Statistica Sinica **21** (2011), 973-999
doi:http://dx.doi.org/10.5705/ss.2009.073

QUASI-LIKELIHOOD ESTIMATION IN STATIONARY AND NONSTATIONARY AUTOREGRESSIVE MODELS WITH RANDOM COEFFICIENTS

Alexander Aue and Lajos Horváth

University of California, Davis and University of Utah

Abstract: We propose a unified quasi-likelihood procedure for the estimation of the unknown parameters of a first-order random coefficient autoregressive, RCA, model that works both for stationary and nonstationary processes. For this procedure, the weak consistency and the asymptotic normality are established under minimal assumptions on the noise sequences. In an empirical study, we highlight the practicality of the quasi-likelihood estimation for applications. As no initial knowledge about the probabilistic properties of the RCA process is required, our theoretical results immediately facilitate the statistical analysis for practitioners. They may, moreover, have an impact on the treatment of the prominent unit-root problems often encountered in econometrics.

Key words and phrases: Nonlinear optimization, nonlinear time series models, unit-roots.

1. Introduction

We develop in this paper a unified estimation theory for random coefficient autoregressive (RCA) processes which works irrespective of stationarity issues. These RCA models are nonlinear, thus extending the class of classical autoregressive (AR) models paramount in linear time series analysis, leading to an increased flexibility in modeling heteroscedasticity often found in data, while still enabling a parsimonious representation. Early contributions in this respect are due to Conlisk (1974, 1976) who dealt with econometric modeling questions, and Andél (1976) who argued that many engineering applications can be adequately described by RCA models. Other successful applications of RCA models to problems in financial, ecological, and longitudinal data analysis can be found in Liu and Tiao (1980), Stenseth et al. (1999), and Rahiala (1999), respectively. Despite this history, RCA models have as of now received less attention than a number of other nonlinear models such as (G)ARCH and threshold models and their variants. In this paper,we argue that it would be worthwhile to give RCA models a closer look and that they are worthy of inclusion in the applied statistician's toolbox. There are several arguments supporting this claim.

First, RCA models allow for a unified estimation procedure which does not depend on the probabilistic structure of the underlying processes. To discriminate between stationary and nonstationary time series can prove difficult in applications, and estimation techniques typically require stationarity as a basic assumption. Under nonstationarity, on the other hand, applied statisticians proceed in their analysis frequently by using the differencing methods based on an ARIMA-type approach. Difficulties arising from (over)differencing the data can be entirely bypassed whenever an RCA framework is appropriate.

Second, even though GARCH-type models have been more popular in the analysis of financial data, it has been pointed out in Tsay (1987) that the ARCH regression model introduced by Engle and Kraft (1983) can be cast into an RCA framework and can therefore be regarded as a special, second-order equivalent case.

Third, even the simple first-order RCA model can prove relevant for applications in econometrics. Owing to the random walk hypothesis (see, for example, Fama (1965), Nelson and Plosser (1982), among many others), the behavior of stock market prices is commonly modeled by an autoregressive time series with autoregressive parameter $\varphi = 1$. Since the value of the autoregressive parameter is exactly on the boundary between stationary and nonstationary AR(1) models, this theory inherits the undesirable side effect of estimation procedures exhibiting significantly different behavior for the cases $\varphi < 1$, $\varphi = 1$, and $\varphi > 1$. This is known as knife edge effect (see Lumsdaine (1996)). If, however, the autoregressive parameter is accompanied by an additional (and maybe marginal) random disturbance, these differences cease to exist as we will show in this paper.

For the sake of clarity in the presentation we focus on first-order RCA models. Similar results for higher orders are to be expected, but would require more technical effort without bringing substantial additional insight into the theory provided for the first-order case. The paper is organized as follows. In Section 2, we review the existing methods concerning the estimation of the parameters determining an RCA process with a focus on the quasi-likelihood approach. The unified estimation theory based on this method is developed in Section 3. The ramifications of these results are discussed in the first two parts of Section 4, while the third part contains a Monte Carlo simulation study which underlines that the novel unified estimation procedure produces satisfactory results also in finite samples of moderate size. In the fourth part of Section 4 we provide applications to the daily trading volume of IBM stock and annual change rates of world GDP. The proofs of all theorems are then given in Sections 5 and 6.

2. Review of Existing Results

We study the first-order random coefficient autoregressive RCA(1) model given by the stochastic difference equations

$$X_j = (\varphi + b_j)X_{j-1} + e_j, \qquad j \in \mathbb{N}, \tag{2.1}$$

where φ is a real number and $\mathbb{N}$ denotes the positive integers. The set of equations (2.1) is initialized with a random variable X_0 which we specify further in the discourse. Throughout, we work with the following requirements on the noise sequences.

Assumption 2.1. *The sequences* $(b_j : j \in \mathbb{N})$ *and* $(e_j : j \in \mathbb{N})$ *are independent sequences of independent and identically distributed random variables with* $E[b_1] = 0$, $E[e_1] = 0$, $0 < \omega^2 = E[b_1^2]$, *and* $0 < \sigma^2 = E[e_1^2]$.

The sequence $(X_j : j \in \mathbb{N})$ defines a nonlinear time series model that has received considerable attention in the literature. Early contributions include Andél (1976), Feigin and Tweedie (1985), Liu and Tiao (1980), Nicholls and Quinn (1980), Quinn and Nicholls (1981), and Robinson (1978).

While most of the previous literature was concerned with weak stationarity, strictly stationary solutions to (2.1) were characterized under a minimal set of assumptions in Aue, Horváth, and Steinebach (2006). Restating their results, we let $\ln^+ x = \max\{\ln x, 0\}$ be the positive part of the natural logarithm. It is then shown that if both $E[\ln^+|\varphi + b_1|]$ and $E[\ln^+|e_1|]$ are finite (and without invoking the moment conditions of Assumption 2.1, (2.1) admits a strictly stationary, nonanticipative solution if and only if

$$-\infty \le E[\ln|\varphi + b_1|] < 0. \tag{2.2}$$

Notice that this statement requires X_0 to possess the strictly stationary and ergodic distribution, so that the sequence $(X_j : j \in \mathbb{N})$ is already properly initialized.

The model (2.1) depends on the unknown parameter vector $\boldsymbol{\theta} = (\varphi, \omega^2, \sigma^2)$. Various estimation procedures for $\boldsymbol{\theta}$ in the strictly stationary case (2.2) have been discussed. A two-step least squares method was covered in Nicholls and Quinn (1980). The same authors also established the large sample properties of the likelihood procedure in Quinn and Nicholls (1981). For other estimation procedures, we refer here only to the contributions by Koul and Schick (1996), Li and Hui (1983), and Rudolph (1998). In this paper, we focus on the quasi-maximum likelihood estimator (QMLE) for $\boldsymbol{\theta}$. It has been shown in Aue, Horváth, and Steinebach (2006) that the log-likelihood function is given by (notice the sign change)

$$\ell_n(\boldsymbol{u}) = \frac{1}{n}\sum_{j=1}^{n} g_j(\boldsymbol{u}), \qquad g_j(\boldsymbol{u}) = -\frac{(X_j - sX_{j-1})^2}{xX_{j-1}^2 + y} - \ln(xX_{j-1}^2 + y), \tag{2.3}$$

where $\boldsymbol{u} = (s, x, y)$ denotes a generic parameter vector. With the suitably chosen parameter space $\Gamma = \{\boldsymbol{u} \in \mathbb{R}^3 \colon s_* \le s \le s^*, x_* \le x \le x^*, y_* \le y \le y^*\}$, where $s_* < s^*$, $0 < x_* < x^*$ and $0 < y_* < y^*$, the QMLE $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}$ is obtained from

$$\ell_n(\hat{\boldsymbol{\theta}}_n) = \sup_{\boldsymbol{u}\in\Gamma} \ell_n(\boldsymbol{u}).$$

Aue, Horváth, and Steinebach (2006) have derived the strong consistency and the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$.

Contributions concerned with the estimation of $\boldsymbol{\theta}$ in the nonstationary case $E[\ln|\varphi + b_1|] \ge 0$ are rare. An exception is the paper by Hwang and Basawa (2005) that discussed the least squares and weighted least squares procedures for the estimation of the parameter φ. The behavior of the QMLE $\hat{\boldsymbol{\theta}}_n$ in the nonstationary case has recently been studied by Berkes, Horváth, and Ling (2009) who prove that the original likelihood procedure does not work under nonstationarity, since $\ell_n(\boldsymbol{u}) \xrightarrow{P} \infty$ for all $\boldsymbol{u} \in \Gamma$. Here and in the following, $\xrightarrow{P}$ indicates convergence in probability. It holds, however, that $\ell_n(\boldsymbol{u}) - \ell_n(\boldsymbol{\theta}) \xrightarrow{P} f(s, x)$, where the limit can be given explicitly. Since $f(s, x)$ is independent of y, the QMLE cannot be employed to produce an estimate of σ^2. This is in accordance with similar results for ARCH and GARCH processes, see Jensen and Rahbek (2004a,b). Consequently, estimation procedures need to be restricted to the parameters φ and ω^2. Let $\boldsymbol{v} = (s, x)$ be the vector consisting of the first two coordinates of $\boldsymbol{u}$. Berkes, Horváth, and Ling (2009) suggested fixing $y > 0$ and maximizing $\ell_n(\boldsymbol{u})$ with respect to $\boldsymbol{v}$ in the restricted parameter space $\Gamma_R(y) = \{\boldsymbol{v} \in \mathbb{R}^2 \colon s_* \le s \le s^*, x_* \le x \le x^*\}$. This leads to the restricted QMLE $\hat{\boldsymbol{\theta}}_n^{(R)}(y)$ given by

$$\ell_n\Big(\hat{\boldsymbol{\theta}}_n^{(R)}(y), y\Big) = \sup_{\boldsymbol{v}\in\Gamma_R(y)} \ell_n(\boldsymbol{v}, y).$$

The weak consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n^{(R)}(y)$ have been established in Berkes, Horváth, and Ling (2009).

As a consequence of the above, the QMLE-based procedures $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n^{(R)}$ cannot be employed for the statistical inference about the unknown parameter $\boldsymbol{\theta}$ without further knowledge of the probabilistic structure of the underlying RCA process $(X_j \colon j \in \mathbb{N})$. In practice, however, it is often hard to discriminate between a stationary and a nonstationary time series. This is in particular true in the transition cases for which $E[\ln|\varphi + b_1|]$ takes a value in the vicinity of zero. It would thus be advantageous to design estimation procedures that work simultaneously irrespective of the particular form of $(X_j : j \in \mathbb{N})$. To derive such a unified estimation theory is the main aim of this paper. As a byproduct, this theory sheds new light on the unit-root phenomenon prominent in the statistical analysis of econometric time series (see, among many others, Phillips and Perron (1998) and Phillips and Xiao (1998)).

