Periodic autoregressive conditional duration

Aknouche, Abdelhakim and Almohaimed, Bader and Dimitrakopoulos, Stefanos

University of Science and Technology Houari Boumediene, Qassim University, Leeds University

8 July 2020

Online at https://mpra.ub.uni-muenchen.de/101696/
MPRA Paper No. 101696, posted 15 Jul 2020 10:22 UTC
Periodic autoregressive conditional duration

Abdelhakim Aknouche*, Bader Almohaimeed**, and Stefanos Dimitrakopoulos***

*Department of Mathematics, College of Science, Qassim University (Saudi Arabia) & Faculty of Mathematics, University of Science and Technology Houari Boumediene (Algeria)
**Department of Mathematics, College of Science, Qassim University, Saudi Arabia.
***Division of Economics, Leeds University, UK

Abstract

We propose an autoregressive conditional duration (ACD) model with periodic time-varying parameters and multiplicative error form. We name this model periodic autoregressive conditional duration (PACD). First, we study the stability properties and the moment structures of it. Second, we estimate the model parameters, using (profile and two-stage) Gamma quasi-maximum likelihood estimates (QMLEs), the asymptotic properties of which are examined under general regularity conditions. Our estimation method encompasses the exponential QMLE, as a particular case. The proposed methodology is illustrated with simulated data and two empirical applications on forecasting Bitcoin trading volume and realized volatility. We found that the PACD produces better in-sample and out-of-sample forecasts than the standard ACD.

Keywords: Positive time series, autoregressive conditional duration, periodic time-varying models, multiplicative error models, exponential QMLE, two-stage Gamma QMLE.
1 Introduction

Recent research in time series analysis tends to avoid transforming original data prior to modeling and prefers to represent them directly through models that take into account the actual support of their distributions. Such an approach parallels to that of generalized linear models (GLM) for independent data (McCullag and Nelder, 1989). In this way, numerous time series models with “specific values” have, recently, received great interest, such as integer-valued models, including count and binary specifications, and positive-valued models.

A well-known model for positive-valued time series data is the autoregressive conditional duration (ACD), introduced by Engel and Russell (1998). Originally designed to model durations between financial events in high-frequency microstructure markets, the ACD model is also useful for modeling a broad range of data, such as regularly-spaced return range series (Chou, 2005), daily realized volatility (Lanne, 2006; Zheng et al, 2015; Aknouche and Francq, 2019) and trading volume (Li, 2019; Aknouche and Francq, 2020). Various generalizations of the ACD model have been proposed to take into account additional facts of positive time series data (Pacurar, 2008; Hautsch, 2012; Bhogal and Variyam, 2019).

As in the case of GARCH models, it has been documented that the high persistence observed in empirical studies utilizing the standard ACD specification, is in fact artificial and can be avoided by considering ACD models with time-varying parameters (Diebold, 1986; Andersen and Bollerslev, 1997; Mikosch and Starica, 2004; Hejer and Veltic 2007; Caporin et al, 2017; Gallo and Ortanto, 2018). In this paper, we extend the literature on time-varying ACD models, by proposing an ACD model, the parameters of which are allowed to evolve periodically over time. We name this model periodic autoregressive conditional duration (PACD).

Such a model aims to represent seasonally varying positive-valued series. The observed process is defined as the product of a unit mean independent and periodically distributed (henceforth ipds) innovation process with the conditional mean of the model having a GARCH-type specification with periodic time-varying parameters. We first study the stability properties of the PACD model, such as the existence of periodically stationary and ergodic solutions with finite moments or log-moments. Such properties are needed in the estimation stage, which is the second contribution of this paper.

To estimate the model parameters, the exponential quasi-maximum likelihood (EQMLE)
is used, since it is well-adapted to the support of the distribution of the data, and it does not require specifying a distribution for the periodically distributed innovation sequence. However, because of the periodicity of that sequence, the EQMLE may be less efficient than the Gamma QMLE (GQMLE) which, in fact, accounts for the periodicity of the model innovation.

Consequently, we propose a two-stage Gamma QMLE (2S-GQMLE) which i) utilizes the EQMLE (or a profile GQMLE) in the first stage, ii) estimates the variance innovations, and then iii) uses the latter as a by-product in the second stage of the computation of the GQMLE. Consistency and asymptotic normality (CAN) of the proposed QMLEs are established and the relative efficiency of the 2S-GQMLE is studied for some specific conditional distributions.

The PACD can be used to model various seasonal positive-valued phenomena (realized volatility, trading volumes and transaction rates). The day-of-the-week pattern may be present in all these phenomena, which means that each day of the week may have its own distribution (Franses and Paap, 2000; Boynton et al, 2009; Tsiakas, 2006; Charles, 2010). In that sense, a time-invariant ACD model for daily data is just an average model that does not take into account the specificities of the underlying measures across days. Other examples of non-financial intraday series that may be characterized by periodicity are wind power and wind speed series (Ambach and Croonenbroeck, 2015; Ambach and Schmid, 2015; Ziel et al, 2016).

Our empirical applications concern Bitcoin trading volume data and the UN realized volatility. Both series are characterized by the day-of-the-week effect and we show that the PACD produces better in-sample and out-of-sample forecasts than the benchmark ACD.

The rest of this paper is outlined as follows. In Section 2 we define the PACD and some special cases of it, and describe the link/relationship between the PACD and the periodic GARCH of Bollerslev and Ghysels (1996). In Section 3 we derive the stability conditions of our model. In Section 4, various Gamma QMLEs are proposed and their asymptotic properties are studied. In section 5 we conduct a simulation study and in section 6 we present the empirical results from two series (Bitcoin trading volume and UN realized volatility). All the proofs are given in the Appendix. A Supplementary material accompanies this paper.
2 Periodic Autoregressive Conditional Duration model

All random variables and processes in this paper are defined on a probability space \((Ω, \mathcal{F}, P)\) and valued in the set of positive real numbers \(\mathbb{R}_+ = (0, \infty)\), which is endowed with the Borel field \(\mathcal{B}(\mathbb{R}_+)\). Let \(S \geq 1\) be a positive integer called the period, and \(\omega_t, \alpha_{t1}, \ldots, \alpha_{tq}, \beta_{t1}, \ldots, \beta_{tp}\) \((p, q \in \mathbb{N} = \{0, 1, \ldots\})\) be positive real parameters \(S\)-periodic over time, i.e. \(\omega_t = \omega_{t+kS}, \alpha_{ti} = \alpha_{t+kS,i} (i = 1, \ldots, q)\) and \(\beta_{tj} = \beta_{t+kS,j} (j = 1, \ldots, p)\) for all integers \(k\) and \(t\). Let also \(\{\xi_t, t \in \mathbb{Z}\}\) be a sequence of positive random variables with \(E(\xi_t) = 1\) for all \(t\), and a finite \(\text{Var}(\xi_t) = \sigma^2_t > 0\).

Assume that \(\{\xi_t, t \in \mathbb{Z}\}\) is iid in the sense that \(\xi_t \overset{D}{=} \xi_{t+S}\) for all \(t\), where \(\overset{D}{=}\) denotes equality in distribution.

A positive-valued stochastic process \(\{Y_t, t \in \mathbb{Z}\}\) is said to be a MEM (multiplicative error model; Engle, 2002) periodic autoregressive conditional duration with orders \(p\) and \(q\) (henceforth \(\text{PACD}(p, q)\)) if \(Y_t\) is given for all \(t \in \mathbb{Z}\) by

\[
Y_t = \psi_t \xi_t \quad (2.1a)
\]

and

\[
\psi_t = \omega_t + \sum_{i=1}^{q} \alpha_{ti} Y_{t-i} + \sum_{j=1}^{p} \beta_{tj} \psi_{t-j} \quad (2.1b)
\]

where the innovation term \(\xi_t\) is independent of \(\psi_{t-j}\) for all \(j \geq 1\). To ensure the almost sure (a.s.) positivity of \(\psi_t\), it is assumed that \(\omega_t > 0, \alpha_{ti} \geq 0, \text{ and } \beta_{tj} \geq 0\), for all \(t \in \mathbb{Z}, i = 1, \ldots, q\) and \(j = 1, \ldots, p\). To emphasize the periodicity of the model, let \(t = nS + v\) for \(n \in \mathbb{Z}\) and \(1 \leq v \leq S\). Then, equation (2.1b) can be written as follows

\[
\psi_{v+nS} = \omega_v + \sum_{i=1}^{q} \alpha_{vi} Y_{v-i+nS} + \sum_{j=1}^{p} \beta_{vj} \psi_{v-j+nS}, \quad n \in \mathbb{Z}, \ 1 \leq v \leq S,
\]

where by season or channel \(v\) \((1 \leq v \leq S)\) we denote the set \(\{\ldots, v-S, v, v+S, v+2S, \ldots\}\) with corresponding parameters \(\omega_v, \alpha_{vi}, \beta_{vi}\) and \(\sigma^2_v = \text{Var}(\xi_{v+nS})\). Let \(\mathcal{F}_t\) be the \(\sigma\)-Algebra generated by \(\{Y_{t-i}, i \geq 0\}\). The conditional mean and conditional variance of the model (2.1) are given respectively by

\[
E(Y_t|\mathcal{F}_{t-1}) = \psi_t \quad (2.2a)
\]
and

$$Var(Y_t|\mathcal{F}_{t-1}) = \sigma_t^2 \psi_t^2.$$  \hfill (2.2b)$$

The PACD model, thus, follows the quadratic variance-to-mean relationship (i.e. the conditional variance is proportional to the squared conditional mean), where $\sigma_t^2 > 0$ is the variance of $\xi_t$ and is $S$-periodic by construction (from the $ipd_S$ property of the innovation sequence $\{\xi_t, t \in \mathbb{Z}\}$). The specification (2.1) is a multiplicative error model (MEM) in the sense of Engle (2002), but the conditional mean equation (2.1b) has rather periodic time-varying coefficients. For $S = 1$, model (2.1) reduces to the standard autoregressive conditional duration (ACD in short) of Engel and Russell (1998). No specification for the distribution of $\{\xi_t, t \in \mathbb{Z}\}$ is imposed apart the semiparametric quadratic variance-to-mean function (2.2b). However, a useful family of conditional distributions satisfying (2.2b) is the Gamma distribution with shape $\frac{1}{\sigma_t^2}$ and scale $\frac{1}{\sigma_t \psi_t}$; that is

$$Y_t|\mathcal{F}_{t-1} \sim \Gamma \left( \frac{1}{\sigma_t^2}, \frac{1}{\sigma_t \psi_t} \right),$$  \hfill (2.3)$$

where $\psi_t$ satisfies (2.1b). In the latter case, the innovation term $\xi_t$ in (2.1) will be marginally Gamma distributed

$$\xi_t \sim \Gamma \left( \frac{1}{\sigma_t^2}, \frac{1}{\sigma_t^2} \right),$$  \hfill (2.4)$$

and the process defined by (2.3) is called Gamma PACD($p,q$). A notable particular case of model (2.3) appears when the variance $\sigma_t^2 \equiv 1$ is constant, so $\xi_t \sim \Gamma (1,1)$, which corresponds to the exponential PACD. As in the time-invariant case, the periodic ACD model can be seen as a squared periodic GARCH (PGARCH) model as proposed by Ghysels and Bollerslev (1996). Indeed, consider the following real-valued PGARCH($p,q$) process given by

