2012) (see e.g. the proof of Corollary 6.2.1 therein), we can bound

$$\begin{aligned} \text{var}(S_{n,y} - S_{n,h\varepsilon}^{(appr)}) &\leq C \sum_{j,k=1}^{\lfloor n/2 \rfloor} |v(u_j)v(u_k)| \\ &\quad \times \left| E[w_{y,j}\overline{w_{y,k}}] \right|^2 + |A_{x,j}|^2 |A_{x,k}|^2 |E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}]|^2 - 2|A_{x,k}|^2 |E[w_{y,j}\overline{w_{h\varepsilon,k}}]|^2 \Big|. \end{aligned}$$

By assumption (3.5)(b), $\log n \Delta_{n,h}^2 = o(S^{1/2})$. Hence, by Lemma 8.4(ii)-(iii),

$$\begin{aligned} |E[w_{y,j}\overline{w_{y,k}}]|^2 &= (|A_{x,j}|^2 E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}] + o(n^{-1}S^{1/2})) (|A_{x,k}|^2 E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}] + o(n^{-1}S^{1/2})), \\ |E[w_{y,j}\overline{w_{h\varepsilon,k}}]|^2 &= |A_{x,j}|^2 |E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}] + o(n^{-1}S^{1/2})|^2. \end{aligned}$$

By Lemma 8.4(i) and 8.3(i), $E|w_{h\varepsilon,j}|^2 = (2\pi)^{-1}(S/n)$ and $|E[w_{h\varepsilon,j}\overline{w_{h\varepsilon,k}}]| = (2\pi n)^{-1}|D_{h^2,n}(u_{j-k})| \leq C|j-k|^{-1}\Delta_{h^2,n}$ for $j \neq k$. Hence,

$$\begin{aligned} \text{var}(S_{n,y} - S_{n,h\varepsilon}^{(appr)}) &\leq C \sum_{j,k=1:j \neq k}^{[n/2]} \left(\frac{\Delta_{h^2,n}}{|j-k|} o\left(\frac{S^{1/2}}{n}\right) + o\left(\frac{S}{n^2}\right) \right) + C \sum_{j=k=1}^{[n/2]} \left(\frac{S}{n} o\left(\frac{S^{1/2}}{n}\right) + o\left(\frac{S}{n^2}\right) \right) \\ &= \log n \Delta_{h^2,n} o(S^{1/2}) + o(S) = o(S), \end{aligned}$$

since $\log n \Delta_{h^2,n} \leq C \log n \Delta_{h,n} = o(S^{1/2})$ by (3.5). This proves (8.63)(a), since $S = O(V^*)$.

Next we prove (8.63)(b). Since $f_x(u) = |A_x(u)|^2/2\pi$, setting $v_1(u) = v(u)f_x(u)$, we can write

$$S_{n,h\varepsilon}^{(appr)} = \frac{1}{n} \sum_{t,s=1}^n \sum_{j=1}^{[n/2]} e^{i(t-s)u_j} v_1(u_j) h_s h_t \varepsilon_s \varepsilon_t = \sum_{t,s=1}^n c_n^*(t,s) \varepsilon_s \varepsilon_t,$$

where $c^*(t,s) = h_t h_s c_n(t-s)$, $c_n(t) := n^{-1} \sum_{j=1}^{[n/2]} v_1(u_j) \cos(tu_j)$, $t = 1, 2, \dots$. The matrix $C_n^* = (c_n^*(t,s))_{t,s=1,\dots,n}$ is a symmetric $n \times n$ matrix with real entries. Hence,

$$\begin{aligned} \text{var}(S_{n,h\varepsilon}^{(appr)}) &= 2 \sum_{s,t=1:t \neq s}^n (c_n^*(t,s))^2 + \text{var}(\varepsilon_0^2) \sum_{t=1}^n (c_n^*(t,t))^2 \\ &= 2\|C_n^*\|^2 + (\text{var}(\varepsilon_0^2) - 2)c_n^2(0)V^*. \end{aligned} \quad (8.64)$$

where $\|C_n^*\|^2 = \sum_{s,t=1}^n (c_n^*(t,s))^2$ and $c_n(0) = n^{-1} \sum_{j=1}^{[n/2]} v_1(u_j) \rightarrow 2\pi \int_0^\pi v_1(u) du$.