3. A Unified Estimation Theory Based on QMLE

There are two options that can be used to build a unified estimation theory for an RCA process given by (2.1). One is based on $\hat{\boldsymbol{\theta}}_n$, originally introduced to estimate the parameters of a strictly stationary version of $(X_j : j \in \mathbb{N})$, the other on $\hat{\boldsymbol{\theta}}_n^{(R)}$, designed for the nonstationary case. In this section, we show that only the first approach leads to the desired goal. The first question we address in this section, however, is concerned with the latter approach. How does the restricted QMLE $\hat{\boldsymbol{\theta}}_n^{(R)}(y)$ behave if the RCA equations (2.1) admit a strictly stationary solution? Recall that it has been shown in Berkes, Horváth, and Ling (2009) that, for any $y > 0$, $\hat{\boldsymbol{\theta}}_n^{(R)}(y) \xrightarrow{P} (\varphi, \omega^2)$ as $n \to \infty$ if $E[\ln|\varphi + b_1|] \geq 0$. In the following, we deal with the stationary counterpart of this statement. To do so, we assume again that X_0 has the distribution of the strictly stationary and ergodic solution.

Theorem 3.1. *Let $(X_j : j \in \mathbb{N})$ be a solution to (2.1) such that $E[\ln|\varphi+b_1|] < 0$. If (φ, ω^2) is an inner point of Γ_R, then $\hat{\boldsymbol{\theta}}_n^{(R)}(\sigma^2) \to (\varphi, \omega^2)$ a.s. as $n \to \infty$. If, on the other hand, $y \neq \sigma^2$, then*

$$P\left(\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n^{(R)}(y) = (\varphi, \omega^2)\right) = 0.$$

The proof of Theorem 3.1 is given in Section 5. Theorem 3.1 states that the estimator $\hat{\boldsymbol{\theta}}_n^{(R)}$ cannot be used to estimate the remaining parameters φ and ω^2 in the stationary case unless the value of σ^2 is known beforehand, which is almost always unrealistic to assume in practice. While a unified estimation theory does not follow from the restricted QMLE, it can nevertheless be employed to construct a novel type of stationarity test for RCA processes. We develop these tests in Section 4 below.

Noticing that the variance parameter σ^2 cannot be estimated consistently under nonstationarity, we study next the behavior of the first two coordinates of the original QMLE $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2}, \hat{\theta}_{n,3})$ without specifying the type of solution to (2.1). This means, in particular, that (2.2) may be violated and it is therefore to be replaced with a milder moment assumption. To phrase this precisely, we set

$\tilde{\boldsymbol{\theta}}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2})$ and $\tilde{\boldsymbol{\theta}} = (\varphi, \omega^2)$ and are interested in deriving the large sample properties of $\tilde{\boldsymbol{\theta}}_n$. The first result establishes weak consistency.

Theorem 3.2. *Let* $(X_j : j \in \mathbb{N})$ *be a solution to* (2.1) *such that* $E[\ln^+ |\varphi + b_1|] < \infty$. *Then* $\tilde{\boldsymbol{\theta}}_n \xrightarrow{P} \tilde{\boldsymbol{\theta}}$ *as* $n \to \infty$.

We give the proof of Theorem 3.2 in Section 5. Next, we focus on the asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$. To begin with, we study only its first coordinate $\hat{\theta}_{n,1}$. In many cases, the main part of the statistical inference is about the autoregressive parameter φ. Without the additional random disturbances $(b_j : j \in \mathbb{N})$—in the classical AR(1) setting—this would lead to the unit root phenomenon which, however, does not exist for RCA time series (see Section 4 below). Introducing the quantities

$$\alpha(\kappa, \gamma) = E\left[\frac{X_1^\kappa}{(\omega^2 X_1^2 + \sigma^2)^\gamma}\right], \qquad \kappa = 0, 1, \ldots, 2\gamma, \qquad \gamma \in \mathbb{N}, \tag{3.1}$$

we obtain a central limit theorem for the the inference about φ.

Theorem 3.3. *Let* $(X_j : j \in \mathbb{N})$ *be a solution to* (2.1) *such that* $E[\ln^+ |\varphi + b_1|] < \infty$. *Then,* $\sqrt{n}(\hat{\theta}_{n,1} - \varphi) \xrightarrow{\mathcal{D}} N(0, \tau^2)$ *as* $n \to \infty$, *where*

$$\tau^2 = \begin{cases} \omega^2 & \text{if } E[\ln |\varphi + b_1|] \geq 0, \\ \dfrac{4}{\alpha^2(2,1)}\left[\omega^2 \alpha(4,2) + \sigma^2 \alpha(2,2)\right] & \text{if } E[\ln |\varphi + b_1|] < 0. \end{cases}$$

Observe that the condition $E[\ln^+ |\varphi + b_1|] < \infty$ is only needed in the stationary case. The asymptotic normality for $\hat{\theta}_{n,1} - \varphi$ is therefore valid without further restrictions on the process $(X_j : j \in \mathbb{N})$. The asymptotic variance τ^2, however, is different in stationary and nonstationary cases. To use Theorem 3.3. for statistical inference, we consequently need to find an estimator $\hat{\tau}_n^2$ for τ^2 that works for both. Let

$$\hat{\alpha}_n(\kappa, \gamma) = \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\kappa}{(\hat{\theta}_{n,2} X_j^2 + \hat{\theta}_{n,3})^\gamma}, \qquad \kappa = 0, 1, \ldots, 2\gamma, \qquad \gamma \in \mathbb{N}.$$

Corollary 3.1. *Suppose that the assumptions of Theorem* 3.3 *are satisfied. Then,* $\sqrt{n}\hat{\tau}_n^{-1}(\hat{\theta}_{n,1} - \varphi) \xrightarrow{\mathcal{D}} N(0,1)$ *as* $n \to \infty$, *where* $\hat{\tau}_n^2 = 4\hat{\alpha}_n^{-2}(2,1)\left[\hat{\theta}_{n,2}\hat{\alpha}_n(4,2) + \hat{\theta}_{n,3}\hat{\alpha}_n(2,2)\right]$.

Note that $\hat{\tau}_n^2$ is consistent for τ^2 in all cases. This seems to be intuitive in the stationary case, where the moment estimators $\hat{\alpha}_n(\kappa, \gamma)$ are computed from the strictly stationary quantities $X_j^\kappa(\hat{\theta}_{n,2} X_j^2 + \hat{\theta}_{n,3})^{-2}$. In the nonstationary case, however, $\hat{\theta}_{n,3}$ is noninformative for σ^2. Here, the exponential growth of $(X_j : j \in \mathbb{N})$ (see Corollary 3.1 in Berkes, Horváth, and Ling (2009)) implies that $\hat{\alpha}_n(2,2)$—and therefore also $\hat{\theta}_{n,3}\hat{\alpha}_n(2,2)$—vanishes asymptotically. All details are provided in Section 6, where we prove both Theorem 3.3 and Corollary 3.1.

In general, it may be difficult to decide whether a given series of observations is a realization of a stationary or nonstationary process. Corollary 3.1, however, implies that statistical inference about the parameter φ can be performed without specifying the type of solution to (2.1). The same method applies to all cases without further knowledge of the probabilistic structure of the observations. In the remainder of this section, we introduce a similarly appealing procedure to estimate jointly the values of φ and ω^2. Here, however, the transition case $E[\ln|\varphi + b_1|] = 0$ has to be excluded from our considerations. Let

$$g_j'(\boldsymbol{\theta}) = \left(\frac{\partial g_j(\boldsymbol{\theta})}{\partial s}, \quad \frac{\partial g_j(\boldsymbol{\theta})}{\partial x}, \quad \frac{\partial g_j(\boldsymbol{\theta})}{\partial y} \right)^T$$

and set $A = E[g_1'(\boldsymbol{\theta})(g_1'(\boldsymbol{\theta}))^T]$ with ^{T} denoting the transpose. We refer to the entries of A as $A_{i,k}$, $i, k = 1, \ldots, 3$.

Theorem 3.4. *Let $(X_j : j \in \mathbb{N})$ be a solution to (2.1) such that $E[\ln|\varphi+b_1|] \neq 0$. Then, $\sqrt{n}\big[\tilde{\boldsymbol{\theta}}_n - (\varphi, \omega^2)\big] \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, \Sigma)$ as $n \to \infty$, where*

$$\Sigma = \begin{cases} \begin{bmatrix} \omega^2 & E[b_1^3] \\ E[b_1^3] & \mathrm{Var}(b_1^2) \end{bmatrix} & \text{if } E[\ln|\varphi + b_1|] > 0, \\[2ex] \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} & \text{if } E[\ln|\varphi + b_1|] < 0, \end{cases}$$

with

$$\Sigma_{1,1} = \frac{4}{\alpha^2(2,1)} \left[\omega^2 \alpha(4,2) + \sigma^2 \alpha(2,2) \right], \tag{3.2}$$

$$\Sigma_{1,2} = \frac{1}{2D\alpha(2,1)} \left[A_{1,2}\alpha(0,2) - A_{1,3}\alpha(2,2) \right] = \Sigma_{2,1}, \tag{3.3}$$

$$\Sigma_{2,2} = \frac{1}{D^2} \left[A_{2,2}\alpha^2(0,2) - 2A_{2,3}\alpha(0,2)\alpha(2,2) + A_{3,3}\alpha^2(2,2) \right], \tag{3.4}$$

and $D = \alpha(4,2)\alpha(0,2) - \alpha^2(2,2)$.

To estimate the covariance matrix Σ in both the stationary and nonstationary environment, we utilize the moment estimators $\hat{\alpha}_n(\kappa, \gamma)$ and estimate the components of the matrix A by

$$\hat{A}_{n,i,k} = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial g_j(\hat{\boldsymbol{\theta}}_n)}{\partial_i}\frac{\partial g_j(\hat{\boldsymbol{\theta}}_n)}{\partial_k}, \qquad i,k = 1,2,3,$$

where ∂_i stands for ∂s, ∂x or ∂y depending on whether $i = 1$, $i = 2$ or $i = 3$. Now we form the covariance estimator $\hat{\Sigma}_n$ by substituting the theoretical values $\alpha(\kappa,\gamma)$ and $A_{i,k}$ in (3.2)–(3.4) by their respective estimated counterparts $\hat{\alpha}_n(\kappa,\gamma)$ and $\hat{A}_{n,i,k}$, and by replacing ω^2 and σ^2 with $\hat{\theta}_{n,2}$ and $\hat{\theta}_{n,3}$. This leads to the following corollary of Theorem 3.4.

Corollary 3.2. *Suppose that the assumptions of Theorem* 3.4 *are satisfied. Then,* $\sqrt{n}\hat{\Sigma}_n^{-1/2}\big[\tilde{\boldsymbol{\theta}}_n - (\varphi,\omega^2)\big] \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, I)$ *as* $n \to \infty$*, where* I *denotes the* 2×2 *identity matrix.*

We give the proofs of Theorem 3.4 and Corollary 3.2 in Section 6 below.

The intuition as to what enables the unified estimation of the parameters φ and ω is the following. First, Theorem 3.4 establishes a central limit theorem for the QMLE estimator $\tilde{\boldsymbol{\theta}}_n$ with the normalizations that are different for the stationary and nonstationary subcases. It is therefore paramount to utilize *one* estimator for the limiting covariance matrix that converges in probability to the right limit in each of the *two* possible scenarios. The proposed moment-based estimator $\hat{\Sigma}_n$ obviously possesses this quality in the stationary case. That it also works under nonstationarity is, as pointed out after Corollary 3.1., a consequence of the exponential growth of the RCA process that cancels undesired effects due to the noninformativeness of the QMLE for σ^2. For more details we refer to the proofs in Section 6.