$$X_t = \sqrt{h_t} \eta_t$$  \hfill (2.5a)$$

and

$$h_t = \omega_t + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$  \hfill (2.5b)$$

where $\{\eta_t, t \in \mathbb{Z}\}$ is an $ipd_S$ sequence with mean zero and unit variance, and the parameters $\omega_t$, $\alpha_i$, and $\beta_j$ are defined as above. It is clear that the squared PGARCH process defined by $Y_t = X_t^2$ ($t \in \mathbb{Z}$) satisfies the PACD equation (2.1) with $\xi_t = \eta_t^2$ and $\psi_t = h_t$. Conversely,
let \( \{Y_t, t \in \mathbb{Z}\} \) be a PACD model given by (2.1), and assume \( \{z_t, t \in \mathbb{Z}\} \) is an independent and identically distributed (iid) sequence uniformly distributed in \( \{-1, 1\} \) (see also Francq and Zakoian, 2019 for the non-periodic case \( S = 1 \)). Assume \( \{z_t, t \in \mathbb{Z}\} \) and \( \{\xi_t, t \in \mathbb{Z}\} \) are independent and define the process \( \{X_t, t \in \mathbb{Z}\} \) by

\[
X_t = z_t \sqrt{Y_t} = \sqrt{h_t} \eta_t
\]

where \( h_t = \psi_t \) satisfies (2.5b) and \( \eta_t = z_t \sqrt{\xi_t} \) is a term of an ipd sequence. Hence \( \{X_t, t \in \mathbb{Z}\} \) is a PGARCH model in the sense of (2.5). Note finally that a PACD model admits a weak periodic ARMA (PARMA) (Lund and Basawa, 2000; Francq et al, 2011). Setting \( Y_t = \psi_t + \varepsilon_t \), the process \( \{Y_t, t \in \mathbb{Z}\} \) may be written in the following PARMA

\[
Y_t = \omega_t + \sum_{i=1}^{\max(p,q)} (\alpha_{ti} + \beta_{tj}) Y_{t-i} + \varepsilon_t - \sum_{j=1}^{p} \beta_{tj} \varepsilon_{t-j}
\]

where

\[
\varepsilon_t = Y_t - E(Y_t|\mathcal{F}_{t-1}) = \psi_t (\xi_t - 1)
\]

is a zero-mean term of a martingale difference sequence with a finite periodic variance \( E(\varepsilon_t^2) = E(\psi_t^2) E(\xi_t - 1)^2 = E(\psi_t^2) \sigma_t^2 \).

A more general PACD, which is not necessarily MEM is defined through a conditional distribution of the form

\[
Y_t|\mathcal{F}_{t-1} \sim F_{\psi_t}
\]

where \( F_{\psi} \) is a cumulative probability distribution (with positive support) with mean \( \psi \), and \( \psi_t \) is given by (2.1b).

### 3 Periodic ergodicity and finite moment conditions

We now give necessary and/or sufficient conditions for model (2.1) to be strictly periodically stationary and \textit{periodically ergodic}. Such properties are recalled in the Supplementary material. We also consider conditions for the existence of finite moments. Combining (2.1a) and (2.1b)
we obtain the following stochastic recurrence equation (SRE)

\[ Y_t = A_t Y_{t-1} + B_t \]  

(3.1)
driven by the ipd* sequence \{(A_t, B_t), t \in \mathbb{Z}\}, where \( Y_t = (Y_t, ..., Y_{t-q+1}, \psi_t, ..., \psi_{t-p+1})' \), \( B_t = (\omega_t, 0_{(q-1)\times 1}, \omega_t, 0_{(p-1)\times 1})' \), and

\[
A_t = \begin{pmatrix}
\alpha_{t1} \xi_t & \cdots & \alpha_{t,q-1} \xi_t & \alpha_{tq} \xi_t & \beta_{t1} \xi_t & \cdots & \beta_{t,p-1} \xi_t & \beta_{tp} \xi_t \\
1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\
\alpha_{t1} & \cdots & \alpha_{t,q-1} & \alpha_{tq} & \beta_{t1} & \cdots & \beta_{t,p-1} & \beta_{tp} \\
0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix},
\]

\(0_{m \times n}\) being the null matrix of dimension \(m \times n\). Let

\[ \gamma^S = \inf \left\{ \frac{1}{n} E \log \|A_nS...A_2A_1\| \; : \; n \geq 1 \right\} \]

be the top Lyapunov exponent associated with the ipd*-driven SRE (3.1) (Aknouche et al., 2020). Let also

\[
\beta_t = \begin{pmatrix}
\beta_{t1} & \cdots & \beta_{t,p-1} & \beta_{tp} \\
1 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 1 & 0
\end{pmatrix},
\]

and denote by \(\rho(A)\) the spectral radius of the squared matrix \(A\), i.e. the maximum modulus of the eigenvalues of \(A\). The following result gives the conditions for equation (3.1) to have a unique strictly periodically stationary and periodically ergodic solution.

**Theorem 3.1** i) Assume \(E(\log(\xi_v)) < \infty\) for all \(1 \leq v \leq S\). A necessary and sufficient condition for model (2.1) to have a unique nonanticipative strictly periodically stationary and
periodically ergodic solution is that
\[ \gamma^S < 0. \] (3.2)

Such a solution is given for all \( t \in \mathbb{Z} \) by
\[ Y_t = \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}, \] (3.3)

where the series in the right hand side of (3.3) converges absolutely almost surely.

ii) If (2.1) admits a strictly periodically stationary solution then
\[ \rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1. \] (3.4)

In the special case where \( p = q = 1 \), the periodic stationarity condition (3.2) is simplified as follows
\[ \sum_{v=1}^{S} E \left( \log (\alpha_v \xi_{v-1} + \beta_v) \right) < 0, \]

while (3.4) reduces to
\[ \prod_{v=1}^{S} \beta_v < 1. \]

Conditions for the existence of moments of the \( PACD(p,q) \) process are given as follows.

**Theorem 3.2** Assume \( E(\xi_v) < \infty \) for all \( 1 \leq v \leq S \). A sufficient condition for the process given by (2.1) to be strictly periodically stationary and periodically ergodic with \( E(Y_t) < \infty \) is that
\[ \rho \left( \prod_{v=0}^{S-1} E(A_{S-v}) \right) < 1. \] (3.5)

Some remarks are in order:
- In the case, where \( S = 1 \), the conditional mean coefficients are time-invariant, that is \( \omega_{tj} = \omega, \alpha_{tj} = \alpha_j \) and \( \beta_{tj} = \beta_j \). Therefore, using a similar device by Chen and An (1998), (3.5) reduces to the following stationarity in mean condition
\[ \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \]
as provided by Engle and Russell (1998).
- When \( p = q = 1 \), the periodic stationarity in mean condition (3.5) is equivalent to the
following condition

$$\prod_{v=1}^{S} (\alpha_v + \beta_v) < 1. \quad (3.6)$$

**Theorem 3.3** i) Under (3.2) there exists $\kappa > 0$ such that for all $1 \leq v \leq S$

$$E(\psi_v^\kappa) < \infty \quad \text{and} \quad E(Y_v^\kappa) < \infty. \quad (3.7)$$

ii) Let $\{Y_t, t \in \mathbb{Z}\}$ be a strictly periodically stationary solution of (2.1) and assume that $E(\xi_v^m) (m \in \mathbb{N}^*)$ is finite for all $1 \leq v \leq S$. A sufficient condition for $E(Y_v^m)$ to be finite (for all $1 \leq v \leq S$) is that

$$\rho \left( \prod_{v=0}^{S-1} E \left( A_{S-v}^{\otimes m} \right) \right) < 1 \quad (3.8)$$

where $A^{\otimes m}$ is the Kronecker product: $A \otimes A \otimes \cdots \otimes A$ with $m$ factors.

In the special case of Gamma PACD with $p = q = 1$, explicit conditions equivalent to (3.8) can be given. These conditions are also necessary for the existence of finite moments.

**Proposition 3.1** The Gamma PACD(1,1) model (2.3) admits a unique nonanticipative periodically ergodic solution $\{Y_t, t \in \mathbb{Z}\}$ such that:

i) $E(Y_v) < \infty$ (1 $\leq v \leq S$) if and only if (3.6) holds.

ii) $E(Y_v^2) < \infty$ (1 $\leq v \leq S$) if and only if $E(\xi_v^2) < \infty$ (1 $\leq v \leq S$), (3.6) and

$$\prod_{v=1}^{S} (\alpha_v^2 E(\xi_{v-1}^2) + 2\alpha_v\beta_v + \beta_v^2) < 1. \quad (3.9)$$

iii) $E(Y_v^3) < \infty$ (1 $\leq v \leq S$) if and only if $E(\xi_v^3) < \infty$ (1 $\leq v \leq S$), (3.6), (3.9) and

$$\prod_{v=1}^{S} \left( E(\xi_{v-1}^3) \alpha_v^3 + 3(\sigma_{v-1}^2 + 1)\alpha_v^2\beta_v + 3\alpha_v\beta_v^2 + \beta_v^3 \right) < 1. \quad (3.10)$$

iv) $E(Y_v^4) < \infty$ (1 $\leq v \leq S$) if and only if $E(\xi_v^4) < \infty$ (1 $\leq v \leq S$), (3.6), (3.9), (3.10) and the following hold

$$\prod_{v=1}^{S} \left( E(\xi_{v-1}^4) \alpha_v^4 + 4(1 + \sigma_{v-1}^2)(1 + 2\sigma_{v-1}^2)\alpha_v^3\beta_v + 6(1 + \sigma_{v-1}^2)\alpha_v^2\beta_v^2 + 4\alpha_v\beta_v^3 + \beta_v^4 \right) < 1. \quad (3.11)$$

For the particular exponential PACD(1,1) model, $Y_t | \mathcal{F}_{t-1} \sim \Gamma \left( 1, \frac{1}{\psi_t} \right)$, just replace in
Proposition 3.1 the moments $E(\xi^2_v)$, $E(\xi^3_v)$ and $E(\xi^4_v)$ by 1, 6 and 24 respectively, and $\sigma^2_v$ by 1 for all $1 \leq v \leq S$.

