Autoregressive conditional proportion: A multiplicative-error model for (0,1)-valued time series

Aknouche, Abdelhakim and Dimitrakopoulos, Stefanos

Department of Mathematics, College of Science, Qassim University, Saudi Arabia., Department of Economics, University of Leeds, UK.

6 December 2021

Autoregressive conditional proportion: A multiplicative-error model for (0,1)-valued time series

Abdelhakim Aknouche

Departement of Mathematics, Qassim University, Saudi Arabia

Stefanos Dimitrakopoulos

Department of Economics, University of Leeds, UK

Abstract

We propose a multiplicative autoregressive conditional proportion (ARCP) model for (0,1)-valued time series, in the spirit of GARCH (generalized autoregressive conditional heteroscedastic) and ACD (autoregressive conditional duration) models. In particular, our underlying process is defined as the product of a (0,1)-valued iid sequence and the inverted conditional mean, which, in turn, depends on past reciprocal observations in such a way that is larger than unity. The probability structure of the model is studied in the context of the stochastic recurrence equation theory, while estimation of the model parameters is performed by the exponential quasi-maximum likelihood estimator (EQMLE). The consistency and asymptotic normality of the EQMLE are both established under general regularity assumptions. Finally, the usefulness of our proposed model is illustrated with simulated and two real datasets.

Keywords: Proportional time series data, Beta-ARMA model, Simplex ARMA, Autoregressive conditional duration, Exponential QMLE.

1 Introduction

Time series models can be classified into two broad categories, equation-based models and (conditional) distribution-based models. In the first category (e.g. Box et *al*, 1994; Brockwell and Davis, 1991; Hamilton, 1994; Francq and Zakoian, 2019), an observed time series sequence is assumed to be a solution to a stochastic difference equation driven by a specific innovation input with a precise structure (iid, martingale difference, white noise, etc.). In the second category (e.g. Zeger and Qaqish, 1987; Li, 1994; Benjamin et *al*, 2003; Zheng et *al*, 2015), a time series is assumed to follow a given conditional distribution, which depends on a stochastic time-varying parameter, representing a specific characteristic, such as the conditional mean, the conditional variance, the conditional kurtosis, and the conditional skewness. In some cases, it is possible to switch from one representation to another (Engle and Russell, 1998; Engle, 2002).

The equation-based approach is quite useful in defining relatively misspecified models, whose probabilistic structure can be easily revealed and their statistical properties can be estimated by distribution-free methods (for example, quasi-maximum likelihood and weighted least squares estimators). On the contrary, distribution-based models have relatively more complex structures (Aknouche and Francq, 2021a) and because their distributions are well-specified, they are usually estimated using maximum likelihood or Bayesian methods. Nevertheless, compared to equation-based models, distribution-based models are adapted more naturally to the support of the data (bounded, positive, integer-valued, binary) and therefore they do not require any prior transformation that could lead to interpretability issues, distributional inadequacy, and biased predictions.

Over the past few years a strand of the statistics and econometrics literature has attempted to model time series having a specific range with the distribution-based approach. This literature deals, mainly, with integer-valued time series models (e.g. Grunwald et al, 2000; Heinen, 2003; McKenzie, 2003; Ferland et *al*, 2006; Davis and Liu, 2016), nonnegative time series models (e.g. Benjamin et *al*, 2003; Zheng et *al*, 2015), and bounded-valued time series models (e.g. Rocha and Cribari-Neto, 2009; Guolo and Varin, 2014; Gorgi and

Koopman, 2021).

Continuous bounded-valued data, are very common in empirical applications. These applications include, among others, proportions, rates, indexes, probabilities, realized correlations and angle measurements (McKenzie, 1985; Paolino, 2001; Kieschnick and McCullough, 2003; Ferrari and Cribari-Neto, 2004; Rocha and Cribari-Neto, 2009; Guolo and Varin, 2014; Zhang et *al*, 2016; Bonat et *al*, 2019; Gorgi and Koopman, 2021). Generally, bounded data exhibit some specific characteristics, such as asymmetry and conditional heteroskedasticity (e.g. Paolino, 2001).

Early research on modeling bounded-valued data focused on Beta regression models with independent data (Ferrari and Cribari-Neto, 2004), adopting the generalized linear model approach (McCullag and Nelder, 1989) aside from the use of the exponential family of distributions (Ferrari and Cribari-Neto, 2004). The logic was to reparametrize the conditional Beta distribution in such a way that it depends on the conditional mean and dispersion parameters and the conditional mean on covariates via an appropriate link function (e.g. logit link).

Later, generalized ARMA models were used for the analysis of bounded time series that still exploited the conditional Beta distribution. In particular, Rocha and Cribari-Neto (2009) proposed a (conditionally) Beta ARMA (βARMA) model that followed a conditional Beta distribution, where the logit link of its conditional mean had an ARMA representation. Guolo and Varin (2014) considered a conditionally Beta distributed process as an inverse Beta integral transformation of the standard Gaussian distribution evaluated at a variable that satisfied an ARMA scheme. Zheng et *al* (2015) set up an alternative Martingale difference formulation of generalized ARMA models that was applied to a Beta ARMA model.

Other extensions of the βARMA model have been put forward to account for some additional characteristics of bounded data, such as long-memory (Fractionally integrated βARMA, Pumi et *al*, 2019) and stochastic seasonality (βSARMA, Bayer et *al*, 2018). More recently, Gorgi and Koopman (2021) proposed a Beta autoregressive conditional mean model (Aknouche and Francq, 2021*b*) which can be seen as a Beta generalized ARMA model with the identity link function. A state-space (parameter-driven) Beta model was proposed by Da Silva and Migon (2016).

It is worth noting that only a few studies have considered the equation-based approach for modeling $(0, 1)$-valued time series data. For instance, an earlier Beta time series model has been proposed by McKenzie (1985) who introduced a marginally distributed Beta process as a solution to an appropriate autoregressive equation with Beta-distributed random coefficients.

However, all the aforementioned papers on $(0, 1)$-valued models are based on the Beta distribution, which is certainly very flexible and versatile, and can take extremely different shapes. However, there are other equally (and perhaps more) flexible distributions that can fit bounded-valued data (Bonat et *al*, 2019), such as the Simplex distribution (cf. Barndorff-Nielsen and Jorgensen, 1991; Kieschnick and McCullough, 2003; Zhang et *al*, 2016), the unit gamma distribution (Grassia, 1977) and the Kumaraswamy distribution (Kumaraswamy, 1980; Mitnik and Bayek, 2013). In fact, the conditional distribution generating the data is rarely known and the choice of a bad distribution can lead to an efficiency loss and even to inconsistency of the maximum likelihood estimator in the case of misspecified models (Aknouche and Francq, 2021*c*). It is, thus, less risky and more prudent to consider models with misspecified distributions (Aknouche and Francq, 2021*c*), for which it is possible to employ exponential family quasi-maximum likelihood-based or weighted least squares estimators that remain consistent, even when the assumed conditional distribution is different from the true one.

Following the equation-based approach, the aim of this paper is to propose a multiplicative error model for $(0, 1)$-valued time series in the spirit of the ACD (autoregressive conditional duration) model (Engle and Russel, 1998) and more generally of the multiplicative error model (MEM, Engle, 2002). To be more specific, the observed process results from the division of a $(0, 1)$-valued iid innovation sequence by a dependent process, which, in turn, is proportional to the inverse conditional mean. The latter is specified as a linear function of its lagged values and lagged values of the inverse observed process with an intercept exceeding one, so that the observed process always lies in the $(0, 1)$ interval. We refer to the resulting specification as the "*autoregressive conditional proportion*" (ARCP) model.

We first study some dynamic properties of it, using the stochastic recurrence equation approach. Then, we propose an exponential quasi-maximum likelihood estimator (EQMLE), for which we establish the consistency, asymptotic normality, and asymptotic efficiency (in the class of all exponential family QMLEs) under mild conditions. The relative optimality of the EQMLE results from the fact that the conditional variance of the model is a quadratic function of the conditional mean. Finally, we illustrate the usefulness of our ARCP model with simulated data and real-life applications.

Compared to the existing bounded-valued time series models, our model has several advantages: It is semiparametric and encompasses many (0,1)-valued conditional distributions, it can deal with higher orders without complicated parametric constraints, it allows for exogenous variables, it has a simple probability structure, its EQMLE has a certain asymptotic optimality, and finally it can be easily extended to multivariate settings.

The structure of the paper is organized as follows. In Section 2 we review two basic $(0, 1)$-valued distributions, we set up our proposed specification, the ARCP model, and also show its connection with other well-known models, such as the ACD and weak ARMA models. In Section 3 we study the existence of a stationary and ergodic solution of the ARCP equation and derive its finite negative moment conditions. In Section 4, we study the asymptotic properties of the exponential QMLE for the conditional mean parameters and we propose a weighted least squares estimator for the innovation variance as well. In Section 5 we present a simulation study in order to assess the performance of the EQMLE, using simulated ARCP series, while in Section 6 we conduct our empirical analysis, using two real data sets. Finally, Section 7 discusses potential extensions of the univariate ARCP model that include, among others, random coefficients, nonlinear conditional means and multivariate settings. A supplementary material containing all the relevant proofs and a third application accompanies this paper.

2 Autoregressive conditional proportion model

2.1 Beta and Simplex distributions

Let $(0,1)$ and $\mathbb{Z} = \{..., -1, 0, 1, ...\}$ be, respectively, the open Simplex interval and the set of integer numbers. In the sequel, all random variables and stochastic processes of interest are defined on a probability space $(\Omega, \mathcal{F}, P)$ and valued in $(0,1)$. A $(0,1)$-valued random variable X is said to have a (standard) Beta distribution ($X \sim \mathcal{B}eta\,(a,b)$) with parameters $a > 0$ and $b > 0$ if its probability density function $f(x; a, b)$ is given by

$$f(x; a, b) = \tfrac{1}{Beta(a,b)} x^{a-1} (1-x)^{b-1} 1_{(0,1)}(x),$$

where 1_B denotes the indicator function of the set B and $Beta(a,b)$ is the standard Beta function. The mean and variance of X are respectively $E(X) = \frac{a}{a+b}$ and $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$. When $a > 1$, the mean of the reciprocal variable is $E\left(\frac{1}{X}\right) = \frac{a+b-1}{a-1}$. If $X \sim \mathcal{B}eta\,(a,b)$ and $c > 0$ then $Z = cX \sim \mathcal{B}eta\,(a,b,c,0)$ is (generalized) Beta distributed with density function

$$f(z; a, b, c) = \tfrac{1}{c^{a+b-1} Beta(a,b)} z^{a-1} (c-z)^{b-1} 1_{(0,c)}(z),$$

so Z is $(0,c)$-valued. Another parametrization of the standard Beta distribution, especially useful for Beta regression-based models (cf. Ferrari and Cribari-Neto, 2004) consists of the $\mathcal{B}eta\,(\phi\mu, \phi(1-\mu))$ distribution, where $\phi > 0$ is called the dispersion parameter while $\mu \in (0,1)$ is the mean parameter. The density $f(x; \phi, \mu)$ thus becomes

$$f(x; \phi, \mu) = \tfrac{1}{Beta(\phi\mu, \phi(1-\mu))} x^{\phi\mu-1} (1-x)^{\phi(1-\mu)-1} 1_{(0,1)}(x).$$

Under the latter representation, the mean and variance of X are given by $EX = \mu$, and $Var(X) = \frac{\mu(1-\mu)}{1+\phi}$. Moreover $EX^{-1} = \frac{\phi-1}{\phi\mu-1}$ provided that $\phi > \mu^{-1}$.

A random variable X is said to have a Simplex distribution (cf. Barndorff-Nielsen and Jorgensen, 1991; Jorgensen, 1997) with mean $\mu \in (0,1)$ and dispersion $\phi > 0$ ($X \sim S^{-}(\mu, \phi)$ in short) if

$$f(x; \phi, \mu) = \tfrac{1}{\sqrt{2\pi\phi x(1-x)^3}} \exp\left(-\tfrac{1}{2\phi} \tfrac{(x-\mu)^2}{x(1-x)\mu^2(1-\mu)^2}\right) 1_{(0,1)}(x).$$

The mean and variance of the Simplex variable are respectively given by $E\left(X\right)=\mu$ and $Var\left(X\right)=\mu\left(1-\mu\right)-\frac{1}{\phi\sqrt{2}}\exp\left(\frac{1}{2\phi^2\mu^2(1-\mu)^2}\right)\Gamma\left(\frac{1}{2},\frac{1}{2\phi^2\mu^2(1-\mu)^2}\right)$, where $\Gamma\left(a,b\right)=\int_b^\infty x^{a-1}e^{-x}dx$ $(a>0, b>0)$ denotes the incomplete Gamma function (cf. Barndorff-Nielsen and Jorgensen, 1991; Kieschnick and McCullough, 2003; Zhang et *al*, 2016; Bonat et *al*, 2019). As is the case with the Beta distribution, the Simplex distribution can also take various shapes, such as the J shape, the U shape, and the inverted J shape. However, unlike the Beta distribution, the Simplex distribution (e.g. $S^-(0.5,0.5)$) is able to accommodate bimodal forms (cf. Espinheira and Silva, 2020).

2.2 MEM-based autoregressive conditional proportion

Let $\{\xi_t, t\in\mathbb{Z}\}$ be an independent and identically distributed (iid) sequence of $(0,1)$-valued random variables, such that $E\xi_t=\mu_0\in(0,1)$ is an arbitrarily fixed known constant, $0<E(\xi_t^{-1})<\infty$, and $Var\left(\xi_t\right)=\sigma_0^2>0$ is unknown. Denote by $\{\mathbf{X}_t, t\in\mathbb{Z}\}$ a sequence of stationary and ergodic (almost surely, $a.s.$) positive covariates such that $\mathbf{X}_t=(X_{1t},...,X_{rt})'$ with $\{\mathbf{X}_t, t\in\mathbb{Z}\}$ and $\{\xi_t, t\in\mathbb{Z}\}$ being independent. A $(0,1)$-valued stochastic process $\{Y_t, t\in\mathbb{Z}\}$ is said to be a (multiplicative-error) *autoregressive conditional proportion* (ARCP) if it is given for all $t\in\mathbb{Z}$ by

$$Y_t=\frac{\xi_t}{\lambda_t} \quad\text{and}\quad \lambda_t=\omega_0+\sum_{i=1}^{q}\alpha_{0i}\frac{1}{Y_{t-i}}+\sum_{j=1}^{p}\beta_{0j}\lambda_{t-j}+\boldsymbol{\pi}_0'\mathbf{X}_{t-1}, \tag{2.1}$$

where, to ensure $\lambda_t>1$ a.s. and therefore $Y_t\in(0,1)$ a.s. for all $t\in\mathbb{Z}$, it is assumed that $\omega_0>1$, $\alpha_{0i}\geq 0$, $\beta_{0j}\geq 0$ $(i=1,...,p,\ j=1,...,q)$ and $\boldsymbol{\pi}_0=(\pi_{01},...,\pi_{0r})'\geq 0$. When $\boldsymbol{\pi}_0=0$, model (2.1) is without covariates. Denote by $\mathcal{F}_t$ the information set available at time t, i.e. the sigma-field generated by $\{Y_u,\mathbf{X}_u, u\leq t\}$. The conditional mean and conditional variance of the process $\{Y_t, t\in\mathbb{Z}\}$ are given, respectively, by

$$E\left(Y_t|\mathcal{F}_{t-1}\right)=\frac{\mu_0}{\lambda_t} \quad\text{and}\quad Var\left(Y_t|\mathcal{F}_{t-1}\right)=\frac{\sigma_0^2}{\lambda_t^2}. \tag{2.2}$$

As expected, the conditional variance is a quadratic function of the conditional mean. In fact, the specification (2.1) is a multiplicative error model (MEM) in the spirit of Engle (2002) (see also Engle and Gallo, 2006; Hautsch, 2012; Cipolini and Gallo, 2021) but the underlying process is $(0,1)$-valued and the inverted mean depends on covariates. The distribution of $\{\xi_t, t \in \mathbb{Z}\}$ is unspecified, whose mean and variance are the parameters of the semiparametric quadratic variance-to-mean relationship given in (2.2). The proposed model is, thus, semi-parametric. A useful family of conditional distributions for ξ_t is the $\mathcal{B}eta(\phi_0\mu_0, \phi_0(1-\mu_0))$ distribution with $E\xi_t = \mu_0$, $E\xi_t^{-1} = \frac{\phi_0-1}{\mu_0\phi_0-1}$ $(\phi_0\mu_0 > 1)$ and $Var(\xi_t) = \frac{\mu_0(1-\mu_0)}{1+\phi_0}$. In this case, the conditional distribution of $Y_t|\mathcal{F}_{t-1}$ is generalized $\mathcal{B}eta\left(\phi_0\mu_0, \phi_0(1-\mu_0), \lambda_t^{-1}, 0\right)$ distributed with a probability density function

$$f_{Y_t|\mathcal{F}_{t-1}}(y_t) = \frac{y_t^{\phi_0\mu_0-1}\left(\lambda_t^{-1}-y_t\right)^{\phi_0(1-\mu_0)-1}\lambda_t^{\phi_0-1}}{Beta(\phi_0\mu_0,\phi_0(1-\mu_0))}, \quad 0 \le y_t \le \lambda_t^{-1}. \tag{2.3}$$

Note that (2.3) should not be confused with the (observation driven) conditional Beta model proposed by Gorgi and Koopman (2021), for which the conditional distribution is Beta distributed, i.e.

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{B}eta\left(c\chi_t, c(1-\chi_t)\right),$$

where the conditional mean χ_t is specified as a function of past observations under the "stationary" constraint $\chi_t \in (0,1)$ for all $t \in \mathbb{Z}$. This constraint can severely restrict the model parameters and therefore prevents the model from capturing some interesting characteristics of bounded data, such as persistence.

Another law for ξ_t we use in this paper is the Simplex distribution $S^-(\mu_0, \phi_0)$ with mean μ_0 and dispersion parameter ϕ_0. Other potential distributions of interest for ξ_t would be the Kumaraswamy distribution and the unit-versions of nonnegative distributions, such as the unit-gamma, the unit-Weibull, the unit-Lindley, etc., (see Grassia, 1977 for the precursor unit-gamma distribution).

In lieu of (2.1), we can think of the following $(0,1)$-valued ACD (autoregressive conditional duration) model, proposed by Engle and Russell (1998)

$$Y_t = \varphi_t\xi_t \text{ and } \varphi_t = \omega_0 + \sum_{i=1}^{q}\alpha_{0i}Y_{t-i} + \sum_{j=1}^{p}\beta_{0j}\varphi_{t-j} + \boldsymbol{\pi}_0'\mathbf{X}_{t-1} \tag{2.4}$$

with a conditional mean $\mu_0\varphi_t$. However, even without covariates, for φ_t (and hence Y_t) to be in the interval $(0,1)$, model (2.4) would require the following constraint (see Ristić et *al*

(2016) for the case of a Binomial INGARCH model and Gorgi and Koopman (2021) for the case of a Beta-observation driven model)

$$\omega_0 + \sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1, \tag{2.5}$$

which involves the intercept $\omega_0 > 0$. Such a constraint makes the model incapable of allowing for high persistence, since under (2.5), parameter values close to the boundary of the stationary region given by

$$\mu_0 \sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j} < 1,$$

are not allowed, especially when μ_0 is close to 1 and ω_0 is quite large. This is the reason for which we rather prefer model (2.1), as it is not restrictively constrained and can accommodate high persistence as well as other interesting characteristics.

