Nonparametric Estimation of Volatility Models with General Autoregressive Innovations

by

J. E. Figueroa-López
Purdue University

M. Levine
Purdue University

Technical Report #11-04

Department of Statistics
Purdue University

December 2011

Nonparametric estimation of volatility models with general autoregressive innovations

José E. Figueroa-López
Michael Levine

Keywords and phrases: autoregressive error process, heteroscedastic, semiparametric estimators, difference based estimation approach

Abstract

We are interested in modeling a zero mean heteroscedastic time series process with autoregressive error process of finite known order p. The model can be used for testing a martingale difference sequence hypothesis that is often adopted uncritically in financial time series against a fairly general alternative. When the argument is deterministic, 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.

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. Consider the following non-parametric volatility model with serially correlated innovations:

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

$$v_t = \phi_1 v_{t-1} + \phi_2 v_{t-2} + \varepsilon_t, \tag{1.2}$$

for $t = 2, \ldots, T$, where $(\varepsilon_t)_{t \geq 2}$ are iid with mean 0 and variance 1. This model can also be interpreted as belonging to the a functional autoregressive model (FAR) class first introduced in Chen and Tsay (1993). Indeed, the process y_t can be re-expressed as

$$y_t = \phi_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1} + \phi_2 \sigma_t \sigma_{t-2}^{-1} y_{t-2} + \sigma_t \varepsilon_t, \tag{1.3}$$

with functional coefficients being $\phi_1 \sigma_t \sigma_{t-1}^{-1}$ and $\phi_2 \sigma_t \sigma_{t-2}^{-1}$. If the AR(2) time series process is stationary, the variance and autocovariances of v are:

$$\gamma_0 := \mathrm{Var}(v_k) = \frac{1-\phi_2}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}, \tag{1.4}$$

$$\gamma_1 := \mathrm{Cov}(v_k, v_{k+1}) = \frac{\phi_1}{1-\phi_2}\gamma_0, \tag{1.5}$$

$$\gamma_j := \mathrm{Cov}(v_k, v_{k+j}) = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, \quad \text{for any} \quad k \in \mathbb{N}. \tag{1.6}$$

The functional autoregressive representation (1.3) suggests that the direct attempt to estimate the conditional variance function σ_t^2 based on the residuals from the regression of y_t on y_{t-1} and y_{t-2} will produce an inconsistent estimator due to the time-dependent nature of the coefficients of y_{t-1} and y_{t-2}. This potential source of inconsistency has often been ignored in the econometric literature (e.g. when estimating ARCH or GARCH models) due to the uncritical adoption of the assumption that errors make up a martingale difference sequence. Note that as a by-product of our analysis, a simple parametric test of the martingale difference hypothesis, i.e., $\phi_1 = 0$ and $\phi_2 = 0$, can be proposed that enables the researcher to avoid this potential pitfall. Estimation of the conditional variance structure is also of interest in several situations. For instance, consider the modified model (1.1):

$$y_t = \mu_t + \sigma_t v_t,$$

where the conditional mean function μ is non-zero; for example, a commonly encountered case is the p^{th} order autoregression $\mu(y_{t-1}, \ldots, y_{t-p}) = \theta_0 + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p}$ with $\theta_p \neq 0$ (see, e.g., Phillips and Xu (2006)). In this situation, there is a need for robust inference concerning the coefficients θ_i, $i = 1, \ldots, p$, in the presence of the specific type of heteroscedasticity. 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. Unlike Phillips and Xu (2006), we consider the conditional variance structure with serially dependent errors, which hasn't been seriously investigated in the literature to the best of our knowledge. Finally, 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 intend to consider both design of robust unit-root tests and the robust inference for autoregression coefficients in the presence of conditional variance structure with serially dependent errors as part of the future research.