4 Gamma quasi-maximum likelihood estimates

Let $Y_1, Y_2, ..., Y_T$ be a series generated from the PACD($p, q$) model, which we can rewrite in the following form

$$Y_{nS+v} = \psi_{nS+v}\xi_{nS+v},$$

$$\psi_{nS+v} = \psi_{nS+v}(\theta_0) = \omega^0_v + \sum_{i=1}^{q} \alpha^0_v Y_{nS+v-i} + \sum_{j=1}^{p} \beta^0_v \psi_{nS+v-j}, \quad 1 \leq v \leq S, \ n \in \mathbb{Z} \quad (4.1)$$

where the true parameter $\theta_0 = (\theta^0_1, \theta^0_2, ..., \theta^0_S)'$ with $\theta^0_v = (\omega^0_v, \alpha^0_v, \beta^0_v)' (1 \leq v \leq S)$ belongs to a parameter space $\Theta \subset \left( (0, \infty) \times [0, \infty)^{(p+q)} \right)^S$. The true innovation variance parameter $\sigma^2_0 = (\sigma^2_{01}, ..., \sigma^2_{0S})'$ with $\sigma^2_{0v} = Var(\xi_{nS+v}) (1 \leq v \leq S)$ also belongs to a parametric space $\Delta \subset \mathbb{R}^S_+$. The sample size $T = NS (N \geq 1)$ is assumed without loss of generality a multiple of $S$. Given initial values $Y_0, ..., Y_{1-q}, \tilde{\psi}_0, ..., \tilde{\psi}_{1-p}$ and a generic parameter $\theta \in \Theta$ define

$$\tilde{\psi}_{nS+v}(\theta) = \omega_v + \sum_{i=1}^{q} \alpha_v Y_{nS+v-i} + \sum_{j=1}^{p} \beta_v \tilde{\psi}_{nS+v-j}(\theta), \ 1 \leq v \leq S, \ n \geq 0, \quad (4.2a)$$

as an observable proxy for $\psi_{nS+v}(\theta)$. The latter is defined as a periodically stationary solution of the following generic model ($\theta \in \Theta$)

$$\psi_{nS+v}(\theta) = \omega_v + \sum_{i=1}^{q} \alpha_v Y_{nS+v-i} + \sum_{j=1}^{p} \beta_v \psi_{nS+v-j}(\theta), \ 1 \leq v \leq S, \ n \in \mathbb{Z}. \quad (4.2b)$$

4.1 Exponential and profile Gamma QMLEs

The true conditional distribution of (4.1) is unknown due to the unpecification of the law of $\xi_v (1 \leq v \leq S)$. Thus, a quasi-maximum likelihood estimate (QMLE) which does not require any precise knowledge of the conditional distribution is suitable for estimating the parameter $\theta_0$ involved in the conditional mean. Among many possible QMLEs, the one computed on the basis of the exponential distribution (EQMLE in short) is especially useful for positive duration data because it reduces to the maximum likelihood estimate when $\xi_v$ is exponentially
distributed (Aknouche and Francq, 2020). A more general QMLE, which can be more efficient than the EQMLE in the periodic time-varying innovation context is the one computed on the basis of the Gamma distribution with arbitrary fixed variance parameters. Let \((\sigma^2_t)\) be fixed known positive numbers, \(S\)-periodic over \(t\), i.e. \(\sigma^2_{t+kS} = \sigma^2_t\), for all \(k \in \{0, ..., N - 1\}\). The profile Gamma likelihood associated with \(\sigma^2 = (\sigma^2_1, ..., \sigma^2_S)^\prime > 0\) is given, ignoring constants, by

\[
\tilde{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{l}_t(\theta),
\]

(4.3a)

\[
\tilde{l}_t(\theta) = \frac{1}{\sigma^2_t} \left( \frac{Y_t}{\nu(\theta)} + \log \tilde{\psi}_t(\theta) \right), \quad t \geq 1.
\]

(4.3b)

The profile Gamma QMLE (GQMLE) \(\hat{\theta}_G\) of \(\theta_0\) is, then, the minimizer of \(\tilde{L}_T(\theta)\) over \(\Theta\),

\[
\hat{\theta}_G = \arg\min_{\theta \in \Theta} \tilde{L}_T(\theta).
\]

(4.4)

When \(\sigma^2 = (1, ..., 1)^\prime\), the GQMLE defined by (4.4) reduces to the EQMLE and is denoted by \(\hat{\theta}_E\) (Aknouche and Francq, 2020).

Let \(\gamma^S(A^0)\) be the top Lyapunov exponent associated with \((A^0_t, t \in \mathbb{Z})\) where the matrix \(A^0_t\) is just \(A_t\) defined in (3.1) with \(\theta_0\) in place of \(\theta\). To establish the strong consistency of \(\hat{\theta}_G\) we need to the following assumptions.

**A1** \(\gamma^S(A^0) < 0\) and \(\forall \theta \in \Theta, \rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1\).

**A2** \(\theta_0 \in \Theta\) and \(\Theta\) is compact.

**A3** The polynomials \(\alpha^0_v(z) = \sum_{i=1}^{q} \alpha^0_{vi} z^i\) and \(\beta^0_v(z) = 1 - \sum_{j=1}^{p} \beta^0_{vj} z^j\) have no common root, \(\alpha^0_v(1) \neq 0\), and \(\alpha^0_{eq} + \beta^0_{vp} \neq 0\) for all \(1 \leq v \leq S\).

**A4** \(\xi_v\) is non-degenerate for all \(1 \leq v \leq S\).

As seen in Section 3, \(\gamma^S(A^0) < 0\) in **A1** ensures periodic stationarity and periodic ergodicity of the PACD model (4.1). The condition \(\rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1\) is imposed for the invertibility of equation (4.2b) for any \(\theta \in \Theta\). The compactness assumption **A2** is standard while **A3** and **A4** are made to guarantee the identifiability of the model.

**Theorem 4.1** Let \(\left(\hat{\theta}_G\right)\) be a sequence of EQMLES defined by (4.3). Under **A1-A4**, \(\hat{\theta}_G \rightarrow \theta_0\) a.s. as \(N \rightarrow \infty\) for all \(\sigma^2 > 0\).
Turn now to the asymptotic normality property of \( \hat{\theta}_G \). The following assumptions are to be considered.

**A5** \( \theta_0 \) belongs to the interior of \( \Theta \).

**A6** The matrices

\[
I (\theta_0, \sigma^2) = \sum_{v=1}^{S} \frac{\sigma_v^2}{\sigma_0^2} E \left( \frac{1}{\psi_v'(\theta_0)} \frac{\partial \psi_v(\theta_0)}{\partial \theta} \frac{\partial \psi_v(\theta_0)}{\partial \theta'} \right), \quad J (\theta_0, \sigma^2) = \sum_{v=1}^{S} \frac{1}{\sigma_v^2} E \left( \frac{1}{\psi_v'(\theta_0)} \frac{\partial \psi_v(\theta_0)}{\partial \theta} \frac{\partial \psi_v(\theta_0)}{\partial \theta'} \right)
\]

are finite, and \( J (\theta_0, \sigma^2) \) is nonsingular for all \( \sigma^2 > 0 \).

**Theorem 4.2** Under **A1-A6** we have

\[
\sqrt{N} \left( \hat{\theta}_G - \theta_0 \right) \xrightarrow{D} \mathcal{N} (0, \Sigma) \quad \text{as} \quad N \to \infty \quad \text{for all} \quad \sigma^2 > 0
\]

(4.6a)

where

\[
\Sigma = J (\theta_0, \sigma^2)^{-1} I (\theta_0, \sigma^2) J (\theta_0, \sigma^2)^{-1}
\]

(4.6b)

is block-diagonal and \( \xrightarrow{D} \) stands for convergence in distribution.

**Remark 4.1**

i) When \( \sigma^2 = (1, ..., 1)' := 1 \), the EQMLE has a covariance matrix in a "sandwich" form and is, in general, not asymptotically efficient unless \( \sigma_0^2 = 1 \) and the conditional distribution is exponential.

ii) For the special exponential PACD(\( p, q \)) model corresponding to \( \text{Var} (\xi_v) = 1 \) for all \( 1 \leq v \leq S \), if we set \( \sigma^2 = 1 \) then \( J (\theta_0, 1) = I (\theta_0, 1) \) and the asymptotic covariance matrix of the EQMLE reduces to \( \Sigma = J (\theta_0, 1)^{-1} \). The EQMLE is thus asymptotically efficient.

iii) If \( \xi_v \) has a constant variance, i.e. \( \sigma_v^2 = \text{Var} (\xi_v) = \sigma_0^2 \) for all \( 1 \leq v \leq S \), then it suffices to take \( \sigma^2 = (1, ..., 1)' \) and apply the EQMLE. We would have \( I (\theta_0, 1) = \sigma_0^2 J (\theta_0, 1) \) and the covariance matrix would be equal to \( \Sigma = \sigma_0^2 J (\theta_0, 1)^{-1} \). In this case, the EQMLE is the best QMLE among all QMLEs belonging to the linear exponential family.

iv) For the non-periodic ACD corresponding to \( S = 1 \) and then \( \sigma_{0v}^2 = \sigma_0^2 \) for all \( 1 \leq v \leq S \), it is natural to take \( \sigma_v^2 = \sigma_0^2 \) for all \( 1 \leq v \leq S \). In this case, the profile likelihood (4.3) would be given by \( \tilde{L}_t (\theta) = \frac{1}{\sigma_0^2} \sum_{t=1}^{T} \left( \frac{Y_t}{\psi_v(\theta)} + \log \left( \tilde{\psi}_t (\theta) \right) \right) \) and the resulting GQMLE then reduces to maximizing \( \frac{1}{T} \sum_{t=1}^{T} \left( \frac{Y_t}{\psi_v(\theta)} + \log \left( \tilde{\psi}_t (\theta) \right) \right) \) which is nothing else but the EQMLE criterion. This
is why, in general, the EQMLE is the most used QMLE for non-periodic ACD even when the latter is strictly (conditionally) Gamma distributed.

v) When the profile variance \( \sigma^2 \) coincides with the true variance \( \sigma_0^2 \) we would have \( J(\theta_0, \sigma_0^2) = I(\theta_0, \sigma_0^2) \) and \( \Sigma = J(\theta_0, \sigma_0^2)^{-1} \), where the GQMLE is the most efficient among all QMLEs belonging to the exponential family. As \( \sigma_0^2 \) is generally unknown, a crucial step is to get a consistent estimate \( \hat{\sigma}_0^2 \) and construct with it an estimated (profile) log-likelihood from which a new Gamma QMLE, called the two-stage Gamma QMLE (2S-GQMLE), is computed. The resulting estimate would have the aforementioned efficiency property.

vi) Theorem 4.1 and 4.2 also hold for the non-MEM PACD (2.7). It suffices to replace the assumptions A1-A4 by the following:

\[ \text{A1'} \] The process \( \{Y_t, t \in \mathbb{Z}\} \) is strictly periodically stationary and periodically ergodic.
\[ \text{A2'} \] \( E(Y_{t+\epsilon}) < \infty \) for some \( \epsilon > 0 \).
\[ \text{A3'} \] \( \psi_t(\theta) = \psi_t(\theta_0) \text{ a.s. } \Rightarrow \theta = \theta_0. \)

### 4.2 Estimating the innovation variances

To estimate the unknown variances \( \sigma_0^2 \) under the MEM constraint recall (2.1)-(2.2) and let

\[ u_t = (Y_t - \psi_t)^2 - \text{Var}(Y_t|\mathcal{F}_{t-1}) = \psi_t^2 ((\xi_t - 1)^2 - \sigma_0^2). \]

Then,

\[ \frac{(Y_t - \psi_t)^2}{\psi_t^2} = \sigma_0^2 + v_t \]  \hspace{1cm} (4.7a)

where \( v_t = \frac{u_t}{\psi_t^2} = (\xi_t - 1)^2 - \sigma_0^2. \) The sequence \( (v_t) \) is thus zero-mean iid with variance \( E((\xi_t - 1)^2 - \sigma_0^2)^2 \), which is finite under the following assumption.

\[ \text{A7} \] \( E(\xi_v^4) < \infty \) for all \( v = 1, \ldots, S \).