The ARCP model (2.1) can be seen as an inverted ACD model, but with $\omega_0 > 1$. Indeed, the process $\{W_t, t \in \mathbb{Z}\}$ given by $W_t = \frac{1}{Y_t}$ $(t \in \mathbb{Z})$ clearly satisfies the ACD equation

$$W_t = \psi_t z_t, \qquad \psi_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} W_{t-i} + \sum_{j=1}^{p} \beta_{0j} \psi_{t-j} + \boldsymbol{\pi}_0' \mathbf{X}_{t-1} \tag{2.6}$$

with $z_t = \xi_t^{-1}$ and $\psi_t = \lambda_t$ (e.g. Engle and Gallo, 2006; Aknouche and Francq, 2021*a*; Aknouche et *al*, 2021*b*). Conversely, if $\{W_t, t \in \mathbb{Z}\}$ is an ACD model of the form (2.6) with $\omega_0 > 1$, the process $\{Y_t, t \in \mathbb{Z}\}$ given by $Y_t = W_t^{-1}$ clearly satisfies (2.1) with $\xi_t = z_t^{-1}$ and $\lambda_t = \psi_t$.

Note finally that any inverted ARCP process is a weak ARMA (autoregressive moving average) model. Letting $\varepsilon_t = Y_t^{-1} - E\left(Y_t^{-1} | \mathcal{F}_{t-1}\right) = \lambda_t \left(\xi_t^{-1} - E\xi_t^{-1}\right)$ be a zero-mean term of a martingale difference sequence, the process $\left\{Y_t^{-1}, t \in \mathbb{Z}\right\}$ can be written in the following weak ARMA form

$$Y_t^{-1} = \left(\omega_0 + \boldsymbol{\pi}_0' \mathbf{X}_{t-1}\right) E\xi_t^{-1} + \sum_{i=1}^{\max(p,q)} \left(\alpha_{0i} E\xi_t^{-1} + \beta_{0i}\right) Y_{t-i}^{-1} + \varepsilon_t - \sum_{j=1}^{p} \beta_{0j} \varepsilon_{t-j}$$

with variance $Var\left(\varepsilon_t\right) = E\left(\lambda_t^2\right) Var\left(\xi_t^{-1}\right)$ that is finite, provided that $E\left(\xi_t^{-2}\right) < \infty$ and $E\left(\lambda_t^2\right) < \infty$.

3 Ergodicity and negative moment conditions

Since model (2.1) is defined through a stochastic equation driven by an iid innovation and stationary and ergodic covariates, its ergodic and (negative) moment properties can be obtained using routinely the stochastic recurrence equation (SRE) approach (Bougerol, 1993). As in Francq and Thieu (2019), assume first that:

A0 $\{(\mathbf{X}_t, t \in \mathbb{Z}\}$ *is strictly stationary and there exists* $\kappa \in (0,1)$ *such that* $E\|\mathbf{X}_t\|^\kappa < \infty$. Define $\mathbf{Y}_t = (Y_t^{-1}, \ldots, Y_{t-q+1}^{-1}, \lambda_t, \ldots, \lambda_{t-p+1})'$, $B_{0t} = \left((\omega_0 + \boldsymbol{\pi}_0'\mathbf{X}_{t-1})\,\xi_t^{-1}, 0_{(q-1)\times 1}, \omega_0, 0_{(p-1)\times 1}\right)'$ ($0_{m\times n}$ being the null matrix of dimension $m \times n$), and

$$A_{0t} = \begin{pmatrix} \alpha_{01}\xi_t^{-1} & \cdots & \alpha_{0,q-1}\xi_t^{-1} & \alpha_{0q}\xi_t^{-1} & \beta_{01}\xi_t^{-1} & \cdots & \beta_{0,p-1}\xi_t^{-1} & \beta_{0p}\xi_t^{-1} \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{01} & \cdots & \alpha_{0,q-1} & \alpha_{0q} & \beta_{01} & \cdots & \beta_{0,p-1} & \beta_{0p} \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then model (2.1) can be cast in the following SRE

$$\mathbf{Y}_t = A_{0t}\mathbf{Y}_{t-1} + B_{0t} \tag{3.1}$$

that is driven by the stationary and ergodic non-negative input sequence $\{(A_{0t}, B_{0t}), t \in \mathbb{Z}\}$ with $\{A_{0t}, t \in \mathbb{Z}\}$ being iid. Note that the matrices A_{0t} and B_{0t} in (3.1) have the same zero structure as the input matrices of the SRE of a GARCH model (e.g. Francq and Zakoian, 2019; Francq and Thieu, 2019). Therefore, the ergodic and moment conditions for model (3.1) are obtained in a similar way as for the GARCH model. Let

$$\gamma(\mathbb{A}_0) = \inf\left\{\tfrac{1}{n}E\log\|A_{0n}\ldots A_{02}A_{01}\|,\ n \geq 1\right\}$$

be the largest Lyapunov exponent associated with the SRE (3.1) (cf. Bougerol and Picard, 1992). Let also

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0,p-1} & \beta_{0p} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{3.2}$$

and denote by $\rho(A)$ the maximum absolute eigenvalue of A. A strict stationarity and ergodicity condition for model (2.1) is as follows:

Theorem 3.1 i) *Assume* $E\left(\log\left(\xi_t^{-1}\right)\right) < \infty$. *A sufficient condition for model* (2.1) *to have a unique nonanticipative strictly stationary and ergodic solution given by*

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{0,t-i} B_{0,t-j}, \quad t \in \mathbb{Z}, \tag{3.3}$$

where the series converges absolutely almost surely, is that

$$\gamma\left(\mathbb{A}_0\right) < 0. \tag{3.4}$$

ii) When $\gamma\left(\mathbb{A}_0\right) \geq 0$, *there exists no stationary solution to* (3.1).

iii) If (2.1) *admits a strictly stationary solution then*

$$\rho\left(\boldsymbol{\beta}_0\right) < 1. \tag{3.5}$$

Note that the existence of a stationary and ergodic sequence $\left\{Y_t^{-1}, t \in \mathbb{Z}\right\}$ entails the stationarity and ergodicity of the process $\{Y_t, t \in \mathbb{Z}\}$. In the special case where $p = q = 1$, a more simple and explicit necessary and sufficient condition for strict stationarity of model (2.1) writes as follows (e.g. Francq and Zakoian, 2019)

$$E\left(\log\left(\alpha_{01}\xi_t^{-1} + \beta_{01}\right)\right) < 0.$$

Since the ARCP process is by construction $a.s.$ bounded, its higher-order positive moments EY_t^m ($m > 0$) are obviously finite. However, the negative moment EY_t^{-m}, which can be infinite, is of interest especially when studying the asymptotic properties of the exponential quasi-maximum likelihood estimate (EQMLE) that will be dealt with in the following Section.

A (negative) first-order stationarity condition for model (2.1) is given by the following result:

Theorem 3.2 *Assume $E\xi_t^{-1} < \infty$. A necessary and sufficient condition for the process given by* (2.1) *to be strictly stationary and ergodic with $EY_t^{-1} < \infty$ is that*

$$\rho\left(E\left(A_{0t}\right)\right) < 1. \tag{3.6}$$

In view of the form of A_t in (3.1) and using a similar device by Francq and Zakoian (2019), condition (3.6) takes the following explicit form

$$E\xi_t^{-1}\sum_{i=1}^{q}\alpha_{0i} + \sum_{j=1}^{p}\beta_{0j} < 1. \tag{3.7}$$

Conditions for the existence of negative moments of the ARCP model (2.1) are given as follows. Let m be a positive integer.

Theorem 3.3 i) *Under* (3.4) *there exists $\kappa > 0$ such that*

$$E\lambda_t^{\kappa} < \infty \quad \text{and} \quad EY_t^{-\kappa} < \infty. \tag{3.8}$$

ii) *Let $\{Y_t, t \in \mathbb{Z}\}$ be a strictly stationary solution of* (2.1) *and assume that $E\xi_t^{-m}$ is finite. A sufficient condition for EY_t^{-m} to be finite is that*

$$\rho\left(E\left(A_{0t}^{\otimes m}\right)\right) < 1, \tag{3.9}$$

where $A^{\otimes m}$ stands for the Kronecker product $A \otimes A \otimes \cdots \otimes A$ of m factors.

4 Exponential quasi-maximum likelihood estimation

4.1 Estimating the conditional mean parameter

Due to the misspecification of the distribution of ξ_t, the true conditional distribution of $Y_t|\mathcal{F}_{t-1}$ in (2.1) is unknown. A quasi-maximum likelihood estimate (QMLE) that ignores the specification of the conditional distribution is, thus, suitable for estimating the parameter θ_0. Since the variance-to-mean equation (2.2) is quadratic, the QMLE computed on the basis of the exponential distribution (Exponential QMLE, EQMLE in short) would have the smallest asymptotic variance among all QMLEs belonging to the exponential family (e.g. Wooldridge, 1999; Aknouche et *al*, 2018; Aknouche and Francq, 2021*a*).

Let $Y_1, Y_2, ..., Y_n$ be a series generated from the ARCP(p,q) model (2.1) with $\lambda_t = \lambda_t(\theta_0)$ where

$$\theta_0 = (\omega_0, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, ..., \beta_{0p}, \boldsymbol{\pi}_0')'$$

is the true (conditional mean) parameter which belongs to a parameter space $\Theta \subset (1,\infty) \times [0,\infty)^{(p+q+r)}$. The true innovation variance $\sigma_0^2 = Var(\xi_t)$ belongs to a parameter space $\Delta \subset (0,\infty)$. For given initial values $Y_0, ..., Y_{1-q}$, $\widetilde{\lambda}_0 \geq 0, ..., \widetilde{\lambda}_{1-p} \geq 0$, $\mathbf{X}_0 \geq 0$ and a generic parameter $\theta = (\omega, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p, \boldsymbol{\pi}')' \in \Theta$ define the sequence $(\widetilde{\lambda}_t(\theta))_{t\geq 0}$ as

$$\widetilde{\lambda}_t(\theta) = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \widetilde{\lambda}_{t-j}(\theta) + \boldsymbol{\pi}'\mathbf{X}_{t-1}, \quad t \geq 0. \tag{4.1}$$

The latter is an observable proxy for $(\lambda_t(\theta))_{t\in\mathbb{Z}}$, which is, under $\sum_{j=1}^{p} \beta_j < 1$, a stationary and ergodic solution of the following generic model ($\theta \in \Theta$)

$$\lambda_t(\theta) = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j}(\theta) + \boldsymbol{\pi}'\mathbf{X}_{t-1}, \quad t \in \mathbb{Z}. \tag{4.2}$$

Based on the exponential distribution with mean $\frac{\mu_0}{\lambda_t(\theta)}$, the logarithmed exponential quasi-likelihood function of any $\theta \in \Theta$ is given by

$$\widetilde{L}_n(\theta) = \tfrac{1}{n} \sum_{t=1}^{n} \widetilde{l}_t(\theta) \quad \text{with} \quad \widetilde{l}_t(\theta) = \tfrac{\widetilde{\lambda}_t(\theta)}{\mu_0} Y_t - \log \widetilde{\lambda}_t(\theta), \tag{4.3}$$

where $\mu_0 = E\xi_t$ is assumed known. The exponential QMLE (EQMLE) $\widehat{\theta}_n$ of θ_0 is the minimizer of $\widetilde{L}_n(\theta)$ over Θ,

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \widetilde{L}_n(\theta). \tag{4.4}$$

Let $\gamma(\mathbb{A}_0)$ be the top Lyapunov exponent associated with $(A_{0t})_t$ in (3.1). Let also $\boldsymbol{\beta}$ be defined as $\boldsymbol{\beta}_0$ in (3.2) with β_j in place of β_{0j} $(j = 1, ..., p)$. To establish the strong consistency of $\widehat{\theta}_n$ we need **A0** and the following assumptions.

A1 $\gamma(\mathbb{A}_0) < 0$ *and* $\forall \theta \in \Theta$, $\rho(\boldsymbol{\beta}) < 1$.

A2 $\theta_0 \in \Theta$ *and* Θ *is compact.*

A3 *The polynomials* $\alpha_{\theta_0}(z) = \sum_{i=1}^{q} \alpha_{0i} z^i$ *and* $\beta_{\theta_0}(z) = 1 - \sum_{j=1}^{p} \beta_{0j} z^j$ *have no common root,* $\alpha_{\theta_0}(1) \neq 0$, *and* $\alpha_{0q} + \beta_{0p} \neq 0$.

A4 ξ_t^{-1} *is non-degenerate, and for all* $i \geq 1$ *the conditional distribution of* ξ_{t-i}^{-1} *given* $\mathcal{F}_{t-1}^{X} = \sigma\{\mathbf{X}_{t-k},\ k \geq 1\}$ *is non-degenerate.*

A5 The variable $\mathbf{c}'\mathbf{X}_t$ is not degenerated for all non-zero vector $\mathbf{c}$ of $\mathbb{R}^r$.

Assumptions **A1**-**A3** are standard. Assumption **A1** ensures the stationarity and ergodicity of the ARCP model (2.1). The condition $\rho(\boldsymbol{\beta}) < 1$ guarantees the invertibility of equation (4.2) for any $\theta \in \Theta$. The assumptions **A3**-**A5** are made to ensure the identifiability of the model. As in Francq and Thieu (2019) and Han and Kristensen (2014), the second part of **A4** means that there is no colinearities between the exogenous covariates $\mathbf{X}_{t-1}$ and the functions of the past inverted process $\left(Y_{t-i}^{-1}\right)_{i\geq 1}$. When there is no covariate, **A5** vanishes and **A4** reduces to: ξ_t^{-1} *is non-degenerate.* As the function $x \mapsto x^{-1}$ is one-to-one on $(0, \infty)$, the non-degeneracy of ξ_t^{-1} in **A4** is equivalent to the non-degeneracy of ξ_t. Let $\underset{n\to\infty}{\overset{a.s.}{\to}}$ and $\underset{n\to\infty}{\overset{\mathcal{D}}{\to}}$ respectively denote the almost sure (a.s.) convergence and the convergence in distribution as $n \to \infty$.

Theorem 4.1 *Let* $(\widehat{\theta}_n)_n$ *be a sequence of EQMLEs defined by* (4.4). *Under* ***A0-A5***,

$$\widehat{\theta}_n \underset{n\to\infty}{\overset{a.s.}{\to}} \theta_0.$$

To study the asymptotic normality of $\widehat{\theta}_n$ let us consider the following supplementary assumption.

A6 θ_0 *belongs to the interior of* Θ.

Theorem 4.2 *Under* ***A0-A6*** *we have*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0, \frac{\sigma_0^2}{\mu_0^2} J(\theta_0)^{-1}\right), \tag{4.5}$$

where

$$J(\theta_0) = E\left(\frac{1}{\lambda_t^2} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'}\right). \tag{4.6}$$

Remark 4.1 As for the Gaussian QMLE of GARCH models, the EQMLE does not require any moment condition on the observed process $\{Y_t, t \in \mathbb{Z}\}$, which is a quite desirable property.

4.2 Estimating the innovation variance

In view of (2.2) define

$$u_t = \left(Y_t - \tfrac{\mu_0}{\lambda_t}\right)^2 - Var\left(Y_t|\mathcal{F}_{t-1}\right) = \lambda_t^{-2}\left((\xi_t - \mu_0)^2 - \sigma_0^2\right)$$

to get the following regression

$$\left(Y_t - \tfrac{\mu_0}{\lambda_t}\right)^2 \lambda_t^2 = \sigma_0^2 + v_t \tag{4.7}$$

where $v_t = u_t\lambda_t^2 = (\xi_t - \mu_0)^2 - \sigma_0^2$ is a member of a zero-mean *iid* sequence with finite variance $E\left((\xi_t - \mu_0)^2 - \sigma_0^2\right)^2 = Var\left((\xi_t - \mu_0)^2\right)$. As $\lambda_t = \lambda_t(\theta_0)$ depends on the unknown parameter θ_0, the regressand in (4.7) is unobservable. Replacing $\lambda_t(\theta_0)$ by $\widehat{\lambda}_t = \lambda_t\left(\widehat{\theta}_n\right)$, where $\widehat{\theta}_n$ is given by (4.4), we get the following estimable regression

$$\left(Y_t - \tfrac{\mu_0}{\widehat{\lambda}_t}\right)^2 \widehat{\lambda}_t^2 = \sigma_0^2 + \widehat{v}_t. \tag{4.8}$$

From (4.8) a feasible weighted least squares estimate of σ_0^2 is

$$\widehat{\sigma}_n^2 = \tfrac{1}{n}\sum_{t=1}^{n}\left(Y_t\widehat{\lambda}_t - \mu_0\right)^2. \tag{4.9}$$

When $\xi_t \sim \mathcal{B}eta(\phi_0\mu_0, \phi_0(1-\mu_0))$, an estimate $\widehat{\phi}_n$ of ϕ_0 can be deduced from $\widehat{\sigma}_n^2$:

$$\widehat{\phi}_n = \tfrac{\mu_0(1-\mu_0)}{\widehat{\sigma}_n^2} - 1.$$

Next, it is shown that $\widehat{\sigma}_n^2$ is consistent and asymptotically Gaussian.

Theorem 4.3 *Under **A0-A6***

$$\widehat{\sigma}_n^2 \underset{n\to\infty}{\overset{a.s.}{\to}} \sigma_0^2 \tag{4.10}$$

and

$$\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}(0, \Lambda), \tag{4.11}$$

where $\Lambda = Var((\xi_t - \mu_0)^2) \in (0, \infty)$.

Note that the existence of $Var\left(\xi_t^2\right)$ is obvious as ξ_t is bounded. A consistent estimate of the limiting variance Λ in (4.11) is given by

$$\widehat{\Lambda}_n = \frac{1}{n}\sum_{t=1}^{n}\left(\left(\widehat{\xi}_t - \mu_0\right)^2 - \widehat{\sigma}_n^2\right)^2, \tag{4.12}$$

where $\widehat{\xi}_t = Y_t\widehat{\lambda}_t$ is the normalized residual of model (2.1). From (4.10) and (4.11), the asymptotic matrix in (4.5) can be estimated. Replacing the true parameter θ_0 by its estimate $\widehat{\theta}_n$, a consistent estimate of $J(\theta_0)$ takes the form

$$\widehat{J}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\lambda}_t^2\left(\widehat{\theta}_n\right)}\frac{\partial\widetilde{\lambda}_t\left(\widehat{\theta}_n\right)}{\partial\theta}\frac{\partial\widetilde{\lambda}_t\left(\widehat{\theta}_n\right)}{\partial\theta'}. \tag{4.13}$$

The asymptotic variance in (4.5) is then estimated by

$$\frac{\widehat{\sigma}_n^2}{\mu_0^2}\widehat{J}_n^{-1}. \tag{4.14}$$

5 Simulation study

In this Section, we examine the finite-sample behavior of the EQMLE with a simulation study. Four instances of model (2.1) are considered. The first one is a mean-inverse stationary ARCP$(1,1)$ model, driven by a Beta distributed innovation $\mathcal{B}eta\left(\phi_0\mu_0, \phi_0\left(1-\mu_0\right)\right)$ with $\mu_0 = 0.9$, $\phi_0 = 1.2$, $\sigma_0^2 \simeq 0.0409$, $E\xi_t^{-1} = 2.5$ and parameter $\theta_0 = (1.3, 0.2, 0.1)'$ such that $E\xi_t^{-1}\alpha_{01} + \beta_{01} = 0.6 < 1$ (cf. Table 5.1). The second case is a strictly stationary but mean-inverse nonstationary ARCP$(1,1)$ model, driven by a Beta distributed innovation with $\mu_0 = 0.9$, $\phi_0 = 1.2$, $\sigma_0^2 \simeq 0.0409$, $E\xi_t^{-1} = 2.5$ and $\theta_0 = (1.1, 0.4, 0.3)'$ so that $E\xi_t^{-1}\alpha_{01} + \beta_{01} = 1.3 > 1$ (cf. Table 5.2). In the third case, the model is an inverse-mean stationary ARCP$(2,1)$ with Beta distributed innovation (cf. Table 5.3) with $\mu_0 = 0.85$, $\phi_0 = 1.5$, $\sigma_0^2 = 0.0510$, $E\xi_t^{-1} \simeq 1.8182$, and $\theta_0 = (1.1, 0.2, 0.1, 0.1)'$ such that $E\xi_t^{-1}\left(\alpha_{01} + \alpha_{02}\right) + \beta_{01} = 0.6454 < 1$. Finally, the fourth case concerns an inverse mean-stationary ARCP$(2,1)$ driven by a Simplex distributed $\left(S^-\left(\mu_0, \phi_0\right)\right)$ innovation with $\mu_0 = 0.5$, $\phi_0 = 2$, $\sigma_0^2 \simeq 0.0393$, $E\xi_t^{-1} \simeq 2.5177$ and $\theta_0 = (1.1, 0.2, 0.1, 0.1)'$ so that $E\xi_t^{-1}\left(\alpha_{01} + \alpha_{02}\right) + \beta_{01} \simeq 0.855\,31$ (cf.