A simpler version of the model (1.1) with AR(1) innovations (i.e. $\phi_2 = 0$) has been studied earlier in Dahl and Levine (2006). They designed a simple and intuitively appealing procedure for estimating both Euclidean parameter ϕ_1 and the function σ_t^2. The key building blocks of their estimation method were the two-lag difference statistic:

$$\eta_t = \frac{y_t - y_{t-2}}{\sqrt{2}}. \tag{1.7}$$

Such a statistic is also often called a pseudoresidual. In order to explain in simple terms the intuition behind their method, let us assume that σ is constant. 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,

$$\mathbb{E}\eta_t^2 = \sigma^2(\gamma_0 - \gamma_2) = \sigma^2, \tag{1.8}$$

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 devise a consistent estimator for the non-constant function σ_t^2 as well. In the case of non-constant σ_t and under a fixed design on the unit interval ($x_t = t/T$, $t = 1, \ldots, T-1$), we have

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

Simple heuristics suggests 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

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

where $\tilde{\varepsilon}_t$ are approximately centered random errors. Dahl and Levine (2006) used local linear estimation to estimate σ_t^2, whose asymptotic properties were also fully characterized. The parameter ϕ_1 was estimated using a weighted least square estimator (WLSE). More specifically, noting that (1.3) with $\phi_2 = 0$ implies

$$\sigma_t^{-1} y_t = \phi_1 \sigma_{t-1}^{-1} y_{t-1} + \varepsilon_t, \quad t = 2, \dots, T, \tag{1.9}$$

it follows that a natural estimator for ϕ_1 is given by

$$\begin{aligned} \hat{\phi}_1 &:= \arg\min_{\phi_1 \in (0,1)} \frac{1}{T} \sum_{t=2}^{T} \left(\hat{\sigma}_t^{-1} y_t - \phi_1 \hat{\sigma}_{t-1}^{-1} y_{t-1} \right)^2 \\ &= \left(\frac{1}{T} \sum_{t=2}^{T} \hat{\sigma}_{t-1}^{-2} y_{t-1}^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=2}^{T} \hat{\sigma}_t^{-1} \hat{\sigma}_{t-1}^{-1} y_t y_{t-1} \right). \end{aligned} \tag{1.10}$$

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 of our estimators is given 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 the previous section, we showed how to apply the two-lag difference statistic (1.7) to estimate the conditional variance function in the model (1.1) in light of (1.8). A natural question is whether there exists any other linear statistic

$$\eta_t := \sum_{i=0}^{m} a_i y_{t-i}, \tag{2.1}$$

such that

$$\mathbb{E}\eta_t^2 \approx \sigma_t^2$$

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 follow from similar arguments. 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., $\phi_2 = 0$). Then, if*

$$\mathbb{E}\eta_t^2 = \sigma^2,$$

for any $\phi_1 \in (-1, 1)$, there exists a $0 \le k \le m - 2$ 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 (2.1) with $a_0 \neq 0$ that can result in $\mathbb{E}\eta_t^2$ being independent of ϕ_1 is the two-lag difference statistic

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

Now we intend to show that the same statistic can be used in the case of AR(2) process for the general model (1.1)-(1.2). The key observation is that for a general AR(2) innovation process and for $\sigma_t^2 \equiv \sigma$, the equation (1.8) simplifies nicely as follows:

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

Indeed, using (1.4)-(1.5),

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

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

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

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:

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

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 for now the 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

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

where

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

We finally obtain

$$\bar{\phi}_2 := (\bar{A}^2 - \bar{B}^2)^{-1} (\bar{A}\bar{C} - \bar{B}^2), \quad \bar{\phi}_1 = \bar{A}^{-1} \bar{B} (1 - \hat{\phi}_2). \tag{2.2}$$

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 algorithm:

1. Estimate the function

$$\sigma_t^{2,bias} := \frac{\sigma^2(x_t)}{1 + \phi_2}, \tag{2.3}$$

   by a non-parametric smoothing method (e.g. local linear regression) applied to the two-lag difference statistics η_t^2 defined in (1.7). 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:

$$\hat{\phi}_2 := (A^2 - B^2)^{-1}(AC - B^2), \quad \hat{\phi}_1 = A^{-1}B(1 - \hat{\phi}_2). \tag{2.4}$$

   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

$$\hat{\sigma}_t^2 := (1 + \hat{\phi}_2)\tilde{\sigma}_t^2. \tag{2.5}$$

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

3 Asymptotics

Let us now consider the estimation problem for the 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 applying local linear regression to the squared-pseudoresiduals η_t^2; such an estimator is inconsistent since, e.g., in the homoscedastic model case (i.e. $\sigma_t^2 \equiv \sigma^2$), $\mathbb{E}\eta_t^2 = \sigma^2/(1+\phi_2)$. However, note that it is expected to be a *consistent* estimator of the quantity $\sigma_t^{2,bias} = \frac{\sigma_t^2}{1+\phi_2}$, as it will be proved in Section 4. We denote

$$\boldsymbol{\sigma}_t = (\sigma_t, \sigma_{t-1}, \sigma_{t-2})', \quad \tilde{\boldsymbol{\sigma}}_t = (\tilde{\sigma}_t, \tilde{\sigma}_{t-1}, \tilde{\sigma}_{t-2})', \quad \boldsymbol{\sigma}_t^{bias} = (\sigma_t^{bias}, \sigma_{t-1}^{bias}, \sigma_{t-2}^{bias})'.$$

