Nonparametric regression with rescaled time series errors

José E. Figueroa-López[†] Michael Levine[*]

Abstract

We consider a heteroscedastic nonparametric regression model with an autoregressive error process of finite known order p. The heteroscedasticity is incorporated using a scaling function defined at uniformly spaced design points on an interval [0,1]. We provide an innovative nonparametric estimator of the variance function and establish its consistency and asymptotic normality. We also propose a semiparametric estimator for the vector of autoregressive error process coefficients that is $\sqrt{T}$ consistent and asymptotically normal for a sample size T. Explicit asymptotic variance covariance matrix is obtained as well. Finally, the finite sample performance of the proposed method is tested in simulations.

Keywords and phrases: autoregressive error process; heteroscedastic; semiparametric estimators; difference-based estimation approach.

1 Introduction

In this manuscript, we consider the estimation of a time series process with a time-dependent conditional variance function and serially dependent errors. More precisely, we assume that there are T observations $\{(x_t, y_t)\}_{t\in\{1,\dots,T\}}$ available that have been generated by the following model:

$$y_t = \sigma_t v_t, \qquad \sigma_t = \sigma(x_t), \tag{1.1}$$

$$v_t = \sum_{j=1}^{p} \phi_j v_{t-j} + \varepsilon_t, \tag{1.2}$$

where $\{\varepsilon_t : -\infty < t < \infty\}$ are independent identically distributed with mean 0 and variance 1, while the autoregressive order is a fixed known integer $p > 0$. We also assume that x_t's form an increasing equally spaced sequence on the interval $[0, 1]$. Then, the model (1.1)-(1.2) can be viewed as a nonparametric regression model with the mean function identically equal to zero and scaled autoregressive time series errors v_t. The simplest case of $p = 1$ was treated earlier in Dahl and Levine(2006). Another interpretation of this model would be as a type of functional autoregressive models (FAR) that were first introduced in Chen and Tsay (1993). Indeed, the process y_t can be re-expressed as

$$y_t = \sum_{j=1}^{p} \phi_j \sigma_t \sigma_{t-j}^{-1} y_{t-j} + \sigma_t \varepsilon_t, \tag{1.3}$$

with functional coefficients being $\phi_j \sigma_t \sigma_{t-j}^{-1}$, $j = 1, \dots, p$.

Nonparametric regression models with autoregressive time series errors have a long history. Hall and Van Keilegom (2003) and, more recently, Shao and Yang (2011) considered a general nonparametric model of the form

$$y_t = \mu_t + \sigma_t v_t, \tag{1.4}$$

with a regression mean process μ_t given by $\mu_t = g(x_t)$ for a smooth trend function g, an error process $\{v_t : -\infty < t < \infty\}$ given as in (1.2), and an error scaling function σ_t given by $\sigma_t \equiv \sigma$, for a (possibly unknown) constant σ. Note that our model has a more general correlation structure since it rescales the AR(p) error process by the conditional variance of the unobserved process $\sigma(x_t)$. Phillips and Xu (2005) also considered the model (1.4) with a general scaling function $\sigma_t = \sigma(x_t)$, but with the autoregressive mean process structure $\mu_t = \theta_0 + \theta_1 y_{t-1} + \ldots + \theta_q y_{t-q}$ and with zero mean martingale difference sequence as an error process $\{v_t : -\infty < t < \infty\}$. As mentioned above, the model considered in this manuscript is an extension of that studied in Dahl and Levine (2006), who treated the particular case $p = 1$.

In the present manuscript, we provide an innovative nonparametric estimator of the variance function $\sigma(x)$ and establish its consistency and asymptotic normality. We also propose a semiparametric estimator for the autoregressive error process coefficients $(\phi_1, \ldots, \phi_p)$, which is $\sqrt{T}$ consistent and asymptotically normal. The estimators of the error covariance structure (determined by $\sigma(x_t)$ and $(\phi_1, \ldots, \phi_p)$) are needed, for example, in order to estimate the variance of the regression mean function μ_t. Another possible application is in using bootstrap methods when constructing confidence bands for the regression mean. For some discussion on this subject see, for example, Hall and Van Keilegom (2003). Our results can also be used for testing the martingale difference sequence hypothesis $H_0 : \phi_1 = \cdots = \phi_p = 0$, that is often uncritically adopted in financial time series, against a fairly general alternative.

Our estimation approach is based on the two-lag difference statistics (pseudoresiduals):

$$\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}}. \tag{1.5}$$

Both Hall and Van Keilegom (2003) and Dahl and Levine (2006) also use two-lag difference statistics but the method in the former paper does not apply in the presence of a non-constant scaling function $\sigma(x_t)$ (see Remark 3.5 below), while the method in the second paper relies heavily on the autocorrelation properties of an AR(1) process and, thus, cannot deal with higher order autocorrelation models for $\{v_t : -\infty < t < \infty\}$ (see justification in the paragraph after Eq. (2.4)). For simplicity, we mostly focus on the case of a regression mean μ_t which is identically equal to zero since our main goal is the estimation of the covariance structure of the process (1.1)-(1.2). Nevertheless, we also analyze the effect of a nonzero mean function and propose a natural correction method in this case (see end of Section 3 for more details). We also show that the addition of a sufficiently smooth non-zero mean function has no effect on the asymptotic properties of $\hat{\boldsymbol{\phi}}$. This fact is explained in more details in Corollary 3.6; see also Hall and Van Keilegom (2003) for some additional heuristics on this subject.

The rest of the paper is organized as follows. In Section 2, we present our estimation approach. The consistency and central limit theorem for estimators of autoregressive coefficients ϕ_1 and ϕ_2 are given in Section 3. The analogous results for the estimator of the variance function $\sigma(x)$ are presented in Section 4. A Monte Carlo simulation study and a real data application of our estimators are then presented in Section 5. In Section 6, we conclude the paper with a discussion section about extensions of our method and some interesting open problems. The proofs of main results are given in the Appendix section.

2 Estimation method based on two-lag differences

In this section, we present our estimation approach. For simplicity, we first illustrate our method for $p = 2$; however, it will work for any finite order p as we explain further at the end of the present section. As it was mentioned in the introduction, Dahl and Levine (2006) proposed an estimation method for the model (1.1)-(1.2) in the simple case of $p = 1$ based on the two-lag difference statistics η_t defined in (1.5). In order to explain in simple terms the intuition behind our method, we recall the approach therein.

Suppose for now that the scaling function σ_t is constant; i.e. $\sigma_t \equiv \sigma$. We denote γ_k the autocovariance of the error process v_t at lag k. Then, note that $\eta_t^2 = \frac{\sigma^2}{2}\left(v_t^2 + v_{t-2}^2 - 2v_t v_{t-2}\right)$ and, therefore,

$$E\,\eta_t^2 = \sigma^2\left(\gamma_0 - \gamma_2\right) \equiv \sigma^2, \tag{2.1}$$

because, under the AR(1) specification,

$$\gamma_2 = \phi_1\gamma_1 = \frac{\phi_1^2}{1-\phi_1^2}.$$

It is now intuitive that η_t^2 can be used to develop a consistent estimator for a non-constant function σ_t^2 as well. Indeed, in the case of non-constant σ_t and under a fixed design on the unit interval (i.e., $x_t := t/T$ for $t = 1, \ldots, T-1$), we have

$$E\,\eta_t^2 = \frac{1}{2}\left(\sigma_t^2\gamma_0 + \sigma_{t-2}^2\gamma_0 - 2\sigma_t\sigma_{t-2}\gamma_2\right).$$

Simple heuristics suggest that the above expression can be accurately approximated by σ_t^2 in large samples for sufficiently large T. This, in turn, suggests turning the original problem (1.1) into a non-parametric regression

$$\eta_t^2 = \sigma^2(x_t) + \tilde{\varepsilon}_t, \tag{2.2}$$

where $\{\tilde{\varepsilon}_t\}_{t=2,\ldots,T}$ are approximately centered random errors. Dahl and Levine (2006) used a local linear estimator $\hat{\sigma}_t^2$ to estimate $\sigma_t^2 := \sigma^2(x_t)$, while the parameter ϕ_1 was estimated using a weighted least square estimator (WLSE). More specifically, noting that (1.3) with $\phi_2 = \cdots = \phi_p = 0$ implies

$$\sigma_t^{-1}y_t = \phi_1\sigma_{t-1}^{-1}y_{t-1} + \varepsilon_t, \quad t = 2, \ldots, T, \tag{2.3}$$

it follows that a natural estimator for ϕ_1 is given by

$$\begin{aligned}\hat{\phi}_1 &:= \arg\min_{\phi_1\in(-1,1)}\frac{1}{T}\sum_{t=2}^{T}\left(\hat{\sigma}_t^{-1}y_t - \phi_1\hat{\sigma}_{t-1}^{-1}y_{t-1}\right)^2 \\ &= \left(\frac{1}{T}\sum_{t=2}^{T}\hat{\sigma}_{t-1}^{-2}y_{t-1}^2\right)^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\hat{\sigma}_t^{-1}\hat{\sigma}_{t-1}^{-1}y_t y_{t-1}\right),\end{aligned} \tag{2.4}$$

where above $\{\hat{\sigma}_t^2\}_{t=1,\ldots,T}$ is the local linear estimator of $\sigma^2(x_t)$ based on non-parametric model (2.2).

The approach described in the previous paragraph will not work for any $p > 1$ because, in that case, $\gamma_0 - \gamma_2 \neq 1$ and, therefore, even the constant variance function σ^2 cannot be estimated consistently. A first natural idea to extend the method in Dahl and Levine (2006) is to look for a linear statistic

$$\eta_t := \sum_{i=0}^{m} a_i y_{t-i}, \tag{2.5}$$

such that

$$E\,\eta_t^2 \approx \sigma_t^2,$$

for sufficiently large sample size T. The following result shows that this is essentially impossible even for the simplest AR(1) case. The impossibility for a general AR(p) model will immediately follow since obviously the AR(1) model can be seen as a degenerate case of the general AR(p) model. The proof of the following result is deferred to the appendix section.