Table 5.4). The analytical expression of $E\xi_t^{-1}$ is not available for the Simplex distribution, but we approximated it by its sample counterpart.

In all instances, the ARCP model is strictly stationary from which we generate $N = 1000$ replications for three sample sizes n (500, 1000, 2000). The EQMLE is computed for each model from all the 1000 replications. Tables 5.1-5.4 report the sample means and sample standard deviations for the parameter estimates. They also display the means of the asymptotic standard errors (ASE) of these estimates, computed from (4.12), (4.13) and (4.14). For the Beta innovation case, the ASE of ϕ_0 is not given.

$\mu_0 = 0.9$	θ_0	ω_0	α_0	β_0	σ_0^2	ϕ_0
n	θ_0	1.3	0.2	0.1	0.0409	1.2
500	Mean	**1.2999**	**0.1987**	**0.1016**	**0.0407**	**1.2558**
	StD	0.0771	0.0170	0.0441	0.0059	0.3390
	ASE	0.0951	0.0167	0.0627	0.0394	-
1000	Mean	**1.3007**	**0.2003**	**0.0991**	**0.0408**	**1.2269**
	StD	0.0500	0.0118	0.0294	0.0040	0.2246
	ASE	0.0657	0.0117	0.0430	0.01927	-
2000	Mean	**1.3001**	**0.1997**	**0.1003**	**0.0409**	**1.2102**
	StD	0.0346	0.0082	0.0202	0.0034	0.1579
	ASE	0.0466	0.0083	0.0305	0.0099	-

Table 5.1. EQMLE results obtained from an inverse-mean stationary ARCP(1, 1) model with $\varepsilon_t \sim \mathcal{B}eta\left(\phi_0\mu_0, \phi_0\left(1-\mu_0\right)\right)$.

$\mu_0 = 0.9$	θ_0	ω_0	α_0	β_0	σ_0^2	ϕ_0
n	θ_0	1.1	0.4	0.3	0.0409	1.2
500	Mean	**1.1238**	**0.4006**	**0.2951**	**0.0410**	**1.2421**
	StD	0.0616	0.02441	0.0338	0.0217	0.3322
	ASE	0.0256	0.0242	0.0352	0.0469	-
1000	Mean	**1.1126**	**0.4001**	**0.2978**	**0.0409**	**1.2206**
	StD	0.0613	0.0167	0.0229	0.0042	0.2337
	ASE	0.0175	0.0170	0.0247	0.0225	-
2000	Mean	**1.1038**	**0.3997**	**0.2996**	**0.0409**	**1.2071**
	StD	0.0432	0.0117	0.0166	0.0028	0.1496
	ASE	0.0103	0.0123	0.0168	0.0115	-

Table 5.2. EQMLE results obtained from a non-inverse mean stationary ARCP$(1,1)$ model with $\varepsilon_t \sim \mathcal{B}eta\left(\phi_0\mu_0, \phi_0\left(1-\mu_0\right)\right).$

$\mu_0 = 0.85$	θ_0	ω_0	α_{01}	α_{02}	β_0	σ_0^2	ϕ_0
n	θ_0	1.1	0.2	0.1	0.1	0.0510	1.5
500	Mean	**1.1148**	**0.1989**	**0.1013**	**0.0931**	**0.0508**	**1.5437**
	StD	0.0923	0.0181	0.0240	0.0639	0.0059	0.3074
	ASE	0.1523	0.0176	0.0312	0.0902	0.0601	-
1000	Mean	**1.1072**	**0.2003**	**0.1009**	**0.0956**	**0.0510**	**1.5143**
	StD	0.0721	0.0123	0.0184	0.0511	0.0041	0.2063
	ASE	0.1120	0.0124	0.0222	0.0558	0.0306	-
2000	Mean	**1.1008**	**0.1999**	**0.0994**	**0.1005**	**0.0509**	**1.5092**
	StD	0.0557	0.0088	0.0143	0.0413	0.0028	0.1365
	ASE	0.0620	0.0087	0.0157	0.0405	0.0151	-

Table 5.3. EQMLE results obtained from the ARCP$(2,1)$ model with $\varepsilon_t \sim \mathcal{B}eta\left(\phi_0\mu_0, \phi_0\left(1-\mu_0\right)\right).$

$\mu_0 = 0.5$	θ_0	ω_0	α_{01}	α_{02}	β_0	σ_0^2
n	θ_0	1.1	0.2	0.1	0.1	0.0393
500	Mean	**1.1572**	**0.1982**	**0.0986**	**0.0978**	**0.0391**
	StD	0.1673	0.0177	0.0294	0.0815	0.0023
	ASE	0.1049	0.0154	0.0338	0.0514	0.0147
1000	Mean	**1.1303**	**0.1987**	**0.1007**	**0.0947**	**0.0392**
	StD	0.1314	0.0120	0.0223	0.0448	0.0016
	ASE	0.0852	0.0121	0.0193	0.0303	0.0098
2000	Mean	**1.1194**	**0.1998**	**0.1011**	**0.0960**	**0.0393**
	StD	0.0972	0.0088	0.0173	0.0011	0.0061
	ASE	0.0750	0.0087	0.0119	0.0117	0.0089

Table 5.4. EQMLE results obtained from the ARCP$(2, 1)$ model with $\varepsilon_t \sim S^-(\mu_0, \phi_0)$, $\mu_0 = 0.5$ and $\phi_0 = 2$.

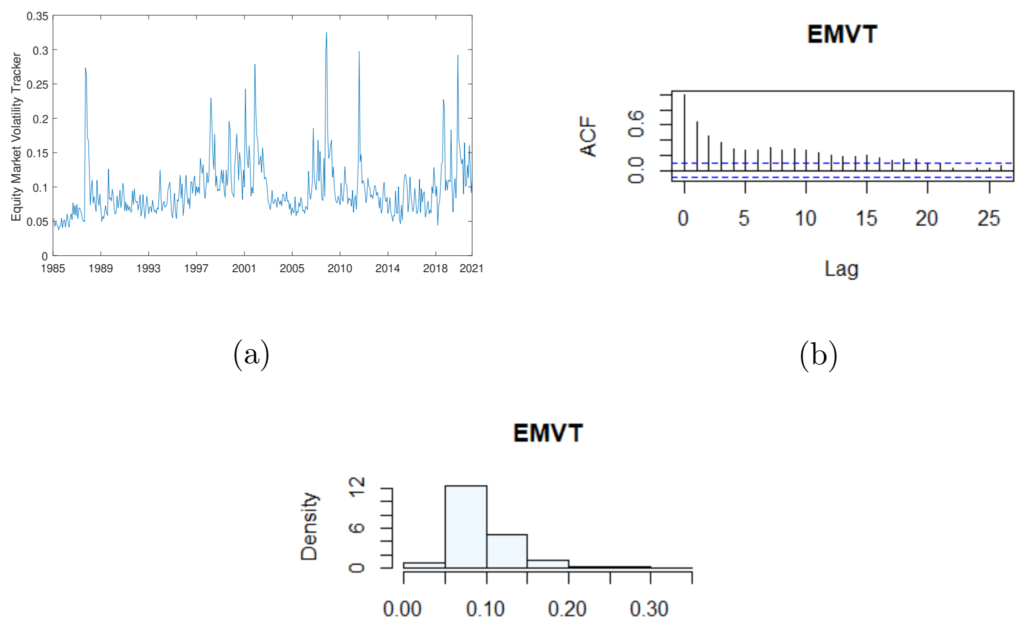
Some general conclusions can be drawn from the Tables 5.1-5.4: i) The EQMLE and WLSE estimate well the conditional mean and the innovation variance parameters, respectively, as indicated by the small bias and the small (empirical and asymptotic) standard errors. ii) The quality of estimates is satisfactory under the strict stationarity region described by (3.4), inside or outside the inverse-mean stationarity region given by (3.6). iii) The estimation results are consistent with the asymptotic theory as the larger the sample size, the better the estimation accuracy (as measured by the bias and the standard errors). Moreover, the ASEs and StDs are quite close to each other. iv) As expected, the EQMLE gives equally good results, irrespective of the distribution (Beta or Simplex) of the innovation and the orders of the model.

The estimation methods were implemented on a desktop with Intel Core i7 using R and, in particular, the nlminb() optimization function. The Beta, generalized Beta, and Simplex distributions were handled under the packages VGAM, Rmutil and Actuar in R.

6 Empirical applications

6.1 Equity Market Volatility Tracker data (EMVT)

The first application concerns fitting the ARCP model to the monthly Equity Market Volatility Tracker (EMVT) series, which is an index that moves with the realized volatility of returns on the S&P 500 (Baker et *al*, 2019). The dataset, divided by 100, was obtained from the FRED website (https://fred.stlouisfed.org) and spans from January, 1, 1985 to June 1, 2021, with a sample size $n = 438$ (see Figure 6.1a). The empirical distribution of the data is right-skewed (see Figure 6.1 (c)).



(a)

(b)

(c)

Figure 6.1. (a) The EMVT series; (b) sample autocorrelation; (c) histogram.

In fitting the ARCP(p, q) model to the EMVT series, the first issue we encountered was to select an appropriate fixed value for μ_0, which was assumed to be known so far. For the standard MEM model (Engle and Russell, 1998) the mean of the innovation (or identifiability parameter) is fixed to unity. In our case, μ_0 cannot take the value 1 because ξ_t is $(0, 1)$-valued, which makes the choice of μ_0 important. To avoid this problem, we construct a grid of nine evenly spaced points in the interval [0.1, 0.9] for μ_0 and then estimate the ARCP$(1,1)$ model, using each grid value of μ_0. The best value is the one that minimizes the mean absolute forecast error ($MAFE$) criterion given by $MAFE = \sum_{t=1}^{n} \left| Y_t - \frac{\mu_0}{\widehat{\lambda}_t} \right|$, where $\widehat{\lambda}_t = \widetilde{\lambda}_t \left(\widehat{\theta}_n \right)$ and $\widetilde{\lambda}_t$ is given by (4.1) with obvious notation. Other criteria, such as the mean square error forecast and QLIKE could be used (e.g. Patton, 2011; Aknouche and Francq, 2021c). The choice of $p = q = 1$ is adopted for simplicity, while it also avoids order selection issues. Table 6.1 reports the $MAFEs$ for the EMVT series, which are calculated for each grid value of μ_0. As it can be seen, the minimum $MAFE$ is achieved at $\mu_0 = 0.5$

μ_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$100MAFE$	2.5991	2.1020	2.0572	2.0145	**2.0140**	2.0141	2.0144	2.0143	2.0240

Table 6.1 Mean absolute forecasts errors (MAFE) obtained from the ARCP $(1,1)$ for the grid values$of \mu$

This median choice for μ_0 has also the advantage of avoiding estimates of ξ_t outside the interval $(0,1)$. The second issue we encountered was to select the orders p and q of the ARCP model. To this end, we used the AIC and BIC under the generalized conditional beta distribution given by (2.3), i.e. when ξ_t is Beta distributed. For the case where the innovation ξ_t is Simplex distributed, the conditional distribution of the model is not available. The values of the log-likelihood, of the AIC and of the BIC, reported in Table 6.2, indicated the orders $p = 1$ and $q = 2$.

Orders (p,q)	$(1,1)$	$(2,1)$	$(1,2)$	$(2,2)$
Log-Likelihood	-249.6596	-249.6597	$\mathbf{-21.97175}$	-248.948
AIC	505.3192	507.3194	**51.94351**	507.896
BIC	516.3289	521.999	**66.62303**	526.2454

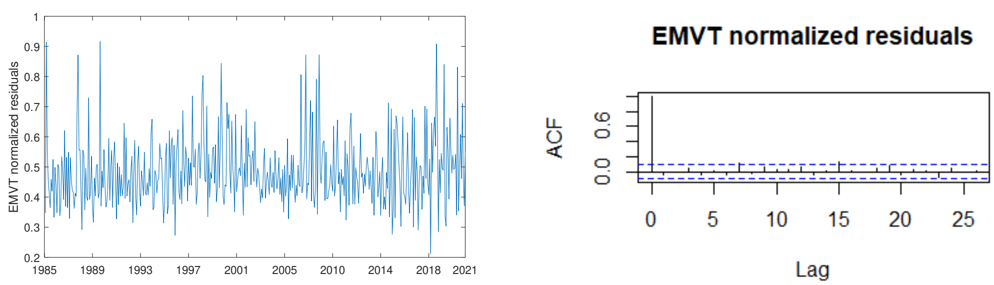
Table 6.2 Order-selection for the ARCP (p,q) under the conditional Generalized Beta distribution.

Note that for orders p and q exceeding 2, the EQML estimates are generally not significant. We, therefore, estimate an ARCP(1,2) model, using the EQML method for the conditional mean parameters and the WLS for the innovation variance. The initial parameter values are set to $\theta^{(0)} = \left(\omega^{(0)}, \alpha_1^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}\right)' = (0.1, 0.1, 0.1, 0.1)'$ and the starting values of the conditional mean equation are fixed to $\lambda_0 = \lambda_{-1} = \omega^{(0)}$ and $Y_0 = 1/\lambda_0$. Table 6.3 displays the parameter estimates of the selected ARCP$(1, 2)$ model, and their ASEs in parentheses. The mean-inverted persistence parameter $E\xi_t^{-1} \sum_{i=1}^{q} \alpha_{0i} + \sum_{j=1}^{p} \beta_{0j}$ is estimated by $Persi = \overline{\xi}_t^{-1} \widehat{\alpha}_1 + \widehat{\beta}_1 + \widehat{\beta}_2 = 0.8107$ (see Table 6.3), where $\overline{\xi}_t^{-1} = \frac{1}{n} \sum_{t=1}^{n} \widehat{\xi}_t^{-1} = 2.1550$ with $\widehat{\xi}_t = Y_t \widehat{\lambda}_t$ being the normalized residual. Thus, the estimated model is inverse-mean stationary with a quite strong persistence. The estimate $\widehat{\phi}_n = 7.8793$ is valid only in case, where the conditional distribution of the model is the generalized Beta distribution.

$\widehat{\omega}$	$\widehat{\alpha}_{1n}$	$\widehat{\beta}_{1n}$	$\widehat{\beta}_{2n}$	$\widehat{\sigma}_n^2$	$\widehat{\phi}_n$	$Persi$	MAFE	MSFE
1.0486 (0.0169)	0.3136 (0.0014)	0.0272 (0.0054)	0.1077 (0.0045)	0.0282 (0.0085)	7.8793	0.8107	0.0198	0.0010

Table 6.3. EQML and WLS estimates for the ARCP$(1, 2)$ model; EMVT series.

Table 6.3 also reports the values of the mean absolute forecast error (MAFE) and the Mean square forecast error (MSFE), where $MSFE = \sum_{t=1}^{n} \left(Y_t - \frac{\mu_0}{\widehat{\lambda}_t}\right)^2$. The plot of the ARCP residual series along with its sample autocorrelation function, both shown in Figure 6.2, imply that the iid assumption is compatible with the innovation sequence. Note that the sample mean $\frac{1}{n} \sum_{t=1}^{n} \widehat{\xi}_t = 0.4871$ is close to $\mu_0 = E(\xi_t) = 0.5$ to which it is a consistent estimate.



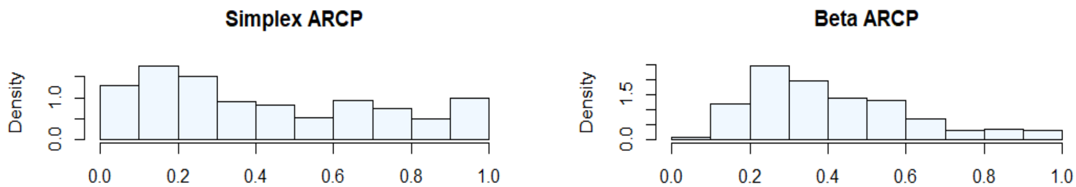
(a) (b)

Figure 6.2. (a) Normalized residuals; (b) sample autocorrelation.

To identify the distribution of the innovation of the ARCP model, Figure 6.3 plots the probability integral transform (PIT) of the normalized residuals $(\widehat{\xi}_t)$ with respect to two models; the Beta distribution $\mathcal{B}eta\left(\widehat{\phi}_n\mu_0, \widehat{\phi}_n(1-\mu_0)\right)$ (where $\widehat{\phi}_n = 7.8793$, see Table 6.3) and the Simplex distribution $S^-(\mu_0, 0.6)$, where 0.6 is an approximate rounded solution to the integral equation

$$\mu_0(1-\mu_0) - \frac{1}{\widehat{\phi}_n\sqrt{2}}\exp\left(\frac{1}{2\widehat{\phi}_n^2\mu^2(1-\mu)^2}\right)\Gamma\left(\frac{1}{2}, \frac{1}{2\widehat{\phi}_n^2\mu_0^2(1-\mu_0)^2}\right) = \widehat{\sigma}_n^2,$$

for $\widehat{\phi}_n$. It can be seen that the residuals have a better compatibility with the Simplex distribution rather than the Beta distribution, as the PIT for the former is near to a straight line.



(a) (b)

Figure 6.3. PIT of the ARCP residuals with respect to: (a) the simplex distribution, (b) the beta distribution. EMVT series.

As a benchmark to our ARCP model, we also fit to the EMVT series a linear conditional-mean version of the Beta observation driven (BOD) model, proposed by Gorgi and Koopman (2021). It is given by

$$\begin{aligned} Y_t|\mathcal{F}_{t-1} &\sim \mathcal{B}eta(c\chi_t, c(1-\chi_t)) \\ \chi_t &= d + aY_{t-1} + b\chi_{t-1} \end{aligned} \tag{5.1}$$

where $c > 0$, $d > 0$, $a \geq 0$, $b \geq 0$ and $d + a + b < 1$. The maximum likelihood estimate (MLE) of the parameter $g = (d, a, b, d)'$ is given by

$$\widehat{g}_n = (0.0228, 0.5466, 0.2194, 120.5648)',$$

with persistence parameter 0.766, which is smaller than the one obtained by the ARCP. Note that the $MAFE = \sum_{t=1}^{n} |Y_t - \widehat{\chi}_t| = 0.0199$ and $MSFE = \sum_{t=1}^{n} (Y_t - \widehat{\chi}_t)^2 = 0.0011$ (where

$\widehat{\chi}_t = \chi_t(\widehat{g}_n))$ corresponding to the BOD model are not significantly different from those obtained from the ARCP model, albeit they are slightly larger. In Figure 6.4 we plotted the EMVT series together with the conditional mean series generated by the ARCP and BOD models. Both generated series tend to have similar behavior over time.

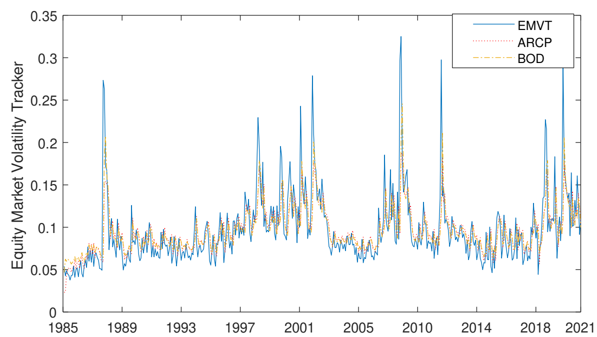


Figure 6.4. The EMVT series and its estimated conditional means generated by the ARCP and the BOD models.