As it was explained before, it seems reasonable to estimate the coefficients ϕ_1 and ϕ_2 using an inconsistent estimator $\tilde{\sigma}_t^2$ first and, then, correct it to obtain the asymptotically consistent estimator $\hat{\sigma}_t^2 = \tilde{\sigma}_t^2(1 + \hat{\phi}_2)$. The following detailed algorithm illustrates our approach to the estimation of the model (1.1):

1. Using the functional autoregression form of the model (1.1), define the least squares estimator $\hat{\boldsymbol{\phi}} := (\hat{\phi}_1, \hat{\phi}_2)$ of $\boldsymbol{\phi} := (\phi_1, \phi_2)$ and establish its consistency.

2. Show that, under additional regularity conditions, $\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}$ as $T \to \infty$.

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

We recall that the functional autoregressive form of (1.1) is

$$\sigma_t^{-1} y_t = \phi_1 \sigma_{t-1}^{-1} y_{t-1} + \phi_2 \sigma_{t-2}^{-1} y_{t-2} + \varepsilon_t.$$

Next, for any $\boldsymbol{\vartheta} = (\vartheta_0, \vartheta_{-1}, \vartheta_{-2})$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$, let

$$\bar{m}_{k,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := \frac{1}{T} \sum_{t=2}^{T} m_{k,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}), \qquad (k = 1, 2), \tag{3.1}$$

where

$$m_{k,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := \vartheta_{-k}^{-1} y_{t-k} [\vartheta_0^{-1} y_t - \vartheta_{-1}^{-1} \varphi_1 y_{t-1} - \vartheta_{-2}^{-1} \varphi_2 y_{t-2}], \tag{3.2}$$

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

$$m_t(\boldsymbol{\vartheta}; \boldsymbol{\varphi}) := (m_{1,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}); m_{2,t}(\boldsymbol{\vartheta}; \boldsymbol{\varphi}))'. \tag{3.3}$$

Note that

$$m_{1,t}(\boldsymbol{\sigma}_t; \boldsymbol{\phi}) = v_{t-1} \varepsilon_t, \quad \text{and} \quad m_{2,t}(\boldsymbol{\sigma}_t; \boldsymbol{\phi}) = v_{t-2} \varepsilon_t, \tag{3.4}$$

and, therefore, neither depends on σ_t in that case. Then, the first order conditions that determine the least-square estimator $\hat{\boldsymbol{\phi}}$ of ϕ are given by

$$\bar{m}_{k,t}(\tilde{\boldsymbol{\sigma}}_t; \hat{\boldsymbol{\phi}}) = 0, \quad (k = 1, 2). \tag{3.5}$$

Hence, the least-squares estimators $\hat{\boldsymbol{\phi}} := (\hat{\phi}_1, \hat{\phi}_2)$ of $\boldsymbol{\phi} = (\phi_1, \phi_2)$ are given as in (2.4).

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

1. ε_t are iid errors with mean zero and variance 1.

2. $\mathbb{E}|\varepsilon_t|^{4+\gamma} < \infty$ for some small $\gamma > 0$.

3. $\sigma_t^2 \in \mathcal{F} := C^2[0, 1]$.