Since \( \psi_t = \psi_t(\theta_0) \) depends on the unknown parameter \( \theta_0 \), the regressand in (4.7a) is unobservable. If we replace \( \theta_0 \) by a consistent estimate, say the GQMLE in (4.4), then we get the following approximate regression but with observable regressand

\[ \frac{(Y_t - \hat{\psi}_t)^2}{\hat{\psi}_t^2} = \sigma_0^2 + \hat{v}_t, \]  \hspace{1cm} (4.7b)
where \( \hat{\psi}_t = \psi_t(\hat{\theta}_G) \). From (4.7b) a feasible OLS estimate (OLSE) of \( \sigma_0^2 \) is given by

\[
\hat{\sigma}_v^2 = \frac{1}{N} \sum_{n=0}^{N-1} \frac{(Y_{v+nS} - \hat{\psi}_{v+nS})^2}{\hat{\psi}_{v+nS}^2}, \text{ for all } v = 1, \ldots, S. \tag{4.8}
\]

The following result shows that the OLSE \( \hat{\sigma}_v^2 \) (1 \( \leq v \leq S \)) is consistent and asymptotically Gaussian.

**Theorem 4.3** Under A1-A4

\[
\hat{\sigma}_v^2 \to \sigma_v^2 \text{ a.s. as } N \to \infty, \text{ for all } v = 1, \ldots, S. \tag{4.9a}
\]

If in addition A7 holds then for all \( v = 1, \ldots, S \)

\[
\sqrt{N} (\hat{\sigma}_v^2 - \sigma_v^2) \overset{D}{\to} \mathcal{N}(0, \Lambda_v) \text{ as } N \to \infty \tag{4.9b}
\]

where \( \Lambda_v = E \left((\xi_v - 1)^2 - \sigma_{0v}^2\right)^2 \).

A consistent estimate of the limiting variance \( \Lambda_v \) in (4.9b) is given by

\[
\hat{\Lambda}_v = \frac{1}{N^2} \sum_{n=0}^{N-1} \left(\left(\hat{\xi}_{v+nS} - 1\right)^2 - \hat{\sigma}_v^2\right)^2, \ v = 1, \ldots, S, \tag{4.10}
\]

where \( \hat{\xi}_{v+nS} = \frac{Y_{v+nS}}{\hat{\psi}_{v+nS}} \) is the residual of model (2.1). With (4.9b) and (4.10), the asymptotic matrices in (4.5) may also be estimated. A consistent estimate of \( \Sigma \) is

\[
\hat{\Sigma} = \hat{J}^{-1} \hat{I}^{-1} \hat{J}^{-1} \tag{4.11}
\]

where

\[
\hat{J} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{v=1}^{S} \frac{\partial^2 \psi_{v+nS}(\hat{\theta}_G)}{\partial \theta \partial \theta} \frac{\partial \psi_{v+nS}(\hat{\theta}_G)}{\partial \theta}, \quad \hat{I} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{v=1}^{S} \frac{\sigma_v^2}{\psi_{v+nS}(\hat{\theta}_G)} \frac{\partial^2 \psi_{v+nS}(\hat{\theta}_G)}{\partial \theta \partial \theta} \frac{\partial \psi_{v+nS}(\hat{\theta}_G)}{\partial \theta}. \]

### 4.3 Two-stage Gamma QMLE

We have seen above that the asymptotic distribution and then the asymptotic efficiency of the profile GQMLE depend on the choice of the profile variance \( \sigma^2 \). To improve the efficiency of the
GQMLE, we can replace in (4.4) the profile variances $\sigma^2$ by the OLS estimates $\hat{\sigma}^2 = (\hat{\sigma}_1^2, ..., \hat{\sigma}_S^2)'$ given by (4.8). The resulting estimate is denoted by 2S-GQMLE and is given by the following steps.

**Algorithm 4.1 Two-stage GQMLE**

i) Fix an arbitrarily $\sigma^2 > 0$, for example $\sigma^2 = (1, ..., 1)'$.

ii) Get the profile GQMLE $\hat{\theta}_G$ from (4.4).

iii) Estimate the variance innovation $\sigma_0^2$ using $\hat{\sigma}^2$ in (4.8).

iv) Consider the 2S-GQMLE as a solution of the following problem

$$\hat{\theta}^*_G = \arg \min_{\theta \in \Theta} N^{-1} \sum_{n=0}^{N-1} \sum_{v=1}^S \left( \frac{Y_{n+v}S}{\sigma_v^2 \psi_{n+v}S(\theta)} + \frac{1}{\sigma_v^2} \log \tilde{\psi}_{n+v}S(\theta) \right).$$  \hspace{1cm} (4.12)

Consistency of asymptotic normality of $\hat{\theta}^*_G$ are a by-product of Theorems 4.1-4.2.

**Corollary 4.1** Under A1-A4

$$\sqrt{N} \left( \hat{\theta}^*_G - \theta_0 \right) \xrightarrow{D} \mathcal{N} \left( 0, J(\theta_0, \sigma^2)^{-1} \right) \text{ as } N \to \infty \text{ for all } \sigma^2 > 0.$$

The latter result shows that whatever the distribution of $(\xi_v)_v$ is, the 2S-GQMLE $\hat{\theta}^*_G$ is asymptotically the most efficient one among all QMLEs belonging to the linear exponential family (cf. Gourieroux et al, 1984; Wooldridge, 1999). In particular, $\hat{\theta}^*_G$ is never asymptotically less efficient than the profile GPQMLE $\hat{\theta}_G$ and therefore than the EQMLE $\hat{\theta}_E$.

## 5 Simulation study

We examine the finite-sample behavior of the Gamma QMLEs, as defined above, using many simulated PACD(1,1) series with sample size $T = 2000$. We consider two distributions for the innovation $\xi_t$ in (2.1), namely i) the exponential distribution $(\xi_t \sim \mathcal{E}(1) \equiv \Gamma(1,1))$ so that $Y_t|\mathcal{F}_{t-1} \sim \Gamma(1,1/\psi_t)$, and ii) the Gamma distribution $(\xi_t \sim \Gamma(\sigma_0^{-2}, \sigma_0^{-2}))$ so that $Y_t|\mathcal{F}_{t-1} \sim \Gamma(\sigma_0^{-2}, \sigma_0^{-2}/\psi_t)$, where $\sigma_0^{-2} (1 \leq v \leq S)$.

For these two cases we take $S = 5$, which is representative of many real daily trading measurements, such as trading volumes and realized volatilities. The true conditional mean parameters, which are reported in Tables 5.1 and 5.2, are chosen so that the PACD model to
be stable in the sense of Section 3, while implying fairly persistent series that are in accordance with the empirical evidence. For each case and for each series we compute the EQMLE and the two-stage GQMLE (2S-GQMLE), using 1000 Monte Carlo replications.

The starting parameter value in the nonlinear optimization routines (4.4) and (4.12) is set to the true value, while the unobservable starting values $Y_0$ and $\psi_0(\theta)$ of the PACD(1,1) equation are set to the intercept $\omega_0^0$. The two-stage GQMLE is calculated, with the EQMLE computed in the first stage.

<table>
<thead>
<tr>
<th>$v$</th>
<th>$\theta^0$</th>
<th>$\omega^0_v$</th>
<th>$\alpha^0_v$</th>
<th>$\beta^0_v$</th>
<th>$\sigma^2_{\theta v}$</th>
<th>$\omega^0_v$</th>
<th>$\alpha^0_v$</th>
<th>$\beta^0_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Mean</td>
<td>0.5</td>
<td>0.6</td>
<td>0.35</td>
<td>1</td>
<td>0.5</td>
<td>0.6</td>
<td>0.35</td>
</tr>
<tr>
<td>Std</td>
<td>0.3163</td>
<td>0.0709</td>
<td>0.0698</td>
<td>0.0968</td>
<td>0.3168</td>
<td>0.0710</td>
<td>0.0699</td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>Mean</td>
<td>0.9</td>
<td>0.4</td>
<td>0.5</td>
<td>1</td>
<td>0.9</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Std</td>
<td>0.3624</td>
<td>0.0723</td>
<td>0.0921</td>
<td>0.1001</td>
<td>0.3628</td>
<td>0.0723</td>
<td>0.0924</td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>Mean</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Std</td>
<td>0.4795</td>
<td>0.0784</td>
<td>0.1010</td>
<td>0.0992</td>
<td>0.4806</td>
<td>0.0783</td>
<td>0.1011</td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>Mean</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>1</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Std</td>
<td>0.4000</td>
<td>0.0635</td>
<td>0.0795</td>
<td>0.1041</td>
<td>0.4007</td>
<td>0.0636</td>
<td>0.0798</td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>Mean</td>
<td>0.7</td>
<td>0.55</td>
<td>0.40</td>
<td>1</td>
<td>0.7</td>
<td>0.55</td>
<td>0.40</td>
</tr>
<tr>
<td>Std</td>
<td>0.3828</td>
<td>0.0726</td>
<td>0.0804</td>
<td>0.0939</td>
<td>0.3832</td>
<td>0.0728</td>
<td>0.0806</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1. EQMLE and 2S-GQMLE results for 1000 PACD(1,1) series with $n = 2000$ generated from the exponential $\Gamma (1, 1/\psi_t)$ distribution.

Means and standard deviations of the estimates $\hat{\theta}_E$ and $\hat{\theta}_G$ over the 1000 replications are reported in Table 5.1 for the exponential PACD(1,1) model and in Table 5.2 for the homolog Gamma PACD(1,1) model. It can be observed from Tables 5.1-5.2 that the results are consistent.
with asymptotic theory. They are indeed almost identical in the exponential case with a slight superiority of the EQMLE over the 2S-GQMLE (cf. Table 5.1). The two estimates are, in fact, asymptotically efficient in this case but the EQMLE is much simpler to compute. For the Gamma PACD model in Table 5.2, $\hat{\theta}^G_C$ outperforms $\hat{\theta}_E$ in terms of bias and variability, as expected. In all cases, the 2S-GQMLE is the least risky one in the misspecification case.

<table>
<thead>
<tr>
<th>$v$</th>
<th>$\theta^0_v$</th>
<th>EQMLE</th>
<th>2S-GQMLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\omega^0_v$</td>
<td>$\alpha^0_v$</td>
<td>$\beta^0_v$</td>
</tr>
<tr>
<td>True</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Mean</td>
<td>0.2099</td>
<td>0.4022</td>
<td>0.4991</td>
</tr>
<tr>
<td>Std</td>
<td>0.1573</td>
<td>0.0408</td>
<td>0.0768</td>
</tr>
<tr>
<td>True</td>
<td>0.9</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Mean</td>
<td>0.8843</td>
<td>0.3045</td>
<td>0.6046</td>
</tr>
<tr>
<td>Std</td>
<td>0.1462</td>
<td>0.0654</td>
<td>0.0891</td>
</tr>
<tr>
<td>True</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Mean</td>
<td>0.3341</td>
<td>0.5030</td>
<td>0.3815</td>
</tr>
<tr>
<td>Std</td>
<td>0.2856</td>
<td>0.1130</td>
<td>0.1240</td>
</tr>
<tr>
<td>True</td>
<td>0.4</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean</td>
<td>0.4040</td>
<td>0.4526</td>
<td>0.4534</td>
</tr>
<tr>
<td>Std</td>
<td>0.2645</td>
<td>0.0702</td>
<td>0.0971</td>
</tr>
<tr>
<td>True</td>
<td>0.5</td>
<td>0.55</td>
<td>0.35</td>
</tr>
<tr>
<td>Mean</td>
<td>0.4997</td>
<td>0.5546</td>
<td>0.3494</td>
</tr>
<tr>
<td>Std</td>
<td>0.2729</td>
<td>0.0921</td>
<td>0.1002</td>
</tr>
</tbody>
</table>

Table 5.2. EQMLE and 2S-GQMLE results for 1000 PACD(1,1) series with $T = 2000$ generated from the Gamma $\Gamma(1/\sigma^2_{0t}, 1/\sigma^2_{0t}\psi_t)$ distribution.