To measure the adequacy of the whole conditional distribution of the BOD model, we display in Figure 6.5 the PIT of the EMVT series with respect to the conditional beta distribution given by (5.1). The EMVT series is less compatible with this conditional distribution than it is with the ARCP model having a Simplex distributed innovation.

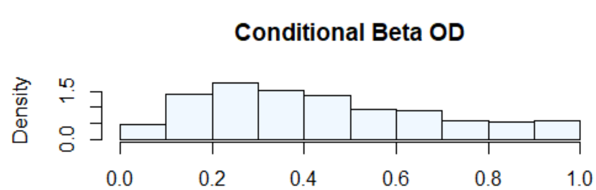


Figure 6.5. PIT of the EMVT series with respect to Beta distribution.

Next, we compare the out-of-sample forecasting performance of the ARCP$(2,1)$ and BOD$(1,1)$ models. We estimate the two models using the first n_f observations of the series, where $1 < n_f < n = 438$. The ARCP model is estimated by EQMLE, while the BOD model is estimated by MLE. We compute the one-step ahead forecasts, $\frac{\mu_0}{\lambda_t}$, from the resulting ARCP model over the period $(n_f+1,...,n)$, using

$$\widehat{\lambda}_t = \widehat{\omega} + \widehat{\alpha}_1 Y_{t-1}^{-1} + \widehat{\beta}_1 \widehat{\lambda}_{t-1} + \widehat{\beta}_2 \widehat{\lambda}_{t-1} \text{ for } t = n_f+1,...,n,$$

and the one-step ahead forecasts, $\widehat{\chi}_t$, for the BOD model, using

$$\widehat{\chi}_t = \widehat{d} + \widehat{a} Y_{t-1} + \widehat{b} \widehat{\chi}_{t-1} \text{ for } t = n_f+1,...,n.$$

For each model we obtain the i) the mean square forecast error, given by MSFE$=\frac{1}{n-n_f}\sum_{t=n_f+1}^{n}(Y_t-\widehat{\eta}_t)^2$, and ii) the mean absolute forecast error, given by MAFE $=\frac{1}{n-n_f}\sum_{t=n_f+1}^{n}|Y_t-\widehat{\eta}_t|$, where $\widehat{\eta}_t = \mu_0 \widehat{\lambda}_t^{-1}$ for the ARCP model and $\widehat{\psi}_t = \widehat{\chi}_t$ for the BOD model.

Table 6.4 displays the computed values of two criteria for the two models and for several truncated series with sample size $n_f \in \{200, 250, 300, 350, 400\}$. The ARCP model yields better out-of-sample predictions, in terms of the MSFE and MAFE values, irrespective of the time-cut n_f. For certain time-cuts n_f (e.g. 400, 350) the difference is significant, as the MAFE value for the BOD is twice the MAFE value for the ARCP.

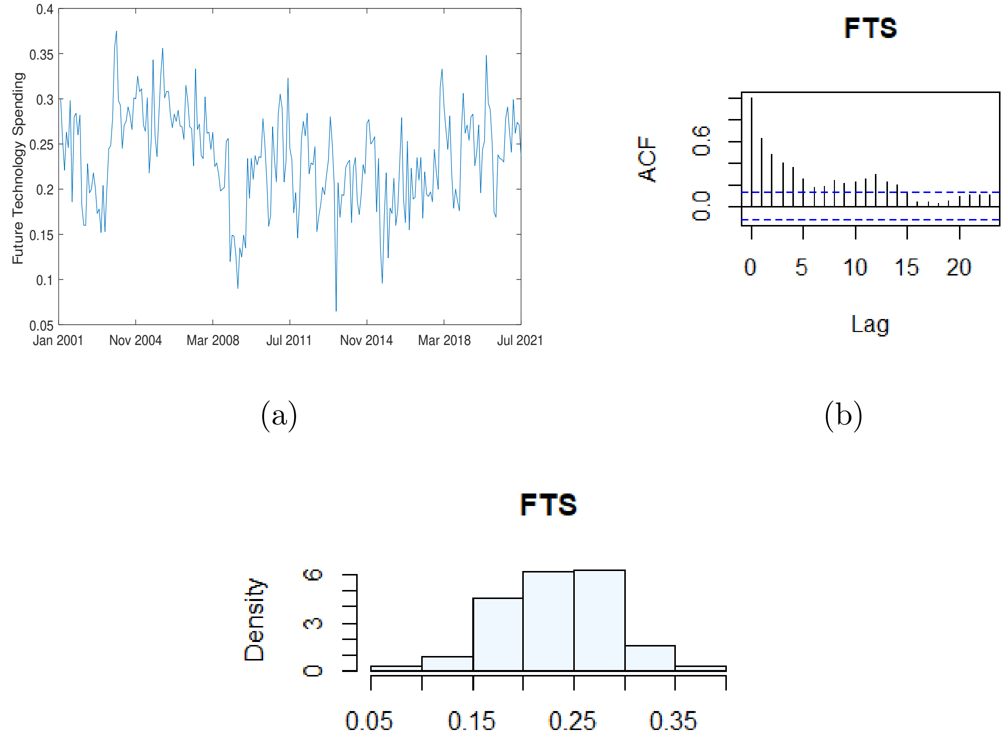
In summary, while the ARCP$(1,2)$ and BOD$(1,1)$ have similar in-sample forecasting performance, the ARCP$(1,2)$ clearly outperforms the benchmark model, in terms of the out-of-sample forecasting criterion.

	n_f	200	250	300	350	400
ARCP$(1,2)$	MSFE	0.0012	0.0013	0.0010	0.0011	0.0021
	MAFE	0.0215	0.0224	0.0212	0.0231	0.0307
BOD$(1,1)$	MSFE	0.0014	0.0018	0.0021	0.0024	0.0037
	MAFE	0.0279	0.0331	0.0375	0.0404	0.0503

Table 6.4. Out-of-sample forecasting performances of ARCP and BOD; EMVT series.

6.2 Future Technology Spending: Percent Expecting Increases for New York (FTS)

In the second example, we fitted the ARCP model to the monthly Future Technology Spending (FTS) series, which represents the expected direction of technology spending, in percent, over the six months ahead for the state of New York. The dataset, downloaded from the FRED database (https://fred.stlouisfed.org), ranges from July 1, 2001 to June 1, 2021, with a total of $n = 240$ observations (see Figure 6.6a).



(a) (b)

(c)

Figure 6.6. (a) FTS series; (b) sample autocorrelation; (c) histogram.

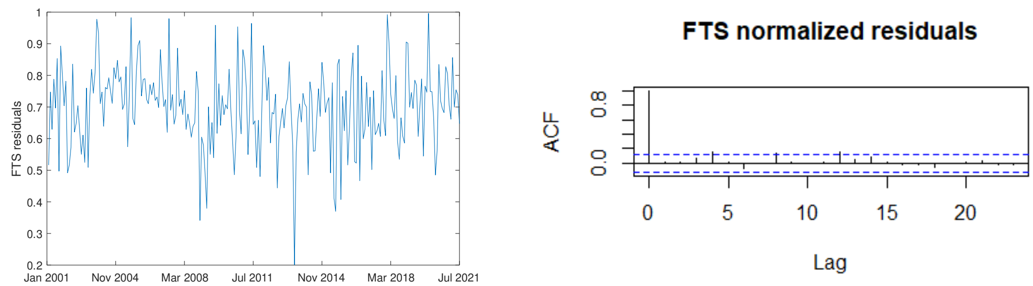
Using the same analysis as in Section 6.1, we set $\mu_0 = 0.7$. We also choose $p = q = 1$, since for larger orders the estimated parameters are not significant. We, thus, estimate an ARCP(1, 1) model with initial parameter value $\theta^{(0)} = \left(\omega^{(0)}, \alpha_1^{(0)}, \beta_1^{(0)}\right)' = (0.1, 0.1, 0.1)'$ and starting values $\lambda_0 = \omega^{(0)}$ and $Y_0 = 1/\lambda_0$. Table 6.5 shows the parameter estimates of the ARCP(1, 1) model, as well as their ASEs in parentheses. The mean-inverted persistence parameter was $Persi = 0.6714$ (see Table 6.5) and the estimated mean innovation was

$\frac{1}{n}\sum_{t=1}^{n}\widehat{\xi}_t = 0.7040$, which is close to $\mu_0 = E(\xi_t) = 0.7$.

$\widehat{\omega}$	$\widehat{\alpha}_{1n}$	$\widehat{\beta}_{1n}$	$\widehat{\sigma}_n^2$	$\widehat{\phi}_n$	$Persi$	MAFE	MSFE
1.0001 (0.0103)	0.2794 (0.0026)	0.2569 (0.0054)	0.0172 (0.0092)	11.1850	0.6714	0.0314	0.0017

Table 6.5. EQML and WLS estimates for the ARCP$(1,1)$ model; FTS series.

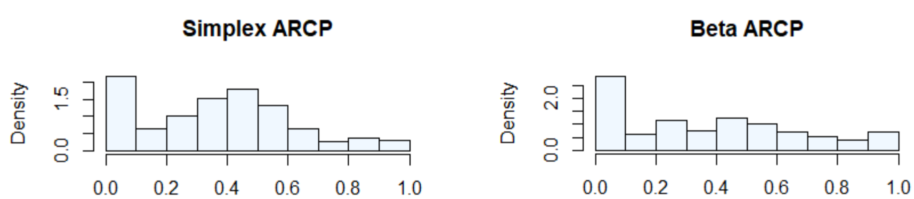
The MAFE and MSFE values are respectively 0.0315 and 0.0017. Figure 6.7 shows the normalized residuals $(\widehat{\xi}_t)$ and its sample autocorrelation function, suggesting that there is no evidence of correlation.



(a) (b)

Figure 6.7. (a) Normalized residuals; (b) residual sample autocorrelation.

The PITs of the normalized residuals $(\widehat{\xi}_t)$ (see Figure 6.8) with respect to the Beta distribution $\mathcal{B}eta\left(\widehat{\phi}_n\mu_0, \widehat{\phi}_n(1-\mu_0)\right)$ (with $\widehat{\phi}_n = 11.1850$, see Table 6.5) and the Simplex distribution $S^-(\mu_0, 6.52)$, imply that these two distributions are not good choices for the dataset in question.



(a) (b)

Figure 6.8. PIT of the ARCP residuals with respect to: (a) the simplex distribution, (b) the beta distribution. FTS series.

In addition, we fitted to the FTS series the linear conditional-mean Beta-observation driven (BOD) model given by (5.1). The MLE is

$$\widehat{g}_n = (0.0463, 0.4939, 0.3058, 94.7839)',$$

with persistence parameter 0.7997, $MAFE = 0.0312$ and $MSFE = 0.0018$. Figure 6.9 presents the EMVT series along with the conditional means obtained from the ARCP and BOD models. Similar comovements are observed for the two produced series.

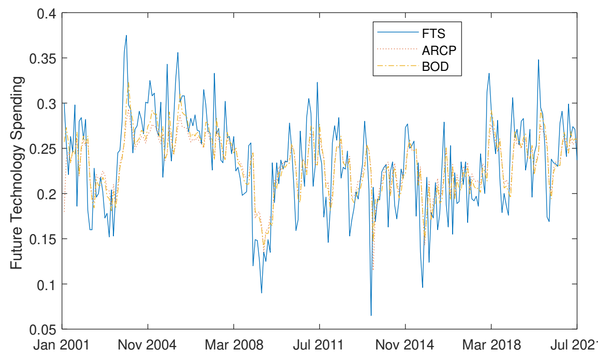


Figure 6.9. FTS series and its estimated conditional means generated by the ARCP and the BOD models.

Figure 6.10 displays the PIT of the FTS series with respect to the BOD distribution given by (5.1). Compared to the aforementioned ARCP distributions, the BOD distribution is not quite consistent with the FTS series. This supports our claim that the semiparametric modeling of $(0,1)$-valued time series data has a clear advantage over fully parametric settings.

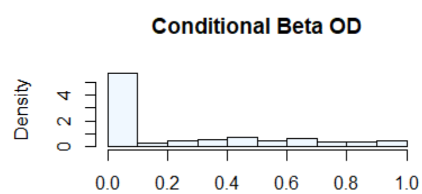


Figure 6.10. PIT of the FTS series with respect to the Beta conditional distribution.

Finally, we compare the out-of-sample forecasting performance of the ARCP$(1,1)$ with that of the BOD$(1,1)$. Table 6.6 reports the MSFE and MAFE values for the two models and for truncation points $n_f \in \{180, 200, 210, 220, 230\}$. The ARCP$(1,1)$ model yields slightly better out-of-sample forecasts regarding the MSFE and MAFE for all chosen values of the time-cut n_f. Overall, the ARCP$(1,1)$ and BOD$(1,1)$ models have similar performances in terms of in-sample and out-of-sample forecasting with a slight advantage of the ARCP$(1,1)$ model.

	n_f	180	200	210	220	230
ARCP$(1,1)$	MSFE	0.0014	0.0015	0.0016	0.0015	0.0013
	MAFE	0.0296	0.0308	0.0292	0.0294	0.0264
BOD$(1,1)$	MSFE	0.0016	0.0017	0.0016	0.0022	0.0018
	MAFE	0.0310	0.0311	0.0310	0.0332	0.0327

Table 6.6. Out-of-sample forecasting performances of ARCP and BOD; FTS series.

7 Conclusion

In this paper, we proposed an MEM equation-based model for $(0,1)$-valued time series. Compared to existing Beta and/or distribution-based models, our ARCP(p,q) formulation has several advantages. First of all, it is semi-parametric, it allows for high orders as well as for exogenous covariates, while numerous $(0,1)$-valued distributions for the innovation can be considered. Furthermore, the multiplicative form of the ARCP simplifies the study of its probability structure, which was carried out using standard tools for GARCH-type models (Francq and Zakoian, 2019). In addition, the estimation of the ARCP model was easily conducted using the exponential QMLE, which is asymptotically optimal in the class of exponential family QMLEs, as our model satisfies the quadratic-variance-to mean relationship.

We applied our model to two economic datasets and showed that it produces better out-of-sample forecasts than the (conditionally) beta observation-driven model, recently proposed by Gorgi and Koopman (2021). A third application of our model to the relative range of the S&P 500 volume (RRVSP500), which appears in the Supplementary material, leads to similar conclusions. The ARCP model was built and assessed in terms of its forecasting performance, without assuming a specific distribution for the innovation. However, the PIT showed that for the three datasets, the Beta distribution is generally a bad choice, whereas the Simplex distribution is a better one, in particular, for the EMVT and the RRVSP500 series.

Although in the set up of our proposed model specification we initially included covariates, we excluded them from our empirical applications. The reason was that we wanted to examine mainly the performance/behavior of our newly presented model without covariates. The case of the ARCP with covariates will be studied in detail in a future paper. The ARCP model could be generalized in many ways. First, nonlinear forms of the reciprocal conditional mean equation in (2.1) could be adopted (e.g. Creal et *al*, 2013). Second, time-varying and random-coefficients formulations of the λ_t equation could be proposed. In particular, threshold, mixture, Markov-mixture, and periodic extensions of (2.1) seem appealing. Third, the model could be easily adapted to model bounded-valued data not necessarily in the interval $(0, 1)$. For singular 0 or 1 values, zero-inflated versions could be introduced. Fourth, the model could be easily adapted to multivariate settings. Given a $(0, 1)^m$-valued $(m \geq 1)$ multivariate innovation $(\boldsymbol{\xi}_t)$, the distribution of which can be constructed using copula, a multivariate extension of the ARCP model (2.1) would have the form (assuming $p = q = 1$ for simplicity)

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\lambda}_t \odot \boldsymbol{\xi}_t \\ \boldsymbol{\lambda}_t &= \boldsymbol{\omega} + \boldsymbol{\alpha} \mathbf{Y}_t^{-1} + \boldsymbol{\beta} \boldsymbol{\lambda}_{t-1} \end{aligned}$$

where $\boldsymbol{\xi}_t = (\xi_{1t}, ..., \xi_{mt})'$, $\boldsymbol{\lambda}_t = (\lambda_{1t}, ..., \lambda_{mt})'$, $\mathbf{Y}_t^{-1} := \left(Y_{1t}^{-1}, ..., Y_{mt}^{-1}\right)'$, $\boldsymbol{\omega} = (\omega_1, ..., \omega_m)'$ with $\omega_i > 1$ $(i = 1, ..., m)$, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being appropriate non-negative matrices (see e.g. Fabrizio et *al*, 2013 for the positive-valued data case). The symbol $\odot$ denotes the Hadamard product. Clearly, $\mathbf{Y}_t \in (0, 1)^m$. Needless to say, most distribution-based models with a beta distribution seem very difficult to extend to multivariate forms.

References

[1] Aknouche, A. and Francq, C. (2021a). Count and duration time series with equal conditional stochastic and mean orders. *Econometric Theory* 37, 248–280.

[2] Aknouche, A. and Francq, C. (2021b). Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *Journal of Econometrics, forthcoming.*

[3] Aknouche, A. and Francq, C. (2021c). Stationarity and ergodicity of Markov-switching positive conditional mean models. *Journal of Time Series Analysis, forthcoming.*

[4] Aknouche, A., Almohaimeed, B. and Dimitrakopoulos, S. (2021). Periodic autoregressive conditional duration. *Journal of Time Series Analysis*, *forthcoming.*

[5] Aknouche, A., Bendjeddou, S. and Touche, N. (2018). Negative binomial quasi-likelihood inference for general integer-valued time series models. *Journal of Time Series Analysis* 39, 192–211.

[6] Baker, S.R., Bloom, N. Davis, S.J. and Kost, K. (2019). Policy news and stock market volatility. Working paper, *Federal Reserve Bank of St. Louis.*

[7] Barndorff-Nielsen, O.E. and Jorgensen, B. (1991). Some parametric models on the Simplex. *Journal of Multivariate Analysis* 39, 106–116.

[8] Bayer, F.M., Cintra, R.J. and Cribari-Neto, F. (2018). Beta seasonal autoregressive moving average models. *Journal of Statistical Computation and Simulation* 88, 2961-2981.

[9] Benjamin, M.A., Rigby, R.A. and Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214–223.

[10] Bonat, W.H., Petterle, R.R., Hinde, J. and Demétrio, C.G.B. (2019). Flexible quasi-Beta regression models for continuous bounded data. *Statistical Modeling* 19, 617–633.

[11] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31, 942–959.

[12] Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Annals of Probability*, **20**, 1714-1730.

[13] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) *Time Series Analysis; Forecasting and Control*. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

[14] Brockwell, P.J. and Davis, R.A. (1991). *Time series: theory and methods*. Springer Science & Business Media.

[15] Cipolini, F. and Gallo, G.M. (2021). Multiplicative Error Models: 20 years on. ArXiv working paper, arXiv:2107.05923.

[16] Creal, D., Koopman, S.J. and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795.

[17] Da Silva, C.Q. and Migon, H.S. (2016). Hierarchical dynamic beta model. *REVSTAT – Statistical Journal* 14, 49–73.

[18] Davis, R.A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* 26, 1673–1707.

[19] Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.

[20] Engle, R. and Gallo. G.M. (2006). A Multiple Indicators Model for Volatility Using Intra-Daily Data. *Journal of Econometrics* 131, 3–27.

[21] Engle, R. and Russell, J. (1998). Autoregressive conditional duration: A new model for irregular spaced transaction data. *Econometrica* 66, 1127–1162.

[22] Espinheira, P.L. and Silva, A.O.S. (2020). Residual and influence analysis to a general class of Simplex regression. *Test* 29, 523–552.

[23] Ferrari, S.L.P. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 799–815.

[24] Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* 27, 923–942.

[25] Francq, C., and Thieu, L. (2019). QML inference for volatility models with covariates. *Econometric Theory* 35, 37–72.

[26] Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition, John Wiley & Sons.

[27] Fabrizio, C., Engle, R.F. and Gallo. G.M. (2013). Semiparametric Vector MEM. *Journal of Applied Econometrics* 28, 1067–1086.