We denote Θ_0 the set of $\boldsymbol{\phi}$ such that the roots of the characteristic equation $1 - \phi_1 z - \phi_2 z^2 = 0$ are greater then $1 + \delta$ in absolute value for some $\delta > 0$. This guarantees causality and stationarity of the AR(2) error process v_t; moreover, it also implies that v_t can be represented as an MA(∞):

$$v_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{3.6}$$

where all of the coefficients

$$K_1 \rho^i \leq |\psi_i| \leq K_2 \rho^i \tag{3.7}$$

for all $i \geq 0$, 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, (3.7) implies that the series $\{\psi_i\}_{i\geq 0}$ is absolutely converging: $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

Remark 3.1. *The coefficients $\{\psi_i\}_{i\geq 0}$ are defined recursively as*

$$\psi_0 = 1, \quad \psi_1 - \psi_0\phi_1 = 0, \quad \psi_2 - \psi_1\phi_1 - \psi_0\phi_2 = 0, \ \ldots \tag{3.8}$$

or, in a more compact form, $\psi_j - \psi_{j-1}\phi_1 - \psi_{j-2}\phi_2 = 0$ if it is assumed that $\psi_j \equiv 0$ for any $j < 0$. 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 $m_t(\boldsymbol{\sigma}_t; \phi)$ (see Andrews (1987), Appendix B for the definition).

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

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

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

Theorem 3.3. *Let Θ be a compact subset of Θ_0 with non-empty interior. Then, the weak consistency of the least-square autoregressive estimator of (2.4) holds true.*

Our next task is to establish asymptotic normality of $\hat{\boldsymbol{\phi}}$. The proof of the following result is deferred to the Appendix B.

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

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

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

$$V := \begin{pmatrix} V_1 & V_2 \\ V_2 & V_1 \end{pmatrix} := \begin{pmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{pmatrix}. \tag{3.9}$$

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). As a reminder, the first step is estimating not σ_t^2 but rather
$$\sigma_t^{2,bias}(x) = \frac{\sigma_t^2(x)}{1+\phi_2},$$
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 $\tilde{\sigma}_t^2(x)$ of $\sigma_t^2(x)$ is defined as the value $\hat{a}$ solving the local least squares problem

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

Since $\tilde{\sigma}_t^2$ estimates $\sigma_t^{2,bias}$ consistently, at the next step we define a consistent estimator of σ_t^2 as follows:

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

where $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \hat{\phi}_2)$ is the least-squares estimator defined in (2.4). We will use $D\sigma_t^2(x)$ and $D^2\sigma_t^2(x)$ to denote the first and the second-order derivatives of the function $\sigma_t^2(x)$, respectively. The following lemma can be proved by following almost verbatim Theorem 3 of Dahl and Levine (2006) and is omitted for brevity.

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

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

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

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

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,* $\hat{\sigma}_t^2$ *is an asymptotically consistent estimator of* σ_t^2 *that is also asymptotically normal with the bias* $(1 + \phi_2)B(\boldsymbol{\phi}, \sigma_t^2)$ *and the variance* $(1 + \phi_2)^2 V(\boldsymbol{\phi}, \sigma_t^2)$*.*

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

5 Numerical results

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. 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 (2.4) 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 (2.5), we compute the 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 in the i^{th} simulation and M is the number of simulations. We use local linear estimators $\hat{\vartheta}(x_t)$ for estimating $\vartheta(x_t)$ in the step 1 of the method outlined above.

Table 2 provides the MSE for the three 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 Table 3, Mn(Sd) stand for "Mean and Standard Deviation". The true parameter values are $\phi_1 = 0.6$ and $\phi_2 = 0.3$. For these parameter values, the asymptotic standard deviation and covariance given in (3.9) 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. 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, Figure 1 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.