6 Empirical applications

6.1 Application to Bitcoin trading volume data

In our application, we fit the PACD(1,1) model to the daily Bitcoin trading volume (BTV). The dataset was obtained from the webpage www.blockchain.com. This series spans from July,
3, 2017 to June, 26, 2020, with a total of $T = 1092 = 7 \times 156$ observations. Figure 6.1 displays the time series plot of the data.

In the context of Bitcoin prices, Mbanga (2019) found evidence of the presence of the day-of-the-week pattern. Our aim here is to show that the Bitcoin volume data are also characterized by the day-of-the-week effect, which implies a period of $S = 7$. Such a case is different from the data usually encountered in non-cryptocurrency returns (such as stocks, exchange rates), which are characterized by a periodicity of $S = 5$, due to the existence of non-trading days at each week (Franses and Paap, 2000; Tsiakas, 2006).

Table 6.1 provides some descriptive statistics for the full sample and for each day of the week separately. The mean of BVT series is clearly different from one day to another. The difference is more pronounced for the Kurtosis and skewness across the days. Also, the estimated kernel densities of the data across the days are visually different (see Supplementary material). In that regard, we suspect that the day-of-the week effect may characterize the Bitcoin trading volume series.

<table>
<thead>
<tr>
<th>Day</th>
<th>Full series</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>40.8394</td>
<td>33.3621</td>
<td>41.3257</td>
<td>42.9304</td>
<td>46.3669</td>
<td>46.7979</td>
<td>44.2866</td>
<td>30.8063</td>
</tr>
<tr>
<td>Std</td>
<td>47.2340</td>
<td>40.9090</td>
<td>41.2095</td>
<td>49.5528</td>
<td>52.3069</td>
<td>57.7127</td>
<td>48.9096</td>
<td>34.3056</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.9736</td>
<td>2.3813</td>
<td>1.9404</td>
<td>3.1176</td>
<td>2.5808</td>
<td>3.8368</td>
<td>2.3894</td>
<td>1.8163</td>
</tr>
</tbody>
</table>

Table 6.1. Day-of-the-week pattern in the BVT series.
We first estimate a standard ACD(1,1) model (i.e. PACD with $S = 1$), using the EQMLE as recommended in Remark 4.1, (iv). This model is used as a competitor to our PACD(1,1). The initial parameter values are set to $\theta^{(0)} = (\omega^{(0)}, \alpha^{(0)}, \beta^{(0)}) = (0.1, 0.3, 0.5)$ and the starting values of the conditional mean equation are fixed to $Y_0 = \psi_0 = \omega^{(0)}$. The estimated parameters and their asymptotic standard errors (ASE) in parentheses, obtained from Theorem 4.2-4.3, are reported in Table 6.2. In particular, the ASE of $\hat{\sigma}^2$ is computed from (4.10). The persistence parameter estimate, $\hat{\alpha} + \hat{\beta} = 0.9865$, indicates a strong persistence in the series, as expected.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\alpha} + \hat{\beta}$</th>
<th>IMSFE</th>
<th>IMAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8293</td>
<td>0.4615</td>
<td>0.5250</td>
<td>0.3351</td>
<td>0.9865</td>
<td>701.9662</td>
<td>14.01243</td>
</tr>
<tr>
<td></td>
<td>(0.0262)</td>
<td>(0.0270)</td>
<td>(0.0296)</td>
<td>(0.0019)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2. EQML estimates for the ACD(1,1); BTV series.

Table 6.2 also displays the in-sample mean square (one-step ahead) forecast error (IMSFE) and the in-sample mean absolute forecast error (IMAFE) given by $\text{IMSFE} = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{\psi}_t)^2$ and $\text{IMAFE} = \frac{1}{T} \sum_{t=1}^{T} |Y_t - \hat{\psi}_t|$, respectively. Unreported sample autocorrelations of the residuals consolidate the validity of the estimated ACD(1,1). Since this model does not take into account the day-of-the-week effect, we fit a 7-periodic PACD(1,1) to the BVT series. To this end, we utilize the 2S-GQMLE by starting from the EQMLE in the first stage with the following initial parameter values for the optimization routine: $\omega^{(0)} = (0.6, 0.25, 0.35, 3, 2, 4, 1.5)$, $\alpha^{(0)} = (0.25, 0.15, 0.1, 0.3, 0.3, 0.4, 0.35)$ and $\beta^{(0)} = (0.7, 0.8, 0.85, 0.4, 0.5, 0.2, 0.45)$. These initial values are arbitrary but we checked that the estimates are robust even with other initial
values.

<table>
<thead>
<tr>
<th>Day</th>
<th>( v )</th>
<th>( \hat{\sigma}_v^2 )</th>
<th>( \hat{\omega}_v )</th>
<th>( \hat{\alpha}_v )</th>
<th>( \hat{\beta}_v )</th>
<th>( \hat{\alpha}_v + \hat{\beta}_v )</th>
<th>IMSFE</th>
<th>IMAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon</td>
<td>1</td>
<td>0.2526 (0.0050)</td>
<td>0.0165 (0.3537)</td>
<td>0.5205 (0.0421)</td>
<td>0.5645 (0.0481)</td>
<td>1.0850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tue</td>
<td>2</td>
<td>0.2716 (0.0102)</td>
<td>2.9945 (0.4361)</td>
<td>0.5423 (0.0379)</td>
<td>0.7173 (0.0430)</td>
<td>1.2597</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed</td>
<td>3</td>
<td>0.3682 (0.0180)</td>
<td>0.3109 (0.2705)</td>
<td>0.1544 (0.0414)</td>
<td>0.8710 (0.0428)</td>
<td>1.0244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thu</td>
<td>4</td>
<td>0.2839 (0.0075)</td>
<td>0.0125 (0.6255)</td>
<td>0.4951 (0.0690)</td>
<td>0.5641 (0.0772)</td>
<td>1.0591</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri</td>
<td>5</td>
<td>0.3299 (0.0092)</td>
<td>0.2889 (0.5360)</td>
<td>0.3663 (0.0655)</td>
<td>0.6358 (0.0714)</td>
<td>1.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sat</td>
<td>6</td>
<td>0.2129 (0.0014)</td>
<td>1.3040 (0.5645)</td>
<td>0.4422 (0.0724)</td>
<td>0.4786 (0.0810)</td>
<td>0.9207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>7</td>
<td>0.2372 (0.0043)</td>
<td>0.3513 (0.4721)</td>
<td>0.4339 (0.0781)</td>
<td>0.2256 (0.0914)</td>
<td>0.6595</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\prod_{v=1}^{7} (\hat{\alpha}_v + \hat{\beta}_v) = 0.9025
\]

Table 6.3. 2S-GQML estimates for the PACD(1,1); BTV series.

The 2S-GQML estimates and their ASEs in parentheses are reported in Table 6.3. We observe that the estimates are quite different across the days. The persistence parameters over the days show locally explosive behaviors except for Saturday and Sunday. However, the whole persistence parameter, \( \prod_{v=1}^{7} (\hat{\alpha}_v + \hat{\beta}_v) = 0.9025 \) (also called the monodromy estimate) is, as expected, considerably smaller than the one given by the estimated standard ACD(1,1). All results have been obtained irrespective of any distributional specification of the models.

Note that the ASE of estimates for the PACD are larger than those obtained for the ACD. This is due to the fact that for the PACD the ASEs are computed for lower channel series with sample size \( \frac{T}{s} = 156 \). To get the same precision as with the ACD we should consider larger series with the sample size multiplied at least by 7. Nevertheless, in term of in-sample forecast ability (IMSFE and IMASE), the PACD model outperforms the standard ACD.

To compare the out-of-sample forecasting performance of the two models, we estimate the two competing models on the basis of the first \( T_f \) observations of the series, where \( 1 < T_f < T \). Then, we compute the one-step ahead forecast on the period \((T_f + 1, ..., T)\) based on

\[
\hat{\psi}_t = \hat{\omega}_t + \hat{\alpha}_t Y_{t-1} + \hat{\beta}_t \hat{\psi}_{t-1} \text{ for } t = T_f + 1, ..., T.
\]
We finally calculate for each model the following three criteria: i) the mean square forecast error \( \text{MSFE} = \frac{1}{T-T_f} \sum_{t=T_f+1}^{T} (Y_t - \hat{\psi}_t)^2 \), ii) the mean absolute forecast error \( \text{MAFE} = \frac{1}{T-T_f} \sum_{t=T_f+1}^{T} |Y_t - \hat{\psi}_t| \), and iii) the mean QLIKE (cf. Patton, 2011; Aknouche and Francq, 2019) \( \text{MQLI} = \frac{1}{T-T_f} \sum_{t=T_f+1}^{T} (\log \hat{\psi}_t + \frac{Y_t}{\hat{\psi}_t}) \).

Table 6.4 shows these computed values of these criteria for the two models and for various truncated series with sample size \( T_f \). It can be observed that irrespective of the chosen sample size, the PACD yields better out-of-sample forecasts with regard to the aforementioned criteria. Overall, the PACD(1, 1) outperforms the ACD(1, 1), both in terms of in-sample and out-of-sample forecasting power.

<table>
<thead>
<tr>
<th>( T_f )</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACD</td>
<td>MSFE</td>
<td>207.8341</td>
<td>165.8957</td>
<td>89.6771</td>
<td>88.3917</td>
<td>94.4021</td>
</tr>
<tr>
<td></td>
<td>MAFE</td>
<td>8.6659</td>
<td>7.8318</td>
<td>6.9127</td>
<td>6.6050</td>
<td>6.9080</td>
</tr>
<tr>
<td></td>
<td>MQLI</td>
<td>3.8983</td>
<td>3.7735</td>
<td>3.6879</td>
<td>3.6052</td>
<td>3.6959</td>
</tr>
<tr>
<td>PACD</td>
<td>MSFE</td>
<td>193.1508</td>
<td>151.0290</td>
<td>79.2446</td>
<td>80.8776</td>
<td>80.4284</td>
</tr>
<tr>
<td></td>
<td>MAFE</td>
<td>8.0187</td>
<td>7.1074</td>
<td>6.1546</td>
<td>5.9031</td>
<td>6.02684</td>
</tr>
<tr>
<td></td>
<td>MQLI</td>
<td>3.8759</td>
<td>3.7495</td>
<td>3.6605</td>
<td>3.5754</td>
<td>3.6635</td>
</tr>
</tbody>
</table>

Table 6.4. Out-of-sample forecasting performance of the PACD and ACD; BVT series.