[28] Grassia, A. (1977). On a family of distributions with argument between 0 and 1 obtained by transformation of the Gamma distribution and derived compound distributions. *Australian Journal of Statistics* 19, 108–14.

[29] Grunwald, G., Hyndman, R., Tedesco, L., and Tweedie, R. (2000). Non-Gaussian conditional linear $AR(1)$ models. *Australian and New Zealand Journal of Statistics* **42**, 479–495.

[30] Gorgi, P. and Koopman, S.J. (2021). Beta observation-driven models with exogenous regressors: A joint analysis of realized correlation and leverage effects. *Journal of Econometrics,* forthcoming.

[31] Guolo, A. and Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to Canada google flu trends. *Annals of Applied Statistics*, 8, 74–88.

[32] Grunwald, G., Hyndman, R., Tedesco, L., and Tweedie, R. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics* **42**, 479–495.

[33] Hamilton J.D. (1994). Time series analysis. Princeton University Press, Princeton, NJ.

[34] Han, H. and Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics* 32, 416–429.

[35] Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Springer-Verlag.

[36] Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *Available at SSRN 1117187*.

[37] Jorgensen, B. (1997). *The theory of dispersion models*. Chapman and Hall, London.

[38] Kieschnick, R. and McCullough, B.D. (2003). Regression analysis of variates observed on (0,1): Percentages, proportions and fractions. *Statistical Modelling* 3, 193–213.

[39] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology* 46, 79–88.

[40] Li, W.K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics* 50, 506–511.

[41] McKenzie, E. (2003). *Discrete variate time series*, in Handbook of statistics. Amsterdam: Elsevier Science.

[42] McKenzie, E. (1985) An autoregressive process for beta random variables. *Management Science* 31, 988–97.

[43] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd edn). London: Chapman and Hall.

[44] Mitnik, P.A. and Baek, S. (2013). The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers* 54, 177–92.

[45] Paolino, P. (2001). Maximum likelihood estimation of models with Beta-distributed dependent variables. *Political Analysis* 9, 325–346.

[46] Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.

[47] Pumi, G., Valk, M., Bisognin, C., Bayer, F.M., and Prass, T.S. (2019). Beta autoregressive fractionally integrated moving average models. *Journal of Statistical Planning and Inference* 200, 196–212.

[48] Ristić, M.M., Weiss, C.H. and Janjić, A.D. (2016). A binomial integer-valued ARCH model. *International Journal of Biostatistics* 12, 1-21.

[49] Rocha, A.V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *TEST* 18, 529–545.

[50] Wooldridge, J.M. (1999). *Quasi-Likelihood Methods for Count Data.* In M.H. Pesaran and P. Schmidt (ed.), *Handbook of Applied Econometrics, Volume 2: Microeconomics, (pp. 35–406).* Oxford: Blackwell.

[51] Zeger, S.L., and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics* 44, 1019–1031.

[52] Zhang, P., Qiu, Z. and Shi, C. (2016). Simplexreg: An R package for regression analysis of proportional data using the Simplex distribution. *Journal of Statistical Software* 71, 1–21.

[53] Zheng, T., Xiao, H. and Chen, R. (2015). Generalized ARMA with martingale difference errors. *Journal of Econometrics* 189, 492–506.

8 Supplementary material: Proofs

8.1 Proof of Theorem 3.1

i) Since by the assumption of the ARCP model $E\xi_t^{-1} < \infty$ we, thus, have $E\log^+\|A_{0t}\| < \infty$ and $E\log^+\|B_t\| < \infty$. By Theorem 2.5 of Bougerol and Picard (1992*b*), equation (3.1) admits a unique nonanticipative strictly stationary and ergodic solution $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ if (3.4) holds. That solution is given by (3.3) where the series therein converges absolutely $a.s.$

Conversely, if model (3.1) admits a nonanticipative strictly stationary solution $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$, then, from the non-negativity of the coefficients of A_{0t} we obtain for all $t > 1$,

$$\mathbf{Y}_0 \geq \sum_{j=0}^{t-1}\prod_{i=0}^{j-1} A_{0,-i}B_{0,-j}, \quad a.s.$$

Hence, the series $\sum\limits_{j=0}^{\infty}\prod\limits_{i=0}^{j-1} A_{0,-i}B_{0,-j}$ converges $a.s.$ and thus $\prod\limits_{i=0}^{j-1} A_{0,-i}B_{0,-j} \underset{j\to\infty}{\overset{a.s}{\to}} 0$, from which condition (3.4) follows from Lemma 3.4 of Bougerol and Picard (1992*a*) as long as $\prod\limits_{i=0}^{j-1} A_{0,-i} \underset{j\to\infty}{\overset{a.s}{\to}} 0$. This holds whenever

$$\prod_{i=0}^{j-1} A_{0,-i}e_m \underset{j\to\infty}{\overset{a.s}{\to}} 0 \text{ for all } 1 \leq m \leq p+q \tag{1}$$

where $(e_m)_{1\leq m\leq p+q}$ is the canonical basis of $\mathbb{R}^{p+q}$. Since A_{0t} has the same zero-structure as the matrix A_t in Bougerol and Picard (1992*b*), condition (1) follows in the same way from their result. See also Pan et *al* (2008) and Francq and Thieu (2019).

ii) By the nonnegativity of $(A_{0t})_t$ we have

$$\gamma(\mathbb{A}_0) \geq \gamma(\boldsymbol{\beta}_0) := \log\rho(\boldsymbol{\beta}_0). \tag{2}$$

If (3.1) admits a strictly stationary solution, then, $\gamma(\mathbb{A}_0) < 0$ and by (2) it follows that

$$\gamma(\boldsymbol{\beta}_0) < 0 \tag{3}$$

which, in turn, implies (3.5). $\square$

8.2 Proof of Theorem 3.2

See the proof of Theorem 3.3 ii) with $m = 1$ for the sufficiency of (3.6). To prove the necessity of (3.6), let $\{Y_t, t \in \mathbb{Z}\}$ be a non-anticipative second-order stationary solution of the $ARCP(p,q)$ equation (2.1). Then, taking expectation in (2.1) we obtain

$$\begin{aligned} EY_t^{-1} &= E\xi_t^{-1}E\lambda_t = E\xi_t^{-1}\left(\omega_0 + \sum_{i=1}^{q}\alpha_{0i}EY_{t-i}^{-1} + \sum_{j=1}^{p}\beta_{0j}EY_{t-j}^{-1}\left(E\xi_{t-j}^{-1}\right)^{-1} + \boldsymbol{\pi}_0' E\mathbf{X}_{t-1}\right) \\ &= E\xi_t^{-1}\left(\omega_0 + \boldsymbol{\pi}_0' E\mathbf{X}_{t-1}\right) + \left(\sum_{i=1}^{q}E\xi_t^{-1}\alpha_{0i} + \sum_{j=1}^{p}\beta_{0j}\right)EY_t^{-1} \end{aligned}$$

so

$$\left(1 - E\xi_t^{-1}\sum_{i=1}^{q}\alpha_{0i} - \sum_{j=1}^{p}\beta_{0j}\right)EY_t^{-1} = E\xi_t^{-1}\left(\omega_0 + \boldsymbol{\pi}_0' E\mathbf{X}_{t-1}\right).$$

Since $E\xi_t^{-1} > 0$ and $\omega_0 > 0$, condition (3.7) (and thus (3.6)) should be satisfied.

8.3 Proof of Theorem 3.3

i) The proof is similar to that of Lemma 2.3 of Berkes et *al* (2003). Let us first show that (3.4) implies the existence of $\delta > 0$ and n_0 such that

$$E\left(\|A_{0,n_0}A_{0,n_0-1}...A_{01}\|^{\delta}\right) < 1. \tag{4}$$

Since $\gamma(\mathbb{A}_0) = \inf_{n\in\mathbb{N}}\left\{\frac{1}{n}E\left(\log\|A_{0n}A_{0,n-1}...A_{01}\|\right)\right\}$ is strictly negative, there exists a positive integer n_0 such that

$$E\left(\log\|A_{0,n_0}A_{0,n_0-1}...A_{01}\|\right) < 0. \tag{5}$$

Using a multiplicative norm and in view of the *iid* property of the sequence $\{A_{0t}, t \in \mathbb{Z}\}$, and the non-negativity of its terms we have

$$\begin{aligned} E\left(\|A_{0,n_0}A_{0,n_0-1}...A_{01}\|\right) &= \|E\left(A_{0,n_0}A_{0,n_0-1}...A_{01}\right)\| \\ &\leq \|E\left(A_{01}\right)\|^{n_0} < \infty. \end{aligned}$$

Let $f(x) = E\left(\|A_{0,n_0}A_{0,n_0-1}...A_{01}\|^{x}\right)$. Since from (5)

$$f'(0) = E\left(\log\|A_{0,n_0}A_{0,n_0-1}...A_{01}\|\right) < 0,$$

the function $f(x)$ decreases in a neighborhood of 0 and since $f(0) = 1$, it follows that there exists $0 < \delta < 1$ such that (4) is satisfied. Now, from (3.3) we have

$$\|\mathbf{Y}_t\| \leq \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \|A_{0,t-i}\| \, \|B_{0,t-j}\| + \|B_{0t}\| \,,$$

and for any $0 < \kappa < 1$, it follows that

$$\|\mathbf{Y}_t\|^{\kappa} \leq \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \|A_{0,t-i}\|^{\kappa} \, \|B_{0,t-j}\|^{\kappa} + \|B_{0t}\|^{\kappa} .$$

Using (4), the independence of $\{\xi_t, t \in \mathbb{Z}\}$, and the finiteness of $E\left(\|B_{0,t-j}\|^{\kappa}\right)$, which stems from assumption **A0**, we finally obtain

$$\begin{aligned} E\|\mathbf{Y}_t\|^{\kappa} &\leq \sum_{j=1}^{\infty} E\left(\prod_{i=0}^{j-1} \|A_{0,t-i}\|^{\kappa}\right) E\left(\|B_{0,t-j}\|^{\kappa}\right) + E\left(\|B_{0t}\|^{\kappa}\right) \\ &\leq E\left(\|B_{0t}\|^{\kappa}\right) \sum_{j=1}^{\infty} \left(E\left(\|A_{0t}\|^{\kappa}\right)\right)^{j} + E\left(\|B_{0t}\|^{\kappa}\right) < \infty \end{aligned}$$

establishing the result.

ii) The proof follows the same lines of Francq and Zakoian (2019, Theorem 2.9). Let

$$A_{t,j} = \begin{cases} \prod\limits_{i=0}^{j-1} A_{0,t-i} & \text{if } j > 0 \\ I_{p+q} & \text{if } j = 0 \end{cases} \quad \text{and } \mathbf{Y}_{t,j} = \begin{cases} A_{t,j} B_{0,t-j} & \text{if } j > 0 \\ B_{0t} & \text{if } j = 0 \end{cases}$$

where I_m denotes the identity matrix of dimension m. It is clear that (3.3) writes as follows

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} \mathbf{Y}_{t,j}$$

where the latter series is defined a.s. in $[0, \infty]$. Using the matrix norm $\|A\| = \sum |a_{ij}|$, the identity $\|A\| \, \|B\| = \|A \otimes B\| = \|B \otimes A\|$, the associativity of $\otimes$, the property $(A \otimes C)(B \otimes D) =$

$AB \otimes CD$, and the iid property of $(\xi_t)_t$ and its independence of $(\mathbf{X}_t)_t$ we obtain

$$\begin{aligned}
E\left(\|\mathbf{Y}_{t,j}\|^m\right) &= E\left(\|\mathbf{Y}_{t,j} \otimes \cdots \otimes \mathbf{Y}_{t,j}\|\right) \\
&= E\left(\|A_{t,j}B_{0,t-j} \otimes \cdots \otimes A_{t,j}B_{0,t-j}\|\right) \\
&= \|E\left(A_{t,j}B_{0,t-j} \otimes \cdots \otimes A_{t,j}B_{0,t-j}\right)\| \\
&= \left\|E\left(A_{t,j}^{\otimes m}B_{0,t-j}^{\otimes m}\right)\right\| \\
&= \left\|E\left(A_{0t}^{\otimes m} \cdots A_{0,t-j+1}^{\otimes m}B_{0,t-j}^{\otimes m}\right)\right\| \\
&= \left\|\left(E\left(A_{0t}^{\otimes m}\right)\right)^j E\left(B_{0,t-j}^{\otimes m}\right)\right\|.
\end{aligned}$$

The matrix norm $\|.\|$ is multiplicative and considering the norm $\|\mathbf{Z}\|_m = \left(E\|\mathbf{Z}\|^m\right)^{\frac{1}{m}}$ it follows that

$$\begin{aligned}
\|\mathbf{Y}_t\|_m &= \left\|\sum_{j=0}^{\infty}\mathbf{Y}_{t,j}\right\|_m \leq \sum_{j=0}^{\infty}\|\mathbf{Y}_{t,j}\|_m \\
&= \sum_{j=0}^{\infty}\left(\left\|\left(E\left(A_{0,t}^{\otimes m}\right)\right)^j E\left(B_{0,t-j}^{\otimes m}\right)\right\|\right)^{\frac{1}{m}} \\
&\leq \sum_{j=0}^{\infty}\left\|\left(E\left(A_{0,t}^{\otimes m}\right)\right)^j\right\|^{\frac{1}{m}}\left\|E\left(B_{0,t-j}^{\otimes m}\right)\right\|^{\frac{1}{m}}
\end{aligned}$$

Since $\rho\left(A_{0t}^{\otimes m}\right) < 1$, the matrix $\left(E\left(A_{0,t}^{\otimes m}\right)\right)^j$ converges to zero at exponential rate when $j \to \infty$. It follows that the latter series converges $a.s.$ and hence $\mathbf{Y}_t$ is a.s. finite.

8.4 Proof of Theorem 4.1

We will prove Theorem 4.1 by using various lemmas below. In what follows, $M > 0$ and $\rho \in (0,1)$ denote generic constants that are not necessarily the same when appearing in different terms. Let

$$l_t(\theta) = \tfrac{\lambda_t(\theta)}{\mu_0}Y_t - \log\lambda_t(\theta), \quad t \in \mathbb{Z}$$

and $L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta)$ be the ergodic counterparts of $\widetilde{l}_t(\theta)$ and $\widetilde{L}_n(\theta)$ respectively.

Lemma 1 *Under **A0**-**A5** we have*

$$\sup_{\theta\in\Theta}\left|L_n(\theta) - \widetilde{L}_n(\theta)\right| \underset{n\to\infty}{\overset{a.s}{\to}} 0.$$

Proof Let us first show that

$$\sup_{\theta\in\Theta}\left|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\right|\leq M\rho^t. \tag{6}$$

Rewrite (4.2) in a vector form as follows

$$\underline{\lambda}_t(\theta)=\boldsymbol{\beta}\underline{\lambda}_{t-1}(\theta)+\underline{\boldsymbol{c}}_t,\ t\in\mathbb{Z} \tag{7}$$

where $\underline{\lambda}_t(\theta)=(\lambda_t(\theta),...,\lambda_{t-p+1}(\theta))'$ and $\underline{\boldsymbol{c}}_t=(\omega+\boldsymbol{\pi}'\mathbf{X}_{t-1}+\sum_{i=1}^{q}\alpha_i Y_{t-i}^{-1},0,...,0)'_{1\times p}$. By **A1** and the assumption **A2** of compactness of Θ it follows that

$$\sup_{\theta\in\Theta}\rho(\boldsymbol{\beta})<1.$$

Iterating (7) gives

$$\underline{\lambda}_t(\theta)=\sum_{k=0}^{t-1}\boldsymbol{\beta}^k\underline{\boldsymbol{c}}_{t-k}+\boldsymbol{\beta}^t\underline{\lambda}_0(\theta)=\sum_{k=0}^{\infty}\boldsymbol{\beta}^k\underline{\boldsymbol{c}}_{t-k},\quad t\in\mathbb{Z}. \tag{8}$$

Denote by $\widetilde{\underline{\lambda}}_t(\theta)$ and $\widetilde{\underline{\boldsymbol{c}}}_t$ the vectors obtained from $\underline{\lambda}_t(\theta)$ and $\underline{\boldsymbol{c}}_t$, respectively, while replacing $\lambda_{t-j}(\theta)$ by $\widetilde{\lambda}_{t-j}(\theta)$ with fixed initial values. That is,

$$\widetilde{\underline{\lambda}}_t(\theta)=\sum_{k=0}^{t-q-1}\boldsymbol{\beta}^k\underline{\boldsymbol{c}}_{t-k}+\sum_{k=t-q}^{t}\boldsymbol{\beta}^k\widetilde{\underline{\boldsymbol{c}}}_{t-k}. \tag{9}$$

From (8) and (9) we, thus, get

$$\sup_{\theta\in\Theta}\left\|\underline{\lambda}_t(\theta)-\widetilde{\underline{\lambda}}_t(\theta)\right\|=\sup_{\theta\in\Theta}\left\|\sum_{k=1}^{q}\boldsymbol{\beta}^{t-k}\left(\underline{\boldsymbol{c}}_k-\widetilde{\underline{\boldsymbol{c}}}_k\right)+\boldsymbol{\beta}^t\left(\underline{\lambda}_0(\theta)-\widetilde{\underline{\lambda}}_0(\theta)\right)\right\|\leq M\rho^t \text{ for all } t,$$

showing (6). Now, using the inequality $\log(x)\leq x-1$ and assumptions **A2**-**A3**, it follows that

$$\begin{aligned}\sup_{\theta\in\Theta}\left|L_n(\theta)-\widetilde{L}_n(\theta)\right| &= \frac{1}{n}\sup_{\theta\in\Theta}\left|\sum_{t=1}^{n}\left(Y_t\left(\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\right)+\log\left(\frac{\widetilde{\lambda}_t(\theta)}{\lambda_t(\theta)}\right)\right)\right|\\ &\leq \frac{1}{n}\sum_{t=1}^{n}\left(Y_t\sup_{\theta\in\Theta}\left|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\right|+\frac{\sup_{\theta\in\Theta}\left|\widetilde{\lambda}_t(\theta)-\lambda_t(\theta)\right|}{\omega}\right)\\ &\leq \frac{M}{n}\sum_{t=1}^{n}\left(Y_t\rho^t+\frac{\rho^t}{\omega}\right).\end{aligned}$$

The existence of EY_t^s $(s > 0)$ (since Y_t^s is bounded) implies, by the Borel-Cantelli lemma, that $\rho^t Y_t \underset{t\to\infty}{\overset{a.s}{\to}} 0$ and the conclusion follows by Césaro's lemma. $\square$

Lemma 2 *Under **A0-A5** there is $t \in \mathbb{Z}$ such that $\lambda_t(\theta) = \lambda_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$.*

Proof From the assumption $\rho(\boldsymbol{\beta}) < 1$ in **A1**, the polynomial

$$\beta_\theta(L) = 1 - \sum_{j=1}^{p} \beta_j L^j$$

is invertible for all $\theta \in \Theta$. Recall that the second equality in (2.1), and (4.2) can be written respectively in the polynomial forms

$$\begin{aligned}\lambda_t(\theta) &= \tfrac{\omega}{\beta_\theta(1)} + \tfrac{1}{\beta_\theta(L)}\boldsymbol{\pi}' L\mathbf{X}_t + \tfrac{\alpha_\theta(L)}{\beta_\theta(L)} Y_t^{-1} \\ \lambda_t(\theta_0) &= \tfrac{\omega_0}{\beta_{\theta_0}(1)} + \tfrac{1}{\beta_{\theta_0}(L)}\boldsymbol{\pi}_0' L\mathbf{X}_t + \tfrac{\alpha_{\theta_0}(L)}{\beta_{\theta_0}(L)} Y_t^{-1}\end{aligned}$$

where $\alpha_\theta(L) = \sum_{i=1}^{q} \alpha_i L^i$ and L is the backshift operator. Assume that

$$\lambda_t(\theta) = \lambda_t(\theta_0) \quad a.s. \quad \text{for some } t \in \mathbb{Z},$$

i.e.,

$$\tfrac{\omega}{\beta_\theta(1)} + \tfrac{\boldsymbol{\pi}' L}{\beta_\theta(L)}\mathbf{X}_t + \tfrac{\alpha_\theta(L)}{\beta_\theta(L)} Y_t^{-1} = \tfrac{\omega_0}{\beta_{\theta_0}(1)} + \tfrac{\boldsymbol{\pi}_0' L}{\beta_{\theta_0}(L)}\mathbf{X}_t + \tfrac{\alpha_{\theta_0}(L)}{\beta_{\theta_0}(L)} Y_t^{-1}.$$

Then,

$$\left(\tfrac{\alpha_\theta(L)}{\beta_\theta(L)} - \tfrac{\alpha_{\theta_0}(L)}{\beta_{\theta_0}(L)}\right) Y_t^{-1} + \left(\tfrac{\boldsymbol{\pi}' L}{\beta_\theta(L)} - \tfrac{\boldsymbol{\pi}_0' L}{\beta_{\theta_0}(L)}\right)\mathbf{X}_t = \tfrac{\omega_0}{\beta_{\theta_0}(1)} - \tfrac{\omega}{\beta_\theta(1)}.$$

If $\frac{\alpha_\theta(L)}{\beta_\theta(L)} \neq \frac{\alpha_{\theta_0}(L)}{\beta_{\theta_0}(L)}$, there exist a sequence of constants $(c_i)_{i\geq 1}$, a sequence of vectors $(\mathbf{d}_k)_{k\geq 1}$ and a constant b such that

$$\sum_{i=1}^{\infty} c_i Y_{t-i}^{-1} + \sum_{k=1}^{\infty} \mathbf{d}_k' \mathbf{X}_{t-k} = b$$

with $c_{i_0} \neq 0$ for some $i_0 \geq 1$. Since $Y_t^{-1} = \lambda_t \xi_t^{-1}$ and $(\xi_t^{-1})_t$ is iid, there exists a constant a such that $a\xi_{t-i_0}^{-1} = c_t$ where c_t is a measurable function of $(\mathbf{X}_{t-k})_{k\geq 1}$. This contradicts **A4** so, $\frac{\alpha_\theta(L)}{\beta_\theta(L)} = \frac{\alpha_{\theta_0}(L)}{\beta_{\theta_0}(L)}$, which by **A3** implies $\alpha_{\theta_0}(L) = \alpha_\theta(L)$ and $\beta_{\theta_0}(L) = \beta_\theta(L)$. Therefore, (9) becomes $(\boldsymbol{\pi} - \boldsymbol{\pi}_0)' \mathbf{X}_{t-1} = \omega_0 - \omega$ and under **A5** it follows that $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ and $\omega = \omega_0$.