4. Further Implications and Empirical Results

4.1. The unit root phenomenon or the lack thereof

Of particular interest is the statistical inference about the deterministic autoregressive parameter φ. Without the additional random perturbations $(b_j : j \in \mathbb{Z})$, the nonlinear model (3.1) reduces to the classical first-order autoregressive, AR(1), model. As we have already pointed out briefly in the previous sections, a voluminous body of literature has been devoted to what is known as the unit root problem or random walk hypothesis: To what extent can econometric time series such as (logarithms of changes in) stock market prices be modeled by AR(1) processes with parameter $\varphi = 1$? Commonly, this question is tackled by using derivatives of the popular test statistics introduced by Dickey and Fuller (1979, 1981) that can, for example, be based on the likelihood procedure. These tests, however, feature low power because they lack the ability to significantly discriminate between the random walk hypothesis and the possible alternatives such as $\varphi \neq 1$, $|\varphi| < 1$ and $\varphi > 1$. Refinements have led to near-integrated settings in which the parameter φ is allowed to tend to unity with increasing sample size (see Phillips (1988), Phillips and Magdalinos (2007) , and Aue and Horváth (2007)).

Our results are in stark contrast to the AR(1) case; there is no "knife-edge" effect for RCA(1) processes. On the contrary, their probabilistic structure is irrelevant for parameter estimation via quasi-likelihood. The sole requirement for the applicability of, say, Theorem 3.3 is that $E[\ln^+|\varphi + b_1|] < \infty$, which is a rather mild condition. In other words, and perhaps surprisingly, the unit root phenomenon does not pertain to the RCA(1) process. As long as the deterministic parameter φ is accompanied by only a moderate random perturbation (that is, a small variance parameter ω^2), it can be consistently estimated even in the vicinity of unity. It is worthwhile mentioning that the processes themselves exhibit different probabilistic structures; see Aue (2008).

4.2. Stationarity tests

While the main aim of this paper is to provide a unified estimation theory independent of stationarity issues of the underlying RCA process, the result of Theorem 3.1 can also be utilized to design a novel type of stationarity test. It has been shown in Nicholls and Quinn (1982) that the RCA equations (2.1) allow for a weakly stationary solution if and only if $\varphi^2 + \omega^2 < 1$. This combined with Assumption 2.1 yields also strict stationarity, so that we henceforth do not make a distiction between the two concepts. To check for stationarity without prior information on the underlying RCA process, it is now natural to check the estimated counterpart $\hat{\theta}_{n,1}^2 + \hat{\theta}_{n,2}$ provided n observations $X_1, \ldots, X_n$ are available. As a straightforward consequence of Corollary 3.2, we obtain the convergence,

$$\frac{\sqrt{n}}{\hat{v}_n}\left[\hat{\theta}_{n,1}^2 + \hat{\theta}_{n,2} - (\varphi^2 + \omega^2)\right] \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as } n \to \infty, \tag{4.1}$$

where $\hat{v}_n^2 = 4\hat{\theta}_{n,1}(\hat{\Sigma}_{n,1,1} + \hat{\Sigma}_{n,1,2}) + \hat{\Sigma}_{n,2,2}$ with the Σ of Corollary 3.2. The hypothesis of stationarity would now be rejected at the asymptotic level $\alpha \in (0,1)$ if $\hat{\theta}_{n,1}^2 + \hat{\theta}_{n,2} > 1 + c_\alpha \hat{v}_n/\sqrt{n}$, where c_α satisfies $\Phi(c_\alpha) = 1 - \alpha$. If the innovations $(e_j\colon j \in \mathbb{N})$ are normal, then condition (2.2) can be rewritten as the integral criterion

$$-\infty \leq \mathcal{I}(\varphi, \omega^2) = \frac{1}{\sqrt{2\pi\omega^2}} \int_{-\infty}^{\infty} \ln|\varphi + z| \exp\Big(-\frac{z^2}{2\omega^2}\Big) dz < 0. \tag{4.2}$$

Using the same strategy as before, one can utilize $\mathcal{I}(\hat\theta_{n,1},\hat\theta_{n,2})$ as a proxy for the true $\mathcal{I}(\varphi,\omega^2)$. Consequently, (4.1) implies that

$$\frac{\sqrt{n}}{\hat w_n}\left[\mathcal{I}(\hat\theta_{n,1},\hat\theta_{n,2})-\mathcal{I}(\varphi,\omega^2)\right]\stackrel{\mathcal{D}}{\longrightarrow} N(0,1)\quad\text{as } n\to\infty, \tag{4.3}$$

where $\hat w_n^2=[\nabla\mathcal{I}(\hat\theta_{n,1},\hat\theta_{n,2})]^T\hat\Sigma_n[\nabla\mathcal{I}(\hat\theta_{n,1},\hat\theta_{n,2})]$. Stationarity would be rejected at the asymptotic level $\alpha\in(0,1)$ if $\mathcal{I}(\hat\theta_{n,1},\hat\theta_{n,2})>c_\alpha\hat w_n/\sqrt{n}$, where c_α satisfies $\Phi(c_\alpha)=1-\alpha$.

4.3. Simulation results

In this subsection, we provide the results of a simulation study that further supports the unified estimation theory established in Section 3. We have performed simulations with the following types of RCA(1) processes $(X_j\colon j\in\mathbb{N})$ given by (2.1) with centered normal noise sequences $(b_j\colon j\in\mathbb{N})$ and $(e_j\colon j\in\mathbb{N})$.

1. $(X_j^{(1)}\colon j\in\mathbb{N})$ is strictly stationary with finite variance. The model parameters are $\varphi=0.5$, $\omega^2=0.25$, and $\sigma^2=1$. To mimic the stationary and ergodic distribution, X_0 was obtained after a burn-in period of 500 repetitions initialized with a standard normal variable.
2. $(X_j^{(2)}\colon j\in\mathbb{N})$ is strictly stationary with infinite variance. The model parameters are $\varphi=0.5$, $\omega^2=1.5$, and $\sigma^2=1$. The initial value of X_0 was obtained as before.
3. $(X_j^{(3)}\colon j\in\mathbb{N})$ is nonstationary. The model parameters are $\varphi=1$, $\omega^2=3$, and $\sigma^2=1$. The initial value is $X_0=0$.
4. $(X_j^{(4)}\colon j\in\mathbb{N})$ is nonstationary. The model parameters are $\varphi=1.5$, $\omega^2=1$, and $\sigma^2=1$. The initial value is $X_0=0$.

It can easily be checked that $(X_j^{(1)}\colon j\in\mathbb{N})$ and $(X_j^{(2)}\colon j\in\mathbb{N})$ are strictly stationary RCA(1) processes with respective finite and infinite variances, while $(X_j^{(3)}\colon j\in\mathbb{N})$ and $(X_j^{(4)}\colon j\in\mathbb{N})$ are nonstationary. Typical sample paths of length $n=1{,}000$ for these four processes are displayed in Figure 1. It can be seen that a wide range of dynamics is covered by these choices.

We use the QMLE procedure to estimate the parameters of the processes $(X_j^{(\ell)}\colon j\in\mathbb{N})$, $\ell=1,\ldots,4$. To get an impression of the form of the likelihood as a function of the unknown model parameters, we first display the likelihood surfaces for the first and third RCA(1) processes using the profile likelihood as given in displays (4.2.7)–(4.2.9) of Nicholls and Quinn (1982). This profile likelihood is obtained by concentrating out one of the variance parameters and is a function of φ and $\rho=\omega^2/\sigma^2$. Figure 2 exhibits that this profile likelihood

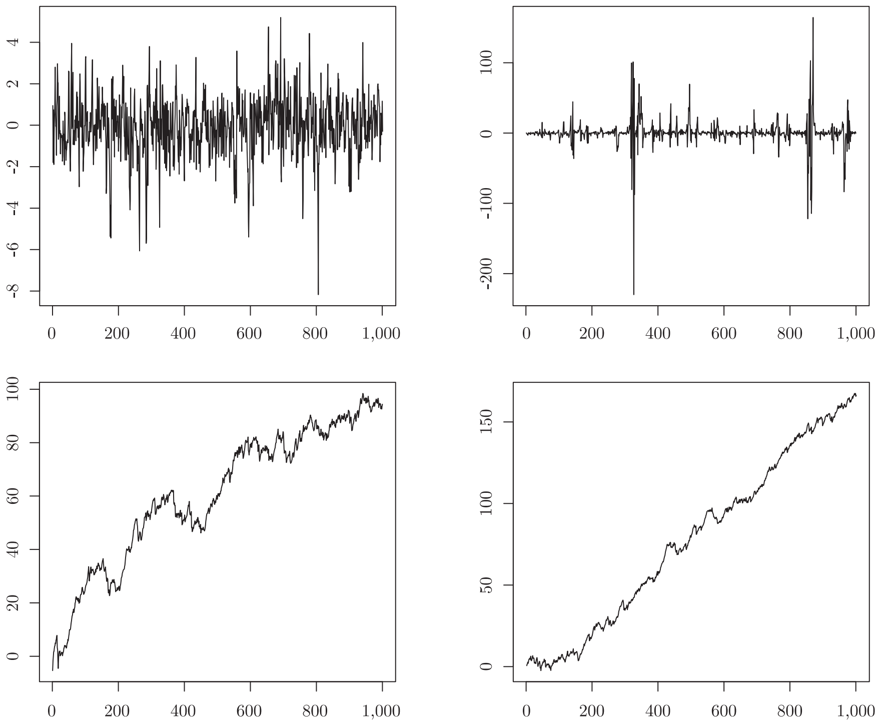

Figure 1. Time series plots of length $n = 1{,}000$ for the RCA(1) processes used in Subsection 4.3. The two processes in the upper panel are the finite variance, strictly stationary $(X_j^{(1)} : j \in \mathbb{N})$ (left) and the infinite variance, strictly stationary $(X_j^{(2)} : j \in \mathbb{N})$ (right). The lower panel displays $(\ln |X_j^{(3)}| : j \in \mathbb{N})$ (left) and $(\ln |X_j^{(4)}| : j \in \mathbb{N})$ (right), indicating the exponential growth of the nonstationary processes.

has a very pronounced peak around the true value of φ, but is rather flat in the direction of the parameter ρ. The likelihood surfaces for the remaining two RCA processes are similar and hence omitted. The foregoing indicates that the numerical nonlinear optimization problem associated with the quasi-likelihood procedure needs a set of reliable initial estimates.