Proposition 2.1. *Consider again the case where $\sigma_t^2 \equiv \sigma^2$ is constant and the error process is AR(1) (i.e., $\phi_2 = \cdots = \phi_p = 0$ in (1.1)-(1.2)). Then, if $E\eta_t^2 = \sigma^2$ for any $\phi_1 \in (-1,1)$, there exists a $0 \leq k \leq m-2$ such that*

$$a_k = \pm\frac{1}{\sqrt{2}}, \quad a_{k+2} = \mp\frac{1}{\sqrt{2}}, \quad a_i = 0, \quad \forall i \neq k, k+2.$$

The previous result shows that the only linear statistic (2.5) with $a_0 \neq 0$ that can result in $E\eta_t^2$ being independent of ϕ_1 is the two-lag difference statistic

$$\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}}.$$

Now we show that the expectation of η_t^2 is indeed independent of ϕ_1 and introduce our new method to estimate the model (1.1)-(1.2) with an AR(2) error process. Recall that (see, for example, Brockwell and Davis (1991)) for a stationary AR(2) process v_t, its variance and autocovariances are:

$$\gamma_0 := \mathrm{Var}(v_k) = \frac{1-\phi_2}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}, \tag{2.6}$$

$$\gamma_1 := \mathrm{Cov}(v_k, v_{k+1}) = \frac{\phi_1}{1-\phi_2}\gamma_0, \tag{2.7}$$

$$\gamma_j := \mathrm{Cov}(v_k, v_{k+j}) = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2}, \quad \text{for any} \quad k \in \mathbb{N},\ j \geq 2. \tag{2.8}$$

Therefore, the mean of squared pseudoresidual of the 2nd order, η_t^2, can be simplified as follows under the assumption of constant variance function $\sigma_t^2 \equiv \sigma$:

$$E\,\eta_t^2 = \sigma^2\left(\gamma_0 - \gamma_2\right) = \frac{\sigma^2}{1+\phi_2}.$$

Indeed, using (2.6)-(2.7),

$$\begin{aligned}
\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 &= \left(\frac{\phi_1^2}{1-\phi_2} + \phi_2\right)\gamma_0 = \frac{\phi_1^2 + (1-\phi_2)\phi_2}{1-\phi_2}\gamma_0, \\
\gamma_0 - \gamma_2 = \gamma_0\left(1 - \frac{\phi_1^2 + (1-\phi_2)\phi_2}{1-\phi_2}\right) &= \gamma_0\left(\frac{(1-\phi_2)^2 - \phi_1^2}{1-\phi_2}\right) \\
&= \left(\frac{1-\phi_2}{(1+\phi_2)((1-\phi_2)^2-\phi_1^2)}\right)\left(\frac{(1-\phi_2)^2-\phi_1^2}{1-\phi_2}\right) = \frac{1}{1+\phi_2}.
\end{aligned}$$

As before, for a general smooth enough function σ_t^2 and under a fixed design on the unit interval ($x_t = t/T$, $t = 1, \ldots, T-1$) with T large enough, we expect that

$$E\,\eta_t^2 \approx \frac{\sigma_t^2}{1+\phi_2},$$

and, hence, we expect to estimate correctly σ_t^2 up to a constant. It turns out that this will suffice to estimate the parameters ϕ_1 and ϕ_2 via weighted least squares (WLSE). Indeed, suppose for now that we know the variance function σ_t^2 and let $\bar{y}_t := \sigma_t^{-1}y_t$. In light of the relationship (1.3), it would then be possible to estimate (ϕ_1, ϕ_2) by the WLSE:

$$(\bar{\phi}_1, \bar{\phi}_2) := \arg\min_{\phi_1,\phi_2}\frac{1}{T}\sum_{t=4}^{T}\left(\bar{y}_t - \phi_1\bar{y}_{t-1} - \phi_2\bar{y}_{t-2}\right)^2.$$

Basic differentiation leads to the following system of normal equations

$$-\sum_{t=4}^{T}\bar{y}_t\bar{y}_{t-1}+\bar{\phi}_1\sum_{t=4}^{T}\bar{y}_{t-1}^2+\bar{\phi}_2\sum_{t=4}^{T}\bar{y}_{t-1}\bar{y}_{t-2}=0,$$
$$-\sum_{t=4}^{T}\bar{y}_t\bar{y}_{t-2}+\bar{\phi}_1\sum_{t=4}^{T}\bar{y}_{t-1}y_{t-2}+\bar{\phi}_2\sum_{t=4}^{T}\bar{y}_{t-2}^2=0.$$

Ignoring negligible edge effects (so that $\sum_{t=4}^{T}\bar{y}_t\bar{y}_{t-1}\approx\sum_{t=4}^{T}\bar{y}_{t-1}\bar{y}_{t-2}$ and $\sum_{t=4}^{T}\bar{y}_t^2\approx\sum_{t=4}^{T}\bar{y}_{t-1}^2$), we can write the above system as

$$\bar{A}\bar{\phi}_1+\bar{B}\bar{\phi}_2-\bar{B}=0,\quad \bar{B}\bar{\phi}_1+\bar{A}\bar{\phi}_2-\bar{C}=0,$$

with

$$\bar{A}:=\sum_{t=4}^{T}\bar{y}_t^2,\quad \bar{B}:=\sum_{t=4}^{T}\bar{y}_t\bar{y}_{t-1},\quad \bar{C}:=\sum_{t=4}^{T}\bar{y}_t\bar{y}_{t-2}.$$

We finally obtain

$$\bar{\phi}_2:=(\bar{A}^2-\bar{B}^2)^{-1}(\bar{A}\bar{C}-\bar{B}^2),\quad \bar{\phi}_1=\bar{A}^{-1}\bar{B}(1-\hat{\phi}_2). \tag{2.9}$$

Obviously, these estimators are not feasible since σ_t^2 is unknown. However, we note that these estimators will not change if instead of σ_t in the definition of $\bar{y}_t$, we use $c\sigma_t$ where c is an arbitrary constant that is independent of t. This fact suggests the following two-step estimation method for the scaling function $\sigma_t=\sigma(x_t)$ and the autocorrelation coefficients (ϕ_1,ϕ_2):

Algorithm 2.1 (AR(2) case).

1. *First, estimate the function*

$$\sigma_t^{2,bias}:=\frac{\sigma^2(x_t)}{1+\phi_2}, \tag{2.10}$$

by a non-parametric smoothing method (e.g. local linear regression) applied to the two-lag difference statistics η_t^2 defined in (1.5). Let $\tilde{\sigma}_t^2$ be the resulting estimator.

2. *Standardize the observations, $\tilde{y}_t:=\tilde{\sigma}_t^{-1}y_t$, and, then, estimate (ϕ_1,ϕ_2) via the WLSE:*

$$\hat{\phi}_2:=(A^2-B^2)^{-1}(AC-B^2),\quad \hat{\phi}_1=A^{-1}B(1-\hat{\phi}_2). \tag{2.11}$$

with

$$A:=\sum_{t=4}^{T}\tilde{y}_t^2,\quad B:=\sum_{t=4}^{T}\tilde{y}_t\tilde{y}_{t-1},\quad C:=\sum_{t=4}^{T}\tilde{y}_t\tilde{y}_{t-2}.$$

3. *Estimate $\sigma_t^2:=\sigma^2(x_t)$ by*

$$\hat{\sigma}_t^2:=(1+\hat{\phi}_2)\tilde{\sigma}_t^2. \tag{2.12}$$

The same method can be easily extended to the case of an arbitrary autoregressive error process AR(p) with $p>2$. Indeed, let $\phi_1,\ldots,\phi_p$ be the coefficients of the AR(p) error process. It can be shown that, for large T, the expectation of the squared pseudoresidual of order 2, η_t^2, is approximately equal to the scaled value of σ_t^2:

$$E\,\eta_t^2\approx\Psi\,\sigma_t^2,$$

where the scaling constant Ψ is an explicit function $\Psi\equiv\Psi(\phi_1,\ldots,\phi_p)$ of $\phi_1,\ldots,\phi_p$. As with the case of $p=2$, since the scaling constant does not depend on the variance function, the following natural extended procedure follows:

Algorithm 2.2 (**General AR(p) case**).

1. *Obtain an estimate of the scaled variance function*

$$\sigma_t^{2,bias} := \sigma^{2,bias}(x_t) := \Psi(\phi_1, \ldots, \phi_p)\sigma^2(x_t),$$

by using a non-parametric smoothing method (e.g., local linear regression) applied to η_t^2. *As before, let* $\tilde{\sigma}_t^2 = \tilde{\sigma}^2(x_t)$ *be the resulting estimator.*

2. *Standardize the observations* $\tilde{y}_t := \tilde{\sigma}_t^{-1} y_t$ *and then estimate* $(\phi_1, \ldots, \phi_p)$ *using the weighted least squares (WLSE):*

$$(\hat{\phi}_1, \ldots, \hat{\phi}_p) := arg\ min_{\phi_1,\ldots,\phi_p} \frac{1}{T} \sum_{t=p+2}^{T} (\tilde{y}_t - \phi_1 \tilde{y}_{t-1} - \ldots - \phi_p \tilde{y}_{t-p})^2 .$$

3. *Estimate* $\sigma_t^2 := \sigma^2(x_t)$ *by*

$$\hat{\sigma}_t^2 := \Psi(\hat{\phi}_1, \ldots, \hat{\phi}_p)^{-1} \tilde{\sigma}_t^2. \tag{2.13}$$

In the next section, we will give a detailed analysis of the consistency and asymptotic properties of the proposed estimators.