Lemma 3 *Under **A0-A5** $E(|l_t(\theta_0)|) < \infty$ and $E(l_t(\theta))$ is minimized at $\theta = \theta_0$.*

Proof By Jensen's inequality and the existence of $E\lambda_t^\kappa$ (cf. Theorem 3.3, i)) we have $E\left|\log\left(\lambda_t\left(\theta_0\right)\right)\right| = \frac{1}{\kappa}E\left|\log\left(\lambda_t\left(\theta_0\right)^\kappa\right)\right| \leq \frac{1}{\kappa}\log E\left(\lambda_t\left(\theta_0\right)^\kappa\right) < \infty$. Since $E\left(Y_t\lambda_t\left(\theta_0\right)\right) = E\left(\xi_t\right) < \infty$, it follows that $E\left|l_t\left(\theta_0\right)\right| < \infty$. Moreover, using the inequality $\log\left(x\right) \leq x - 1$, we have

$$\begin{aligned} E\left(l_t\left(\theta_0\right) - l_t\left(\theta\right)\right) &= E\left(\tfrac{\lambda_t(\theta_0)}{\mu_0}Y_t - \log\lambda_t\left(\theta_0\right) - \left(\tfrac{\lambda_t(\theta)}{\mu_0}Y_t - \log\lambda_t\left(\theta\right)\right)\right) \\ &= E\left(\left(\lambda_t\left(\theta_0\right) - \lambda_t\left(\theta\right)\right)\tfrac{Y_t}{\mu_0} + \log\tfrac{\lambda_t(\theta)}{\lambda_t(\theta_0)}\right) \\ &\leq E\left(\left(\lambda_t\left(\theta_0\right) - \lambda_t\left(\theta\right)\right)\tfrac{Y_t}{\mu_0} - \tfrac{\lambda_t(\theta)-\lambda_t(\theta_0)}{\lambda_t(\theta_0)}\right) \\ &= E\left(\tfrac{\lambda_t(\theta)-\lambda_t(\theta_0)}{\lambda_t(\theta_0)}\right) + E\left(\tfrac{\lambda_t(\theta_0)-\lambda_t(\theta)}{\lambda_t(\theta)}\right) = 0 \end{aligned}$$

with equality iff $\lambda_t\left(\theta\right) = \lambda_t\left(\theta_0\right)$ a.s. and by **Lemma 2** iff $\theta = \theta_0$. It follows that $\forall\theta \neq \theta_0$, $E\left(l_t\left(\theta_0\right) - l_t\left(\theta\right)\right) \in \left[-\infty, 0\right]$ showing that $E\left(l_t\left(\theta\right)\right)$ is minimized at θ_0.

Lemma 4 *Under the assumptions of Theorem 4.1, there exists for any $\theta \neq \theta_0$ a neighborhood $\mathcal{V}\left(\theta\right)$ such that*

$$\lim_{n\to\infty}\inf\inf_{\overline{\theta}\in\mathcal{V}(\theta)}\widetilde{L}_n\left(\overline{\theta}\right) > E\left(l_t\left(\theta_0\right)\right), \text{ a.s.}$$

Proof Let $\mathcal{V}_k\left(\theta\right)$ be the open ball of center θ and radius $1/k$, where $\theta \in \Theta$ and k is a positive integer. In view of **Lemma 1** we have

$$\begin{aligned} \lim_{n\to\infty}\inf\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}\widetilde{L}_n\left(\overline{\theta}\right) &\geq \lim_{n\to\infty}\inf\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}L_n\left(\overline{\theta}\right) - \lim_{n\to\infty}\sup\sup_{\theta\in\Theta}\left|L_n\left(\theta\right) - \widetilde{L}_n\left(\theta\right)\right| \\ &\geq \lim_{n\to\infty}\inf\frac{1}{n}\sum_{t=1}^{n}\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}l_t\left(\overline{\theta}\right). \end{aligned}$$

The ergodic theorem for the stationary sequence $\left\{l_t\left(\overline{\theta}\right)\right\}_t$ with $E\left(l_t\left(\overline{\theta}\right)\right) \in \mathbb{R}\cup\left\{\infty\right\}$ (cf, Billingsley 1995, p. 495) entails

$$\lim_{n\to\infty}\inf\frac{1}{n}\sum_{t=1}^{n}\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}l_t\left(\overline{\theta}\right) = E\left(\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}l_t\left(\overline{\theta}\right)\right).$$

By the Beppo-Levi's theorem (e.g. Billingsley, 1995, p. 219) it follows that

$$E\left(\inf_{\overline{\theta}\in\mathcal{V}_k(\theta)\cap\Theta}l_t\left(\overline{\theta}\right)\right) \to E\left(l_t\left(\theta\right)\right) \text{ as } k \to \infty$$

and by **Lemma 3** the result follows.

Now, the proof of Theorem 4.1 follows by standard compactness arguments (cf. Francq and Zakoian, 2004; Francq and Zakoian, 2019, p. 192), using **Lemmas** 2-4. □

8.5 Proof of Theorem 4.2

The proof of Theorem 4.2 is based on a Taylor expansion of $\frac{\partial \widetilde{L}_n(\theta)}{\partial \theta}$ at θ_0, which by **A6** and the strong consistency of $\widehat{\theta}_n$, yields

$$0 = \sqrt{n}\frac{\partial \widetilde{L}_n(\widehat{\theta}_n)}{\partial \theta} = \sqrt{n}\frac{\partial L_n(\theta_0)}{\partial \theta} + \sqrt{n}\frac{\partial^2 L_n(\theta^*)}{\partial \theta \partial \theta'}\left(\widehat{\theta}_n - \theta_0\right) + \sqrt{n}\left(\frac{\partial \widetilde{L}_n(\widehat{\theta}_n)}{\partial \theta} - \frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta}\right) \tag{10}$$

where θ^* is between $\widehat{\theta}_n$ and θ_0. The derivatives $\frac{\partial L_n(\theta)}{\partial \theta}$ and $\frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'}$ are given by

$$\frac{\partial L_n(\theta)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t}{\mu_0} - \frac{1}{\lambda_t(\theta)}\right)\frac{\partial \lambda_t(\theta)}{\partial \theta} \tag{11}$$

$$\frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t}{\mu_0}\frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} - \frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial \theta \partial \theta'} + \frac{1}{\lambda_t^2(\theta)}\frac{\partial \lambda_t(\theta)}{\partial \theta}\frac{\partial \lambda_t(\theta)}{\partial \theta'}\right). \tag{12}$$

In view of (10), the asymptotic normality result (4.6) follows whenever the following lemmas are established.

Lemma 5 *Under **A0-A6***

$$E_{\theta_0}\left\|\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\right\| < \infty, \quad E_{\theta_0}\left\|\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\right\| < \infty.$$

Proof From (11) we have

$$\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'} = \left(\frac{\xi_t}{\mu_0} - 1\right)^2 \frac{1}{\lambda_t^2(\theta_0)}\frac{\partial \lambda_t(\theta_0)}{\partial \theta}\frac{\partial \lambda_t(\theta_0)}{\partial \theta'}$$

and since ξ_t is independent of $\lambda_t = \lambda_t(\theta_0)$, to establish the lemma it, thus, suffices to show that

$$E\left\|\frac{1}{\lambda_t(\theta_0)}\frac{\partial \lambda_t(\theta_0)}{\partial \theta \partial \theta'}\right\| < \infty,\ E\left\|\frac{1}{\lambda_t(\theta_0)}\frac{\partial^2 \lambda_t(\theta_0)}{\partial \theta \partial \theta'}\right\| < \infty,\ E\left\|\frac{1}{\lambda_t^2(\theta_0)}\frac{\partial \lambda_t(\theta_0)}{\partial \theta}\frac{\partial \lambda_t(\theta_0)}{\partial \theta'}\right\| < \infty. \tag{13}$$

Using (8) and similarly to Francq and Zakoian (2004) we have

$$\frac{\partial \underline{\lambda}_t(\theta)}{\partial \omega} = \sum_{k=0}^{\infty}\boldsymbol{\beta}^k\underline{\mathbf{1}},\ \frac{\partial \underline{\lambda}_t(\theta)}{\partial \alpha_i} = \sum_{k=0}^{\infty}\boldsymbol{\beta}^k\underline{\mathbf{Y}}_{t-k-i},\ \frac{\partial \underline{\lambda}_t(\theta)}{\partial \beta_j} = \sum_{k=1}^{\infty}\sum_{i=1}^{k}\boldsymbol{\beta}^{i-1}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{k-i}\underline{\boldsymbol{c}}_{t-k},\ \frac{\partial \underline{\lambda}_t(\theta)}{\partial \pi_l} = \sum_{k=0}^{\infty}\boldsymbol{\beta}^k\underline{\mathbf{X}}_{t-k-1}^{(l)} \tag{14}$$

where $\underline{\mathbf{1}} = (1, 0, ..., 0)'$, $\underline{\mathbf{Y}}_t = \left(Y_t^{-1}, 0, ..., 0\right)'$, $\underline{\mathbf{X}}_t^{(l)} = (X_{l,t-k-1}, 0, ..., 0)'$, and $\boldsymbol{\beta}^{(j)}$ is a $p \times p$ matrix with 1 in the entree $(1, j)$ and zeros elsewhere. From (14), it follows that $\frac{\partial \lambda_t(\theta)}{\partial \omega}$ is bounded and since $\lambda_t(\theta) \geq \omega$, the variable $\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \omega}$ is also bounded and has moment of any order, in particular at $\theta = \theta_0$. Moreover, $\alpha_i \frac{\partial \underline{\lambda}_t(\theta)}{\partial \alpha_i} = \sum_{k=0}^{\infty} \boldsymbol{\beta}^k \alpha_i \underline{\mathbf{Y}}_{t-k-i} \leq \sum_{k=0}^{\infty} \boldsymbol{\beta}^k \underline{\boldsymbol{c}}_{t-k} = \underline{\lambda}_t(\theta)$ so

$$\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \alpha_i} \leq \frac{1}{\alpha_i} \tag{15}$$

and thus $\frac{1}{\lambda_t^2(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \alpha_i}$ has also finite moments of all orders at $\theta = \theta_0$. Similarly,

$$\pi_l \frac{\partial \underline{\lambda}_t(\theta)}{\partial \pi_l} = \sum_{k=0}^{\infty} \boldsymbol{\beta}^k \pi_l \underline{\mathbf{X}}_{t-k-1}^{(l)} \leq \sum_{k=0}^{\infty} \boldsymbol{\beta}^k \underline{\boldsymbol{c}}_{t-k} = \underline{\lambda}_t(\theta)$$

hence,

$$\frac{1}{\lambda_t(\theta)} \frac{\partial \underline{\lambda}_t(\theta)}{\partial \pi_l} \leq \frac{1}{\pi_l} \tag{16}$$

and thus $\frac{1}{\lambda_t^2(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \pi_l}$ has also finite moments of any order.

Now, since $\beta_j \boldsymbol{\beta}^{(j)} \leq \boldsymbol{\beta}$, the third equality in (14) entails

$$\beta_j \frac{\partial \underline{\lambda}_t(\theta)}{\partial \beta_j} \leq \sum_{k=1}^{\infty} \sum_{i=1}^{k} \boldsymbol{\beta}^{i-1} \boldsymbol{\beta}^{(j)} \boldsymbol{\beta}^{k-i} \underline{\boldsymbol{c}}_{t-k} = \sum_{k=1}^{\infty} k \boldsymbol{\beta}^k \underline{\boldsymbol{c}}_{t-k}. \tag{17}$$

Since by **A1**, $\left\|\boldsymbol{\beta}^k\right\| \leq M\rho^k$ for all k, by Theorem 3.3 Y_t^{-1} has a finite moment of order $\kappa \in (0,1)$, and by **A0** the moment $E\left\|\mathbf{X}_t\right\|^\kappa$ is finite, then, the same moment of the variable $\underline{\boldsymbol{c}}_t(1) = \omega + \boldsymbol{\pi}'\mathbf{X}_{t-1} + \sum_{i=1}^{q} \alpha_i Y_{t-i}^{-1}$ is finite. In view of (8), $\lambda_t(\theta) \geq \omega + \boldsymbol{\beta}^k(1,1)\, \underline{\boldsymbol{c}}_{t-k}(1)$ and using the inequality $\frac{x}{x+1} \leq x^\kappa$ and (17) it follows that for all $\theta \in \Theta$

$$\begin{aligned} E\left(\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \beta_j}\right) &\leq \frac{1}{\beta_j} \sum_{k=1}^{\infty} E\left(\frac{k\boldsymbol{\beta}^k(1,1)\underline{\boldsymbol{c}}_{t-k}(1)}{\omega + \boldsymbol{\beta}^k(1,1)\underline{\boldsymbol{c}}_{t-k}(1)}\right) \leq \frac{1}{\beta_j} \sum_{k=1}^{\infty} \frac{1}{\omega^\kappa} k E\left(\boldsymbol{\beta}^k(1,1)\, \underline{\boldsymbol{c}}_{t-k}(1)\right)^\kappa \\ &\leq \frac{M^\kappa}{\omega^\kappa \beta_j} E\left(\underline{\boldsymbol{c}}_{t-k}(1)^\kappa\right) \sum_{k=1}^{\infty} k \rho^{\kappa k} \leq \frac{M}{\beta_j}. \end{aligned} \tag{18}$$

In view of **A6** we have $\beta_j > 0$ for all j, so with $\theta = \theta_0$ the first expectation in (13) exists.

Turn, now, to the second expectation in (13). Using (14), we have

$$\frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial \omega^2} = \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial \omega \partial \alpha_i} = \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial \omega \partial \pi_l} = 0,\ \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial \omega \partial \beta_j} = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \boldsymbol{\beta}^{i-1} \boldsymbol{\beta}^{(j)} \boldsymbol{\beta}^{k-i} \underline{\mathbf{1}}. \tag{19}$$

Since $\rho(\boldsymbol{\beta}) < 1$, it follows that

$$\beta_j \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial \omega \partial \beta_j} \leq \sum_{k=1}^{\infty} k \boldsymbol{\beta}^k \underline{\mathbf{1}} < \infty$$

implying $\frac{\partial^2 \lambda_t(\theta)}{\partial\omega\partial\beta_j} \leq \frac{M}{\beta_j}$ is bounded and admits moments of all orders, in particular at $\theta = \theta_0$. The same holds for $\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\omega\partial\beta_j}$ and hence for $\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\omega\partial\theta}$. In view of the second and fourth equalities in (14) we have

$$\frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\alpha_i\partial\alpha_{i'}} = \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\alpha_i\partial\pi_l} = 0, \ \frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\alpha_i\partial\beta_j} = \sum_{k=1}^{\infty}\sum_{i=1}^{k} \boldsymbol{\beta}^{i-1}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{k-i}\underline{\mathbf{Y}}_{t-k-i}. \tag{20}$$

By the same argument used to show (18) we obtain

$$E\left(\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\alpha_i\partial\beta_j}\right) \leq \frac{M}{\beta_j}$$

which, in turn, implies $E\left(\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\alpha_i\partial\theta}\right)$ is finite. From the third equality in (14) and using again $\beta_j\boldsymbol{\beta}^{(j)} \leq \boldsymbol{\beta}$ we have

$$\begin{aligned}
\beta_j\beta_{j'}\frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\beta_j\beta_{j'}} &= \beta_j\beta_{j'}\sum_{k=2}^{\infty}\left[\sum_{i=2}^{k}\sum_{m=1}^{i-1}\boldsymbol{\beta}^{m-1}\boldsymbol{\beta}^{(j')}\boldsymbol{\beta}^{i-1-m}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{k-i} + \sum_{i=1}^{k-1}\boldsymbol{\beta}^{i-1}\boldsymbol{\beta}^{(j)}\sum_{m=1}^{i-1}\boldsymbol{\beta}^{m-1}\boldsymbol{\beta}^{(j')}\boldsymbol{\beta}^{k-i-m}\right]\underline{\boldsymbol{c}}_{t-k} \\
&\leq \sum_{k=2}^{\infty}\left[\sum_{i=2}^{k}(i-1)\,\boldsymbol{\beta}^k + \sum_{i=1}^{k-1}(k-i)\,\boldsymbol{\beta}^k\right]\underline{\boldsymbol{c}}_{t-k} \\
&= \sum_{k=2}^{\infty} k\,(k-1)\,\boldsymbol{\beta}^k\underline{\boldsymbol{c}}_{t-k}.
\end{aligned} \tag{21}$$

The same argument used to show (18) entails

$$E\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\beta_j\beta_{j'}} \leq \frac{M}{\beta_j\beta_{j'}}.$$

In view of the third and fourth equality in (14) it follows that

$$\frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\beta_j\partial\pi_l} = \sum_{k=1}^{\infty}\sum_{l=1}^{k}\boldsymbol{\beta}^{l-1}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{k-l}\underline{\mathbf{X}}_{t-k-1}^{(l)}$$

and the same argument used to show (18) implies

$$E\left(\frac{1}{\lambda_t(\theta)}\frac{\partial^2 \lambda_t(\theta)}{\partial\pi_l\partial\beta_j}\right) \leq \frac{M}{\beta_j} \text{ for all } \theta \in \Theta.$$

Since in view of the fourth equality in (14), $\frac{\partial^2 \underline{\lambda}_t(\theta)}{\partial\pi_l\partial\pi_{l'}} = 0$ it follows that the second expectation in (13) is finite.