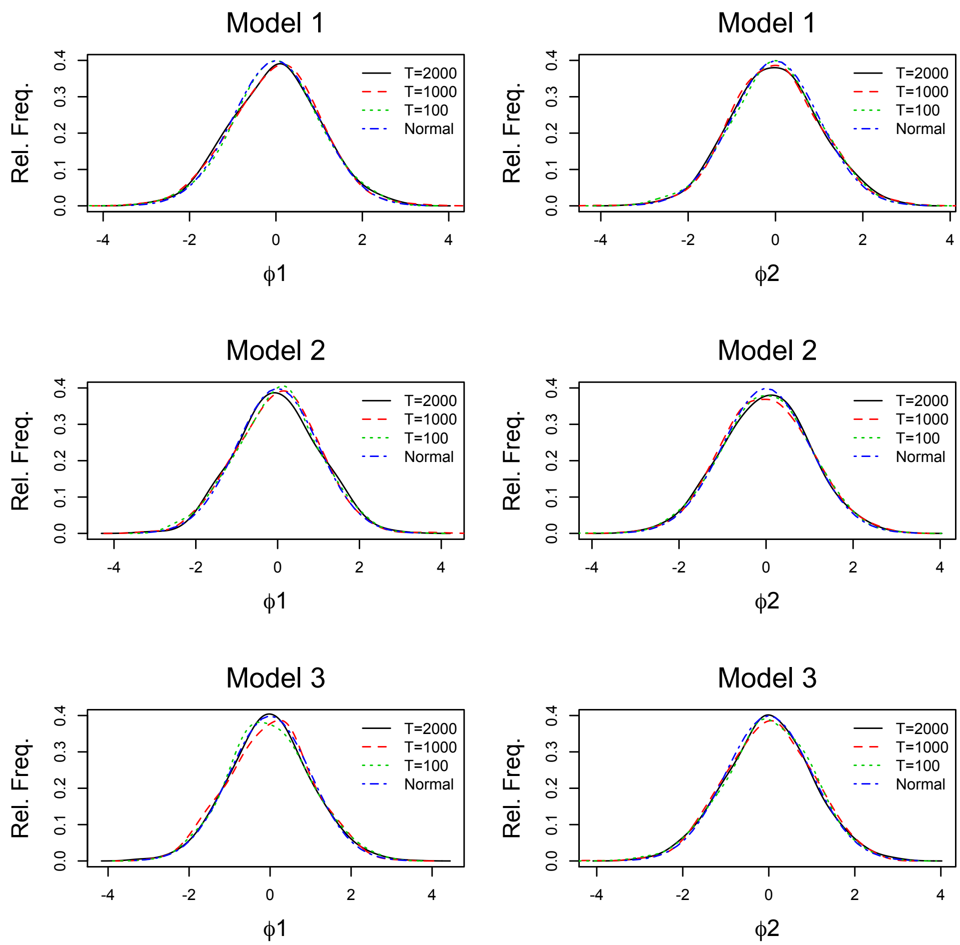

Figure 1: Finite-sample sampling densities in comparison with the standard normal density under the three alternative variance function specifications of Table 1. The number of Monte Carlo replications is 1000.

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.

Model	T=100	T=1000	T=2000
1	0.0241	0.0013	0.0007
2	0.0388	0.0020	0.0009
3	0.0626	0.0026	0.0011

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

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 order more then 1; thus, the method proposed in this manuscript represents a qualitatively new procedure. For the sake of simplicity, we only show in detail how to handle the case of autoregressive error process of order 2; however, this 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 above mentioned AR(p) error process. It can be shown that the expectation of the squared pseudoresidual of order 2, η_t^2, in that case is the scaled value of σ_t^2 where the scaling constant is an explicit function $\Psi \equiv \Psi(\phi_1, \ldots, \phi_p)$ of $\phi_1, \ldots, \phi_p$; since the scaling constant does not depend on the variance function, the following procedure can be suggested:

1. Obtain the estimate of the scaled variance function

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

Model	$T=100$		$T=1000$		$T=2000$	
	Mn(Sd) $\hat{\phi}_1$	Mean(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)

Table 3: Sampling Means and Standard Deviations under alternative variance function specifications and sample sizes with 1000 Monte Carlo replications and 10-fold cross-validation for bandwidth selection. True parameters are $\phi_1 = 0.6$ and $\phi_2 = 0.3$.

by using a non-parametric smoothing method (e.g. local linear regression) applied to η_t^2. Let $\hat{\Theta}_t^2 = \hat{\Theta}^2(x_t)$ be the resulting estimator.

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

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

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

$$\hat{\sigma}_t^2 := \Psi(\hat{\phi}_1, \ldots, \hat{\phi}_p) \hat{\Theta}_t^2. \tag{6.1}$$

Although technically more complicated, the same asymptotic results can be obtained in the general case of $p > 2$ in a straightforward manner with explicit expressions for asymptotic variances of all estimators.

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 unit-root testing for an autoregressive conditional mean of order $l \geq 1$ and some conditionally heteroscedastic error process. For the conditional mean of the form $y_t = \theta_0 + \theta_1 y_{t-1} + \ldots + \theta_l y_{t-l}$, Phillips and Xu (2006) addressed that issue by conducting asymptotic analysis of least squares estimates of the coefficients θ_k, $k = 1, \ldots, l$ 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 (2006) 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). We believe, therefore, that the asymptotic analysis of the least squares estimates of the coefficients θ_k under the same assumptions on the error process as in (1.1) is an important topic for future research. Such an analysis could yield several different robust estimators of the coefficients θ_k as well as corresponding robust unit root tests. 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.