3 Asymptotics

Let us now consider the asymptotic properties of the estimation procedure described at the end of the previous section in the context of the general heteroscedastic process (1.1-1.2). We will use $\tilde{\sigma}_t^2$ to denote the inconsistent estimator of σ_t^2 that is obtained by applying local linear regression to the squared-pseudoresiduals η_t^2. As explained above, such an estimator is inconsistent since, e.g., even for the homoscedastic model (i.e., $\sigma_t^2 \equiv \sigma^2$), $E\eta_t^2 = \sigma^2 \Psi(\phi_1, \ldots, \phi_p)$. However, note that it is expected to be a *consistent* estimator of the quantity $\sigma_t^{2,bias} = \sigma_t^2 \Psi(\phi_1, \ldots, \phi_p)$, as it will be formally proved in Section 4 below. Throughout, we denote

$$\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_T)', \qquad \tilde{\boldsymbol{\sigma}} = (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_T)', \qquad \boldsymbol{\sigma}^{bias} = (\sigma_1^{bias}, \ldots, \sigma_T^{bias})'. \tag{3.1}$$

As explained before, it seems reasonable to estimate the coefficients $\phi_1, \phi_2, \ldots, \phi_p$ using an inconsistent estimator $\tilde{\sigma}_t^2$ first and, then, correct it to obtain the asymptotically consistent estimator:

$$\hat{\sigma}_t^2 = \Psi(\hat{\phi}_1, \ldots, \hat{\phi}_p)^{-1} \tilde{\sigma}_t^2.$$

The following detailed algorithm illustrates our approach to obtaining the asymptotic properties of the proposed estimators:

1. Using the functional autoregression form of the model (1.1), define the least squares estimator $\hat{\boldsymbol{\phi}} := (\hat{\phi}_1, \ldots, \hat{\phi}_p)$ of $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_p)$ and establish its consistency (i.e. $\hat{\boldsymbol{\phi}} \xrightarrow{\mathbb{P}} \boldsymbol{\phi}$ as $T \to \infty$), under additional conditions.

2. Define an asymptotically consistent estimator $\hat{\sigma}_t^2$ and establish its consistency and asymptotic normality.

For convenience, we recall here the functional autoregressive form of (1.1):

$$\sigma_t^{-1} y_t = \phi_1 \sigma_{t-1}^{-1} y_{t-1} + \ldots + \phi_p \sigma_{t-p}^{-1} y_{t-p} + \varepsilon_t.$$

Next, for any $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_T)$ and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)$, let

$$\bar{m}_k(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := \frac{1}{T} \sum_{t=p}^{T} m_{k,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}), \qquad (k = 1, 2, \ldots, p), \tag{3.2}$$

where

$$m_{k,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := \vartheta_{t-k}^{-1} y_{t-k} [\vartheta_t^{-1} y_t - \vartheta_{t-1}^{-1} \varphi_1 y_{t-1} - \ldots - \vartheta_{t-p}^{-1} \varphi_p y_{t-p}], \tag{3.3}$$

where (y_t) is generated by the model (1.1) with true parameters $(\sigma_t, \boldsymbol{\phi})$. Denote

$$m_t(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := (m_{1,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}); \ldots; m_{p,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}))'. \tag{3.4}$$

Note that

$$m_{k,t}(\boldsymbol{\sigma}; \boldsymbol{\phi}) = v_{t-k} \varepsilon_t, \qquad (k = 1, \ldots, p), \tag{3.5}$$

and, therefore, none of them depends on $\{\sigma_t\}$. Then, the first order conditions that determine the least-square estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ are given by

$$\bar{m}_k(\tilde{\boldsymbol{\sigma}}; \hat{\boldsymbol{\phi}}) = 0, \quad (k = 1, \ldots, p). \tag{3.6}$$

A few preliminary results are needed to establish consistency for $\hat{\boldsymbol{\phi}}$. Throughout, we assume that the data generating process (1.1)-(1.2) satisfies the following conditions:

1. $\{\varepsilon_t\}$ are independent identically distributed (i.i.d.) errors with mean zero and variance 1.
2. $E|\varepsilon_t|^{4+\gamma} < \infty$ for some small $\gamma > 0$.
3. $\sigma^2(\cdot) \in \mathcal{F} := C^2[0,1]$, the class of continuous functions on $[0,1]$ with continuous second derivatives in $(0,1)$.

We denote Θ_0 the set of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ such that the roots of the characteristic equation $1 - \phi_1 z - \ldots - \phi_p z^p = 0$ are greater than $1 + \delta$ in absolute value for some $\delta > 0$. Such a condition on $\boldsymbol{\phi}$ guarantees causality and stationarity of the AR(p) error process v_t defined in (1.2); moreover, it also implies that v_t can be represented as the MA(∞) process

$$v_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{3.7}$$

where all of the coefficients $\{\psi_i\}_{i \geq 0}$ satisfy

$$K_1 \rho^i \leq |\psi_i| \leq K_2 \rho^i, \tag{3.8}$$

for some finite ρ such that $\frac{1}{1+\delta} < \rho < 1$, and positive constants K_1 and K_2 (see, e.g., Brockwell and Davis (1991), Chapter 3). In particular, (3.8) implies that the series $\{\psi_i\}_{i \geq 0}$ is absolutely converging: $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

Remark 3.1. *The coefficients* $\{\psi_i\}_{i\geq 0}$ *satisfy the recursive system of equations*

$$\psi_j - \sum_{0<k\leq j} \phi_k \psi_{j-k} = 0, \tag{3.9}$$

with $\psi_0 \equiv 1$ *and* $\phi_k = 0$ *for* $k > p$. *Note that this implies (by induction) that* ψ_j *is a continuously differentiable function of* $\boldsymbol{\phi}$ *for any* $j \geq 0$.

Our first task is to show the weak consistency of the least-square estimator $\hat{\boldsymbol{\phi}}$. As pointed earlier in Dahl and Levine (2006), the estimator $\hat{\boldsymbol{\phi}}$ is an example of a MINPIN semiparametric estimator (i.e., an estimator that minimizes a criterion function that may depend on a Preliminary Infinite Dimensional Nuisance Parameter estimator). MINPIN estimators have been discussed in great generality in Andrews (1994). We first establish the following uniform weak law of large numbers (LLN) for $m_t(\boldsymbol{\sigma};\phi)$ (see Andrews (1987), Appendix A for the definition).

Lemma 3.2. *Suppose that* $\Theta \subset \Theta_0$ *is a compact set with non-empty interior. Then, as* $T \to \infty$,

$$\sup_{\phi\in\Theta} \left| \frac{1}{T} \sum_{t=p}^{T} m_t(\boldsymbol{\sigma};\boldsymbol{\phi}) \right| \xrightarrow{\mathbb{P}} 0.$$

We are ready to show our main consistency result. The proof is presented in Appendix A.

Theorem 3.3. *Let* Θ *be a compact subset of* Θ_0 *with non-empty interior. Then, the least squares estimator* $\hat{\boldsymbol{\phi}}$ *converges to the true* $\boldsymbol{\phi}$ *in probability.*

Our next task is to establish asymptotic normality of $\hat{\boldsymbol{\phi}}$.

Theorem 3.4. *Let all of the assumptions of Theorem 3.3 hold and, in addition,*

1. $m_t(\boldsymbol{\vartheta};\boldsymbol{\phi})$ *is twice continuously differentiable in* $\boldsymbol{\phi}$ *for any fixed* $\boldsymbol{\vartheta}$*;*
2. *The matrix*
$$M \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=p}^{T} E \frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma};\boldsymbol{\phi})$$
exists uniformly over $\mathcal{F} \times \Theta$ *and is continuous at* $(\boldsymbol{\sigma}^{bias}, \boldsymbol{\phi})$ *with respect to any pseudo-metric on* $\mathcal{F} \times \Theta$ *for which* $(\tilde{\boldsymbol{\sigma}}, \hat{\boldsymbol{\phi}}) \to (\boldsymbol{\sigma}^{bias}, \boldsymbol{\phi})$. *Furthermore, the matrix* M *is invertible.*

Then, $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} N(0, V)$ *where*

$$V := \Gamma^{-1} \qquad \textit{with} \qquad \Gamma := [\gamma_{|i-j|}]_{i,j=1,\dots,p} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-2} & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-3} & \gamma_{p-2} \\ \cdots & \cdots & \cdots & \cdots & & \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_1 & \gamma_0 \end{pmatrix}. \tag{3.10}$$

For simplicity, the proof, that can be found in the Appendix A, is only shown for the case $p = 2$; in that case, the covariance matrix takes the form:

$$V := \begin{pmatrix} V_1 & V_2 \\ V_2 & V_1 \end{pmatrix} := \begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}. \tag{3.11}$$

Remark 3.5. *The following observations are appropriate at this point:*

1. *The matrix Γ of (3.10) is non-singular if $\gamma_0 > 0$ and $\gamma_l \to 0$ as $l \to \infty$ (see, e.g. Brockwell and Davis (1991), Proposition 5.1.1). This means that it is non-singular for any causal AR(p) process.*

2. *In the more general model (1.4) with smooth mean function $\mu_t = g(x_t)$ and constant scaling function $\sigma_t \equiv \sigma$ (possibly unknown), Hall and Van Keilegom (2003) and Shao and Yang (2011) proposed estimators for the autocorrelation coefficients that are $T^{-1/2}$ consistent with the same asymptotic variance covariance matrix as in (3.10). Hence, despite the fact that the scaling factor $\sigma(x_t)$ may be non-constant in our setting, our estimators of autoregressive coefficients possess the parametric rate of convergence as if the conditional variance was known.*

3. *It is also important to emphasize that the methods in Hall and Van Keilegom (2003) won't apply in the presence of a varying scaling function $\sigma_t = \sigma(x_t)$. Indeed, the method therein first provides the following estimates for the autocovariances of the error process:*

$$\hat{\gamma}_0 := \frac{1}{m_2 - m_1 + 1} \sum_{m=m_1}^{m_2} \frac{1}{2(T-m)} \sum_{i=m+1}^{T} \left\{(D_m y)_i\right\}^2, \tag{3.12}$$

$$\hat{\gamma}_j := \hat{\gamma}_0 - \frac{1}{2(T-j)} \sum_{i=j+1}^{T} \left\{(D_j y)_i\right\}^2, \tag{3.13}$$

   where $(D_j y)_i := y_i - y_{i-j}$ and, as before, T is the number of observation. Here, m_1, m_2 are positive integer-valued sequences converging to ∞ in such a way that $m_2 - m_1 \to \infty$, $m_1/\log T \to \infty$, and $m_2 = o(T^{1/2})$. Then, the method proceeds to estimate the coefficients $\boldsymbol{\phi} := (\phi_1, \ldots, \phi_p)'$ using Yule-Walker equations $\Gamma\boldsymbol{\phi} = \boldsymbol{\gamma}$ (see Section 8.1 in Brockwell and Davis (1991)):

$$(\hat{\phi}_1, \ldots, \hat{\phi}_p)' := \hat{\Gamma}^{-1}(\hat{\gamma}_1, \ldots, \hat{\gamma}_p)',$$

   where $\hat{\Gamma}$ is the $p \times p$ matrix having $\hat{\gamma}_{|i-j|}$ as its $(i,j)^{th}$ entry ($i, j = 1, \ldots, p$). Such an approach clearly will not work if the scaling function σ_t is not constant.