6.2 Application to the UN realized volatility

The second dataset is the daily UN realized volatility (RV) that covers the sample period from January 04, 1999 to December, 31, 2008 with a total of \( T = 2489 \) observations. The plot of the index series is displayed in Figure 6.2.

Table 6.5 reports some descriptive statistics concerning the whole series and the subseries corresponding to the five trading days. It can be easily seen that these statistics strongly indicate that the distributions of realized volatility are significantly different across the trading days. This is also confirmed by the estimated kernel density of each trading day (Supplementary...
material). These facts suggest using a 5-periodic PACD(1,1) model for these data.

<table>
<thead>
<tr>
<th>Day</th>
<th>Full series</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>2489</td>
<td>469</td>
<td>511</td>
<td>514</td>
<td>504</td>
<td>491</td>
</tr>
<tr>
<td>Mean</td>
<td>1.3085</td>
<td>1.1674</td>
<td>1.2528</td>
<td>1.3028</td>
<td>1.3631</td>
<td>1.4511</td>
</tr>
<tr>
<td>Std</td>
<td>1.7699</td>
<td>1.4919</td>
<td>1.4628</td>
<td>1.5890</td>
<td>1.9073</td>
<td>2.2648</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>47.1196</td>
<td>26.3686</td>
<td>13.9103</td>
<td>23.4364</td>
<td>37.6800</td>
<td>55.2210</td>
</tr>
<tr>
<td>Skewness</td>
<td>5.1689</td>
<td>3.9488</td>
<td>2.9500</td>
<td>3.7675</td>
<td>5.0477</td>
<td>6.0455</td>
</tr>
</tbody>
</table>

Table 6.5. Day-of-the-week pattern in the UN-RV series.

As a reference model, we first estimate a standard ACD(1,1) for the data. Table 6.6 shows the EQML estimates and their asymptotic standard errors in parenthesis. The results signal a high persistence near to instability.

<table>
<thead>
<tr>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\alpha} + \hat{\beta}$</th>
<th>IMSFE</th>
<th>IMAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0109</td>
<td>0.2849</td>
<td>0.7084</td>
<td>0.2841</td>
<td>0.9933</td>
<td>1.1122</td>
<td>0.4842</td>
</tr>
</tbody>
</table>

Table 6.6. EQML estimates for the ACD(1,1); UN-RV series.

Table 6.7 displays the 2S-GQML estimates of the PACD(1,1) based on the UN-RV series. These estimates are quite different across the days and are all significant. Also, the persistence parameter, given by $\prod_{v=1}^{7}(\hat{\alpha}_v + \hat{\beta}_v) = 0.8897$, is significantly smaller than that obtained from the ACD. The ASEs of the estimates for the PACD are smaller than in the first application, since the series here is quite longer. Moreover, the PACD model outperforms the standard
ACD, according to the IMSFE and IMASE criteria.

\[
\pi^2 v \hat{\sigma}^2_v \quad \hat{\omega}_v \quad \hat{\alpha}_v \quad \hat{\beta}_v \quad \hat{\alpha}_v + \hat{\beta}_v \quad \text{IMSE} \quad \text{MAFE}
\]

<table>
<thead>
<tr>
<th>Day</th>
<th>(v)</th>
<th>(\hat{\sigma}^2_v)</th>
<th>(\hat{\omega}_v)</th>
<th>(\hat{\alpha}_v)</th>
<th>(\hat{\beta}_v)</th>
<th>(\hat{\alpha}_v + \hat{\beta}_v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon</td>
<td>1</td>
<td>0.3805 (0.0082)</td>
<td>0.0164 (0.0098)</td>
<td>0.2374 (0.0390)</td>
<td>0.6702 (0.0417)</td>
<td>0.9076</td>
</tr>
<tr>
<td>Tue</td>
<td>2</td>
<td>0.2597 (0.0017)</td>
<td>0.0154 (0.0084)</td>
<td>0.3320 (0.0291)</td>
<td>0.6903 (0.0348)</td>
<td>1.0222</td>
</tr>
<tr>
<td>Wed</td>
<td>3</td>
<td>0.2552 (0.0012)</td>
<td>0.0015 (0.0088)</td>
<td>0.4023 (0.0327)</td>
<td>0.6521 (0.0354)</td>
<td>1.0544</td>
</tr>
<tr>
<td>Thu</td>
<td>4</td>
<td>0.2670 (0.0019)</td>
<td>0.0126 (0.0078)</td>
<td>0.3052 (0.0359)</td>
<td>0.7065 (0.0388)</td>
<td>1.0117</td>
</tr>
<tr>
<td>Fri</td>
<td>5</td>
<td>0.2682 (0.0023)</td>
<td>0.0884 (0.0132)</td>
<td>0.4138 (0.0390)</td>
<td>0.4851 (0.0442)</td>
<td>0.8989</td>
</tr>
</tbody>
</table>

All \(\prod_{v=1}^7 (\hat{\alpha}_v + \hat{\beta}_v)\) 0.8897 1.0570 0.4757

Table 6.7. 2S-GQML estimates for the PACD(1,1); UN-RV series.

We finally compare the out-of-sample forecasting performance of the two models, using the same devices as before. From Table 6.8 it can be concluded that for all truncated series (with sample size \(T_f\)), the PACD gives more accurate forecasts, in terms of the MSFE and MAFE values. Regarding the mean QLIKE criterion, the PACD is clearly the best one (except for \(T_f = 1100\) and \(T_f = 1200\), where the models are almost comparable).

<table>
<thead>
<tr>
<th>(T_f)</th>
<th>1100</th>
<th>1200</th>
<th>1500</th>
<th>1600</th>
<th>1800</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACD</td>
<td>MSFE</td>
<td>1.0803</td>
<td>1.1501</td>
<td>1.4973</td>
<td>1.6645</td>
<td>2.1461</td>
</tr>
<tr>
<td></td>
<td>MAFE</td>
<td>0.3662</td>
<td>0.3644</td>
<td>0.4219</td>
<td>0.4548</td>
<td>0.5509</td>
</tr>
<tr>
<td></td>
<td>MQLI</td>
<td>0.4485</td>
<td>0.4534</td>
<td>0.5903</td>
<td>0.6690</td>
<td>0.9126</td>
</tr>
<tr>
<td>PACD</td>
<td>MSFE</td>
<td>0.9729</td>
<td>1.0299</td>
<td>1.3543</td>
<td>1.4932</td>
<td>1.9275</td>
</tr>
<tr>
<td></td>
<td>MAFE</td>
<td>0.3656</td>
<td>0.3564</td>
<td>0.4094</td>
<td>0.4353</td>
<td>0.5189</td>
</tr>
<tr>
<td></td>
<td>MQLI</td>
<td>0.4564</td>
<td>0.4547</td>
<td>0.5893</td>
<td>0.6686</td>
<td>0.9074</td>
</tr>
</tbody>
</table>

Table 6.8. Out-of-sample forecasting values for the PACD and ACD (UN-RV series).
7 Conclusion and future research

A GARCH-like model for positive-valued data with seasonal behavior has been proposed. The model consists of an ACD model with parameters evolving periodically over time. In our methodology for studying and building such a model, we considered QML estimates that are distribution free and are consistent and asymptotically Gaussian under general conditions. In particular, our proposed two-stage Gamma QMLE takes into account the periodicity of the innovation sequence, giving more accurate results compared to the exponential QMLE. The proposed estimates are also consistent and asymptotically normal for more general non-MEM forms.

The model was applied to two daily financial series with different periods ($S = 7$ and $S = 5$) in an attempt to capture the day-of-the-week effect. A third application to daily S&P 500 volumes (with $S = 5$) is displayed in the Supplementary material and leads to the same conclusions.

Our model can be applied to other data frequencies, such as monthly data with $S = 12$ and quarterly data with $S = 4$. Moreover, it may also be utilized as an approximate model for count time series data with large values, such as the daily number of transactions in a market.

Although our model is named periodic ACD in reference to the ACD proposed by Engel and Russell (1998), it is not recommended to model intraday durations, which are rather characterized by a (stochastic) time-varying period, due to the irregularly-spaced nature of durations. Furthermore, modeling intraday positive-valued series generally requires very large periods for which the estimation of the parameters becomes very challenging.

For models with large periods, some basis functions for reducing the number of parameters, such as Fourier approximation (Bollerslev et al., 2000; Rossi and Fantazani, 2015; Bracher and Held, 2017), periodic B-splines (Ziel et al., 2015) or periodic wavelets (Ziel et al., 2016) could be adapted to our model. These issues are left for future research.
Proofs

Proof of Theorem 3.1 i) The $ipd_S$-driven SRE (3.1) can be embedded in the following system of $S$ SREs

$$Y_{nS+v} = A_{nS+v}Y_{(n-1)S+v} + B_{nS+v}, \quad n \in \mathbb{Z}, \quad v \in \{0, \ldots, S-1\},$$

(A.1)

where $A_{nS+v} = \prod_{i=0}^{S-1} A_{nS+v-i}$ and $B_{nS+v} = \sum_{j=0}^{S-1} \prod_{i=0}^{j-1} A_{nS+v-i} B_{nS+v-j}$ are such that the sequence $\{(A_{nS+v}, B_{nS+v}), n \in \mathbb{Z}\}$ is iid for all $v \in \{0, \ldots, S-1\}$. The standard top Lyapunov exponent $\gamma^{(S)}_v$ associated with the iid-driven SRE (A.1) is given for all $v \in \{0, \ldots, S-1\}$ by (cf. Bougerol and Picard, 1992a)

$$\gamma^{(S)}_v = \inf \left\{ \frac{1}{n} \log \| A_{nS+v} A_{(n-1)S+v} \ldots A_{S+v} \|, \quad n \geq 1 \right\}$$

(A.2)

$$= \lim_{n \to \infty} \frac{1}{n} \log \| A_{nS+v} A_{nS+v-1} \ldots A_{v+1} \| \quad a.s.$$  

Since $E \xi_v < \infty$ it follows that $E \log^+ \| A_v \| < \infty$ and $E \log^+ \| B_v \| < \infty$ for all $0 \leq v \leq S-1$. Therefore, by Theorem 2.5 of Bougerol and Picard (1992a), equation (A.1) admits a unique nonanticipative strictly stationary and ergodic solution $\{Y_{nS+v}, n \in \mathbb{Z}\}$ under $\gamma^{(S)}_v < 0$. That solution is given for all $v \in \{0, \ldots, S-1\}$ by

$$Y_{nS+v} = \sum_{j=0}^{\infty} j^{-1} \prod_{i=0}^{j-1} A_{(n-i)S+v} B_{(n-j)S+v}, \quad n \in \mathbb{Z}, v \in \{0, \ldots, S-1\}$$