In the strictly stationary case, selecting initial estimators does not pose a major problem as one can choose the least squares estimates (LSE) provided in Nicholls and Quinn (1980). These are strongly consistent and asymptotically normal provided that RCA(1) process possesses finite fourth order moments and finite eighth order moments, respectively. Even though this requirement is not satisfied in the case of $(X_j^{(2)} : j \in \mathbb{N})$, we used the LSE to initialize the QMLE procedure. In the nonstationary cases, it is only known from Hwang and Basawa (2005) that the weighted least squares estimator may be utilized to obtain consis-

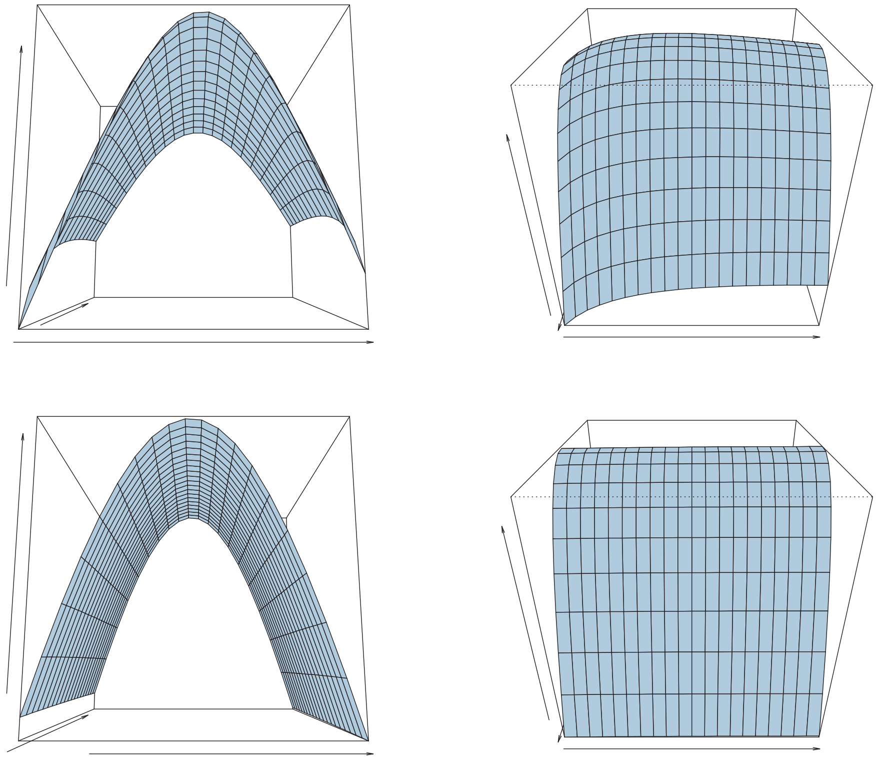

Figure 2. Profile likelihood surfaces of the processes $(X_j^{(1)} : j \in \mathbb{N})$ (upper panel) and $(X_j^{(3)} : j \in \mathbb{N})$ (lower panel) with $n = 1{,}000$. The left panels are viewed from the φ direction, the right panels from the $\rho = \omega^2/\sigma^2$ direction. The true parameter values are always in the middle of the respective axes.

Table 1. The likelihood estimators of the unknown parameters for the processes $(X_j^{(\ell)} : j \in \mathbb{N})$, $\ell = 1, \ldots, 4$. The simulated processes were of length $n = 1{,}000$ and the estimated values were based on $N = 1{,}000$ repetitions, with standard deviations given in brackets. Note that for $\ell = 3, 4$, the variance parameter σ^2 could be estimated.

	φ		ω^2		σ^2	
ℓ	true	estimated	true	estimated	true	estimated
1	0.5000	0.4985 (0.0344)	0.2500	0.2483 (0.0516)	1.0000	0.9991 (0.0865)
2	0.5000	0.5010 (0.0569)	1.5000	1.4913 (0.0991)	1.0000	1.0043 (0.1070)
3	1.0000	0.9998 (0.0572)	3.0000	2.9965 (0.1379)	1.0000	1.4067 (1.9154)
4	1.5000	1.4991 (0.0318)	1.0000	0.9954 (0.0448)	1.0000	1.3283 (1.4056)

tent estimates of φ. For practical applications, however, this estimator requires $\sqrt{n}$-consistent estimators for the variance parameters ω^2 and σ^2 as well. These are not available in the literature and we propose to use the following estimators.

First note that

$$\hat{\rho}_{n,1} = \frac{1}{0.6n} \sum_{j=.4n+1}^{n} \frac{(X_j - X_{j-1})^2}{X_{j-1}^2} \qquad \text{and} \qquad \hat{\rho}_{n,2} = \frac{1}{0.6n} \sum_{j=.4n+1}^{n} \frac{X_j^2}{X_{j-1}^2}$$

are consistent estimators for $(\varphi - 1)^2 + \omega^2$ and $\varphi^2 + \omega^2$, respectively. Therefore, one can utilize the initial estimators $\tilde{\theta}_{n,1} = (\hat{\rho}_{n,2} - \hat{\rho}_{n,1} + 1)/2$ and $\tilde{\theta}_{n,2} = \hat{\rho}_{n,2}^2 - \tilde{\theta}_{n,1}^2$ in the nonstationary case. Notice that we have used only the last 60% of the observations to compute $\hat{\rho}_{n,1}$ and $\hat{\rho}_{n,2}$ due to superior performance in the simulation study. The consistency of all estimators follows immediately from the exponential growth of nonstationary RCA(1) processes.

Table 1 gives the summary statistics for the Monte Carlo study conducted with simulated versions of $(X_j^{(\ell)} : j \in \mathbb{N})$, $\ell = 1, \ldots, 4$, each of which have length $n = 1{,}000$. The results are based on $N = 1{,}000$ repetitions. The parameters φ and ω^2 were in all four cases very reasonably estimated. As predicted by the unified estimation theory, the quality of the estimation was unaffected by the nature of the RCA(1) process. The form of the histograms for φ and ω^2 (not shown here) indicate that normality was approximately approached for the finite sample size under consideration. It can also be conjectured from the simulations that the speed of convergence to the limiting normal law is indeed faster under nonstationarity than under stationarity. For the processes given by $\ell = 1, 2$, the third parameter σ^2 was consistently estimated, which is again according to the theory. For nonstationary RCA(1) processes, the likelihood is noninformative for σ^2. This is clearly reflected in the estimated values for $\ell = 3, 4$ which were respectively 41% and 33% too large.

4.4. Applications

In this section, we discuss two applications of the RCA(1) likelihood procedure to illustrate its performance on data. Due to space constraints, we only highlight the main features without presenting a more detailed analysis.

We considered the daily log volume of IBM stock for the ten year trading period 01/01/1993–12/31/2002, of sample size $n = 2{,}520$. The corresponding time series plot is provided in the first panel of Figure 3. The same data has been analyzed in Wang and Ghosh (2009) in the context of Bayesian estimation techniques. An application of the RCA(1) likelihood procedure initialized with the least squares estimators yields the parameter estimates

$$\hat{\varphi} = \hat{\theta}_{n,1} = 0.5306, \qquad \hat{\omega}^2 = \hat{\theta}_{n,2}^2 = 0.3088, \qquad \text{and} \qquad \hat{\sigma}^2 = \hat{\theta}_{n,3} = 0.1211.$$

These estimates imply that $\hat{\varphi}^2 + \hat{\omega}^2 = 0.5904$, so that the daily log volume can be regarded as stationary and the estimator $\hat{\sigma}^2$ as informative. An application

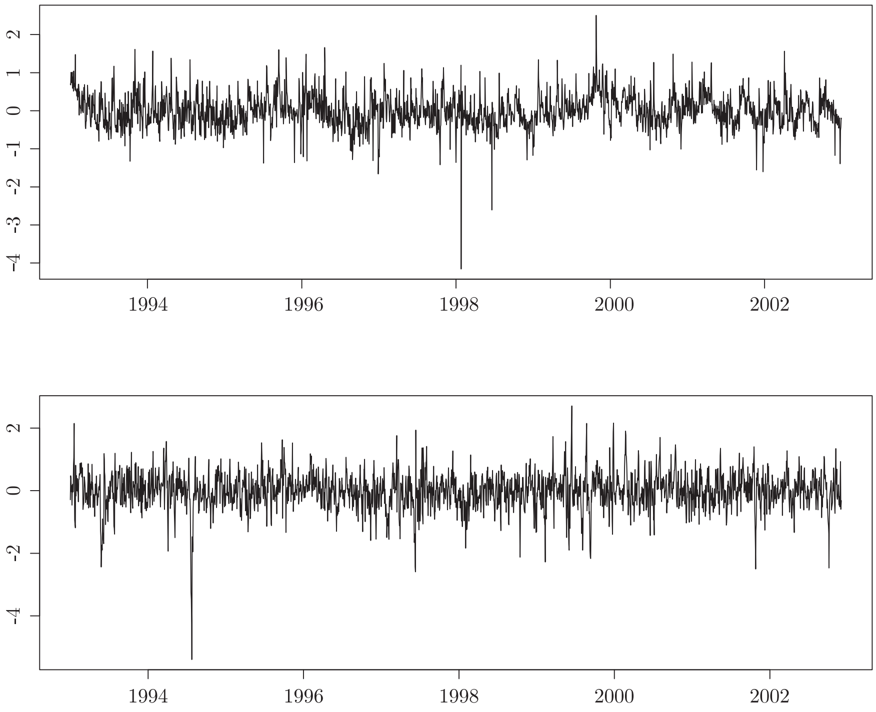

Figure 3. Ten years of detrended daily log volume of the IBM stock (upper panel), and ten years of simulated data using the RCA(1) fit (lower panel).

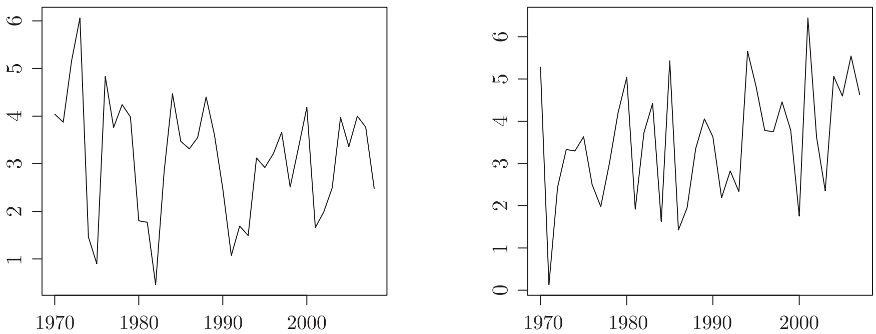

Figure 4. Annual changes in world GDP (left panel) and the corresponding simulated RCA(1) data (right panel).

of the stationarity test based on (4.1) obviously confirms this assertion. One can also use the same limit result to obtain asymptotic confidence intervals for $\hat{\varphi}^2 + \hat{\omega}^2$. For a given significance level $\alpha \in (0,1)$, these have the form $\hat{\varphi}^2 + \hat{\omega}^2 \pm z_{1-\alpha/2}\hat{v}_n/\sqrt{n}$, where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution and $\hat{v}_n^2 = 4\hat{\varphi}^2\hat{\Sigma}_{n,1,1} + 4\hat{\varphi}\hat{\Sigma}_{n,1,2} + \hat{\Sigma}_{n,2,2}$. The estimators $\hat{\Sigma}_{n,i,k}$

are the sample counterparts of the $\Sigma_{n,i,k}$ in equations (3.2)–(3.4). These can be computed utilizing the equalities of derivatives related to the log-likelihood

$$\frac{\partial g_j(\boldsymbol{u})}{\partial s} = 2Y_{j-1}(\boldsymbol{u},1,1)(X_j - sX_{j-1}),$$
$$\frac{\partial g_j(\boldsymbol{u})}{\partial x} = Y_{j-1}(\boldsymbol{u},2,2)(X_j - sX_{j-1})^2 - Y_{j-1}(\boldsymbol{u},2,1),$$
$$\frac{\partial g_j(\boldsymbol{u})}{\partial y} = Y_{j-1}(\boldsymbol{u},0,2)(X_j - sX_{j-1})^2 - Y_{j-1}(\boldsymbol{u},0,1),$$

where the quantitites $Y_{j-1}(\boldsymbol{u},\kappa,\gamma)$ is defined in (5.1). The latter equations can be verified with displays (18)–(20) in Aue, Horváth, and Steinebach (2006). The foregoing leads then to the 90% confidence interval

$$0.5904 \pm 1.6448\sqrt{\frac{1.9936}{2520}} = (0.5441, 0.6366)$$

for $\varphi^2 + \omega^2$. An application of Corollary 3.1 gives in a similar fashion the 90% confidence interval $(0.4942, 0.5670)$ for the autoregressive parameter φ. The second panel of Figure 3 contains 2,500 simulated data points using the RCA(1) specifications. The time series plot exhibits a high degree of similarity with the original data plot, displaying occasional bursty periods. While we do not pursue further diagnostic checking of the model fit, we conclude that the RCA(1) approach seems to be leading to a satisfactory outcome. Wang and Ghosh (2009) found (somewhat) different parameter estimates, namely $\hat{\varphi} = 0.871$, $\hat{\omega}^2 = 0.165$, and $\hat{\sigma}^2 = 0.565$. This leads consequently to $\hat{\varphi}^2 + \hat{\omega}^2 = 0.924$, which is much closer to the boundary than our estimate. Simulated data using the parameter estimates of Wang and Ghosh (2009), however, appeared to perform worse than our set of estimates.