We shall show that

$$\|C_n^*\|^2 = 2^{-1}(2\pi \int_0^\pi v_1^2(u) du) V^* + o(V^*), \quad (8.65)$$

which together with (8.64) and (8.63) implies

$$\text{var}(S_{n,y}) = \text{var}(S_{n,h\varepsilon}^{(appr)})(1 + o(1)) = AV^* + o(V^*) \quad (8.66)$$

and proves Theorem 8.1(ii). It remains to verify (8.65). Write

$$\|C_n^*\|^2 = \sum_{t,s=1}^n h_t^2 h_s^2 c_n^2(t-s) = \sum_{t,s=1}^n h_t^4 c_n^2(t-s) + \sum_{t,s=1}^n h_t^2 (h_t^2 - h_s^2) c_n^2(t-s) =: q_{n,1} + q_{n,2}.$$

We will show that

$$(a) \quad q_{n,1} = \pi \left(\int_0^\pi v_1^2(u) du \right) V^* (1 + o(1)), \quad (b) \quad q_{n,2} = o(S). \quad (8.67)$$

By assumption, $S \leq CV^*$. Hence, (8.67) implies (8.65). We have

$$q_{n,1} = n^{-2} \sum_{j,k=1}^{[n/2]} v_1(u_j) v_1(u_k) \sum_{t=1}^n h_t^4 \left(\sum_{s=1}^n \cos((t-s)u_j) \cos((t-s)u_k) \right).$$

For $j + k < n$, it holds $\sum_{s=1}^n \cos((t-s)u_j) \cos((t-s)u_k) = (n/2)1(j=k)$, see (6.2.32) in Giraitis, Koul and Surgailis (2012). Hence, as $n \rightarrow \infty$,

$$q_{n,1} = (1/2)n^{-2} \sum_{j=1}^{[n/2]-1} v_1^2(u_j) V^*(n/2) + O(n^{-1}V^*) = \pi \left(\int_0^\pi v_1^2(u) du \right) V^*(1 + o(1)),$$

because $n^{-1} \sum_{j=1}^{[n/2]-1} v_1^2(u_j) \rightarrow 2\pi \int_0^\pi v_1^2(u) du$.

It remains to evaluate $q_{n,2}$. Let $L > 1$ be fixed. Then,

$$|q_{n,2}| \leq \sum_{t,s=1}^n |h_t^2(h_t^2 - h_s^2)c_n^2(t-s)| \leq \sum_{t,s=1:|t-s| \geq L}^n [\dots] + \sum_{t,s=1:|t-s| < L}^n [\dots] = q_{n,2;1} + q_{n,2;2}.$$

First we show that

$$|c_n(t-s)| \leq C|t-s|^{-1}, \quad 0 < |t-s| < n. \quad (8.68)$$

Indeed, by the same argument as in (8.55),

$$\left| \sum_{j=1}^{[n/2]} v_1(u_j) e^{iu_j} \right| \leq C|u|^{-1} \left\{ \sum_{j=1}^{[n/2]-1} |v_1(u_j) - v_1(u_{j+1})| + |v_1(u_{[n/2]})| \right\} \leq C|u|^{-1}, \quad 0 < |u| \leq \pi,$$

because $v_1(u)$ has bounded derivative and therefore $|v_1(u_j) - v_1(u_{j+1})| \leq Cn^{-1}$.

Since $\cos((t-s)u_j) = \cos(u_{t-s}j) = (e^{iu_{t-s}j} + e^{-iu_{t-s}j})/2$, this implies

$$|c_n(t-s)| = n^{-1} \left| \sum_{j=1}^{[n/2]} v_1(u_j) \cos((t-s)u_j) \right| \leq Cn^{-1} |u_{t-s}| \leq C|t-s|^{-1}$$

which proves (8.68).

By (8.68) and assumption $|h_t| \leq C$, $t \geq 1$,

$$q_{n,2;1} \leq C \sum_{t,s=1:|t-s| \geq L}^n h_t^2 |t-s|^{-2} \leq CL^{-1}S.$$

On the other hand, using $|c_n(t-s)| \leq Cn^{-1} \sum_{j=1}^{[n/2]} 1 \leq C$ and $|h_t^2(h_t^2 - h_s^2)| \leq C|h_t - h_s|$, we obtain

$$q_{n,2;2} \leq C \sum_{t,s=1:|t-s| < L}^n |h_t - h_s| \leq CL \sum_{j=1}^{n-1} |h_j - h_{j+1}| \leq CL\Delta_{h,n},$$

since $|h_t - h_s| \leq \sum_{j=t-L}^{t+L} |h_j - h_{j+1}|$ for $|t-s| < L$. Hence, for any $L > 1$,

$$q_{n,2} \leq C(L^{-1}S + L\Delta_{h,n}) = o(S)$$

since by (3.5)(b), $\Delta_{h,n} = o(S)$. This completes the proof of (8.68)(b).