4. *Shao and Yang's (2011) approach consists of first estimating the unknown mean function $g(x_t)$ using a B-spline smoother $\hat{g}(x_t)$ and, then, constructing estimators for the coefficients $(\hat{\phi}_1, \ldots, \hat{\phi}_p)'$ using the estimated residuals $r_t := y_t - \hat{g}(x_t)$, via the Yule-Walker equations as in Hall and Van Keilegom (2003). The autocovariances γ_j are estimated using the sample autocovariances $\hat{\gamma}_j := \frac{1}{T}\sum_{t=1}^{T-k} r_t r_{t+k}$. Again, the Yule-Walker estimation approach will not work if σ_t is not constant.*

It is interesting to ask ourselves as to what happens if we consider the more general model (1.4) with a non-trivial mean function $\mu_t = g(x_t)$. In this case, in order to estimate $\boldsymbol{\phi}$, the following algorithm is an option:

Algorithm 3.1.

1. *Estimate the mean function μ_t using, say, a local linear regression method applied to the observations $\{y_t\}_{t\in\{1,\ldots,T\}}$ (see Section 4 below for more details on this method). Let $\hat{\mu}_t$ be the resulting estimator.*

2. *Compute the centered observations $\hat{y}_t = y_t - \hat{\mu}_t$ and define the new pseudoresiduals $\hat{\eta}_t = \frac{\hat{y}_t - \hat{y}_{t-2}}{\sqrt{2}}$.*

3. *Again, obtain a biased estimator of σ_t^2 applying local linear regression to $\hat{\eta}_t^2$ and then follow* (2.11) *and* (2.12).

It turns out that, if the mean function μ_t can be estimated consistently, $\boldsymbol{\phi}$ can be estimated at the same rate of convergence as before. Note that the mean function $\mu_t = g(\frac{t}{T})$ here depends on the number of observations T. This helps us separate "trend", or a "low-frequency component", represented by the μ_t, from the zero-mean error process, that plays the role of the "high-frequency component". A more precise formulation is given in the following result, whose proof is outlined in the Appendix A.

Corollary 3.6. *Let us assume that $g''(\cdot)$ exists and is continuous on $(0,1)$. Also, let K be a kernel function with bounded support that integrates up to 1 (i.e. it is a proper density) and whose first moment is equal to 0. Select a sequence $h := h_T$ such that $h \to \infty$ and $hT^{-1} \to 0$, and let $\hat{\mu}_t$ be the corresponding sequence of local linear estimators that use the kernel $\bar{K}_h(x) := \frac{h}{T}K(\frac{h}{T}x)$. Note that it is $\frac{h}{T}$ that plays the role of bandwidth here and not just h. Then, the estimator $\hat{\boldsymbol{\phi}}$ obtained using the Algorithm 3.1 still satisfies the two assertions of Theorem 3.4.*

4 Variance function estimation

Estimating the variance function $\sigma_t^2(x)$ is very similar to how it was done in Dahl and Levine (2006). For simplicity, we will illustrate the idea only for $p = 2$. As a reminder, the first step of our proposed method is estimating not σ^2 but rather

$$\sigma^{2,bias}(x) = \frac{\sigma^2(x)}{1+\phi_2},$$

by the local linear regression applied to the squared-pseudoresiduals η_t^2. As in Dahl and Levine (2006), we assume that the kernel $K(u)$ is a two-sided proper density second order kernel on the interval $[-1,1]$; this means that

1. $K(u) \geq 0$ and $\int K(u)\,du = 0$
2. $\mu_1 = \int uK(u)\,du = 0$ and $\mu_2 \equiv \sigma_K^2 = \int u^2 K(u)\,du \neq 0$.

We also denote $R_K = \int K^2(u)\,du$. Then, the inconsistent estimator $\tilde{\sigma}^2(x)$ of $\sigma^2(x)$ is defined as the value $\hat{a}$ solving the local least squares problem

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{t=3}^{T} (\eta_t^2 - a - b(x_t - x))^2 K_m(x_t - x),$$

where as usual $K_m(x) := m^{-1}K(x/m)$. Since $\tilde{\sigma}^2(x)$ estimates $\sigma^{2,bias}(x)$ consistently, at the next step we define a consistent estimator of $\sigma^2(x)$ as follows:

$$\hat{\sigma}^2(x) = \tilde{\sigma}^2(x)(1+\hat{\phi}_2), \tag{4.1}$$

where $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \hat{\phi}_2)$ is the least-squares estimator defined in (2.11). The following lemma can be proved by following almost verbatim Theorem 3 of Dahl and Levine (2006) and is omitted for brevity. Below, we will use $D\sigma^2(x)$ and $D^2\sigma^2(x)$ to denote the first and the second-order derivatives of the function $\sigma^2(x)$, respectively.

Lemma 4.1. *Under the above assumptions* (1)-(2) *on the kernel and assumptions* (1)-(3) *on the data generating process* (1.1) *introduced in Section 3, $\tilde{\sigma}^2(x)$ is a consistent estimator of $\sigma^{2,bias}(x)$ Moreover,*

$$\frac{\tilde{\sigma}^2(x) - \sigma^{2,bias}(x) - B(\boldsymbol{\phi}, x)}{V^{1/2}(\boldsymbol{\phi}, x)} \xrightarrow{d} N(0,1),$$

where the bias $B(\boldsymbol{\phi}, x)$ and variance $V(\boldsymbol{\phi}, x)$ of $\tilde{\sigma}^2(x)$ are such that

$$B(\boldsymbol{\phi}, x) = \left\{ \frac{m^2\sigma_K^2}{2} \left[D^2\sigma^2(x)/4 - \gamma_2 (D\sigma^2(x))^2/\sigma^2(x) \right] + o(m^2) + O(T^{-1}) \right\}$$
$$V(\boldsymbol{\phi}, x) = R_K C(\phi_1, \phi_2)\sigma^4(x)(Tm)^{-1} + o(Tm^{-1}).$$

and the above constant $C(\phi_1, \phi_2)$ depends only on ϕ_1 and ϕ_2.

Now we are ready to state the main result of this section.

Theorem 4.2. *Under the same assumptions as in Lemma 4.1, the estimator $\hat{\sigma}^2(x)$ introduced in (4.1) is an asymptotically consistent estimator of $\sigma^2(x)$ that is also asymptotically normal with the bias $(1 + \phi_2)B(\boldsymbol{\phi}, x)$ and the variance $(1 + \phi_2)^2 V(\boldsymbol{\phi}, x)$.*

Proof. By the Slutsky's theorem, we have $\frac{\tilde{\sigma}^2(x) - \sigma^{2,bias}(x) - B(\boldsymbol{\phi}, x)}{\sqrt{V(\boldsymbol{\phi}, x)}}(1 + \hat{\phi}_2) \xrightarrow{d} (1 + \phi_2)\zeta$ with $\zeta \sim N(0, 1)$. This means that $\hat{\sigma}^2$ is a consistent estimator of σ_t^2 with the bias $(1 + \phi_2)B(\boldsymbol{\phi}, x)$ and the variance $(1 + \phi_2)^2 V(\boldsymbol{\phi}, x)$. $\square$

5 Numerical results

5.1 Simulation study

In this part we review the finite-sample performance of the proposed estimators. In order to do this, we consider three model specifications given in Table 1 of Appendix B. The variance function specifications are the same as those in Dahl and Levine (2006). The specification of σ_t^2 in Model 1 is a leading example in econometrics/statistics and can generate ARCH effects if $x_t = y_{t-1}$. Model 2 is adapted from Fan and Yao (1998). In particular, the choice of σ_t^2 is identical to the variance function in their Example 2. The variance function in Model 3 is from Härdle and Tsybakov (1997).

We take a fixed design $x_t = t/T$ for $t = 0, \ldots, T$ and compute the WLSE estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ of (2.11) for the previously mentioned variance function specifications and three different samples sizes, $T = 100$, $T = 1000$, and $T = 2000$. In order to assess the performance of the estimator (2.12), we compute the Mean Squared Error (MSE) defined by

$$MSE(\hat{\sigma}) := \frac{1}{M} \sum_{i=1}^{M} \frac{1}{T} \sum_{t=1}^{T} (\hat{\sigma}_{t,i}^2 - \sigma_t^2)^2,$$

where $\hat{\sigma}_{t,i}^2$ is the estimated variance function at x_t in the i^{th} simulation and M is the number of simulations. We use a local linear estimator $\tilde{\sigma}_t^2$ for estimating the biased variance function $\sigma_t^{2,bias}$ of (2.10), as in the step 1 of the Algorithm 2.1 in Section 2. The selection of the method's bandwidth was carried out by a 10-fold cross-validation method (see, e.g., Section 8.3.5 in Fan and Yao (2003) for further information).

Table 2 in Appendix B provides the MSE for the above three model specifications and sample sizes, while Table 3 shows the sampling mean and standard errors for the estimators $\hat{\phi}_1$ and $\hat{\phi}_2$. In the Table 3 of Appendix B, Mn(Sd) stand for "Mean(Standard Deviation)". We also consider three different parameter settings:

$$(1)\ (\phi_1, \phi_2) = (0.6, 0.3), \qquad (2)\ (\phi_1, \phi_2) = (0.6, -0.6), \qquad (3)\ (\phi_1, \phi_2) = (0.98, -0.6).$$

For the true parameter values $(\phi_1, \phi_2) = (0.6, 0.3)$, the asymptotic standard deviation and covariance given in (3.10) take the values:

$$\sqrt{V_1} = 0.953, \qquad V_2 = -0.780.$$

In particular, the above standard error should be compared with the asymptotic theoretical standard deviation $\sqrt{V_1/T}$ from Theorem 3.4. For the sample sizes 100, 1000, and 2000, $\sqrt{V_1/T}$ takes the values 0.0953, 0.0301, and 0.0213, which match the sampling standard deviations of Table 3 in Appendix B. The results show clear improvement for increasing sample sizes; Models 2 and 3 seem to be a little easier to estimate than Model 1.