To prove the finiteness of the third expectation in (13) note first that as established above, $\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \omega}$ is bounded and by (15) and (16), $\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \alpha_i}$ and $\frac{1}{\lambda_t(\theta)} \frac{\partial \lambda_t(\theta)}{\partial \pi_l}$ are bounded too for all $\theta \in \Theta$. Hence $E \left\| \frac{1}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta_i} \frac{\partial \lambda_t(\theta_0)}{\partial \theta_i'} \right\|$ is finite for $i \in \{1, ..., q+1\} \cup \{p+q+2, ..., p+q+r+1\}$. Using the same arguments to prove (18), the inequality $\frac{x}{1+x} \leq x^{\frac{\kappa}{2}}$ for all $x \geq 0$, and Minkowski's inequality we have

$$\sqrt{E \left(\frac{1}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \beta_j} \right)^2} \leq \frac{1}{\beta_{0j}} \sum_{k=1}^{\infty} k \sqrt{E \left(\frac{\boldsymbol{\beta}_0^k(1,1) \underline{\boldsymbol{c}}_{0,t-k}(1)}{\omega_0} \right)^\kappa} < \infty,$$

where $\underline{\boldsymbol{c}}_{0,t-k}(1)$ is just $\underline{\boldsymbol{c}}_{t-k}(1)$ with θ_0 in place θ. Finally, the finiteness of the third expectation in (13) follows from the Cauchy-Schwarz inequality.

Lemma 6 *Under* ***A0-A6*** $J(\theta_0)$ is invertible and

$$Var\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \right) := E\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right) = \frac{\sigma_0^2}{\mu_0^2} J(\theta_0).$$

Proof From the integrability of the score $\frac{\partial l_t(\theta_0)}{\partial \theta}$ (cf. **Lemma 5**) and the independence between $\xi_t = \lambda_t(\theta_0) Y_t$ and $\lambda_t(\theta_0)$, we have

$$E\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \right) = E\left(\frac{\xi_t}{\mu_0 \lambda_t(\theta_0)} - \frac{1}{\lambda_t(\theta_0)} \right) = E\left(\frac{1}{\lambda_t(\theta_0)} \left(\frac{\xi_t}{\mu_0} - 1 \right) \right) = 0.$$

Hence

$$\begin{aligned} Var\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \right) &= E\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right) \\ &= E\left(\left(\frac{\xi_t}{\mu_0} - 1 \right)^2 \frac{1}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \\ &= \frac{\sigma_0^2}{\mu_0^2} E\left(\frac{1}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \end{aligned}$$

and the latter expectation exists in view of (13).

To establish the invertibility of $J(\theta_0)$ assume that $J(\theta_0)$ is singular. Then, there exists a non-null vector $\mu = (\mu_0, \mu_1, ..., \mu_{p+q+r})' \in \mathbb{R}^{p+q+r+1}$ satisfying $\mu' E\left(\frac{1}{\lambda_t^2(\theta_0)} \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \lambda_t(\theta_0)}{\partial \theta'} \right) \mu = 0$, that is $E\left(\frac{1}{\lambda_t^2(\theta_0)} \left(\mu' \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right)^2 \right) = 0$. The latter equality holds iff $\frac{1}{\lambda_t^2(\theta_0)} \left(\mu' \frac{\partial \lambda_t(\theta_0)}{\partial \theta} \right)^2 = 0$ a.s., which, in turn, is true iff $\mu' \frac{\partial \lambda_t(\theta_0)}{\partial \theta} = 0$ a.s. Now, by the stationarity of $\frac{\partial \lambda_t(\theta_0)}{\partial \theta}$ (i.e. $J(\theta_0) := E\left(\frac{1}{\lambda_{t-j}^2(\theta_0)} \frac{\partial \lambda_{t-j}(\theta_0)}{\partial \theta} \frac{\partial \lambda_{t-j}(\theta_0)}{\partial \theta'} \right)$ so if $J(\theta_0)$ is singular, then, $\mu' \frac{\partial \lambda_{t-j}(\theta_0)}{\partial \theta} = 0$ a.s.) we have

$$
\begin{aligned}
0 &= \mu' \tfrac{\partial \lambda_t(\theta_0)}{\partial \theta} \\
&= \mu' \left(1, Y_{t-1}^{-1}, ..., Y_{t-q}^{-1}, \lambda_{t-1}, ..., \lambda_{t-p}, X_{1,t-1}, ..., X_{r,t-1}\right) + \sum_{j=1}^{p} \beta_j \mu' \tfrac{\partial \lambda_{t-j}(\theta_0)}{\partial \theta} \\
&= \mu' \left(1, Y_{t-1}^{-1}, ..., Y_{t-q}^{-1}, \lambda_{t-1}, ..., \lambda_{t-p}, X_{1,t-1}, ..., X_{r,t-1}\right).
\end{aligned}
$$

It is, thus, clear that $\mu_1 = 0$, otherwise Y_{t-1}^{-1} would be measurable with respect to $\xi_{t-2}^{-1}, \xi_{t-3}^{-1}, ...$ (in fact Y_{t-1}^{-1} is measurable with respect to $\xi_{t-1}^{-1}, \xi_{t-2}^{-1}, ...$). The same argument shows that $\mu_{1+q} = 0$ and so on, to get $\mu_2 = ... = \mu_q = 0$ and $\mu_{2+q} = ... = \mu_{p+q} = 0$. By **A4**, a similar argument to show **Lemma 2** entails $\mu_{p+q+1} = ... = \mu_{p+q+r} = 0$. Thus, $\mu = 0$, unless there exists an ARCP$(p-1, q-1)$ representation which is impossible under **A4**. This establish the result.

Lemma 7 *Under the same assumptions of Theorem 4.2, there exists a neighborhood $\mathcal{V}(\theta_0)$ of θ_0 such that, for all $i, j, k \in \{1, ..., p+q+r+1\}$,*

$$E \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \tfrac{\partial^3 l_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| < \infty.$$

Proof Differentiating (12) we obtain

$$
\begin{aligned}
\tfrac{\partial^3 l_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} &= \left(\tfrac{Y_t \lambda_t(\theta)}{\mu_0} - 1 \right) \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial^3 \lambda_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} + \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_i} \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial^2 \lambda_t(\theta)}{\partial \theta_j \partial \theta_k} + \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t^2(\theta)}{\partial \theta_i \partial \theta_j} \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_k} \\
&\quad + \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t^2(\theta)}{\partial \theta_i \partial \theta_k} \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_j} - \tfrac{2}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_i} \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_j} \tfrac{1}{\lambda_t(\theta)} \tfrac{\partial \lambda_t(\theta)}{\partial \theta_k}. \qquad (22)
\end{aligned}
$$

Let us first show that $Y_t \lambda_t(\theta) = \xi_t \frac{\lambda_t(\theta)}{\lambda_t(\theta_0)}$ is uniformly integrable in a neighborhood $\mathcal{V}(\theta_0)$ of θ_0. Let Θ^* be a compact included in Θ, whose interior contains θ_0. For all $\delta > 0$ there exists a neighborhood $\mathcal{V}(\theta_0) \subset \Theta^*$ such that $\boldsymbol{\beta}_0 \leq (1+\delta) \boldsymbol{\beta}$. From (8) we have

$$\lambda_t(\theta) = \omega \sum_{k=0}^{\infty} \boldsymbol{\beta}^k (1,1) + \sum_{k=0}^{\infty} \boldsymbol{\beta}^k (1,1) \boldsymbol{\pi}' \mathbf{X}_{t-k-1} + \sum_{i=1}^{q} \alpha_i \sum_{k=0}^{\infty} \boldsymbol{\beta}^k (1,1) Y_{t-k-i}^{-1}.$$

Using the inequality $\frac{x}{1+x} \leq x^{\kappa}$ for all $x \geq 0$ and all $\kappa \in (0,1)$, and the fact that

$\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{1}{\alpha_i} < \infty$ (since $\mathcal{V}(\theta_0) \subset \Theta^*$) we obtain

$$\begin{aligned}
&\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\lambda_t(\theta)}{\lambda_t(\theta_0)} \\
\leq\ & \sup_{\theta\in\mathcal{V}(\theta_0)} \left(\frac{\omega}{\omega_0} \sum_{k=0}^{\infty} \boldsymbol{\beta}^k(1,1) + \sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)\boldsymbol{\pi}'\mathbf{X}_{t-k-1}}{\omega_0 + \boldsymbol{\beta}_0^k(1,1)\boldsymbol{\pi}_0'\mathbf{X}_{t-k-1}} + \sum_{i=1}^{q} \alpha_i \sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)Y_{t-k-i}^{-1}}{\omega_0 + \alpha_{0i}\boldsymbol{\beta}_0^k(1,1)Y_{t-k-i}^{-1}} \right) \\
\leq\ & M + \sup_{\theta\in\mathcal{V}(\theta_0)} \left(\sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)\|\boldsymbol{\pi}\|\|\mathbf{X}_{t-k-1}\|}{\omega_0 + \boldsymbol{\beta}_0^k(1,1)\|\boldsymbol{\pi}_0\|\|\mathbf{X}_{t-k-1}\|} + \sum_{i=1}^{q} \frac{\alpha_i}{\alpha_{0i}} \sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)}{\boldsymbol{\beta}_0^k(1,1)} \left(\frac{\alpha_i \boldsymbol{\beta}_0^k(1,1)Y_{t-k-i}^{-1}}{\omega_0} \right)^{\kappa} \right) \\
\leq\ & M + \sup_{\theta\in\mathcal{V}(\theta_0)} \left(\sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)\|\boldsymbol{\pi}\|}{\boldsymbol{\beta}_0^k(1,1)\|\boldsymbol{\pi}_0\|} \left(\frac{\boldsymbol{\beta}_0^k(1,1)\|\boldsymbol{\pi}_0\|\|\mathbf{X}_{t-k-1}\|}{\omega_0} \right)^{\kappa} + \sum_{i=1}^{q} \frac{\alpha_i}{\alpha_{0i}} \sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k(1,1)}{\boldsymbol{\beta}_0^k(1,1)} \left(\frac{\alpha_{0i}\boldsymbol{\beta}_0^k(1,1)Y_{t-k-i}^{-1}}{\omega_0} \right)^{\kappa} \right) \\
\leq\ & M + M \sum_{k=0}^{\infty} (1+\delta)^k \rho^{k\kappa} \|\mathbf{X}_{t-k-1}\|^{\kappa} + M \sum_{i=1}^{q} \frac{\alpha_i}{\alpha_{0i}} \sum_{k=0}^{\infty} (1+\delta)^k \rho^{k\kappa} Y_{t-k-i}^{-\kappa}. \qquad (23)
\end{aligned}$$

It is possible to choose κ, say $\kappa = \frac{1-\rho^{\kappa}}{2\rho^{\kappa}}$, so that $E\left(Y_t^{-\kappa}\right) < \infty$ and thus $E\left(\sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\lambda_t(\theta)}{\lambda_t(\theta_0)}\right) < \infty$. From (23), taking κ so that $E\left(Y_t^{-2\kappa}\right) < \infty$ and $E\|\mathbf{X}_t\|^{2\kappa} < \infty$, the triangle inequality yields

$$\begin{aligned}
&\left\| \sup_{\theta\in\mathcal{V}(\theta_0)} Y_t \lambda_t(\theta) \right\|_2 = \sigma \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\lambda_t(\theta)}{\lambda_t(\theta_0)} \right\|_2 \\
\leq\ & \sigma M + \sigma M \sum_{k=0}^{\infty} (1+\delta)^k \rho^{k\kappa} \left\| \|\mathbf{X}_{t-k-1}\|^{\kappa} \right\|_2 + Mq \sum_{k=0}^{\infty} (1+\delta)^k \rho^{k\kappa} \left\| Y_{t-k-i}^{-\kappa} \right\|_2 < \infty. \qquad (24)
\end{aligned}$$

Now, we consider the second factor in the first term of the right hand side of (22). Differentiating (19), (20), and (21), while using the same argument to show inequality (15), we get

$$\sup_{\theta\in\Theta^*} \frac{1}{\lambda_t(\theta)} \frac{\partial^3 \lambda_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \leq M,$$

where the derivative is taken with respect to at least one parameter different from the β_j. Similarly to (21), we have for all $\kappa \in (0,1)$

$$\begin{aligned}
\beta_{j_1}\beta_{j_2}\beta_{j_3} \frac{\partial^3 \lambda_t(\theta)}{\partial\beta_{j_1}\partial\beta_{j_2}\partial\beta_{j_3}} &\leq \sum_{k=3}^{\infty} k(k-1)(k-2)\boldsymbol{\beta}^k(1,1)\underline{\boldsymbol{c}}_{t-k}(1) \\
\sup_{\theta\in\Theta^*} \frac{1}{\lambda_t(\theta)} \beta_{j_1}\beta_{j_2}\beta_{j_3} \frac{\partial^3 \lambda_t(\theta)}{\partial\beta_{j_1}\partial\beta_{j_2}\partial\beta_{j_3}} &\leq M \sup_{\theta\in\Theta^*} \frac{1}{\omega^{\kappa}} \beta_{j_1}\beta_{j_2}\beta_{j_3} \sum_{k=3}^{\infty} k(k-1)(k-2)\rho^{k\kappa} \left(\sup_{\theta\in\Theta^*} \underline{\boldsymbol{c}}_{t-k}(1) \right)^{\kappa}.
\end{aligned}$$

Since by **A0** and Theorem 3.3, $E\left(\sup_{\theta\in\Theta^*}\underline{\boldsymbol{c}}_{t-k}(1)\right)^{2\kappa}<\infty$ for some $\kappa>0$, we thus obtain

$$E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial^3\lambda_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|^2<\infty. \tag{25}$$

More generally, it can be shown, using the same argument to establish (25), that

$$E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial^3\lambda_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|^d<\infty$$

for all integer d. In view of (24) and (25) and using the Cauchy-Schwarz inequality we obtain

$$E\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\left(\frac{Y_t\lambda_t(\theta)}{\mu_0}-1\right)\frac{1}{\lambda_t(\theta)}\frac{\partial^3\lambda_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<\infty.$$

The remaining terms in the right-hand side of (22) can be handled in a similar way. In particular, it can be easily shown using the latter argument that for all integer d,

$$E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial^2\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}\right|^d<\infty \text{ and } E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\right|^d<\infty. \tag{26}$$

By Hölder's inequality we finally get

$$E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_j}\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_k}\right|\leq\max_i E\sup_{\theta\in\Theta^*}\left|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\right|^3<\infty,$$

establishing the lemma.

Lemma 8 *Under **A1**-**A6***

$$i)\ \sqrt{n}\left\|\frac{\partial L_n(\theta_0)}{\partial\theta}-\frac{\partial\widetilde{L}_n(\theta_0)}{\partial\theta}\right\|\underset{n\to\infty}{\overset{p}{\to}}0,\quad ii)\ \sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\widetilde{L}_n(\theta)}{\partial\theta\partial\theta'}\right\|\underset{n\to\infty}{\overset{p}{\to}}0$$

for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0.

Proof i) From (9), the derivatives of $\widetilde{\underline{\lambda}}_t(\theta)$ are given by

$$\begin{aligned}
\frac{\partial\widetilde{\underline{\lambda}}_t(\theta)}{\partial\omega} &= \sum_{k=0}^{t-q-1}\boldsymbol{\beta}^k\underline{\mathbf{1}}+\sum_{k=t-q}^{t}\boldsymbol{\beta}^k\frac{\partial\widetilde{\underline{\boldsymbol{c}}}_{t-k}}{\partial\omega}\\
\frac{\partial\widetilde{\underline{\lambda}}_t(\theta)}{\partial\alpha_i} &= \sum_{k=0}^{t-q-1}\boldsymbol{\beta}^k\underline{\mathbf{Y}}_{t-k-i}+\sum_{k=t-q}^{t}\boldsymbol{\beta}^k\frac{\partial\widetilde{\underline{\boldsymbol{c}}}_{t-k}}{\partial\omega}\\
\frac{\partial\widetilde{\underline{\lambda}}_t(\theta)}{\partial\beta_j} &= \sum_{k=0}^{t-q-1}\sum_{i=1}^{k}\boldsymbol{\beta}^{i-1}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{k-i}\underline{\boldsymbol{c}}_{t-k}+\sum_{k=t-q}^{t}\boldsymbol{\beta}^k\\
\frac{\partial\widetilde{\underline{\lambda}}_t(\theta)}{\partial\pi_l} &= \sum_{k=0}^{t-q-1}\boldsymbol{\beta}^k\underline{\mathbf{X}}_{t-k-1}^{(l)}+\sum_{k=t-q}^{t}\boldsymbol{\beta}^k\frac{\partial\widetilde{\underline{\boldsymbol{c}}}_{t-k}}{\partial\omega}.
\end{aligned}$$

The second derivatives are computed in a similar way. Now, we have

$$\begin{aligned}
\frac{\partial l_n(\theta)}{\partial\theta_i}-\frac{\partial\widetilde{l}_n(\theta)}{\partial\theta_i} &= \left(\frac{Y_t\lambda_t(\theta)}{\mu_0}-1\right)\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}-\left(\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\frac{1}{\widetilde{\lambda}_t(\theta)}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_i}\\
&= \left(\frac{Y_t\lambda_t(\theta)}{\mu_0}-\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}\right)\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}+\left(\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\left(\frac{1}{\lambda_t(\theta)}-\frac{1}{\widetilde{\lambda}_t(\theta)}\right)\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\\
&\quad+\left(\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\frac{1}{\widetilde{\lambda}_t(\theta)}\left(\frac{\partial\lambda_t(\theta)}{\partial\theta_i}-\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_i}\right).
\end{aligned}$$

By (22), the compactness of Θ (cf. **A2**) and the fact that $\rho(\boldsymbol{\beta})<1$ (cf. **A1**) we have *a.s.*

$$\begin{aligned}
\left|\frac{1}{\widetilde{\lambda}_t(\theta)}-\frac{1}{\lambda_t(\theta)}\right| &\leq \frac{M\rho^t}{\lambda_t(\theta)},\quad \frac{\lambda_t(\theta)}{\widetilde{\lambda}_t(\theta)}\leq(1+M)\rho^t\\
\sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta}\right\| &< M\rho^t,\quad \sup_{\theta\in\Theta}\left\|\frac{\partial^2\lambda_t(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\widetilde{\lambda}_t(\theta)}{\partial\theta\partial\theta'}\right\|<M\rho^t.
\end{aligned}$$

Hence, using the latter inequalities we obtain

$$\left|\frac{\partial l_n(\theta_0)}{\partial\theta_i}-\frac{\partial\widetilde{l}_n(\theta_0)}{\partial\theta_i}\right|\leq M\rho^t\left(1+\frac{\xi_t}{\mu_0}\right)\left|1+\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_i}\right|$$

so

$$\sqrt{n}\left|\frac{\partial L_n(\theta_0)}{\partial\theta}-\frac{\partial\widetilde{L}_n(\theta_0)}{\partial\theta}\right|\leq\frac{M}{\sqrt{n}}\sum_{t=1}^n\rho^t\left(1+\frac{\xi_t}{\mu_0}\right)\left|1+\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_i}\right|.$$

Using (13) and Markov's inequality, it follows that for all $\varepsilon>0$,

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\rho^t\left(1+\frac{\xi_t}{\mu_0}\right)\left|1+\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_i}\right|>\varepsilon\right)\\
\leq\ &\frac{2}{\varepsilon}\left(1+E\left|1+\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta_i}\right|\right)\frac{1}{\sqrt{n}}\sum_{t=1}^n\rho^t\to0\text{ as }n\to\infty
\end{aligned}$$

from which part i) of the lemma follows.