(A.3)

which is exactly (3.3), where the series in equality (A.3) converges absolutely a.s. This shows that $\{Y_t, t \in \mathbb{Z}\}$ is the unique causal strictly periodically stationary and periodically ergodic solution of (3.1). By a sandwiching argument, it is easily seen that for all $v \in \{0, \ldots, S-1\}$

$$\gamma^{(S)}_v = \lim_{n \to \infty} \frac{1}{n} \log \| A_{nS+v} A_{nS+v-1} \ldots A_{v+1} \| = \lim_{n \to \infty} \frac{1}{n} \log \| A_{nS} A_{nS-1} \ldots A_1 \| := \gamma^{(S)}.$$  

ii) If model (3.1) admits a nonanticipative strictly periodically stationary solution $\{Y_t, t \in \mathbb{Z}\}$ then from the non-negativity of the coefficients of $A_i$ in (3.1) it follows that for all $k > 1$,

$$Y_v \geq \sum_{j=0}^{k-1} \prod_{i=0}^{j-1} A_{v-i} B_{v-j}, \quad a.s.,$$
so the series $\sum_{j=0}^{\infty} \prod_{i=0}^{j-1} A_{v-i} B_{v-j}$ converges $a.s.$ Therefore $\prod_{i=0}^{j-1} A_{v-i} B_{v-j} \to 0$ $a.s.$ as $j \to \infty$, from which we have to show that $\prod_{i=0}^{j-1} A_{v-i} \to 0$ $a.s.$ as $j \to \infty$. This holds whenever

$$\lim_{j \to \infty} \prod_{i=0}^{j-1} A_{v-i} e_m = 0, \text{a.s. for all } 1 \leq m \leq r,$$

(A.4)

where $r = p + q$ and $(e_m)_{1 \leq m \leq r}$ is the canonical basis of $\mathbb{R}^r$. Since $A_t$ has the same zero-structure as the matrix $A_t$ in Bougerol and Picard (1992b), then (A.4) follows from their results using the same argument.

ii) By the nonnegativity of $\{A_t, t \in \mathbb{Z}\}$ we have

$$\gamma^S (A) \geq \gamma^S (\beta) := \log \rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right).$$

(A.5)

If (3.1) admits a strictly periodically stationary solution then $\gamma^S (A) < 0$. In view of (A.5), it follows that

$$\gamma^S (\beta) < 0$$

(A.6)

which in turn implies (3.4). □

**Proof of Theorem 3.2** Theorem 3.2 is a particular case of Theorem 3.3, ii). □

**Proof of Theorem 3.3** i) The proof is similar to the one of Lemma 2.3 of Berkes et al (2003). See Supplementary material.

ii) Define $\{\tilde{Y}_t, t \in \mathbb{Z}\}$ by

$$\begin{cases} \tilde{Y}_t = A_t \tilde{Y}_{t-1} + B_t & t \geq 1 \\ \tilde{Y}_t = 0 & t \leq 0, \end{cases}$$

(A.7)

and let $Y^{(v)}$ $(0 \leq v \leq S - 1)$ be a random variable having the same distribution as the term $Y_{nS+v}$ of the unique periodically stationary solution given by (3.1). By construction $\tilde{Y}_{nS+v} \xrightarrow{d} Y^{(v)}$ as $n \to \infty$. Let $m = 2$. From the weak convergence theory (Billingsley, 1968), to show that $E \left( vec \left( Y^{(v)} Y^{(v)' \beta} \right) \right)$ is finite for all $v$, it is sufficient to show that $\lim \inf_{n \to \infty} E \left( vec \left( \tilde{Y}_{nS+v} \tilde{Y}_{nS+v}' \right) \right) < \infty$ for all $v$. Set $V_{nS+v} = E \left( vec \left( \tilde{Y}_{nS+v} \tilde{Y}_{nS+v}' \right) \right)$. From (A.7) we get the following first-order
$S$-periodic difference equation

$$V_{nS+v} = E \left( A^v \otimes 2 \right) V_{nS+v-1} + \left[ E(A_v \otimes B_v) + E(B_v \otimes A_v) \right] E \left( \tilde{Y}_{nS+v} \right) + \text{vec}(E(B_v B'_v)) \quad (A.8)$$

where $E \left( A^v \otimes 2 \right)$, $E(A_t \otimes B_t)$ and $\text{vec}(E(B_t B'_t))$ are finite $S$-periodic matrices over $t$. Since, the last two terms of the right-hand side of $(A.8)$ are bounded, it follows that $\lim_{n \to \infty} V_{nS+v}$ exists for every $1 \leq v \leq S$ as long as (3.8) holds, from which follows the proof for $m = 2$. For general $m$, the proof is similar. □

Before giving the proof of Proposition 3.1, we need to state the following well-known result on linear ordinary periodic difference equations. Let

$$y_t = a_t y_{t-1} + b_t, \quad t \in \mathbb{Z}, \quad (A.9)$$

be an ordinary difference equation with $S$-periodic positive coefficients $a_t = a_{t+S} > 0$ and $b_t = b_{t+S} > 0$ for all $t \in \mathbb{Z}$.

**Lemma 1** The real-valued ordinary difference equation $(A.9)$ admits a unique solution $\{y_t, t \in \mathbb{Z}\}$ if and only if

$$\prod_{v=1}^{S} a_v < 1.$$ 

**Proof of Proposition 3.1** It is well-known that if $Y_t|\mathcal{F}_{t-1} \sim \Gamma \left( \frac{1}{\sigma^2_t}, \frac{1}{\sigma_t \psi_t} \right)$ then the conditional moments up to the fourth order are given by

$$E(Y_t|\mathcal{F}_{t-1}) = \psi_t \quad (A.10a)$$

$$E(Y_t^2|\mathcal{F}_{t-1}) = (1 + \sigma^2_t) \psi_t^2 \quad (A.10b)$$

$$E(Y_t^3|\mathcal{F}_{t-1}) = (1 + \sigma^2_t) (1 + 2\sigma^2_t) \psi_t^3 \quad (A.10c)$$

$$E(Y_t^4|\mathcal{F}_{t-1}) = (1 + \sigma^2_t) (1 + 2\sigma^2_t) (1 + 3\sigma^2_t) \psi_t^4. \quad (A.10d)$$

In view of $(A.10)$ it turns out that the conditional moment of $Y_t$ of order $i$ is a polynomial in $\psi_t$ with degree $i$ ($i = 1, 2, 3, 4$). Hence $E(Y_t^i) < \infty$ if and only if $E(\psi_t^i) < \infty$ ($i = 1, 2, 3, 4$), conditions for which are given as follows.

i) Expanding $E(\psi_t)$ using (2.1b) with $p = q = 1$ and $(A.10a)$, we find the following linear
**S-periodic difference equation**

\[
E(\psi_t) = (\alpha_t + \beta_t) E(\psi_{t-1}) + \omega_t, \quad t \in \mathbb{Z}. \tag{A.11}
\]

By Lemma 1, there is a unique solution of (A.11) if and only if (3.6) holds.

ii) For the existence of the second moments \( E(Y_v^2) \) \((1 \leq v \leq S)\), expanding \( E(\psi_t^2) \) using (2.1b) and (3.10a)-(3.10b), we find the following linear periodic difference equation

\[
E(\psi_t^2) = (\alpha_t^2 E(\xi_{t-1}^2) + 2\alpha_t\beta_t + \beta_t^2) E(\psi_{t-1}^2) + K_t^{(1)}, \quad t \in \mathbb{Z}, \tag{A.12}
\]

where

\[
K_t^{(1)} = (2\alpha_t\omega_t + 2\beta_t\omega_t) E(\psi_{t-1}) + \omega_t^2
\]

is finite if and only if \( E(\psi_{t-1}) < \infty \), and thus if and only if (3.6) holds. By Lemma 1, there exists a unique solution to (A.12) if and only if (3.6) and (3.9) are satisfied.

iii) Expanding \( E(\psi_t^3) \) using (2.1b) and (A.10a)-(A.10c) we get the following linear periodic difference equation

\[
E(\psi_t^3) = (\alpha_t^3 E(\xi_{t-1}^3) + 3\alpha_t^2\beta_t (\sigma_{t-1}^2 + 1) + \beta_t^3 + 3\alpha_t\beta_t^2) E(\psi_{t-1}^2) + K_t^{(2)}, \quad t \in \mathbb{Z} \tag{A.13}
\]

where

\[
K_t^{(2)} = (3\alpha_t^2\omega_t E(\xi_{t-1}^2) + 6\alpha_t\beta_t\omega_t + 3\beta_t^2\omega_t) E(\psi_{t-1}^2) + (3\beta_t\omega_t^2 + 3\alpha_t\omega_t^2) E(\psi_{t-1}) + \omega_t^3
\]

is finite if and only if \( E(\psi_{t-1}^2) < \infty \) and \( E(\psi_{t-1}) < \infty \). By Lemma 1, equation (A.13) admits a unique solution if and only if (3.6), (3.9) and (3.10) hold.

iv) Expanding \( E(\psi_t^4) \) using (2.1b) and (A.10a)-(A.10d) we get the following linear periodic difference equation

\[
E(\psi_t^4) = (\alpha_t^4 E(\xi_{t-1}^4) + 4\alpha_t^3\beta_t (1 + \sigma_{t-1}^2) (1 + 2\sigma_{t-1}^2) + 6\alpha_t^2\beta_t^2 (1 + \sigma_{t-1}^2) + 4\alpha_t\beta_t^3 + \beta_t^4) E(\psi_{t-1}^4) + K_t^{(3)} \tag{A.14}
\]
where

\[
K_t^{(3)} = 4\alpha^3_t \omega_t E (Y_{t-1}^3) + 12\alpha^2_t \beta_t \omega_t E (Y_{t-1}^2 \psi_{t-1}) + 12\alpha_t \beta_t^2 \omega_t E (Y_{t-1} \psi_{t-1}^2) + 4\beta_t^3 \omega_t E (\psi_{t-1}^3) + 6\alpha^2_t \omega_t^2 E (Y_{t-1}^2) + 12\alpha_t \beta_t \omega_t^2 E (Y_{t-1} \psi_{t-1}) + 6\beta_t^2 \omega_t^2 E (\psi_{t-1}^2)
\]

is finite under (3.6), (3.9) and (3.10). By Lemma 1, equation (A.14) admits a unique positive solution if and only if (3.6), (3.9), (3.10) and (3.11) hold.

Proof of Theorem 4.1 Theorem 4.1 will be proved by showing several lemmas below. In what follows \( M > 0 \) and \( \rho \in (0, 1) \) stand for constants that are not necessarily the same when appearing in different terms. Let \( L_T (\theta) \) and \( l_t (\theta) \) be defined in the same way as \( \tilde{L}_T (\theta) \) and \( \tilde{l}_t (\theta) \) in (4.3a) and (4.3b), respectively, with \( \psi_t (\theta) \) in place of \( \tilde{\psi}_t (\theta) \).