Finally, again for $\phi_1 = 0.6$ and $\phi_2 = 0.3$, Figure 1 below shows the sampling densities for $\hat{\phi}_1$ and $\hat{\phi}_2$ corresponding to each of the three models and three sample sizes T. No severe small sample biases seem to be present in any of the pictures.

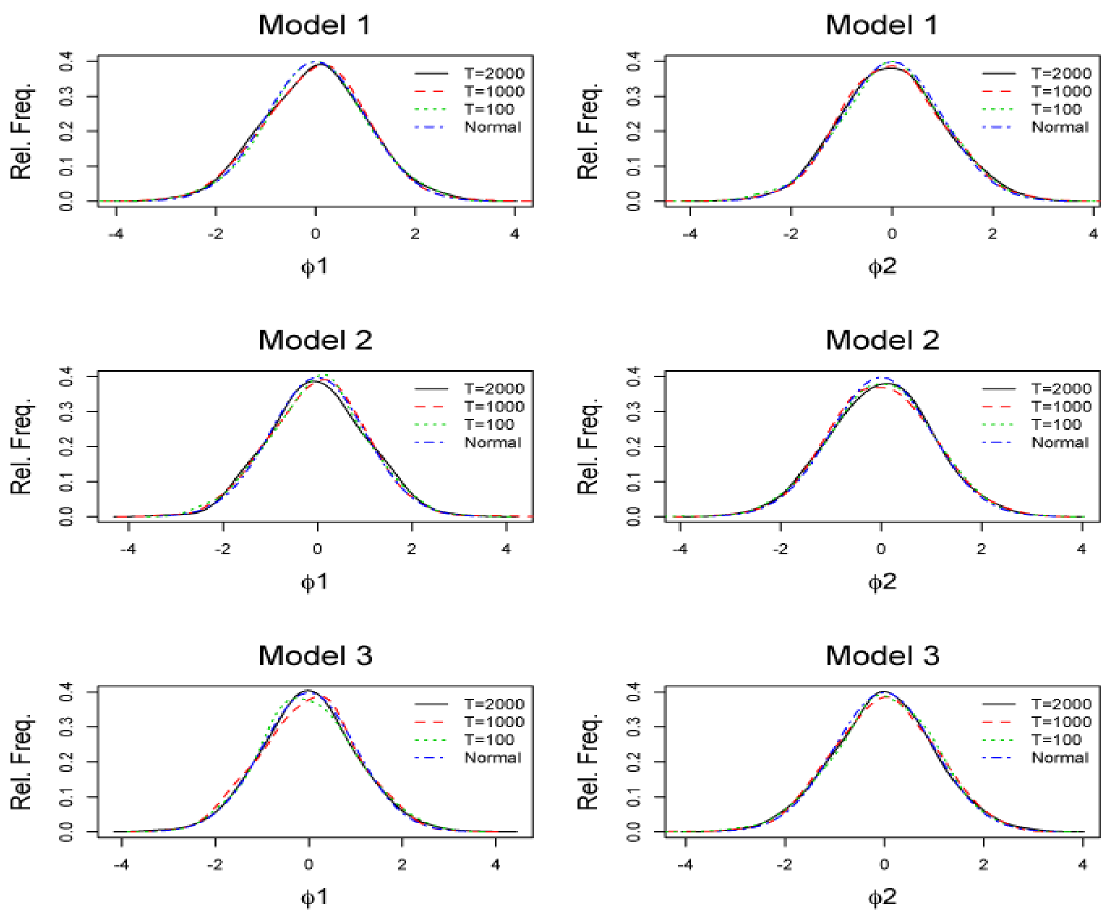

Figure 1: Sampling densities in comparison with the standard normal density under the three alternative variance function specifications of Table 1 of Appendix B. The true parameter values are $\phi_1 = 0.6$ and $\phi_2 = 0.3$. The number of Monte Carlo replications is 1000.

5.2 Real data application

We now apply our method to the first differences of annual global surface air temperatures in Celsius from 1880 through 1985[1]. This data set has been extensively analyzed in the literature using different versions of the general nonparametric regression model (1.4). For instance, Hall and Van Keilegom (2003) fixed $\sigma_t \equiv 1$ and estimated the coefficients of both the AR(1) and AR(2) models for $\{v_t\}_{t\in\{1,\dots,T\}}$, directly from the observations $\{y_t\}_{t\in\{1,\dots,T\}}$ before estimating the trend function $\mu_t = g(x_t)$ (see Remark 3.5 above for further information about their method). They reported point estimates of $\hat{\phi}_1 = 0.442$ and $\hat{\phi}_2 = -0.068$. Assuming again $\sigma_t \equiv 1$ as in Hall and Van Keilegom (2003), Shao and Yang (2011) first used linear B-splines to estimate the trend function g, and then applied a Yule-Walker type estimation method to the residuals $\hat{y}_t = y_t - \hat{\mu}_t$, under an AR(1) model specification (see Remark 3.5). They found a point estimate of $\hat{\phi}_1 = 0.386$ with a standard error 0.090.

The left panel of Figure 2 shows the scatter plot of the temperature differences against time and the estimated mean $\hat{\mu}_t = \hat{g}(x_t)$ using a simple local linear estimator applied to the observations (y_t). We then apply our estimation procedure to the differentials $\hat{y}_t = y_t - \hat{\mu}_t$, assuming an AR(2) model specification for the error process $\{v_t\}_{t\in\{1,\dots,T\}}$. The point estimates for ϕ_1 and ϕ_2 are respectively $\hat{\phi}_1 = 0.4332$ and $\hat{\phi}_2 = -0.0615$ with a standard error of 0.08 (see (3.11)). The resulting estimated variance function is shown in the right panel of Figure 2. As observed there, $\hat{\sigma}_t^2$ exhibits an interesting V-shaped pattern with minimum around the year 1945. Also, the estimated values $\hat{\phi}_1$ and $\hat{\phi}_2$ are consistent with those reported by Hall and Van Keilegom (2003).

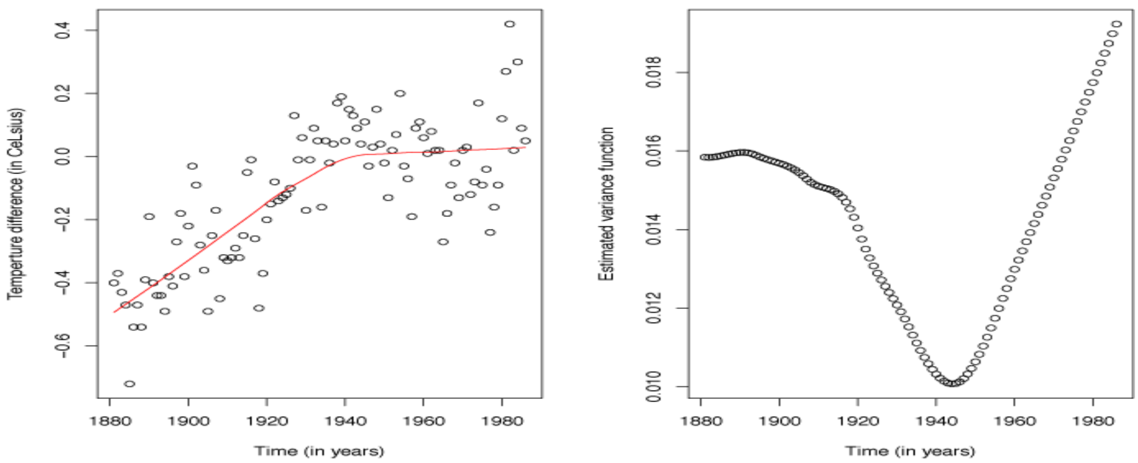

Figure 2: Differences of annual global surface air temperatures in Celsius from 1880 through 1985.

6 Discussion

In this manuscript, we propose a method for estimation of the variance structure of the scaled autoregressive process' unknown coefficients and scale variance function σ_t^2. This method is being proposed to extend earlier results of Dahl and Levine (2006), where the analogous problem for the specific case of error autoregressive process of order 1 was solved. The direct generalization of the method of Dahl and Levine (2006) does not seem to be possible in the case of the autoregressive process of order more than 1; thus, the method proposed in this manuscript represents a qualitatively new procedure.

[1]This data set was obtained from the web site http://robjhyndman.com/TSDL/meteorology/.

There is a number of interesting issues left unanswered here that we plan to address in the future research on this subject. Although we only examined the model (1.1) with the conditional mean equal to zero, an important practical issue is often the robust estimation of the parametrically specified conditional mean when the error process is conditionally heteroscedastic. As an example, an autoregressive conditional mean of order $l \geq 1$ is commonly assumed. Although standard regression procedures can be made robust in this situation by using the heteroscedasticity-consistent (HC) covariance matrix estimates as suggested in Eicker (1963) and White (1980), there may be advantages in considering alternative methods that take a specific covariance structure into account. Such methods are likely to provide more efficient estimators. This has been done in, for example, Phillips and Xu (2005) who addressed that issue by conducting asymptotic analysis of least squares estimates of the conditional mean coefficients μ_t, $t = 1, \ldots, T$ (see (1.4)) under the assumption of strongly mixing martingale difference error process and a non-constant variance function. Our setting is not a special case of Phillips and Xu (2005) since for our error process $E(v_t|\mathcal{F}_{t-1}) \neq 0$ (where $\mathcal{F}_t = \sigma(v_s, s \leq t)$ is the natural filtration). This case, to the best of our knowledge, has not been considered in the literature before. We believe, therefore, that the asymptotic analysis of the least squares estimates of the coefficients μ_t under the same assumptions on the error process as in (1.1) is an important topic for future research.

Also, being able to estimate the exact heteroscedasticity structure is important in econometrics in order to design unit-root tests that are robust to violations of the homoscedasticity assumption. The size and power properties of the standard unit-root test can be affected significantly depending on the pattern of variance changes and when they occur in the sample; an extensive study of possible heteroscedasticity effects on unit-root can be found in Cavaliere (2004). We also intend to consider the design of robust unit-root tests under the serial innovations specification as part of our future research.