ii) From (12) we have

$$\begin{aligned}
&\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial^2L_n(\theta)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\widetilde{L}_n(\theta)}{\partial\theta_i\partial\theta_j}\right\|\leq n^{-1}\sum_{t=1}^n\sup_{\theta\in\mathcal{V}(\theta_0)}\left|\left(\frac{Y_t\lambda_t(\theta)}{\mu_0}-\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}\right)\frac{1}{\lambda_t(\theta)}\frac{\partial^2\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}\right.\\
&+\left(\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\left[\left(\frac{1}{\lambda_t(\theta)}-\frac{1}{\widetilde{\lambda}_t(\theta)}\right)\frac{\partial^2\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}+\frac{1}{\widetilde{\lambda}_t(\theta)}\left(\frac{\partial\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}-\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_i\partial\theta_j}\right)\right]\\
&+2\left(\frac{Y_t\lambda_t(\theta)}{\mu_0}-\frac{Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}\right)\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_j}\\
&+\left(\frac{2Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\left[\left(\frac{1}{\lambda_t(\theta)}-\frac{1}{\widetilde{\lambda}_t(\theta)}\right)\frac{\partial\lambda_t(\theta)}{\partial\theta_i}+\frac{1}{\widetilde{\lambda}_t(\theta)}\left(\frac{\partial\lambda_t(\theta)}{\partial\theta_i}-\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_i}\right)\right]\\
&\left.+\left(\frac{2Y_t\widetilde{\lambda}_t(\theta)}{\mu_0}-1\right)\frac{1}{\widetilde{\lambda}_t(\theta)}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_i}\left[\left(\frac{1}{\lambda_t(\theta)}-\frac{1}{\widetilde{\lambda}_t(\theta)}\right)\frac{\partial\lambda_t(\theta)}{\partial\theta_j}+\frac{1}{\widetilde{\lambda}_t(\theta)}\left(\frac{\partial\lambda_t(\theta)}{\partial\theta_j}-\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\theta_j}\right)\right]\right|\\
&\leq Kn^{-1}\sum_{t=1}^n\rho^t\left[\sup_{\theta\in\mathcal{V}(\theta_0)}\left(\frac{Y_t\lambda_t(\theta)}{\mu_0}+1\right)\left(1+\frac{1}{\lambda_t(\theta)}\frac{\partial^2\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}+\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_j}\right)\right].
\end{aligned}$$

By (24), (26) and Hölder's inequality it is easily seen that there exists a neighborhood $\mathcal{V}(\theta_0)$ such that

$$E \sup_{\theta\in\mathcal{V}(\theta_0)} \left(\frac{Y_t\lambda_t(\theta)}{\mu_0}+1\right)\left(1+\frac{1}{\lambda_t(\theta)}\frac{\partial^2\lambda_t(\theta)}{\partial\theta_i\partial\theta_j}+\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_i}\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta_j}\right)<\infty.$$

Finally, the second part of the lemma follows from Markov's inequality. □

Lemma 9 *Under **A0-A6***

$$\sqrt{n}\frac{\partial L_n(\theta_0)}{\partial\theta} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0,\frac{\sigma_0^2}{\mu_0^2}J(\theta_0)\right).$$

Proof Note that in view of **Lemma 6**, $\sqrt{n}\frac{\partial L_n(\theta_0)}{\partial\theta}=\sum_{t=1}^{n}\frac{1}{\sqrt{n}}\frac{\partial l_t(\theta_0)}{\partial\theta}$ is a term of a square integrable ergodic Martingale with respect to its natural filtration. The ergodic theorem, thus, entails

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\xi_t}{\mu_0}-1\right)^2\frac{1}{\lambda_t^2(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\theta'} \underset{n\to\infty}{\overset{a.s.}{\to}} \frac{\sigma_0^2}{\mu_0^2}J(\theta_0),$$

and the result thus follows from the martingale central limit theorem (e.g. Billingsley, 1995). □

Lemma 10 *Under **A0-A6***

$$\frac{\partial^2 L_n(\theta^*)}{\partial\theta\partial\theta'} \underset{n\to\infty}{\overset{a.s}{\to}} J(\theta_0)$$

for any θ^* between $\widehat{\theta}_n$ and θ_0.

Proof A Taylor expansion of $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\theta_i\partial\theta_j}$ at θ_0 gives

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\theta_i\partial\theta_j}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}+\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta'}\left(\frac{\partial^2 l_t(\theta^{**})}{\partial\theta_i\partial\theta_j}\right)(\theta^{**}-\theta_0)$$

where θ^{**} is between θ^* and θ_0. In view of the strong convergence of θ^{**} to θ_0, the ergodic theorem and **Lemma 7**, we have a.s.

$$\begin{aligned}\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta'}\frac{\partial^2 l_t(\theta^{**})}{\partial\theta_i\partial\theta_j}\right\| &\le \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\mathcal{V}_k(\theta_0)}\left\|\frac{\partial}{\partial\theta'}\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|\\ &= E\sup_{\theta\in\mathcal{V}_k(\theta_0)}\left\|\frac{\partial}{\partial\theta'}\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j}\right\|<\infty.\end{aligned}$$

As $\|\theta^* - \theta_0\| \underset{n\to\infty}{\overset{a.s}{\to}} 0$, we obtain $\frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta'} \frac{\partial^2 l_t(\theta^{**})}{\partial\theta_i\partial\theta_j} (\theta^{**} - \theta_0) \underset{n\to\infty}{\overset{a.s}{\to}} 0$. Moreover, the ergodic theorem entails

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j} \underset{n\to\infty}{\overset{a.s}{\to}} J(\theta_0)(i,j),$$

from which the result follows while using Slutsky's lemma. □

8.6 Proof of Theorem 4.3

i) Let $U_t(\theta) = (Y_t\lambda_t(\theta) - \mu_0)^2$ and denote by $o_{a.s.}(1)$ a term converging almost surely to 0 as $n \to \infty$. If we can show that

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} U_t\left(\widehat{\theta}_n\right) = \frac{1}{n}\sum_{t=1}^{n} (Y_t\lambda_t(\theta_0) - \mu_0)^2 + o_{a.s.}(1)$$

then, the result (4.10) would follow from standard arguments. Now a Taylor expansion of $\frac{1}{n}\sum_{t=1}^{n} U_t(\widehat{\theta}_n)$ around θ_0 yields

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} U_t(\theta_0) + \left(\widehat{\theta}_n - \theta_0\right)\frac{1}{n}\sum_{t=1}^{n} \frac{\partial U_t(\theta^*)}{\partial\theta}$$

where θ^* is between $\widehat{\theta}_n$ and θ_0, and

$$\frac{\partial U_t(\theta)}{\partial\theta} = 2Y_t\lambda_t(\theta)(Y_t\lambda_t(\theta) - \mu_0)\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}.$$

From (24) and (26), we already have

$$E\left(\sup_{\theta\in\mathcal{V}(\theta_0)} Y_t\lambda_t(\theta)\right) < \infty,\ E\left(\sup_{\theta\in\mathcal{V}(\theta_0)} \left\|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|\right) < \infty,$$

for some neighborhood $\mathcal{V}(\theta_0)$ of θ_0. Hence, by the ergodic theorem, the consistency of $\widehat{\theta}_n$ (and hence of θ^*), and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\limsup_{n\to\infty}\left\|n^{-1}\sum_{t=1}^{n}\frac{\partial U_t(\theta^*)}{\partial\theta}\right\| &\leq \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{\partial U_t(\theta)}{\partial\theta}\right\| \\ &= E\left(\sup_{\theta\in\mathcal{V}(\theta_0)}|Y_t\lambda_t(\theta)(Y_t\lambda_t(\theta) - \mu_0)|\sup_{\theta\in\mathcal{V}(\theta_0)}\left\|\frac{1}{\lambda_t(\theta)}\frac{\partial\lambda_t(\theta)}{\partial\theta}\right\|\right) < \infty.\end{aligned}$$

Thus

$$\left(\widehat{\theta}_n - \theta_0\right) n^{-1} \sum_{t=1}^{n} \frac{\partial U_t(\theta^*)}{\partial \theta} = o_{a.s.}(1),$$

establishing (4.10).

ii) Set $V_t(\theta) = (Y_t \lambda_t(\theta) - \mu_0)^2 - \sigma_0^2 = U_t(\theta) - \sigma_0^2$ and denote by $o_p(1)$ a term converging in probability to 0 as $n \to \infty$. If we can show that

$$\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_t\left(\widehat{\theta}_n\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left((Y_t \lambda_t(\theta_0) - \mu_0)^2 - \sigma_0^2\right) + o_p(1) \tag{27}$$

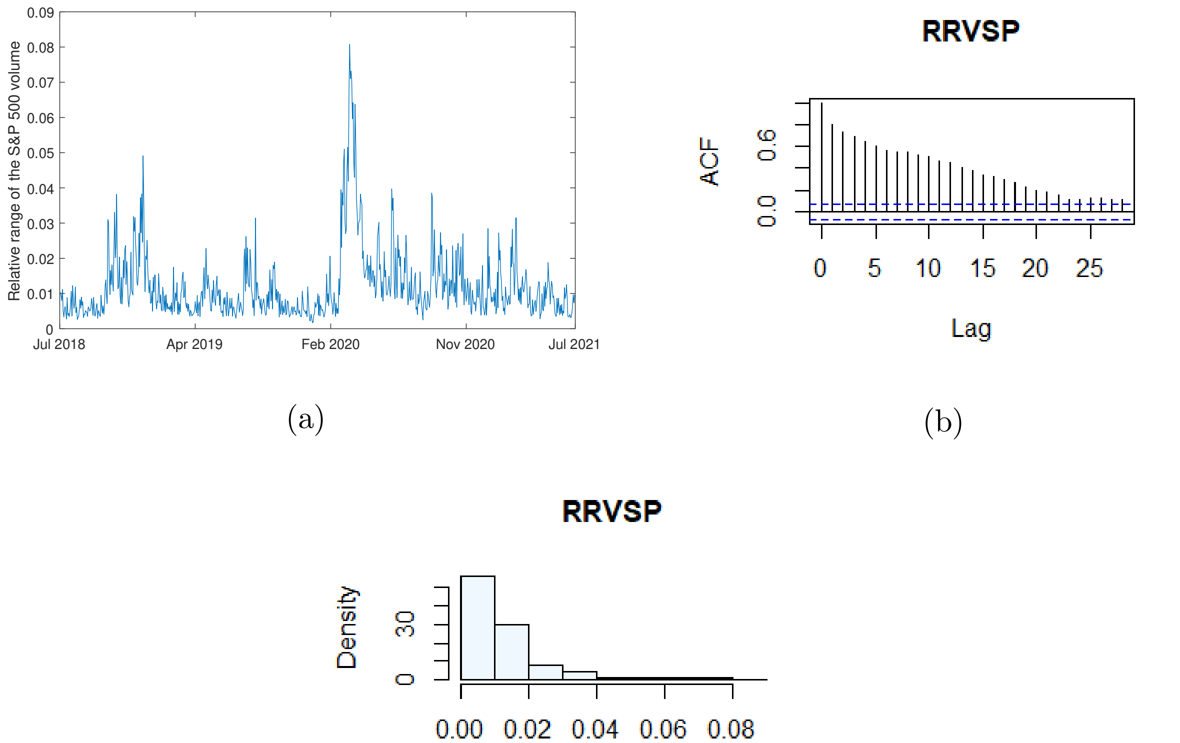
then, the result (4.10) would follow from the standard central limit theorem for iid sequences. A Taylor expansion of $\frac{1}{\sqrt{n}} \sum\limits_{t=1}^{n} V_t(\widehat{\theta}_n)$ around θ_0 yields

$$\begin{aligned} \sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_t(\theta_0) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial V_t(\theta^*)}{\partial \theta'} \left(\widehat{\theta}_n - \theta_0\right) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_t(\theta_0) + \frac{1}{n} \sum_{t=1}^{n} \frac{\partial U_t(\theta^*)}{\partial \theta'} \sqrt{n} \left(\widehat{\theta}_n - \theta_0\right) \end{aligned}$$

where θ^* is between $\widehat{\theta}_n$ and θ_0. By the $\sqrt{n}$-consistency of $\widehat{\theta}_n$ it follows that (27) holds. □

9 Supplementary material: Application to the relative range of the S&P 500 volume

We finally apply the ARCP model to the daily relative range of the S&P 500 volume (RRVSP). Specifically, the series is defined by $RRSP_t = \frac{H_t - L_t}{H_t}$, where H_t and L_t are, respectively, the high and low levels of the S&P 500 volume in day t. The relative range can be seen as a volatility measure (Chou, 2005). The dataset was obtained from Yahoo Finance, and runs from July, 2, 2018 to July, 9, 2021 with a total of $n = 761$ observations (Figure S.1).



(a) (b)

(c)

Figure S.1. (a) RRVSP series; (b) sample autocorrelation; (c) histogram.

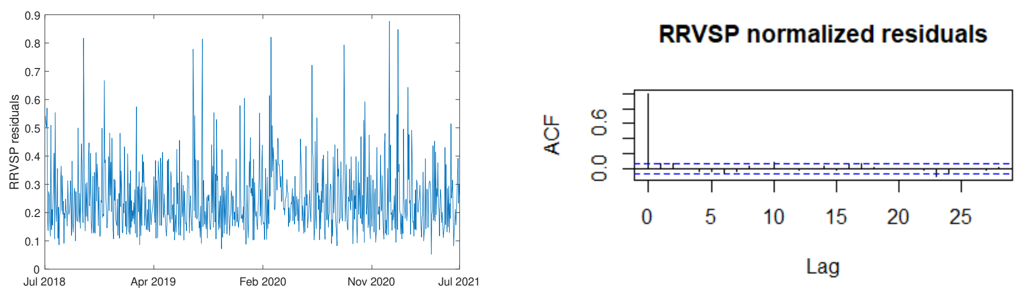
The best value of μ_0 was found to be $\mu_0 = 0.25$. The optimal (p, q) values were found to be (1,2), as they produced the largest log-likelihood (-172.1088) and the smallest BIC (370.7032) and AIC (352.2176). We, thus, estimate an ARCP$(1, 2)$ model with initial parameter value $\theta^{(0)} = \left(\omega^{(0)}, \alpha_1^{(0)}, \beta_1^{(0)}\right)' = (0.1, 0.1, 0.1, 0.1)'$ and starting values $\lambda_0 = \lambda_{-1} = \omega^{(0)}$ and $Y_0 = 1/\lambda_0$. The mean-inverted persistence parameter in Table S.1 was found to be $Persi = 0.9627$ (see. Table S.1), implying a strong persistence, while the estimated mean innovation

$\frac{1}{n}\sum_{t=1}^{n}\widehat{\xi}_t = 0.2526$ is very close to $\mu_0 = E(\xi_t) = 0.25$.

$\widehat{\omega}$	$\widehat{\alpha}_{1n}$	$\widehat{\beta}_{1n}$	$\widehat{\beta}_{2n}$	$\widehat{\sigma}_n^2$	$\widehat{\phi}_n$	$Persi$	MAFE	MSFE
1.1606 (0.0239)	0.1069 (0.0175)	0.2785 (0.0208)	0.1646 (0.0046)	0.0147 (0.0019)	11.7373	0.9627	0.0042	0.00003

Table S.1. EQML and WLS estimates for the ARCP$(1,2)$ model; RRVSP series.

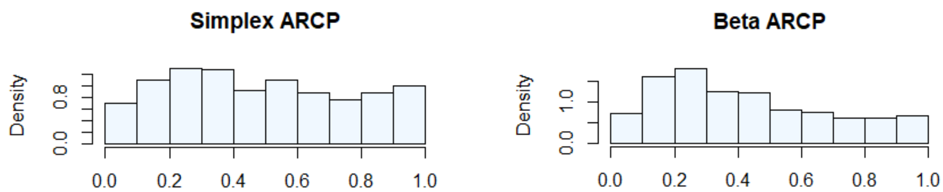
The MAFE and MSFE are respectively 0.0042 and 0.00003. The normalized residual series $(\widehat{\xi}_t)$ and its sample autocorrelation, in Figure S.2, are in favour of the iid assumption for the innovation sequence.



(a) (b)

Figure S.2. (a) Normalized residuals; (b) residual sample autocorrelation.

The PIT of the normalized residuals $(\widehat{\xi}_t)$ (see Figure S.3) with respect to the Beta distribution $\mathcal{B}eta\left(\widehat{\phi}_n\mu_0, \widehat{\phi}_n(1-\mu_0)\right)$ (with $\widehat{\phi}_n = 11.1850$, see Table S.1) does not favor the usage of this distribution for the RRVSP data. In contrast, the residual PIT with respect to the Simplex distribution $S^-(\mu_0, 2.3)$ suggests a quite good fit of the latter distribution to the RRVSP data.



(a) (b)

Figure S.3. PIT of the ARCP residuals with respect to: (a) the simplex distribution, (b) the beta distribution. RRVSP series.

We fitted a linear conditional-mean Beta observation driven (BOD) model given by (5.1) to the RRVSP series. The MLE is

$$\widehat{g}_n = (0.0013, 0.4423, 0.4533, 462.9364)',$$

with persistence parameter 0.8956, $MAFE = 0.0042$ and $MSFE = 0.00003$. Figure S.4 shows the RRVSP series together with the conditional means generated by the ARCP and BOD models. The ARCP and BOD models have visibly similar conditional mean behaviors.

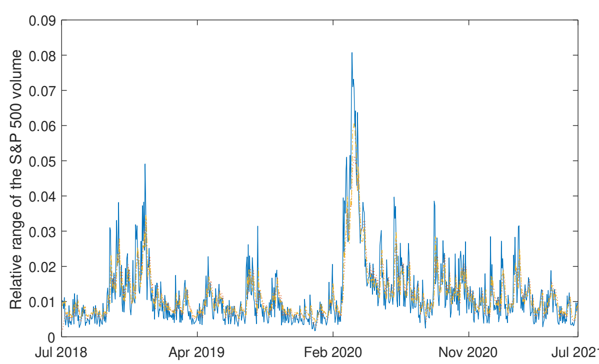


Figure S.4. RRVSP series and its estimated conditional means generated by the ARCP and the BOD models.

Figure S.5 displays the PIT of the series RRVSP with respect to the BOD distribution given by (5.1). By inspecting this figure we conclude that the BOD distribution is not really a proper choice for the RRVSP series.

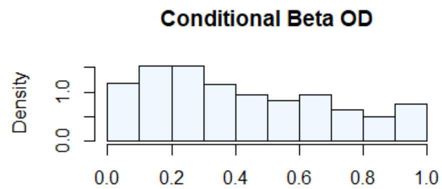


Figure S.5. PIT of the RRVSP series with respect to the Beta conditional distribution.

Finally, in terms of the out-of-sample forecasting performance, both the ARCP$(1, 2)$ and the BOD$(1, 1)$ yielded very similar MSFE and MAFE values (Table S.2).

	n_f	300	400	500	600	640
ARCP$(1, 2)$	MSFE	0.00005	0.0006	0.00004	0.00001	0.00002
	MAFE	0.0045	0.0051	0.0042	0.0035	0.0035
BOD$(1, 1)$	MSFE	0.00004	0.00005	0.00004	0.00002	0.00002
	MAFE	0.0044	0.0050	0.0042	0.0036	0.0035

Table S.2. Out-of-sample forecasting performances of ARCP and BOD; RRVSP series.

References

[1] Berkes, I., Horvath, L. and Kokoskza, P. (2003). GARCH processes: Structure and estimation. *Bernoulli*, 9, 201-227.

[2] Billingsley, P. (1995). *Probability and measure*. John Wiley and Sons, second edition.

[3] Bougerol, P. and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes. *Annals of Probability*, **20**, 1714-1730.

[4] Bougerol, P. and Picard, N. (1992b). Stationarity of GARCH processes and some nonnegative time series. *Journal of Econometrics*, **52**, 115-127.

[5] Chou, R.Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) Model. *Journal of Money, Credit, and Banking* 37: 561–582.

[6] Francq, C. and Zakoïan, J.M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, **10**, 605-637.

[7] Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition, John Wiley & Sons.

[8] Pan, J., Wang, H. and Tong, H. (2008). Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics,* **142**, 352-378.