A 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.

B 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:

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

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 (3.4)-(3.6), we define

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

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

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

$$m_{k,t*}(\boldsymbol{\sigma}_t;\rho) = \inf\{m_{k,t}(\boldsymbol{\sigma}_t;\boldsymbol{\phi}) : \boldsymbol{\phi} \in B(\boldsymbol{\phi}_0;\rho)\}. \tag{B.2}$$

The following two statements hold:

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

(b) Both $m^*_{k,t}(\boldsymbol{\sigma}_t;\rho)$ and $m_{k,t*}(\boldsymbol{\sigma}_t;\rho)$ 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 $d(\tilde{\boldsymbol{\phi}},\boldsymbol{\phi}) \le \tau$ implies

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

where B_t is a non-negative random variable (that may depend on $\boldsymbol{\phi}$) and $\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} \mathbb{E}\, B_t < \infty$, while $h : R^+ \to R^+$ is a nonrandom function such that $h(y) \downarrow h(0) = 0$ as $y \downarrow 0$.

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 (3.4) and (3.6),

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

for any $\phi \in B(\boldsymbol{\phi}_0;\rho)$, we find that the supremum of $m_{1,t}(\boldsymbol{\phi})$ taken over a ball $B(\boldsymbol{\phi}_0;\rho)$ also exists and is a random variable. Indeed, each of the summands in (B.3) is a continuous function of $\boldsymbol{\phi}$ as we already established earlier; the convergence in mean squared to $m_{1,t}(\boldsymbol{\phi})$ is uniform in $\boldsymbol{\phi}$ due to (3.7) and, therefore, $m_{1,t}(\boldsymbol{\phi})$ is continuous in $\boldsymbol{\phi}$ as well. That, in turn, implies the existence of $m^*_{1,t}(\boldsymbol{\sigma}_t;\rho)$; the existence of $m_{1,t*}(\boldsymbol{\sigma}_t;\rho)$; is established in exactly the same way. Moreover, the pointwise WLLNs for both $\sup_B m_{k,t}(\boldsymbol{\phi})$ and $\inf_B m_{k,t}(\boldsymbol{\phi})$ are also clearly satisfied since $\mathbb{E}\, m_{k,t}(\boldsymbol{\phi};\rho) \equiv 0$ for any $\phi \in B(\boldsymbol{\phi};\rho) \subset \Theta$.

2. We now show A4 for $m_{1,t}(\boldsymbol{\phi})$ ($m_{2,t}(\boldsymbol{\phi})$ can be treated analogously). Denote $m^*_{1,t}(\boldsymbol{\phi}) = \varepsilon_t \sum_{i=0}^{\infty} \psi^*_i \varepsilon_{t-1-i}$ 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 for $m^*_{1,t}(\boldsymbol{\phi})$. Let us start by noting that

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

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

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

Then,

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

Now we need to treat the second multiplicative term. First, recalling (3.8), it is easy to conclude, by induction, that the coefficients ψ_j are continuously differentiable functions of ϕ_1 and ϕ_2 for any $\boldsymbol{\phi} \in B(\boldsymbol{\phi}, \rho) \subset \Theta$. Therefore, using (3.7), 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, properly speaking, a metric measuring the distance between $\boldsymbol{\phi} = (\phi_1, \phi_2)$ and $\boldsymbol{\phi}^* = (\phi^*_1, \phi^*_2)$ since it is not symmetric with respect to its arguments 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; only $d(\tilde{\boldsymbol{\phi}}, \boldsymbol{\phi}) \downarrow 0$ as $\tilde{\boldsymbol{\phi}} \to \boldsymbol{\phi}$ is needed.

□

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 $m_t(\boldsymbol{\sigma}_t; \boldsymbol{\phi}) \equiv 0$. It remains to show the other conditions therein.