Finally, yet another interesting topic of future research is a possible extension of these results to the case of a more general ARMA(p,q) error process. One of the possibilities may be using the difference-based pseudoresiduals again to construct an inconsistent estimator of the variance function σ_t^2 first. Indeed, since the scaling constant will only be dependent on the coefficients of the ARMA (p,q) error process, the MINPIN estimators of the coefficients of the error process based on such an inconsistent variance estimator will be unaffected. Therefore, estimation of the coefficients of the error process will proceed in the same way as for usual ARMA processes. The final correction of the nonparametric variance estimator also appears to be straightforward.

Acknowledgements

José E. Figueroa-López has been partially supported by the NSF grant DMS-0906919. Michael Levine has been partially supported by the NSF grant DMS-0805748. The authors gratefully acknowledge the constructive comments provided by two anonymous referees, which significantly contributed to improve the quality of the manuscript.

A Proofs

Proof of Proposition 2.1. It is easy to see that

$$\eta_t^2 = \sigma^2 \left(\gamma_0 \sum_{j=0}^{m} a_j^2 + 2 \sum_{i=1}^{m} \sum_{j=0}^{m-i} a_j a_{j+i} \gamma_i \right).$$

Recalling that for an AR(1) time series,

$$\gamma_0 = \frac{1}{1-\phi_1^2}, \quad \text{and} \quad \gamma_i = \phi_1^i \gamma_0,$$

it follows that

$$\left(\frac{1}{1-\phi_1^2} \sum_{j=0}^{m} a_j^2 + \frac{2}{1-\phi_1^2} \sum_{i=1}^{m} \sum_{j=0}^{m-i} a_j a_{j+i} \phi_1^i \right) = 1.$$

This can be written as the following polynomial of ϕ_1:

$$\sum_{j=0}^{m} a_j^2 - 1 + \phi_1^2 (1 + 2 \sum_{j=0}^{m-2} a_j a_{j+2}) + 2 \sum_{i=1, i\neq 2}^{m} \sum_{j=0}^{m-i} a_j a_{j+i} \phi_1^i = 0.$$

Then, we get the following system of equations:

$$(i) \sum_{j=0}^{m} a_j^2 = 1, \quad (ii)\ 1 + 2 \sum_{j=0}^{m-2} a_j a_{j+2} = 0,$$
$$(iii) \sum_{j=0}^{m-i} a_j a_{j+i} = 0, \quad \forall i \in \{1, 3, \ldots, m\}.$$

Suppose that $a_0 \neq 0$. Then, equation (iii) for $i = m$, implies that $a_0 a_m = 0$ and, hence, $a_m = 0$. Equation (iii) for $i = m-1$ yields $a_0 a_{m-1} + a_1 a_m = 0$ and thus $a_{m-1} = 0$. By induction, it follows that $a_m = a_{m-1} = \cdots = a_3 = 0$. Plugging in (i-iii),

$$a_0^2 + a_1^2 + a_2^2 = 1, \quad 1 + 2a_0 a_2 = 0, \quad a_0 a_1 + a_1 a_2 = 0,$$

which admits as unique solution

$$a_0 = \pm \frac{1}{\sqrt{2}}, \quad a_1 = 0, \quad a_2 = \mp \frac{1}{\sqrt{2}}.$$

If $a_0 = 0$, but $a_1 \neq 0$, one can similarly prove that the only solution is

$$a_1 = \pm \frac{1}{\sqrt{2}}, \quad a_3 = \mp \frac{1}{\sqrt{2}}, \quad a_i = 0, \text{ otherwise.}$$

The statement of the proposition can be obtained by induction in k. □

Proof of Lemma 3.2. This can be done by appealing to Theorem 1 in Andrews (1987). First, using representations (3.5)-(3.7), we define

$$W_t := (\varepsilon_t, \varepsilon_{t-1}, \ldots) \in \mathbb{R}^{\mathbb{N}}, \qquad q_{t,k}(W_t, \boldsymbol{\phi}) := \varepsilon_t \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-k-i}.$$

It remains to verify the assumptions A1-A3 of Theorem 1 in Andrews (1987). As stated in Corollary 2 of Andrews (1987), one can check its condition A4 therein instead of condition A3 since A4 implies A3. We now state these three conditions:

A1. Θ is a compact set.

A2. Let $B(\boldsymbol{\phi}_0;\rho) \subset \Theta$ be an open ball around $\boldsymbol{\phi}_0$ of radius ρ and let

$$m^*_{k,t}(\boldsymbol{\sigma};\rho) = \sup\{m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi}) : \boldsymbol{\phi} \in B(\boldsymbol{\phi}_0;\rho)\}, \tag{A.1}$$
$$m_{k,t*}(\boldsymbol{\sigma};\rho) = \inf\{m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi}) : \boldsymbol{\phi} \in B(\boldsymbol{\phi}_0;\rho)\}. \tag{A.2}$$

Then, the following two conditions are satisfied:

(a) All of $m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$, $m^*_{k,t}(\boldsymbol{\sigma};\rho)$, and $m_{k,t*}(\boldsymbol{\sigma};\rho)$ are random variables for any $\boldsymbol{\phi} \in \Theta$, any t and any sufficiently small ρ;

(b) Both $m^*_{k,t}(\boldsymbol{\sigma};\rho)$ and $m_{k,t*}(\boldsymbol{\sigma};\rho)$ satisfy pointwise weak laws of large numbers for any sufficiently small ρ.

A4. For each $\boldsymbol{\phi} \in \Theta$ there is a constant $\tau > 0$ such that $d(\tilde{\boldsymbol{\phi}}, \boldsymbol{\phi}) \leq \tau$ implies

$$||m_t(\boldsymbol{\sigma};\tilde{\boldsymbol{\phi}}) - m_t(\boldsymbol{\sigma};\boldsymbol{\phi})|| \leq B_t h(d(\tilde{\boldsymbol{\phi}}, \boldsymbol{\phi}))$$

where B_t is a nonnegative random variable (that may depend on $\boldsymbol{\phi}$) and $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T EB_t < \infty$, while $h : R^+ \to R^+$ is a nonrandom function such that $h(y) \downarrow h(0) = 0$ as $y \downarrow 0$. Above, $|| \cdot ||$ denotes the standard Euclidean norm in $\mathbb{R}^p$.

Since condition A1 above is assumed as a hypothesis, we only need to work with conditions A2 and A4. The verification of these is done through the following two steps:

1. Let $\rho > 0$ small enough such that $B(\boldsymbol{\phi}_0;\rho) \subset \Theta$. Then, recalling the representation (3.5) and (3.7),

$$m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi}) = \varepsilon_t \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1-i}, \tag{A.3}$$

we find that the supremum of $m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$, taken over $\boldsymbol{\phi} \in B(\boldsymbol{\phi}_0;\rho)$, also exists and is a random variable. Indeed, each of the summands in (A.3) is a continuous function of $\boldsymbol{\phi}$ as we already established earlier, the convergence in mean squared to $m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ is uniform in $\boldsymbol{\phi}$ due to (3.8) and, therefore, $m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ is continuous in $\boldsymbol{\phi}$ as well. That, in turn, implies the existence of $m^*_{1,t}(\boldsymbol{\sigma};\rho)$. The existence of $m_{1,t*}(\boldsymbol{\sigma};\rho)$ is established in exactly the same way. Moreover, the pointwise WLLNs for both $\sup_{\boldsymbol{\phi}\in B} m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ and $\inf_{\boldsymbol{\phi}\in B} m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ are also clearly satisfied since $E\, m_{k,t}(\boldsymbol{\sigma};\boldsymbol{\phi}) \equiv 0$ for any $\boldsymbol{\phi} \in B(\boldsymbol{\phi};\rho) \subset \Theta$.

2. We now show condition A4 above for $m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ ($m_{2,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ can be treated analogously). Denote $m^*_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi}) = \varepsilon_t \sum_{i=0}^{\infty} \psi^*_i \varepsilon_{t-1-i}$ where ψ^*_i correspond to the MA(∞) representation of the AR(2) series with the parameter vector $\boldsymbol{\phi}^* = (\boldsymbol{\phi}^*_1, \boldsymbol{\phi}^*_2) \in \Theta$. For the sake of brevity we will use m^*_1 and m_1 for $m^*_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$ and $m_{1,t}(\boldsymbol{\sigma};\boldsymbol{\phi})$, respectively. Note that

$$\begin{aligned} |m^*_1 - m_1| &= |\varepsilon_t \sum_{i=0}^{\infty} \psi^*_i \varepsilon_{t-1-i} - \varepsilon_t \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-1-i}| \\ &\leq \sqrt{\sum_{i=0}^{\infty} \psi_i^2 \varepsilon_{t-1-i}^2 \varepsilon_t^2} \sqrt{\sum_{i=0}^{\infty} \left(\frac{\psi^*_i - \psi_i}{\psi_i}\right)^2}. \end{aligned}$$

Let $B_t := \sqrt{\sum_{i=0}^{\infty} \psi_i^2 \varepsilon_{t-1-i}^2 \varepsilon_t^2}$ and

$$d(\boldsymbol{\phi}^*, \boldsymbol{\phi}) := \sum_{i=0}^{\infty} \left(\frac{\psi^*_i - \psi_i}{\psi_i}\right)^2.$$

Then,

$$\sup_T \frac{1}{T}\sum_{t=2}^{T} EB_t \le \sup_T \frac{1}{T}\sum_{t=2}^{T}\sqrt{E\left(\sum_{i=0}^{\infty}\psi_i^2\varepsilon_{t-1-i}^2\varepsilon_t^2\right)} = \sqrt{\sum_{i=0}^{\infty}\psi_i^2} < \infty.$$

Now we need to treat the second multiplicative term. First, recalling (3.9), it is easy to conclude, by induction, that the coefficients ψ_j are continuously differentiable functions of $\boldsymbol{\phi} = (\phi_1, \phi_2)$ for any $\boldsymbol{\phi} \in B(\boldsymbol{\phi}, \rho) \subset \Theta$. Therefore, using (3.8), one can easily establish that

$$\lim_{\boldsymbol{\phi}^* \to \boldsymbol{\phi}} d(\boldsymbol{\phi}^*, \boldsymbol{\phi}) = 0.$$

Note that the quantity $\sum_{i=0}^{\infty}\left(\frac{\psi_i^* - \psi_i}{\psi_i}\right)^2$ is not a metric and, therefore, the verification of Assumption A4 seems in doubt at first sight. However (see Andrews (1992)), the fact that Assumption A4 implies Assumption A3 does not need the argument $d(\tilde{\boldsymbol{\phi}}, \boldsymbol{\phi})$ of the function h to be a proper metric but only that $d(\tilde{\boldsymbol{\phi}}, \boldsymbol{\phi}) \to 0$ as $\tilde{\boldsymbol{\phi}} \to \boldsymbol{\phi}$.