(iii) We shall show that

$$\|C_n^*\|_{sp} = o(\|C_n^*\|). \quad (8.69)$$

It is well known, see e.g. Theorem 6.2.2 in Giraitis, Koul and Surgailis (2012), that (8.69) implies

$$(\text{var}(S_{n,h\varepsilon}^{(appr)}))^{-1/2}(S_{n,h\varepsilon}^{(appr)} - E[S_{n,h\varepsilon}^{(appr)}]) \rightarrow_D \mathcal{N}(0, 1)$$

which together with (8.63) and (8.66) proves Theorem 8.1(iii).

We proceed to the proof of (8.69). Let $x \in R^n$ be such that $\|x\|^2 = 1$. By definition, $\|C^*\|_{sp}^2 = \max_{\|x\|=1} \|C_n^*x\|^2$. Let $L > 1$. Then,

$$\|C_n x\|^2 = \sum_{t=1}^n h_t^2 \left(\sum_{s=1}^n h_s^2 c_n(t-s)x_s \right)^2 \leq 2(r_{n,1} + r_{n,2})$$

where $r_{n,1} = \sum_{t=1}^n h_t^2 \left(\sum_{s=1: |t-s| \geq L}^n h_s^2 c_n(t-s)x_s \right)^2$, $r_{n,2} = \sum_{t=1}^n h_t^2 \left(\sum_{s=1: |t-s| < L}^n h_s^2 c_n(t-s)x_s \right)^2$. To bound $r_{n,1}$, from (8.68) and $h_s^2 \leq C$ we obtain

$$r_{n,1} \leq C \sum_{t=1}^n h_t^2 \left(\sum_{s=1: |t-s| \geq L}^n |t-s|^{-1} |x_s| \right)^2 \leq C \sum_{t=1}^n h_t^2 \sum_{s=1: |t-s| \geq L}^n |t-s|^{-2} \|x\|^2 \leq CL^{-1}S.$$

In turn, bounding $|c_n(t-s)| \leq C$, $h_t^2 \leq C$, $h_s^2 \leq C$ and using the inequality $(a_1 + \dots + a_L)^2 \leq L(a_1^2 + \dots + a_L^2)$, we obtain

$$r_{n,2} \leq C \sum_{t=1}^n \left(\sum_{s=1: |t-s| < L}^n |x_s| \right)^2 \leq CL \sum_{t=1}^n \sum_{s=1: |t-s| < L}^n x_s^2 \leq CL \sum_{s=1}^n x_s^2 \left(\sum_{t=1: |t-s| < L}^n 1 \right) \leq CL^2 \|x\|^2 = CL^2.$$

Hence, for any $L > 1$, it holds $\|C_n x\|^2 \leq C(L^{-1}S + L^2)$. Since $S \rightarrow \infty$, this implies

$$\|C_n x\|^2 = o(S) = o(V^*) = o(\|C_n^*\|^2)$$

which completes the proof of (8.69) and Theorem 8.1(iii). \square

REFERENCES

- Altman, N., 1993. Estimating error correlation in non-parametric regression. *Statist. & Prob. Letters* 18, 213-218.
- Anderson, T. W., Walker, A. M., 1964. On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Statist.* 35, 1296-1303.
- Beran, J., Feng, Y., 2002. SEMIFAR models - a semiparametric approach to modelling trends, long range dependence and nonstationarity. *Comp. Statist. Data. Analysis* 40, 393-419.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.
- Bai, J., Perron, P., 2003. Computation and Analysis of Multiple Structural Change Models *J. Appl. Econometrics* 18, 1-22.