The first part of Assumption C(b) of the Theorem A is immediately satisfied because $m(\boldsymbol{\sigma}_t; \boldsymbol{\phi}) \equiv 0$ for any $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_t$. 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 $m(\boldsymbol{\sigma}_t; \boldsymbol{\phi}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} E\, m_t(\boldsymbol{\sigma}_t; \boldsymbol{\phi})$ is finite. In our case, since both martingale difference sequences $v_{t-1}\varepsilon_t$ and $v_{t-2}\varepsilon_t$ have mean zero, clearly $\sup_{\Theta\times\mathcal{F}} ||m(\boldsymbol{\sigma}_t; \boldsymbol{\phi})|| = 0 < \infty$ and the Assumption C(c) is satisfied. The Assumption C(d) is true because Θ is compact, the functional

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

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: $\hat{\boldsymbol{\phi}} \xrightarrow{p} \boldsymbol{\phi}$. □

Proof of Theorem 3.4. Recall that $\tilde{\sigma}_t$ is the inconsistent estimator of σ_t that, however, 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 $\boldsymbol{\sigma}_t = (\sigma_t, \sigma_{t-1}, \sigma_{t-2})^{'}$, $\tilde{\boldsymbol{\sigma}}_t = (\tilde{\sigma}_t, \tilde{\sigma}_{t-1}, \tilde{\sigma}_{t-2})^{'}$, $\boldsymbol{\sigma}_t^{bias} = (\sigma_t^{bias}, \sigma_{t-1}^{bias}, \sigma_{t-2}^{bias})^{'}$; then, for some generic argument $\boldsymbol{\vartheta} = (\vartheta_0, \vartheta_{-1}, \vartheta_{-2})$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ we have

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

with $m_t(\boldsymbol{\vartheta}, \boldsymbol{\varphi})$ given as in (3.3). Since the vector valued function $m_t(\boldsymbol{\vartheta}_t; \boldsymbol{\varphi})$ is twice continuously differentiable, we can use a Taylor expansion around $(\tilde{\boldsymbol{\sigma}}_t; \boldsymbol{\phi})$:

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

for some $\boldsymbol{\phi}^*$ that lies on the straight line connecting $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$. Then, due to the first order conditions (3.5) and the existence of an invertible matrix

$M = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} E\,\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}_t;\boldsymbol{\phi})$, we obtain that

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

In our case,

$$\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}_t,\boldsymbol{\phi}) = \left(\begin{array}{cc} \frac{\partial m_{1,t}}{\partial \phi_1} & \frac{\partial m_{1,t}}{\partial \phi_2} \\ \frac{\partial m_{2,t}}{\partial \phi_1} & \frac{\partial m_{2,t}}{\partial \phi_2} \end{array}\right)\Bigg|_{(\boldsymbol{\sigma}_t,\boldsymbol{\phi})} = \left(\begin{array}{cc} -v_{t-1}^2 & -v_{t-1}v_{t-2} \\ -v_{t-1}v_{t-2} & -v_{t-2}^2 \end{array}\right),$$

In light of (1.4), it follows that

$$M = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E\frac{\partial m_t}{\partial \boldsymbol{\phi}}(\boldsymbol{\sigma}_t;\boldsymbol{\phi}) = -\left(\begin{array}{cc} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{array}\right) =: S.$$

Let $m_T^*(\boldsymbol{\vartheta},\boldsymbol{\varphi}) = \frac{1}{T}\sum_{t=2}^{T} E\, m_t(\boldsymbol{\vartheta},\boldsymbol{\varphi})$ and define the empirical process

$$\nu_T(\boldsymbol{\sigma}_t) = \sqrt{T}\,[\bar{m}_T(\boldsymbol{\sigma}_t;\boldsymbol{\phi}) - m_T^*(\boldsymbol{\sigma}_t;\boldsymbol{\phi})] \tag{B.4}$$

Clearly,

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

Now, using (3.4),

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

Hence, using the CLT for martingale difference sequences (see Billingsley (1961)), $\sqrt{T}\bar{m}_T(\boldsymbol{\sigma}_t;\boldsymbol{\phi})$ is asymptotically normal $N(0,S)$ with the covariance matrix

$$\left(\begin{array}{cc} Ev_{t-1}^2\varepsilon_t^2 & Ev_{t-1}v_{t-2}\varepsilon_t^2 \\ Ev_{t-1}v_{t-2}\varepsilon_t^2 & Ev_{t-2}^2\varepsilon_t^2 \end{array}\right) = \left(\begin{array}{cc} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{array}\right) = -S.$$