□

Proof of Theorem 3.3. Our proof of consistency will rely on the Theorem A1 of Andrews (1994) with $W_t = (y_t, y_{t-1}, y_{t-2})'$. We will simply verify that the sufficient conditions of Theorem A1 are true. The first assumption C(a) follows from Lemma 3.2 taking $m_t(\boldsymbol{\sigma}; \boldsymbol{\phi}) \equiv 0$. It remains to show the other conditions therein.

The first part of Assumption C(b) of the Theorem A is immediately satisfied because $m(\boldsymbol{\sigma}; \boldsymbol{\phi}) \equiv 0$ for any $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$. Since $\tilde{\sigma}_t^2$ is a local linear regression estimator, it is clear that it is twice continuously differentiable as long as the kernel function $K()$ is twice continuously differentiable; thus, the second part of of the Assumption C(b) is also true. The Assumption C(c) is true if the Euclidean norm of $m(\boldsymbol{\sigma}; \boldsymbol{\phi}) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} E\, m_t(\boldsymbol{\sigma}; \boldsymbol{\phi})$ is finite. Hereafter, we denote the euclidian norm by $||\cdot||$. In our case, since both martingale difference sequences $v_{t-1}\varepsilon_t$ and $v_{t-2}\varepsilon_t$ have mean zero, clearly $\sup_{\Theta\times\mathcal{F}} ||m(\boldsymbol{\sigma}; \boldsymbol{\phi})|| = 0 < \infty$ and the Assumption C(c) is satisfied. The Assumption C(d) is true because Θ is compact, the functional

$$d_t = m_t^{'}(\boldsymbol{\sigma}; \boldsymbol{\phi}) m_t(\boldsymbol{\sigma}; \boldsymbol{\phi})/2,$$

is continuous in $\boldsymbol{\phi}$ and the Hessian matrix $\frac{\partial^2 d_t}{\partial \boldsymbol{\phi}^2}$ is positive definite (can be verified). All of the above allows us to conclude that the weak consistency holds: $\hat{\boldsymbol{\phi}} \xrightarrow{\mathbb{P}} \boldsymbol{\phi}$. □

Proof of Theorem 3.4. Recall that $\tilde{\sigma}_t$ is the inconsistent estimator of σ_t which, nevertheless, estimates the quantity $\sigma_t^{bias} = \frac{\sigma_t}{\sqrt{1+\phi_2}}$ consistently (see Lemma 4.1 above); it is corrected to obtain $\hat{\sigma}_t = \tilde{\sigma}_t(1 + \hat{\phi}_2)$. We also recall the notation (3.1) and define, for some generic argument $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_T)$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$,

$$\bar{\mathbf{m}}_T(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := \frac{1}{T}\sum_{t=2}^{T} m_t(\boldsymbol{\vartheta}; \boldsymbol{\varphi}),$$

with $m_t(\boldsymbol{\vartheta}, \boldsymbol{\varphi})$ given as in (3.4). Since the vector valued function $m_t(\boldsymbol{\vartheta}; \boldsymbol{\varphi})$ is twice continuously differentiable, we can use a Taylor expansion around $(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi})$ to get

$$\sqrt{T}\,\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\sigma}}; \hat{\boldsymbol{\phi}}) = \sqrt{T}\,\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi}) + \frac{\partial}{\partial \boldsymbol{\phi}}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi}^*)\sqrt{T}\,(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}), \tag{A.4}$$

for some $\boldsymbol{\phi}^*$ that lies on the straight line connecting $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$. Note that, due to the first order conditions (3.6), the left-hand side of the equation (A.4) cancels out. Hence, using the consistency of $\hat{\boldsymbol{\phi}}$ for $\boldsymbol{\phi}$ (Theorem 3.3) and the assumption 2 in the statement of the theorem, we obtain that

$$\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) = -[\bar{M}^{-1} + o_p(1)]\sqrt{T}\,\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi}),$$

provided that the matrix

$$\bar{M} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} E\,\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) \tag{A.5}$$

exists and is invertible. In order to verify the latter condition, note that in our case,

$$\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) = \left.\begin{pmatrix} \frac{\partial m_{1,t}}{\partial \phi_1} & \frac{\partial m_{1,t}}{\partial \phi_2} \\ \frac{\partial m_{2,t}}{\partial \phi_1} & \frac{\partial m_{2,t}}{\partial \phi_2} \end{pmatrix}\right|_{(\boldsymbol{\sigma}^{bias}, \boldsymbol{\phi})} = (1+\phi_2)^{-1}\begin{pmatrix} -v_{t-1}^2 & -v_{t-1}v_{t-2} \\ -v_{t-1}v_{t-2} & -v_{t-2}^2 \end{pmatrix},$$

In light of (2.6), it follows that

$$\bar{M} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} E\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) = -(1+\phi_2)^{-1}\begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{pmatrix},$$

which, in particular, is clearly invertible.

Next, let $\mathbf{m}_T^*(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) = \frac{1}{T}\sum_{t=2}^{T} E\,m_t(\boldsymbol{\vartheta}; \boldsymbol{\varphi})$ and define the empirical process

$$\nu_T(\boldsymbol{\vartheta}) = \sqrt{T}\,[\bar{\mathbf{m}}_T(\boldsymbol{\vartheta}; \boldsymbol{\phi}) - \mathbf{m}_T^*(\boldsymbol{\vartheta}; \boldsymbol{\phi})] \tag{A.6}$$

Clearly,

$$\sqrt{T}\bar{\mathbf{m}}_T(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi}) = \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) + \nu_T(\tilde{\boldsymbol{\sigma}}) - \nu_T(\boldsymbol{\sigma}^{bias}) + \sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi})$$

Now, using (3.5),

$$\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) = \frac{1}{\sqrt{T}}\sum_{t=2}^{T} m_t(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi}) = \frac{1}{1+\phi_2}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T} v_{t-1}\varepsilon_t, \frac{1}{\sqrt{T}}\sum_{t=2}^{T} v_{t-2}\varepsilon_t\right).$$

Hence, using the CLT for martingale difference sequences (see Billingsley (1961)), $\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\sigma}^{bias}; \boldsymbol{\phi})$ is asymptotically normal $N(0, (1+\phi_2)^{-2}S)$, where

$$S := \begin{pmatrix} Ev_{t-1}^2\varepsilon_t^2 & Ev_{t-1}v_{t-2}\varepsilon_t^2 \\ Ev_{t-1}v_{t-2}\varepsilon_t^2 & Ev_{t-2}^2\varepsilon_t^2 \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{pmatrix}.$$

Note that $\bar{M} = -(1+\phi_2)^{-1}S$. Then, if we can show that

$$\nu_T(\tilde{\boldsymbol{\sigma}}) - \nu_T(\boldsymbol{\sigma}^{bias}) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\phi}) \xrightarrow{\mathbb{P}} 0, \tag{A.7}$$

our task is over and we can say that $\sqrt{T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} N(0, V)$ with

$$\begin{aligned} V &= \bar{M}^{-1}((1+\phi_2)^{-2}S)(\bar{M}^{-1})' = (-(1+\phi_2)^{-1}S)^{-1}((1+\phi_2)^{-2}S)(-(1+\phi_2)^{-1}S')^{-1} \\ &= S^{-1} = \frac{1}{\gamma_0^2 - \gamma_1^2}\begin{pmatrix} \gamma_0 & -\gamma_1 \\ -\gamma_1 & \gamma_0 \end{pmatrix}. \end{aligned}$$

A simple computation using (2.6)-(2.7) leads to (3.10). We show (A.7) through the following two steps:

(1) First, denoting the indicator function of an event A by $\mathbf{1}_A$, we have, for any $\varepsilon > 0$,

$$\sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}};\boldsymbol{\phi}) = \sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}};\boldsymbol{\phi})\mathbf{1}_{\{\rho(\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}^{bias})>\varepsilon\}} + \sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}};\boldsymbol{\phi})\mathbf{1}_{\{\rho(\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}^{bias})\le\varepsilon\}}.$$

The first term on the right is $o_p(1)$ due to consistency of $\tilde{\boldsymbol{\sigma}}$ as an estimator of $\boldsymbol{\sigma}^{bias}$. Similarly, the second term therein is $o_p(1)$ due to the fact that $E\,m_t(\boldsymbol{\sigma}^{bias};\boldsymbol{\phi}) = 0$ and $Em_t(\boldsymbol{\vartheta};\boldsymbol{\varphi})$ is uniformly continuous in $\boldsymbol{\vartheta}$. We then conclude that $\sqrt{T}\mathbf{m}_T^*(\tilde{\boldsymbol{\sigma}};\boldsymbol{\phi}) \xrightarrow{\mathbb{P}} 0$