As a second application, we considered data consisting of annual changes in world GDP between 1970 and 2008, with sample size $n = 39$. The corresponding time series plot is displayed in the left panel of Figure 4. As before, we apply the least squares initialized RCA(1) likelihood procedure to the data to obtain the parameter estimates

$$\hat{\varphi} = \hat{\theta}_{n,1} = 0.3274, \qquad \hat{\omega}^2 = \hat{\theta}^2_{n,2} = 0.8376, \qquad \text{and} \qquad \hat{\sigma}^2 = \hat{\theta}_{n,3} = 0.8563.$$

In the GDP case, we therefore have $\hat{\varphi}^2 + \hat{\omega}^2 = 0.9448$. For comparison, we have included in the right panel of Figure 4 a typical realization of an RCA(1) process using the GDP specifications. A visual inspection confirms that the simulated data seems to capture the main features of the original data.

5. The Consistency Proofs

Proof of Theorem 3.1. Suppose that $(X_j : j \in \mathbb{N})$ is the strictly stationary solution to (2.1). Following the arguments in Aue, Horváth, and Steinebach (2006), we obtain that $\sup_{\boldsymbol{u}\in\Gamma} \left|\ell_n(\boldsymbol{u}) - \ell(\boldsymbol{u})\right| \to 0$ a.s. as $n \to \infty$, where $\ell(\boldsymbol{u}) = E[\ell_n(\boldsymbol{u})]$. Since $\boldsymbol{u} = (\boldsymbol{v}, y)$, it holds with probability one that, for any $y > 0$, $\sup_{\boldsymbol{v}\in\Gamma_R} \ell_n(\boldsymbol{u}) \to \sup_{\boldsymbol{v}\in\Gamma_R} \ell(\boldsymbol{u})$ and consequently $\ell_n\big(\hat{\boldsymbol{\theta}}_n^{(R)}(y), y\big) \to \sup_{\boldsymbol{v}\in\Gamma_{R(y)}} \ell(\boldsymbol{u})$ a.s. as $n \to \infty$.

Observe next that $\ell_n(\boldsymbol{u}) = [g_1(\boldsymbol{u}) + \ldots + g_n(\boldsymbol{u})]/n$. The partial derivatives of the $g_j(\boldsymbol{u})$'s with respect to s and x are at (18) and (19) of Aue, Horváth, and Steinebach (2006), respectively. Utilizing these derivatives and the Dominated Convergence Theorem, we obtain that $\partial\ell(\boldsymbol{u})/\partial s = 2(\varphi - s)E\left[Y_1(\boldsymbol{u}, 2, 1)\right]$ and

$$\frac{\partial\ell(\boldsymbol{u})}{\partial x} = (\varphi - s)^2 E[Y_1(\boldsymbol{u}, 4, 2)] + \omega^2 E[Y_1(\boldsymbol{u}, 4, 2)] + \sigma^2 E[Y_1(\boldsymbol{u}, 2, 2)] - E[Y_1(\boldsymbol{u}, 2, 1)],$$

where we have used the notation

$$Y_j(\boldsymbol{u}, \kappa, \gamma) = \frac{X_j^\kappa}{(xX_j^2 + y)^\gamma}, \qquad \kappa = 0, 1, \ldots, 2\gamma, \qquad \gamma \in \mathbb{N}. \tag{5.1}$$

If (φ, ω^2) were indeed the location of the maximum, the partial derivatives of $\ell(\boldsymbol{u})$ with respect to s and x evaluated at this point must vanish. (Notice that (φ, ω^2) is assumed to be an inner point of the parameter space Γ_R.) This is clearly the case for the partial derivative with respect to s. On the other hand, the partial derivative with respect to x is

$$\left.\frac{\partial\ell(\boldsymbol{u})}{\partial x}\right|_{\boldsymbol{v}=(\varphi,\omega^2)} = E\left[\frac{X_1^2(\omega^2 X_1^2 + \sigma^2)}{(\omega^2 X_1^2 + y)^2}\right] - E\left[\frac{X_1^2}{\omega^2 X_1^2 + y}\right] = E\left[\frac{\sigma^2 - y}{(\omega^2 X_1^2 + y)^2}\right]$$

which is nonzero whenever y and σ^2 do not coincide. We consequently have with probability one that

$$\sup_{\boldsymbol{v}\in\Gamma_R} \ell(\boldsymbol{u}) > \ell(\varphi, \omega^2, y) \qquad \text{for all } y \neq \sigma^2. \tag{5.2}$$

Assume that $\hat{\boldsymbol{\theta}}_{n_k}^{(R)}(y) \to (\varphi, \omega^2)$ with probability one along a subsequence $(n_k : k \in \mathbb{N})$. Then $\ell_n(\hat{\boldsymbol{\theta}}_{n_k}^{(R)}(y), y)$ needs to converge with probability one to both $\sup_{\boldsymbol{v}\in\Gamma_R} \ell(\boldsymbol{u})$ and $\ell(\varphi, \omega^2, y)$, which is a contradiction in view of the strict inequality in (5.2). The proof of Theorem 3.1 is complete.

Proof of Theorem 3.2. In the strictly stationary case $E[\ln|\varphi + b_1|] < 0$, the result follows immediately from Theorem 2 in Aue, Horváth, and Steinebach (2006). For the rest of the proof take $E[\ln|\varphi + b_1|] \geq 0$. In the nonstationary case, it has been shown in Lemma 4.2 of Berkes, Horváth, and Ling (2009) that $\ell_n(\boldsymbol{u}) \to \infty$ for all $\boldsymbol{u} \in \Gamma$, but that

$$\sup_{\boldsymbol{u}\in\Gamma} \left|\ell_n(\boldsymbol{u}) - \ell_n(\boldsymbol{\theta}) - f(s,x)\right| \xrightarrow{P} 0 \quad \text{as } n \to \infty, \tag{5.3}$$

where the limit

$$f(s,x) = \ln \frac{\omega^2}{x} + 1 - \frac{\omega^2}{x} - \frac{(\varphi - s)^2}{x}$$

has a unique maximum in the point $(s,x) = (\varphi, \omega^2)$. It is, moreover, independent of the particular value of y. Using the Skorohod-Dudley-Wichura Representation Theorem (see Theorem 4 of Shorack and Wellner (1986)). we can define the quantities $\bar{\ell}_n(\boldsymbol{u})$ and $\bar{\boldsymbol{\theta}}_n$ as $\sup_{\boldsymbol{u}\in\Gamma} \bar{\ell}_n(\boldsymbol{u}) = \bar{\ell}_n(\bar{\boldsymbol{\theta}}_n)$ with the additional requirement that, for each n, $(\ell_n(\boldsymbol{u}), \hat{\boldsymbol{\theta}}_n \colon \boldsymbol{u} \in \Gamma)$ and $(\bar{\ell}_n(\boldsymbol{u}), \bar{\boldsymbol{\theta}}_n \colon \boldsymbol{u} \in \Gamma)$ coincide in distribution and such that (5.3) holds almost surely (instead of only in probability). To prove the assertion of Theorem 3.2, we show that the first two coordinates of $\bar{\boldsymbol{\theta}}_n$, denoted by $(\bar{\theta}_{n,1}, \bar{\theta}_{n,2})$, converge almost surely to (φ, ω^2). To do so, assume the contrary. Then there is a (random) subsequence $(n_k \colon k \in \mathbb{N})$ such that

$$(\bar{\theta}_{n,1}, \bar{\theta}_{n,2}) \to (\theta_1, \theta_2) \neq (\varphi, \omega^2) \quad \text{as } k \to \infty. \tag{5.4}$$

Since (5.3) is assumed to hold almost surely, it follows that $\sup_{\boldsymbol{u}\in\Gamma} \left[\ell_{n_k}(\bar{\boldsymbol{\theta}}_{n_k}) - \ell_{n_k}(\boldsymbol{\theta})\right] \to \sup_{\boldsymbol{u}\in\Gamma} f(s,x)$ as $k \to \infty$. Since $f(s,x)$ has its unique maximum in (φ, ω^2), the convergence in (5.4) produces a contradiction. The proof of Theorem 3.2 is hence complete.

6. The Asymptotic Normality Proofs

6.1. The proofs of Theorems 3.3 and 3.4

Consider the stationary case with $E[\ln|\varphi + b_1|] < 0$. Since the QMLE $\tilde{\boldsymbol{\theta}}$ consists of the first two coordinates of the full QMLE $\hat{\boldsymbol{\theta}}$ of Aue, Horváth, and Steinebach (2006), we can base our analysis on their results.

Lemma 6.1. *If* $(X_j \colon j \in \mathbb{N})$ *satisfies* $-\infty \leq E[\ln|\varphi + b_1|] < 0$, *then* $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}) \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, \Sigma)$ *as* $n \to \infty$, *where* Σ *is defined (for* $E[\ln|\varphi+b_1|] < 0$*) in Theorem* 3.4.

Proof. It follows from Theorem 3 in Aue, Horváth, and Steinebach (2006) (note the misprint in the covariance matrix) that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ is asymptotically normal with mean vector zero and covariance matrix $H^{-1}AH^{-1}$, where A has been defined before Theorem 3.4 and

$$H = \begin{bmatrix} 2\alpha(2,1) & 0 & 0 \\ 0 & \alpha(4,2) & \alpha(2,2) \\ 0 & \alpha(2,2) & \alpha(0,2) \end{bmatrix}.$$

Consequently, Σ is given by the upper (left) 2×2 submatrix in $H^{-1}AH^{-1}$. To verify that Σ as defined in Theorem 3.4 and this submatrix indeed conincide, one utilizes the inversion formula for a partitioned matrix (see Seber and Lee (2003)), to first compute H^{-1} and then performs the matrix multiplication. The proof is complete.