- Brown, B.M., 1971. Martingale central limit theorems. *Ann. Math. Statist.* 42, 59-66.
- Cavaliere, G., Nielsen, M.O., Taylor, A.M.R., 2017. Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. *J. Econometrics* 198, 165–188.
- Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. *Ann. Statist.* 25, 1–37.
- Dahl, C.M., Levine, M., 2006. Nonparametric estimation of volatility models with serially dependent innovations. *Statist. & Prob. Letters* 76, 2007-2016.
- Deschamps, P.J., 2008. Comparing smooth transition and Markov switching autoregressive models of US unemployment. *J. Appl. Econometrics*, 23, 435–462.
- van Dijk, D., Franses, P.H., R. Paap, R., 2002. A nonlinear long memory model, with an application to US unemployment. *J. Econometrics* 110, 135–165.
- Figuroa-López, J.E. and Levine, M., 2013. Nonparametric regression with rescaled time series errors. *J. Time Series Anal.* 34, 345-361.
- Fox, R., Taqqu, M.S., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* 14, 517-532.
- Gasser, T., Mueller, H.G., 1979. Kernel estimation of regression functions. In "Smoothing Techniques for Curve Estimation". In *Lect. Notes Math.*, vol. 757, Springer Verlag, pp. 23–68.
- Giraitis, L., Hidalgo, J., Robinson, P.M., 2001. Gaussian estimation of parametric spectral density with unknown pole. *Ann. Statist.* 29, 987-1023.
- Giraitis, L., Koul, H.L., Surgailis, D., 2012. *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.
- Golan, A., Perloff, J.M., 2004. Superior forecasts of the U.S. unemployment rate using a nonparametric method. *Rev. Econ. Statist.* 86, 443–438.
- Guo, H., Koul, H.L., 2007. Nonparametric regression with heteroscedastic long memory errors. *J. Statist. Plan. Infer.* 137, 379–404.
- Hall, P., Van Keilegom, I., 2003. Using difference-based methods for inference in non-parametric regression with time series errors. *J. Royal Stat. Soc. Ser. B* 65, 443-456.
- Hannan, E.J., 1973. The asymptotic theory of linear time series models. *J. Appl. Probab.* 10, 130 – 145.
- Harvey, A.C., Robinson, P. M., 1988. Efficient estimation of nonstationary time series regression. *J. Time Ser. Anal.* 9. 201-214.
- Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Ann. Statist.* 39, 3152–3181.

- Hurvich, C., Lang, G., Soulier, P., 2005. Estimation of long memory in the presence of a smooth nonparametric trend. *JASA* 100, 853–871.
- Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components. *J. Time Ser. Anal.* 31, 37–49.
- Künsch, H. R., 1986. Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* 23, 1025–1030.
- McCloskey, A., 2010. Parameter estimation robust to low-frequency contamination with applications to ARMA, GARCH and stochastic volatility models. Unpublished Manuscript, Department of Economics Brown University.
- McCloskey, A., 2013. Estimation of the long-memory stochastic volatility model parameters that is robust to level shifts and deterministic trends. *J. Time Series Anal.* 34, 285–301.
- McCloskey, A., Hill, J.B., 2017. Parameter estimation robust to low-frequency contamination. *J. Bus. Econ. Statist.* 345, 598–610.
- McCloskey, A., Perron, P., 2013. Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory* 29, 1196–1237.
- Montgomery, A.L., Zarnowitz, V, Tsay, R. S., Tiao, G. C., 1998. Forecasting the U.S. Unemployment Rate. *JASA* 92. 478–486.
- Parzen, E., 1963. On spectral analysis with missing observations and amplitude modulation. *Sankhya A* 25, 383-392.
- Phillips, P.C.B., Xu, K., 2005. Inference in autoregression under heteroscedasticity. *J. Time Series Anal.* 27, 289–308.
- Priestley, M. B., Chao, K., 1972. Non-parametric function fitting. *J. Rol. Statist. Soc. Ser. B.* 34, 385-392.
- Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. In: *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.) Springer-Verlag, New York, pp. 253–264.
- Robinson, P.M., 1997. Large sample inference for nonparametric regression with dependent errors. *Ann. Statist.* 25, 2054–2083.
- Shao, Q., Yang, L.J., 2011. Autoregressive coefficient estimation in nonparametric analysis. *J. Time Series Anal.* 32, 587–597.
- Truong, Y.K., 1991. Nonparametric curve estimation with time series errors. *J. Statist. Plan. Inference.* 28, 167–183.
- Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *JASA* 95, 1229-1243.

- Xu, K.-L., Phillips, P.C.B., 2008. Adaptive estimation of autoregressive models with time-varying variances. *J. Econometrics* 142, 265-280.
- Zhao, Z., Zhang, Y., Li, R., 2014. Non-parametric estimation under strong dependence. *J. Time Series Anal.* 35, 4-15.
- Zhu, K., 2016. Statistical inference for autoregressive models under heteroscedasticity of unknown form. Preprint, Hong Kong Univ.