(2) We can show that $\nu_T(\tilde{\boldsymbol{\sigma}}) - \nu_T(\boldsymbol{\sigma}^{bias}) \xrightarrow{\mathbb{P}} 0$ using the argument of Andrews (1994), pp. 48-49, that requires, first, establishing stochastic equicontinuity of $\nu_T(\boldsymbol{\vartheta})$ at $\boldsymbol{\sigma}^{bias}$ and then deducing the required convergence in probability. To do this, note first that

$$T^{-1/2}\sum_{t=2}^{T}\left(v_{t-1}^2 - E\,v_{t-1}^2\right) = O_p(1), \quad T^{-1/2}\sum_{t=2}^{T} v_{t-k}v_{t-1-k} = O_p(1), \tag{A.8}$$

for $k = 0, 1$. The empirical process $\nu_T(\boldsymbol{\vartheta})$ has its values in R^2; examining its first coordinate $\nu_T^1(\boldsymbol{\vartheta})$, one easily obtains $\nu_T^1(\boldsymbol{\vartheta}) = T^{-1/2}\sum_{t=2}^{T} S_t^{'}\tau_t(\boldsymbol{\vartheta})$, where

$$S_t = \left(y_t y_{t-1}, \phi_1 y_{t-1}^2, \phi_2 y_{t-1}y_{t-2}\right)', \qquad \tau_t(\boldsymbol{\vartheta}) = (\vartheta_t^{-1}\vartheta_{t-1}^{-1}, -\vartheta_{t-1}^{-2}, -\vartheta_{t-1}^{-1}\vartheta_{t-2}^{-1})'.$$

Then, denoting the standard Euclidean norm in $\mathbb{R}^3$ by $||\cdot||$, for $\eta > 0$ and $\delta > 0$, we have

$$\begin{aligned}
&\limsup_{T\to\infty} P\left(\sup_{\rho(\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}^{bias})<\delta} |\nu_T^1(\tilde{\boldsymbol{\sigma}}) - \nu_T^1(\boldsymbol{\sigma}^{bias})| > \eta\right) \\
&= \limsup_{T\to\infty} P\left(\sup_{\rho(\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}^{bias})<\delta} \left|T^{-1/2}\sum_{t=2}^{T}(S_t' - E\,S_t')(\tilde{\tau}_t(\tilde{\boldsymbol{\sigma}}) - \tau_t(\boldsymbol{\sigma}^{bias}))\right| > \eta\right) \\
&\le \limsup_{T\to\infty} P\left(\sup_{\rho(\tilde{\boldsymbol{\sigma}},\boldsymbol{\sigma}^{bias})<\delta} \left\|T^{-1/2}\sum_{t=2}^{T}(S_t - E\,S_t)\right\| > \frac{\eta}{\delta}\right) < \varepsilon,
\end{aligned}$$

for $\delta > 0$ small enough, since $T^{-1/2}\sum_{t=2}^{T}(S_t - E\,S_t) = O_p(1)$ due to (A.8). Using exactly the same argument, one easily obtains stochastic equicontinuity of the second coordinate $\nu_T^2(\boldsymbol{\vartheta})$ at $\boldsymbol{\sigma}^{bias}$; therefore, needed convergence in probability for $\nu_T(\tilde{\boldsymbol{\sigma}}) - \nu_T(\boldsymbol{\sigma}^{bias})$ follows from the convergence in both coordinates separately and the Slutsky's theorem. □

Proof of Corollary 3.6. The following is a sketch of the proof for the case of $p = 2$. The conditions given in the corollary are not the least restrictive; they guarantee consistency of the estimator $\hat{\mu}_t \equiv g(x_t)$ and the order $O((h/T)^2)$ for the bias; see Theorem 6.1 in Fan and Yao (2003) for details. Note that the model (1.4) can be represented in the functional autoregressive form

$$\sigma_t^{-1}y_t = f_t(\boldsymbol{\phi}) + \phi_1\sigma_{t-1}^{-1}y_{t-1} + \phi_2\sigma_{t-2}^{-1}y_{t-2} + \varepsilon_t$$

where the new mean function is $f_t(\boldsymbol{\phi}) = \sigma_t^{-1}g(x_t) - \phi_1\sigma_{t-1}^{-1}g(x_{t-1}) - \phi_2\sigma_{t-2}^{-1}g(x_{t-2})$. Recall that functions $m_{k,t}(\boldsymbol{\phi}) \equiv v_{t-k}\varepsilon_t$, $k = 1, 2$ do not depend on σ_t; therefore, (3.5) is still true and the Lemma 3.2 is also true as well. This implies that the Theorem 3.3 is true as well. Finally , when trying to prove Theorem 3.4, one quickly discovers that the matrix $\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\phi})$ is now different; for example, $\frac{\partial m_{1,t}}{\partial \phi_2} = \gamma_0 - g(x_{t-1})\sigma_{t-1}^{-1}v_{t-1}$. Nevertheless, since the second term of the above has the mean zero, its expectation is the same as before and the same is true of the other four elements of the matrix; thus, matrix M stays the same. Finally, the vector $S_t = ((y_t - \mu_t(y_t - \mu_{t-1}), \phi_1(y_{t-1} - \mu_{t-1})^2, \phi_2(y_{t-1} - \mu_{t-1})(y_{t-2} - \mu_{t-2}))'$ is now different; nevertheless, it is easy to verify that $T^{-1/2}\sum_{t=2}^{T}(S_t - E\,S_t) = O_p(1)$ again; thus the final result is still valid. □

References

Andrews, D.W.K. (1987) "Consistency in nonlinear econometric models: a generic uniform law of large numbers", Econometrica 55, 1465-1471.

Andrews, D.W.K. (1994) "Asymptotics for semiparametric econometric models via stochastic equicontinuity", Econometrica 62, 43-72.

Billingsley, P. (1961) "The Lindeberg-Levy Theorem for Martingales", in Proceedings of American Mathematical Society, Vol 12, 788-792.

Brockwell, P., and Davis, R. (1991) Time series: theory and methods. Springer.

Cavaliere, G. (2004) "Unit root tests under time-varying variance shifts", Econometric Reviews 23, 259-92.

Chen, R. and Tsay, R.S. (1993) "Functional coefficient autoregressive models", Journal of American Statistical Association, 88, 298–308.

Dahl, C. and Levine, M. (2006) "Nonparametric estimation of volatility models with serially dependent innovations", Statistics and Probability Letters, 76(18), 2007–2016.

Eicker, B. (1963) "Limit theorems for regression with unequal and dependant errors", Annals of Mathematics and Statistics 34, 447-456.

Fan, J. and Yao, Q. (1998) "Efficient estimation of conditional variance functions in stochastic regression", Biometrika, 85, 645-660.

Fan, J. and Yao, Q. (2003) Nonlinear time series - nonparametric and parametric methods. Springer.

Hall, P. and Van Keilegom, I. (2003) "Using difference-based methods for inference in nonparametric regression with time series errors", Journal of the Royal Statistical Society (B), 85(2), 443–456.

Härdle, W. and Tsybakov, A. (1997) "Local polynomial estimators of the volatility function in nonparametric autoregression", Journal of Econometrics, 81, 223-242.

Phillips, P.C.B., and Xu, K.-L. (2005) "Inference in Autoregression under heteroscedasticity", Journal of Time Series Analysis 27, 289-308.

Shao, Q. and Yang, L. (2011) "Autoregressive coefficient estimation in nonparametric analysis", Journal of Time Series Analysis 32, 587-597.

White, H. (1980) "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity", Econometrica 48, 817-838.

B Tables

Model	Specifications
1	$\sigma_t^2 = 0.5x_t^2 + 0.1$
2	$\sigma_t^2 = 0.4\exp(-2x_t^2) + 0.2$
3	$\sigma_t^2 = \varphi(x_t + 1.2) + 1.5\varphi(x_t - 1.2)$

Table 1: Alternative data generating processes. $\varphi(\cdot)$ denotes the standard normal probability density.

	$\phi_1 = 0.6, \quad \phi_2 = 0.3$			$\phi_1 = 0.6, \quad \phi_2 = -0.6$			$\phi_1 = 0.98, \quad \phi_2 = -0.6$		
Model	T=100	T=1000	T=2000	T=100	T=1000	T=2000	T=100	T=1000	T=2000
1	0.0241	0.0013	0.0007	0.0220	0.0024	0.0014	0.0271	0.0024	0.0014
2	0.0388	0.0020	0.0009	0.0404	0.0025	0.0018	0.0552	0.0037	0.0018
3	0.0626	0.0026	0.0011	0.0576	0.0019	0.0019	0.0660	0.0041	0.0019

Table 2: Mean Squared Errors (MSE) of $\hat{\sigma}^2(x)$ under the variance function specifications of Table 1 with 1000 Monte Carlo replications and a 10-fold cross-validation for bandwidth selection.

$\phi_1 = 0.6, \quad \phi_2 = 0.3$						
Model	$T = 100$		$T = 1000$		$T = 2000$	
	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$
1	0.559(0.114)	0.298(0.110)	0.594(0.029)	0.301(0.030)	0.595(0.021)	0.301(0.021)
2	0.565(0.105)	0.296(0.098)	0.594(0.029)	0.301(0.030)	0.598(0.021)	0.299(0.021)
3	0.566(0.103)	0.294(0.099)	0.598(0.030)	0.298(0.0302)	0.597(0.022)	0.300(0.021)
$\phi_1 = 0.6, \quad \phi_2 = -0.6$						
	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$
1	0.554(0.111)	-0.535(0.104)	0.595(0.026)	-0.591(0.024)	0.597(0.017)	-0.595(0.017)
2	0.561(0.100)	-0.542(0.099)	0.597(0.025)	-0.595(0.025)	0.596(0.017)	-0.596(0.018)
3	0.564(0.100)	-0.546(0.097)	0.596(0.025)	-0.594(0.025)	0.597(0.017)	-0.596(0.017)
$\phi_1 = 0.98, \quad \phi_2 = -0.6$						
	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$	Mn(Sd) $\hat{\phi}_1$	Mn(Sd) $\hat{\phi}_2$
1	0.894(0.140)	-0.517(0.118)	0.972(0.026)	-0.591(0.025)	0.975(0.017)	-0.594(0.018)
2	0.919(0.108)	-0.535(0.099)	0.973(0.025)	-0.592(0.025)	0.976(0.017)	-0.596(0.017)
3	0.917(0.102)	-0.535(0.098)	0.974(0.026)	-0.594(0.026)	0.976(0.017)	-0.596(0.018)

Table 3: Sampling Means and Standard Deviations under under the variance function specifications of Table 1 with 1000 Monte Carlo replications.

†Department of Statistics, 250 N. University St., Department of Statistics, Purdue University, West Lafayette, IN 47907-2066. E-mail: figueroa@purdue.edu

*Department of Statistics, 250 N. University St., Department of Statistics, Purdue University, West Lafayette, IN 47907-2066. E-mail: mlevins@purdue.edu.