In the nonstationary case, $\hat{\boldsymbol{\theta}}_n$ may not converge in probability to $\boldsymbol{\theta}$, since asymptotically the quasi-likelihood procedure does not contain information on the value of σ^2, or $\hat{\boldsymbol{\theta}}_n$ may also be on the boundary of the admissible parameter set Γ. The first two coordinates subsumed under $\tilde{\boldsymbol{\theta}}_n$, however, do converge. This means that the partial derivatives of the log-likelihood function $\ell_n(\boldsymbol{u})$ with respect to s and x vanish in $\hat{\boldsymbol{\theta}}_n$. Let

$$\ell_n'(\boldsymbol{u}) = \left(\frac{\partial \ell_n(\boldsymbol{u})}{\partial s}, \frac{\partial \ell_n(\boldsymbol{u})}{\partial x}, \frac{\partial \ell_n(\boldsymbol{u})}{\partial y}\right)^T = \left(\ell_{n,1}'(\boldsymbol{u}), \ell_{n,2}'(\boldsymbol{u}), \ell_{n,3}'(\boldsymbol{u})\right)^T.$$

We work with the Mean Value Theorem for $\ell_{n,i}'(\boldsymbol{u})$, $i = 1, 2, 3$, about the true parameter vector $\boldsymbol{\theta}$. It holds that

$$\ell_{n,i}'(\hat{\boldsymbol{\theta}}_n) = \ell_{n,i}'(\boldsymbol{\theta}) + \boldsymbol{d}_{n,i}^T(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}), \qquad i = 1, 2, 3, \tag{6.1}$$

where $\boldsymbol{d}_{n,i}$ consists of the derivatives of $\ell_{n,i}'(\boldsymbol{u})$ with respect to s, x, and y in a random point $\boldsymbol{\theta}_{n,i}^*$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}_{n,i}^*\| \leq \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n\|$. Introducing the matrix

$$A_n = \begin{bmatrix} \boldsymbol{d}_{n,1}^T \\ \boldsymbol{d}_{n,2}^T \\ \boldsymbol{d}_{n,3}^T \end{bmatrix} = \begin{bmatrix} \tilde{A}_n & \boldsymbol{a}_n \\ \boldsymbol{a}_n^T & b_n \end{bmatrix}$$

and $\tilde{\ell}_n'(\boldsymbol{u}) = (\ell_{n,1}'(\boldsymbol{u}), \ell_{n,2}'(\boldsymbol{u}))^T$, (6.1) leads to the two-dimensional system

$$\boldsymbol{0} = \tilde{\ell}_n'(\boldsymbol{\theta}) + \tilde{A}_n(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}) + \boldsymbol{a}_n(\hat{\theta}_{n,3} - \sigma^2). \tag{6.2}$$

Our first result concerns the behavior of the 2×2 matrix $\tilde{A}_n$ as n increases. We write $\tilde{A}_n = (\tilde{A}_{n,i,j} \colon i, j = 1, 2)$ and $\tilde{A}^{-1} = (\tilde{A}_{n,i,j}^{(-1)} \colon i, j = 1, 2)$.

Lemma 6.2. *If* $(X_j \colon j \in \mathbb{N})$ *satisfies* $E[\ln|\varphi + b_1|] \geq 0$, *then*

$$\tilde{A}_n \xrightarrow{P} -\begin{bmatrix} \frac{2}{\omega^2} & 0 \\ 0 & \frac{1}{\omega^4} \end{bmatrix} = \tilde{A} \quad \text{as } n \to \infty. \tag{6.3}$$

This implies that $\tilde{A}_n^{-1} = \tilde{A}_n^{-1}(\boldsymbol{\theta}_n^*)$ *exists with probability tending to one and*

$$\tilde{A}_n^{-1} \xrightarrow{P} -\begin{bmatrix} \frac{\omega^2}{2} & 0 \\ 0 & \omega^4 \end{bmatrix} = \tilde{A}^{-1} \quad \text{as } n \to \infty. \tag{6.4}$$

Moreover,

$$\tilde{A}_{n,1,2} = \tilde{A}_{n,2,1} = \frac{\theta^*_{n,1} - \varphi}{\omega^2} + o_P(1)|\theta^*_{n,1} - \varphi| + o_P\left(\frac{1}{\sqrt{n}}\right) \qquad \text{as } n \to \infty, \tag{6.5}$$

$$\tilde{A}^{(-1)}_{n,1,2} = \tilde{A}^{(-1)}_{n,2,1} = \frac{2(\varphi - \theta^*_{n,1})}{\omega^4} + o_P(1)|\theta^*_{n,1} - \varphi| + o_P\left(\frac{1}{\sqrt{n}}\right) \ \text{as } n \to \infty. \tag{6.6}$$

Proof. Since $\tilde{\boldsymbol{\theta}}_n \to \tilde{\boldsymbol{\theta}}$ in probability as $n \to \infty$, we study more generally the properties of the second derivatives of $\ell_n(\boldsymbol{u})$ on the parameter set $\Gamma_n = \{\boldsymbol{u} \in \mathbb{R}^3 \colon |s - \varphi| \le \epsilon_n, |x - \omega^2| \le \epsilon_n, y_* \le y \le y^*\}$, where $\epsilon_n \searrow 0$ as $n \to \infty$. We start with $\partial^2 \ell_n(\boldsymbol{u})/\partial s^2$. Computing the second derivative of the log-likelihood and equation (2.3), it holds that

$$\frac{\partial^2 \ell_n(\boldsymbol{u})}{\partial s^2} = -\frac{2}{n}\sum_{j=1}^n \frac{X^2_{j-1}}{xX^2_{j-1} + y}.$$

Now

$$\sup_{\boldsymbol{u}\in\Gamma_n}\left|\frac{\partial^2 \ell_n(\boldsymbol{u})}{\partial s^2} + \frac{2}{\omega^2}\right| \le \sup_{\boldsymbol{u}\in\Gamma_n}\left|\frac{2}{n}\sum_{j=1}^n \frac{y}{(xX^2_{j-1} + y)x}\right| + \sup_{\boldsymbol{u}\in\Gamma_n}\left|\frac{2}{\omega^2} - \frac{2}{x}\right| = o_P(1),$$

since $X^2_{j-1} \to \infty$ in probability (see Corollary 3.1 in Berkes, Horváth, and Ling (2009)). We have further that

$$\begin{aligned}\frac{\partial^2 \ell_n(\boldsymbol{u})}{\partial s \partial x} &= -\frac{\varphi - s}{n}\sum_{j=1}^n Y_{j-1}(\boldsymbol{u},4,2) - \frac{1}{n}\sum_{j=1}^n b_j Y_{j-1}(\boldsymbol{u},4,2) - \frac{2}{n}\sum_{j=1}^n e_j Y_{j-1}(\boldsymbol{u},3,2)\\ &= B_{n,1}(\boldsymbol{u}) + B_{n,2}(\boldsymbol{u}) + B_{n,3}(\boldsymbol{u}).\end{aligned}$$

Now

$$\begin{aligned}&\sup_{\boldsymbol{u}\in\Gamma}\left|\frac{1}{n}\sum_{j=1}^n Y_{j-1}(\boldsymbol{u},4,2) - \frac{1}{x^2}\right|\\ &\quad\le \frac{1}{x_* n}\sum_{j=1}^n \frac{X^2_{j-1}}{(x_* X^2_{j-1} + y_*)^2} + \frac{(y^*)^2}{(x_*)^2 n}\sum_{j=1}^n \frac{1}{(x_* X^2_{j-1} + y_*)^2} = o_P(1)\end{aligned}$$

and therefore

$$\sup_{\boldsymbol{u}\in\Gamma_n}\left|B_{n,1}(\boldsymbol{u}) - \frac{s - \varphi}{x^2}\right| = o_P(1)\sup_{\boldsymbol{u}\in\Gamma_n}\left|\frac{s - \varphi}{x^2}\right| \qquad (n \to \infty). \tag{6.7}$$

Next, write

$$\sqrt{n}B_{n,2}(\boldsymbol{u}) = \frac{1}{x^2\sqrt{n}}\sum_{j=1}^{n} b_j - \frac{2}{x\sqrt{n}}\sum_{j=1}^{n} b_jY_{j-1}(\boldsymbol{u},2,2) - \frac{y^2}{x^2\sqrt{n}}\sum_{j=1}^{n} b_jY_{j-1}(\boldsymbol{u},0,2)$$
$$= \tilde{B}_{n,1}(\boldsymbol{u}) + \tilde{B}_{n,2}(\boldsymbol{u}) + \tilde{B}_{n,3}(\boldsymbol{u}). \tag{6.8}$$

Clearly, $\sup_{\boldsymbol{u}\in\Gamma_n}|\tilde{B}_{n,1}(\boldsymbol{u})| = \mathcal{O}_P(1)$. For $n \geq 1$, let $Z_n(x,y) = x\tilde{B}_{n,2}(\boldsymbol{u})/2$. Then, direct computations show that $E[Z_n(x,y)] = 0$ and $\mathrm{Var}(Z_n(x,y)) \to 0$ as $n \to \infty$, using the independence of b_j and X_{j-1}. Consequently, we obtain for any admissible x and y that

$$Z_n(x,y) \xrightarrow{P} 0 \quad \text{as } n \to \infty. \tag{6.9}$$

For $x_1 < x_2$ and $y_1 < y_2$ denote by $\Delta_{(x_1,y_1)}^{(x_2,y_2)}Z_n(\boldsymbol{u})$ the increment of $Z_n(\boldsymbol{u})$ over the rectangle determined by the coordinates (x_1, y_1) and (x_2, y_2). Then,

$$E\left[\left(\Delta_{(x_1,y_1)}^{(x_2,y_2)}Z_n(\boldsymbol{u})\right)^2\right] = \frac{\omega^2}{n}\sum_{j=1}^{n} E\left[\left(\Delta_{(x_1,y_1)}^{(x_2,y_2)}Y_{j-1}(\boldsymbol{u},2,2)\right)^2\right]$$
$$\leq C(x_2-x_1)^2(y_2-y_1)^2,$$

where $C > 0$ is a suitable constant. Hence, Bickel and Wichura (1971) yields that the sequence $(Z_n(\boldsymbol{u})\colon n \in \mathbb{N})$ is tight which, in combination with (6.9), implies that $|Z_n(x,y)| = o_P(1)$ uniformly in $\boldsymbol{u} \in \Gamma$. We conclude that

$$\sup_{\boldsymbol{u}\in\Gamma_n}|\tilde{B}_{n,2}(\boldsymbol{u})| = o_P(1) \quad \text{as } n \to \infty. \tag{6.10}$$

By analogous reasoning, one shows that $\sup_{\boldsymbol{u}\in\Gamma_n}|\tilde{B}_{n,3}(\boldsymbol{u})| = o_P(1)$, resulting in

$$\sup_{\boldsymbol{u}\in\Gamma_n}|B_{n,2}(\boldsymbol{u})| = o_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{as } n \to \infty. \tag{6.11}$$

Similar arguments apply to the remaining term $B_{n,3}(\boldsymbol{u})$ and (6.5) is established. By (6.5) we have immediately that $\tilde{A}_{n,1,2} = \tilde{A}_{n,2,1} \to 0$ in probability as $n \to \infty$.