If we can show that

$$\nu_T(\tilde{\boldsymbol{\sigma}}_t) - \nu_T(\boldsymbol{\sigma}_t) \xrightarrow{p} 0 \quad \text{and} \quad \sqrt{T}m_T^*(\tilde{\boldsymbol{\sigma}}_t;\boldsymbol{\phi}) \xrightarrow{p} 0, \tag{B.5}$$

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

$$V = M^{-1}S(M^{-1})' = S^{-1}S(S^{-1})' = S^{-1} = \frac{1}{\gamma_0^2 - \gamma_1^2}\left(\begin{array}{cc} \gamma_0 & -\gamma_1 \\ -\gamma_1 & \gamma_0 \end{array}\right).$$

A simple computation using (1.4)-(1.5) leads to (3.9). We show (B.5) through the following two steps:

(1) First, note that for any $\varepsilon > 0$,

$$\sqrt{T} m_T^*(\tilde{\boldsymbol{\sigma}}_t;\boldsymbol{\phi}) = \sqrt{T} m_T^*(\tilde{\boldsymbol{\sigma}}_t;\boldsymbol{\phi}) 1(\rho(\tilde{\boldsymbol{\sigma}}_t, \boldsymbol{\sigma}_t^{bias}) > \varepsilon) + \sqrt{T} m_T^*(\tilde{\boldsymbol{\sigma}}_t;\boldsymbol{\phi}) 1(\rho(\tilde{\boldsymbol{\sigma}}_t, \boldsymbol{\sigma}_t^{bias}) \leq \varepsilon).$$

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

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

$$T^{-1/2} \sum_{t=2}^{T} v_{t-1}^2 = O_p(1), \quad T^{-1/2} \sum_{t=2}^{T} v_{t-k} v_{t-1-k} = O_p(1), \tag{B.6}$$

for $k = 0, 1$. The empirical process $\nu_T(\boldsymbol{\sigma}_t)$ has its values in R^2; examining its first coordinate $\nu_T^1(\boldsymbol{\sigma}_t)$, one easily obtains $\nu_T^1(\boldsymbol{\sigma}_t) = S_t^{'}\tau$ where $S_t = (y_t y_{t-1}, \phi_1 y_{t-1}^2, \phi_2 y_{t-1} y_{t-2})^{'}$ while $\tau = (\sigma_t^{-1}\sigma_{t-1}^{-1}, -\sigma_{t-1}^{-2}, -\sigma_{t-1}^{-1}\sigma_{t-2}^{-1})$. Then, for any $\eta > 0$, $\delta > 0$, we have

$$\begin{aligned} &\lim_{T\to\infty} P\left(\sup_{\rho(\tilde{\sigma}_t, \sigma_t^{bias})<\delta} |\nu_T^1(\hat{\boldsymbol{\sigma}}_t) - \nu_T^1(\boldsymbol{\sigma}_t)| \geq \eta \right) \\ &= \lim_{T\to\infty} P\left(\sup_{\rho(\tilde{\sigma}_t, \sigma_t^{bias})<\delta} |T^{-1/2} \sum_{t=2}^{T} (S_t - E\, S_t)(\hat{\tau} - \tau)| > \eta \right) \\ &= P\left(\sup_{\rho(\tilde{\sigma}_t, \sigma_t^{bias})<\delta} ||T^{-1/2} \sum_{t=2}^{T} (S_t - E\, S_t)|| > \frac{\eta}{\delta} \right) \overset{T\to\infty}{\to} 0 \end{aligned}$$

since $T^{-1/2} \sum_{t=2}^{T} (S_t - E\, S_t) = O_p(1)$ due to (B.6). Using exactly the same argument, one easily obtains stochastic equicontinuity of the second coordinate $\nu_T^2(\boldsymbol{\sigma}_t)$; therefore, needed convergence in probability for $\nu_T(\tilde{\boldsymbol{\sigma}}_t) - \nu_T(\boldsymbol{\sigma}_t)$ follows from the convergence in both coordinates separately and the Slutsky's theorem. □

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

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

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

Eicker, B. (1963) "Limit theorems for regression with unequal and dependant errors", Annals of Mathematics and Statistics 34, 447-56

Phillips, P.C.B., and Xu, K.-L. (2006) "Inference in Autoregression under heteroscedasticity", Journal of Time Series Analysis 27, 289-308

White, H. (1980) "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity", Econometrica 48, 817-38

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