The last quantity to consider is $\tilde{A}_{n,2,2}(\boldsymbol{u})$. It can be shown that

$$\frac{\partial^2\ell_n(\boldsymbol{u})}{\partial x^2} = -\frac{2}{n}(\varphi-s)^2\sum_{j=1}^{n}Y_{j-1}(\boldsymbol{u},6,3) - \frac{4}{n}(\varphi-s)\sum_{j=1}^{n}b_jY_{j-1}(\boldsymbol{u},6,3)$$
$$-\frac{2}{n}\sum_{j=1}^{n}b_j^2Y_{j-1}(\boldsymbol{u},6,3) - \frac{4}{n}\sum_{j=1}^{n}(\varphi-s+b_j)Y_{j-1}(\boldsymbol{u},5,3)$$
$$-\frac{2}{n}\sum_{j=1}^{n}e_j^2Y_{j-1}(\boldsymbol{u},4,3) + \frac{1}{n}\sum_{j=1}^{n}Y_{j-1}(\boldsymbol{u},4,2)$$
$$= C_{n,1}(\boldsymbol{u}) + \ldots + C_{n,6}(\boldsymbol{u}).$$

Using the arguments previously established in this proof, we can conclude that, for $i = 1, 2, 4, 5$, $C_{n,i}(\boldsymbol{u})$ converges in probability to 0 uniformly in $\boldsymbol{u} \in \Gamma_n$ as $n \to \infty$. It remains to investigate the asymptotic behavior of $C_{n,3}(\boldsymbol{u}) + C_{n,6}(\boldsymbol{u})$. To this end, observe first that

$$\begin{aligned}\sup_{\boldsymbol{u}\in\Gamma_n}\left|C_{n,3}(\boldsymbol{u}) + \frac{2}{\omega^4}\right| &\leq \sup_{\boldsymbol{u}\in\Gamma_n}\left|C_{n,3} + \frac{2\omega^2}{n}\sum_{j=1}^{n} Y_{j-1}(\boldsymbol{u}, 6, 3)\right| \\ &\quad + \sup_{\boldsymbol{u}\in\Gamma_n}\left| - \frac{2\omega^2}{n}\sum_{j=1}^{n} Y_{j-1}(\boldsymbol{u}, 6, 3) + \frac{2}{\omega^4}\right| \\ &= o_P(1).\end{aligned}$$

For the first supremum, this follows similarly as above. For the second supremum, it can be seen from Taylor expansions for the terms $Y_{j-1}(\boldsymbol{u}, 6, 3)$ that $[Y_0(\boldsymbol{u}, 6, 3) + \ldots + Y_{n-1}(\boldsymbol{u}, 6, 3)]/n$ converges in probability to $-1/\omega^4$ uniformly in $\boldsymbol{u} \in \Gamma_n$. In the same fashion, one shows that $\sup_{\boldsymbol{u}\in\Gamma_n}\left|C_{n,6}(\boldsymbol{u}) - 1/\omega^4\right| = o_P(1)$ as $n \to \infty$. The proof of (6.3) is now complete. By (6.3) and (6.5), the inverse formula for a symmetric 2×2 matrix implies (6.6).

We return to the two-dimensional equation system given in (6.2). Since the inverse $\tilde{A}_n^{-1}$ exists with probability approaching one according to Lemma 6.2, we can solve (6.2) for $\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}$. This yields

$$\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}} = -\tilde{A}_n^{-1}\tilde{\ell}_n'(\boldsymbol{\theta}) - \tilde{A}_n^{-1}\boldsymbol{a}_n(\hat{\theta}_{n,3} - \sigma^2). \tag{6.12}$$

It is our goal to establish the asymptotic normality in this section. We do so first for the first coordinate in (6.12). The result is formulated as the next lemma.

Lemma 6.3. *If* $(X_j : j \in \mathbb{N})$ *satisfies* $E[\ln|\varphi + b_1|] \geq 0$, *then* $\sqrt{n}(\hat{\theta}_{n,1} - \varphi) \xrightarrow{\mathcal{D}} N(0, \omega^2)$ *as* $n \to \infty$.

Proof. The first coordinate of (6.12) reads

$$\begin{aligned}\hat{\theta}_{n,1} - \varphi &= -\tilde{A}_{n,1,1}^{(-1)}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial s} - \tilde{A}_{n,1,2}^{(-1)}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial x} - \left[\tilde{A}_{n,1,1}^{(-1)}a_{n,1} + \tilde{A}_{n,1,2}^{(-1)}a_{n,2}\right](\hat{\theta}_{n,3} - \sigma^2) \\ &= D_{n,1} + D_{n,2} + D_{n,3},\end{aligned} \tag{6.13}$$

where $\boldsymbol{a}_n = (a_{n,1}, a_{n,2})^T$. The proof is given in three steps. In the first step, we show that the asymptotic normality is due to the term $\sqrt{n}D_{n,1}$. In the remaining two steps, we show that $\sqrt{n}D_{n,2}$ and $\sqrt{n}D_{n,3}$ vanish in the limit.

Step 1: Utilizing the partial derivative of $g_j(\boldsymbol{\theta})$ with respect to s and the definition of the log-likelihood function in (2.3), we get

$$\sqrt{n}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial s} = \frac{2}{\sqrt{n}}\sum_{j=1}^{n} b_j Y_{j-1}(\boldsymbol{\theta}, 2, 1) + \frac{2}{\sqrt{n}}\sum_{j=1}^{n} e_j Y_{j-1}(\boldsymbol{\theta}, 1, 1) = E_{n,1} + E_{n,2}.$$

Easy computations using the independence of e_j and $Y_{j-1}(\boldsymbol{\theta},1,1)$ imply that $\mathrm{Var}(E_{n,2}) \to 0$, and therefore that $E_{n,2} \xrightarrow{P} 0$ as $n \to \infty$. On the other hand, it holds that

$$E_{n,2} = \frac{2}{\sqrt{n}\omega^2}\sum_{j=1}^{n} b_j - \frac{2\sigma^2}{\sqrt{n}\omega^2}\sum_{j=1}^{n}\frac{b_j}{\omega^2 X_{j-1}^2+\sigma^2}.$$

The second term on the right-hand side of the latter array is negligible because it has zero mean and its variance converges to zero. The first term on the right-hand side clearly converges in distribution to a normal random variable with zero mean and variance $4/\omega^2$. This, combined with $\tilde{A}_{n,1,1}^{(-1)} \xrightarrow{P} \omega^2/2$ (see Lemma 6.2), implies that $\sqrt{n}D_{n,1}$ converges in distribution to a centered normal variate with variance ω^2.

Step 2: To prove that $\sqrt{n}D_{n,2}$ does not contribute to the limit distribution, it suffices to study the partial derivative of $\ell_n(\boldsymbol{\theta})$ with respect to x. Now

$$\begin{aligned}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial x} &= \frac{1}{n}\sum_{j=1}^{n}(b_j^2-\omega^2)Y_{j-1}(\boldsymbol{\theta},4,2) + \frac{2}{n}\sum_{j=1}^{n} b_j e_j Y_{j-1}(\boldsymbol{\theta},3,2)\\ &\quad + \frac{1}{n}\sum_{j=1}^{n} e_j^2 Y_{j-1}(\boldsymbol{\theta},2,2) + \frac{1}{n}\sum_{j=1}^{n}\left[\omega^2 Y_{j-1}(\boldsymbol{\theta},4,2) - Y_{j-1}(\boldsymbol{\theta},2,1)\right]\\ &= F_{n,1} + \ldots + F_{n,4}.\end{aligned}$$

Notice that elementary calculations yield $E[F_{n,1}] = 0$, $\mathrm{Var}(F_{n,1}) \to 0$, $E[F_{n,2}] = 0$ and $\mathrm{Var}(F_{n,2}) \to 0$, as well as $E[F_{n,3}] \to 0$. This results in $F_{n,i} = o_P(1)$ for $i = 1,2,3$ as $n \to \infty$. Moreover, the nonpositive $F_{n,4}$ can be transformed into $F_{n,4} = -(\sigma^2/n)\sum_{j=1}^{n} Y_{j-1}(\boldsymbol{\theta},2,2)$, thus showing that $E[F_{n,4}] \to 0$ and consequently $F_{n,4} = o_P(1)$ as $n \to \infty$. This completes the second part of the proof.

Step 3: That $\sqrt{n}D_{n,3} = o_P(1)$ is implied by (a) $\sqrt{n}a_{n,1} = o_P(1)$ and (b) $a_{n,2} = o_P(1)$. As for (a), observe that $a_{n,1} = \partial^2\ell_n(\boldsymbol{\theta}_n^*)/(\partial y \partial s)$. We can write

$$\sqrt{n}\frac{\partial^2 \ell_n(\boldsymbol{u})}{\partial y \partial s} = \frac{2}{\sqrt{n}}\sum_{j=1}^{n}(\varphi - s + b_j)Y_{j-1}(\boldsymbol{u},2,2) + \frac{2}{\sqrt{n}}\sum_{j=1}^{n} e_j Y_{j-1}(\boldsymbol{u},1,2).$$

Repeating the arguments that led to (6.10), we get that both sums on the right-hand side of the latter equation converge to zero in probability uniformly in $\boldsymbol{u} \in \Gamma_n$. The proof of (b) is similar and therefore omitted.

Combining the three steps of this lemma with (6.6) and (6.13) yields

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{n,1}-\varphi) &= -\tilde{A}_{n,1,1}^{(-1)}\sqrt{n}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial s} - \sqrt{n}(\theta_{n,1}^*-\varphi)o_P(1) - \tilde{A}_{n,1,1}^{(-1)}\sqrt{n}a_{n,1}(\tilde{\theta}_{n,3}-\sigma^2)\\ &= -\tilde{A}_{n,1,1}^{(-1)}\sqrt{n}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial s} + o_P(1),\end{aligned}$$

where the last equality sign follows from $\sqrt{n}|\theta^*_{n,1} - \varphi| \leq \sqrt{n}|\hat{\theta}_{n,1} - \varphi| = \mathcal{O}_P(1)$ and Step 3 above. Lemma 6.3 is established.

Proof of Theorem 3.3. The assertion follows from Lemmas 6.1 and 6.3.

In going back to display (6.12), we study the joint behavior of both coordinates of $\tilde{\boldsymbol{\theta}}_n$ and state the following analogue of Lemma 6.3. Notice that the transition case $E[\ln|\varphi + b_1|] = 0$ is excluded here.

Lemma 6.4. *If* $(X_j : j \in \mathbb{N})$ *satisfies* $E[\ln|\varphi + b_1|] > 0$*, then* $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}) \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, \Sigma)$ *as* $n \to \infty$*, where* Σ *is (for* $E[\ln|\varphi + b_1|] > 0$*) defined in Theorem 3.4.*

Proof. The proof is given in two steps. In the first step, we show that the term $-\tilde{A}_n^{-1}\tilde{\ell}'_n(\boldsymbol{\theta})$ in the decomposition (6.12) induces the asymptotic normality. In the second step, it is shown that the remainder term $-\tilde{A}_n^{-1}\boldsymbol{a}_n(\hat{\theta}_{n,3} - \sigma^2)$ is negligible.

Step 1: The first coordinate has already been studied in Lemma 6.3, where it was shown that

$$\sqrt{n}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial s} = \frac{2}{\sqrt{n}\omega^2}\sum_{j=1}^n b_j + o_P(1) \quad \text{as } n \to \infty.$$

The same arguments also give

$$\sqrt{n}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial x} = \frac{1}{\sqrt{n}\omega^4}\sum_{j=1}^n (b_j^2 - \omega^2) + o_P(1) \quad \text{as } n \to \infty.$$

The Multivariate Central Limit Theorem then implies that the vector $\ell'_n(\boldsymbol{\theta})$ converges in distribution to a bivariate normal random variate with mean zero and covariance matrix

$$\Omega = \begin{bmatrix} \frac{4}{\omega^2} & \frac{2E[b_1^3]}{\omega^6} \\ \frac{2E[b_1^3]}{\omega^6} & \frac{\mathrm{Var}(b_1^2)}{\omega^8} \end{bmatrix}.$$

Using the matrix $\tilde{A}^{-1}$ defined in Lemma 6.2, it is now easily established that $\tilde{A}^{-1}\Omega\tilde{A}^{-1} = \Sigma$. Thus, $-\tilde{A}_n^{-1}\tilde{\ell}'_n(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, \Sigma)$ as $n \to \infty$, and the first step of the proof is complete.

Step 2: The first coordinate of the remainder term $-\tilde{A}_n^{-1}\boldsymbol{a}_n(\hat{\theta}_{n,3} - \sigma^2)$ has shown to be negligible in the proof of Lemma 6.3, Step 3. The second coordinate reads as $-\big(\tilde{A}^{-1}_{n,2,1}a_{n,1} + \tilde{A}^{-1}_{n,2,2}a_{n,2}\big)(\hat{\theta}_{n,3} - \sigma^2)$. It follows from the proof of Lemma 6.2 that $\tilde{A}^{-1}_{n,2,1} = \mathcal{O}_P(1/\sqrt{n})$ and, from Step 3 in the proof of Lemma 6.3, that $a_{n,1} = o_P(1)$ as $n \to \infty$. Hence, the first term in the last display disappears in the limit. As for the second, it holds that $\tilde{A}^{-1}_{n,2,2} \xrightarrow{P} \omega^4$. Thus, we need to prove that $a_{n,2} = o_P(1/\sqrt{n})$ to complete the proof. To this end, notice that, if $\kappa < 2\gamma$ and with probability one, $\sup_{\boldsymbol{u}\in\Gamma_n}\sum_{j=1}^n b_j^2 Y_{j-1}(\boldsymbol{u},\kappa,\gamma) < \infty$, $\sup_{\boldsymbol{u}\in\Gamma_n}\sum_{j=1}^n |b_j e_j| Y_{j-1}(\boldsymbol{u},\kappa,\gamma) < \infty$ and $\sup_{\boldsymbol{u}\in\Gamma_n}\sum_{j=1}^n e_j^2 Y_{j-1}(\boldsymbol{u},\kappa,\gamma) < \infty$. This implies that

$$\sup_{\boldsymbol{u}\in\Gamma_n}\frac{a_{n,2}}{\sqrt{n}} = \sup_{\boldsymbol{u}\in\Gamma_n}\frac{1}{\sqrt{n}}\left|\sum_{j=1}^n \frac{\partial^2 g_j(\boldsymbol{u})}{\partial x \partial y}\right| \xrightarrow{P} 0 \quad \text{as } n\to\infty,$$

and the proof of Lemma 6.4 is complete.

Proof of Theorem 3.4. The assertion follows from Lemmas 6.1 and 6.4.

6.2. The proofs of Corollaries 3.1 and 3.2

In the remaining subsection, we establish the consistency of the covariance estimator $\hat{\Sigma}_n$ that has been used in Corollary 3.2. Since its first entry is equal to the estimator $\hat{\tau}_n^2$ of Corollary 3.1, both corollaries are immediate consequences of the next two lemmas.

Lemma 6.5. *If $(X_j\colon j\in\mathbb{N})$ satisfies $E[\ln+|\varphi+b_1|]<\infty$, then $\hat{\tau}_n^2$ is a weakly consistent estimator for τ^2, where $\hat{\tau}_n^2$ and τ^2 are defined in Corollary 3.1 and Theorem* 3.3.

Proof. We need to show that $\hat{\tau}_n^2 \xrightarrow{P} \tau^2$ holds both in the stationary and nonstationary environment. Let $E[\ln|\varphi+b_1|]<0$. It follows from an application of the ergodic theorem that, uniformly in $\boldsymbol{u}\in\Gamma_n$ and with probability one, $(1/n)\sum_{j=1}^n Y_j(\boldsymbol{u},\kappa,\gamma) \to E[Y_1(\boldsymbol{u},\kappa,\gamma)]$ as $n\to\infty$. Since the limit is continuous and $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}$ (even with probability one; see Theorem 2 in Aue, Horváth, and Steinebach (2006)), we conclude that $\alpha_n(\kappa,\gamma) \xrightarrow{P} \alpha(\kappa,\gamma)$. This proves consistency in the stationary case.

Let $E[\ln|\varphi+b_1|]>0$ and note that

$$\frac{1}{n}\sum_{j=1}^n Y_{j-1}(\boldsymbol{u},2,1) = \frac{1}{x} + \frac{1}{n}\sum_{j=1}^n\left(\frac{X_j^2}{xX_{j-1}^2+y} - \frac{1}{x}\right) = \frac{1}{x} - \frac{y}{x}\frac{1}{n}\sum_{j=1}^n Y_j(\boldsymbol{u},0,1),$$

where the last sum converges in probability to zero uniformly in $\boldsymbol{u}\in\Gamma_n$. Since $\hat{\theta}_{n,2} \xrightarrow{P} \omega^2$, we get $\hat{\alpha}_n(2,1) \xrightarrow{P} 1/\omega^2$ as $n\to\infty$. Similarly one concludes that $\hat{\alpha}_n(4,2) \xrightarrow{P} 1/\omega^4$ and $\hat{\alpha}_n(2,2) \xrightarrow{P} 0$ as $n\to\infty$. Thus $\hat{\tau}_n^2 \xrightarrow{P} \omega^2$ in the nonstationary case. This completes the proof.

Lemma 6.6. *If $(X_j\colon j\in\mathbb{N})$ satisfies $E[\ln|\varphi+b_1|]\neq 0$, then $\hat{\Sigma}_n$ is a weakly consistent estimator for the covariance matrix Σ defined in Theorem* 3.4.

Proof. The proof is similar to the proof of the preceding lemma, so we only give an outline here. One verifies, with the Ergodic Theorem, that, uniformly in $\boldsymbol{u}$,

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\partial g_j(\boldsymbol{u})}{\partial_i}\frac{\partial g_j(\boldsymbol{u})}{\partial_k}\xrightarrow{P} E\left[\frac{\partial g_1(\boldsymbol{u})}{\partial_i}\frac{\partial g_1(\boldsymbol{u})}{\partial_k}\right] \quad \text{as } n\to\infty$$

in the stationary case. Since the limit is continuous in $\boldsymbol{u}$ and since $\hat{\boldsymbol{\theta}}_n \stackrel{P}{\to} \boldsymbol{\theta}$, it follows that $\hat{A}_{n,i,k} \stackrel{P}{\to} A_{i,k}$. In the nonstationary case, similar arguments as those applied in Step 2 of the previous proof lead to the conclusion. Details are omitted.

Acknowledgement

This research was supported in part by National Science Foundation grants DMS 0604670, DMS 0652420 and DMS 0905400.

References

Andél, J. (1976). Autoregressive series with random parameters. *Mathematische Operationsforschung und Statistik, Series Statistics* **7**, 735-741.

Aue, A. (2008). Near-integrated random coefficient autoregressive time series. *Econom. Theory* **24**, 1343-1372.

Aue, A. and Horváth, L. (2007). A limit theorem for mildly explosive autoregression with stable errors. *Econom. Theory* **23**, 201-220.

Aue, A., Horváth, L. and Steinebach, J. (2006). Estimation in random coefficient autoregressive models. *J. Time Ser. Anal.* **27**, 61-76.

Berkes, I., Horváth, L. and Ling, S. (2009). Estimation in nonstationary random coefficient autoregressive models. *J. Time Ser. Anal.* **30**, 395-416.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656-1670.

Conlisk, J. (1974). Stability in a random coefficient model. *Internat. Econom. Rev.* **15**, 529-533.

Conlisk, J. (1976). A further note on stability in a random coefficient model. *Internat. Econom. Rev.* **17**, 759-764.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427-431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica* **49**, 1057-1072.

Engle, R. F. and Kraft, D. F. (1983). Multiperiod forecast error variances on inflation estimated from ARCH models. In *Appl. Time Ser. Econom. Data* (edited by A. Zellner), 293-302. U.S. Bureau of the Census, Washington DC.

Fama, E. F. (1965). The behavior of stock-market prices. *J. Business* **38**, 34-105.

Feigin, P. D. and Tweedie, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6**, 1-14.

Hwang, S. Y. and Basawa, I. V. (2005). Explosive random coefficient AR(1) processes and related asymptotics for least squares estimation. *J. Time Ser. Anal.* **26**, 807-824.

Jensen, S. T. and Rahbek, A. (2004a). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**, 641-646.

Jensen, S. T. and Rahbek, A. (2004b). Asymptotic inference for nonstationary GARCH. *Econom. Theory* **20**, 1203-1226.

Koul, H. L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* 24, 1025-1052.

Li, W. K. and Hui, Y. V. (1983). Estimation of random coefficient autoregressive process: An empirical Bayes approach. *J. Time Ser. Anal.* **4**, 89-94.

Liu, L. M. and Tiao, G. C. (1980). Random coefficient first-order autoregressive models. *J. Econom.* **13**, 305-325.

Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **64**, 575-596.

Nelson, C. R. and Plosser, C. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *J. Monetary Economics* **10**, 139-162.

Nicholls, D. F. and Quinn, B. G. (1980). The estimation of random coefficient autoregressive models. *J. Time Ser. Anal.* **1**, 37-46.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Springer-Verlag, New York.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica* **56**, 1021-1043.

Phillips, P. C. B. and Magdalinos, T. (2007). Limit Theory for Moderate Deviations From a Unit Root. *J. Econometrics* **136**, 115-130.

Phillips, P. C. B. and Perron, P. (1998). Testing for a unit root in time series regression. *Biometrika* **75**, 335-346.

Phillips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing. *J. Econom. Surveys* **12**, 423-470.

Quinn, B. G. and Nicholls, D. F. (1981). The estimation of random coefficient autoregressive models II. *J. Time Ser. Anal.* **2**, 185-203.

Rahiala, M. (1999). Random coefficient autoregressive models for longitudinal data. *Biometrika* **86**, 718-722.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scand. J. Statist.* **5**, 163-168.

Rudolph, A. (1998). A central limit theorem for random coefficient autoregressive models and ARCH/GARCH models, *Adv. Appl. Probab.* **30**, 113-121.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis.* John Wiley, Hoboken, NY.

Shorack, G. R., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley, New York.

Stenseth, N. C., Falck, W., Chan, K. S., Bjornstad, O. N., O'Donoghue, M., Tong, H., Boonstra, R., Boutin, S., Krebs, C. J. and Yoccoz, N. G. (1999). From ecological patterns to ecological processes: Phase- and density-dependencies in the Canadian lynx cycle," *Proc. Natl. Acad. Sci. USA* **95**, 15430-15435.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *J. Amer. Statist. Assoc.* **82**, 590-604.

Wang, D. and Ghosh, S. K. (2009). Bayesian analysis of random coefficient autoregressive models. Institute of Statistics Mimeo Series #2566, North Carolina State University.

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.

E-mail: alexaue@wald.ucdavis.edu

Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, UT 84112, USA.

E-mail: horvath@math.utah.edu

(Received March 2009; accepted January 2010)