Lemma 1 Under A1 and A2 we have

\[
\sup_{\theta \in \Theta} \left| L_T (\theta) - \tilde{L}_T (\theta) \right| \rightarrow 0 \ a.s. \ as \ T \rightarrow \infty.
\]

Proof Rewrite (4.2b) in a vector form as follows

\[
\tilde{\psi}_t = \beta_t \tilde{\psi}_{t-1} + \alpha_t, \ t \in \mathbb{Z}, \quad (A.15)
\]

where \( \tilde{\psi}_t = (\psi_t (\theta), \psi_{t-1} (\theta), ..., \psi_{t-p+1} (\theta))' \) and \( \alpha_t = \left( \omega_t + \sum_{i=1}^{q} \alpha_t Y_{t-i}, 0, ..., 0 \right)' \). By A1 and the assumption A2 of compactness of \( \Theta \) it follows that

\[
\sup_{\theta \in \Theta} \rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1. \quad (A.16)
\]

Iterating (A.15) gives

\[
\tilde{\psi}_t = \sum_{k=0}^{t-1} \prod_{i=0}^{k-1} \beta_{t-i} \tilde{\alpha}_{t-k} + \prod_{i=0}^{t} \beta_{t-i} \tilde{\psi}_0, \ t \in \mathbb{Z}.
\]

Denote by \( \tilde{\psi}_t \) and \( \tilde{\alpha}_t \) the vectors obtained from \( \tilde{\psi}_t \) and \( \tilde{\alpha}_t \), respectively, while replacing \( \psi_{t-j} (\theta) \) by \( \tilde{\psi}_{t-j} \) with fixed initial values. From (A.15) and (A.16) we thus get

\[
\sup_{\theta \in \Theta} \left\| \tilde{\psi}_t - \tilde{\psi}_0 \right\| = \sup_{\theta \in \Theta} \left\| \sum_{k=t-q}^{t-1} \prod_{i=0}^{k-1} \beta_{t-i} (\tilde{\alpha}_{t-k} - \tilde{\alpha}_{t-k}) + \prod_{i=0}^{t-1} \beta_{t-i} (\tilde{\psi}_0 - \tilde{\psi}_0) \right\| \leq M \rho^t. \quad (A.17)
\]
Using the inequality $|\log \frac{y}{x}| \leq \frac{|y-x|}{\min\{y,x\}}$ for positive $x$ and $y$ (cf. Francq and Zakoian, 2019) it follows that

$$
\sup_{\theta \in \Theta} \left| L_T (\theta) - \tilde{L}_T (\theta) \right| \leq \frac{T}{T} \sum_{t=1}^{T} \sup_{\theta \in \Theta} \left[ \left| \frac{\psi_t-\psi_t(\theta)}{\psi_t(\theta)} \right| Y_t + \left| \log \left( \frac{\psi_t(\theta)}{\psi_t(\theta)} \right) \right| \right] \\
\leq \max_{1 \leq v \leq S} \sup_{\theta \in \Theta} \left( \frac{\omega^2_{v}}{\sigma^2_{v}} \right) \frac{M}{T} \sum_{t=1}^{T} \rho^t Y_t + \max_{1 \leq v \leq S} \sup_{\theta \in \Theta} \left( \frac{\omega^2_{v}}{\sigma^2_{v}} \right) \frac{M}{T} \sum_{t=1}^{T} \rho^t .
$$

The existence of $E(Y_t^\delta)$ (cf. Theorem 3.3, i)) implies, by the Borel-Cantelli lemma, that $\rho^t Y_t \rightarrow 0$ a.s. and the conclusion follows by Césaro’s lemma. □

**Lemma 2** Under A1-A4 there is $t \in \mathbb{Z}$ such that $\psi_t (\theta) = \psi_t (\theta_0)$ a.s. if and only if $\theta = \theta_0$.

**Proof** From the assumption $\rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1$ in A1, the polynomials $(\beta_v (L))_v$ are invertible for all $1 \leq v \leq S$ and all $\theta \in \Theta$. Assume $\psi_t (\theta) = \psi_t (\theta_0)$ a.s. for some $t \in \mathbb{Z}$. Using the second equality in (4.1) and (4.2b) we have

$$
\left( \frac{\alpha_v (L)}{\beta_v (L)} - \frac{\alpha_v^0 (L)}{\beta_v^0 (L)} \right) Y_{v+nS} = \left( \frac{\omega^0_v}{\rho^0_v (L)} - \frac{\omega_v}{\rho_v (L)} \right) \text{ for all } 1 \leq v \leq S.
$$

If $\frac{\alpha_v (L)}{\beta_v (L)} \neq \frac{\alpha_v^0 (L)}{\beta_v^0 (L)}$ for some $1 \leq v \leq S$ then there exists a deterministic periodic time-varying combination of $Y_{t-j}$, $j \geq 1$. This contradicts A4 which assumes $(\xi_t, t \in \mathbb{Z})$ non-degenerate, since by (2.6) we have $Y_t = E(Y_t | \mathcal{F}_{t-1}) + \psi_t (\xi_t - 1)$. Therefore,

$$
\frac{\alpha_v (z)}{\beta_v (z)} = \frac{\alpha_v^0 (z)}{\beta_v^0 (z)} \forall |z| \leq 1 \text{ and } \frac{\omega^0_v}{\rho^0_v (L)} - \frac{\omega_v}{\rho_v (L)} \text{ for all } 1 \leq v \leq S,
$$

and by the assumption A3 of no common roots between $\alpha_v^0 (z)$ and $\beta_v^0 (z)$ it follows that $\alpha_v (z) = \alpha_v^0 (z)$, $\beta_v (z) = \beta_v^0 (z)$ and $\omega_v = \omega_v^0$ for all $1 \leq v \leq S$. □

**Lemma 3** Under A1

$$
\sum_{v=1}^{S} E \left( l_v (\theta_0) \right) < \infty,
$$

and $\sum_{v=1}^{S} E \left( l_v (\theta) \right)$ is minimized at $\theta = \theta_0$.

**Proof** By Jensen’s inequality and Theorem 3.3, ii) we have

$$
\sum_{v=1}^{S} E \left( \log \psi_v (\theta_0) \right) = \frac{1}{\delta} \sum_{v=1}^{S} E \left( \log \psi_v (\theta_0)^\delta \right) \leq \frac{1}{\delta} \sum_{v=1}^{S} \log E \left( \psi_v (\theta_0)^\delta \right) < \infty.
$$

30
Hence
\[ \sum_{v=1}^{S} E(l_v(\theta_0)) = \sum_{v=1}^{S} \frac{1}{\sigma_v} E[\xi_v + \log \psi_v(\theta_0)] = \sum_{v=1}^{S} \frac{1}{\sigma_v} + \sum_{v=1}^{S} \frac{1}{\sigma_v} E(\log \psi_v(\theta_0)) < \infty. \]

Using the inequality \( \log(x) \leq x - 1 \) we have for all \( \theta \in \Theta \)
\[ \sum_{v=1}^{S} E(l_v(\theta)) - \sum_{v=1}^{S} E(l_v(\theta_0)) = \sum_{v=1}^{S} \frac{1}{\sigma_v} E \left[ \log \left( \frac{\psi_v(\theta)}{\psi_v(\theta_0)} \right) + \frac{\psi_v(\theta_0)}{\psi_v(\theta)} - 1 \right] \]
\[ \geq \sum_{v=1}^{S} \frac{1}{\sigma_v} E \left[ \log \frac{\psi_v(\theta)}{\psi_v(\theta_0)} + \log \frac{\psi_v(\theta_0)}{\psi_v(\theta)} \right] = 0, \quad (A.18) \]
showing that \( \sum_{v=1}^{S} E(l_v(\theta)) \) is minimized at \( \theta_0 \).

**Lemma 4** For any \( \theta \neq \theta_0 \) there exists a neighborhood \( \mathcal{V}(\theta) \) such that
\[ \lim_{N \to \infty} \inf_{\theta \in \mathcal{V}(\theta)} \tilde{L}_N S(\theta) > \frac{1}{S} \sum_{v=1}^{S} E(l_v(\theta_0)). \]

**Proof** For all \( \theta \in \Theta \) and any positive integer \( k \), let \( \mathcal{V}_k(\theta) \) be the open ball of center \( \theta \) and radius \( 1/k \). In view of Lemma 1 we have
\[ \lim_{N \to \infty} \inf_{\theta \in \mathcal{V}_k(\theta)} \tilde{L}_N S(\theta) \geq \lim_{N \to \infty} \inf_{\theta \in \mathcal{V}_k(\theta)} \frac{1}{S} \sum_{v=1}^{S} \inf_{\theta \in \mathcal{V}_k(\theta)} l_{v+nS}(\theta). \]

By the ergodic theorem for the stationary sequence \( \left\{ \sum_{v=1}^{S} l_{v+nS}(\theta) \right\} \) with \( E \left( \sum_{v=1}^{S} l_{v+nS}(\theta) \right) \in \mathbb{R} \cup \{\infty\} \) (cf, Billingsley 1995, p. 495) it follows that
\[ \lim_{N \to \infty} \inf_{\theta \in \mathcal{V}_k(\theta)} l_{v+nS}(\theta) = \frac{1}{S} \sum_{v=1}^{S} E \left( \inf_{\theta \in \mathcal{V}_k(\theta)} l_v(\theta) \right). \]

Beppo-Levi’s theorem (e.g. Billingsley, 1995, p. 219) yields
\[ \frac{1}{S} \sum_{v=1}^{S} E \left( \inf_{\theta \in \mathcal{V}_k(\theta)} l_{v+nS}(\theta) \right) \to \frac{1}{S} \sum_{v=1}^{S} E(l_v(\theta)) \text{ as } k \to \infty, \]
and by (A.18) the result follows. \( \Box \)
Proof of Theorem 4.1 The proof of the theorem is completed by standard compactness arguments using Lemmas 2-4. □

Proof of Theorem 4.2 The proof of Theorem 4.2 is based on a Taylor expansion of $\frac{\partial L_T(\theta)}{\partial \theta}$ at $\theta_0$ which, by A5 and the strong consistency of $\hat{\theta}_G$, yields

$$0 = \sqrt{N} \frac{\partial L_T(\hat{\theta}_G)}{\partial \theta} = \sqrt{N} \frac{\partial L_T(\theta_0)}{\partial \theta} + \sqrt{N} \frac{\partial^2 L_T(\theta^*)}{\partial \theta \partial \theta} (\hat{\theta}_G - \theta_0) + \sqrt{N} \left( \frac{\partial L_T(\hat{\theta}_G)}{\partial \theta} - \frac{\partial L_T(\hat{\theta}_G)}{\partial \theta} \right)$$

where $\theta^*$ is between $\hat{\theta}_G$ and $\theta_0$. The derivatives $\frac{\partial L_T(\theta)}{\partial \theta}$ and $\frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta}$ are given by

$$\frac{\partial L_T(\theta)}{\partial \theta} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial l(Y_t|\theta)}{\partial \theta} \quad \text{and} \quad \frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta} = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{\psi_1(\theta)} \frac{\partial \psi_1(\theta)}{\partial \theta} + \frac{2}{\psi_1(\theta)} - 1 \right) \frac{1}{\sigma^2_1(\theta)} \frac{\partial \psi_1(\theta)}{\partial \theta} \frac{\partial \psi_1(\theta)}{\partial \theta}.$$

Therefore, the asymptotic normality result (4.6) follows whenever the following lemmas are established.

Lemma 5 Under A1-A2

i) $\sup_{\theta \in V(\theta_0)} \sqrt{N} \left\| \frac{\partial L_T(\theta)}{\partial \theta} - \frac{\partial L_T(\theta)}{\partial \theta} \right\| \xrightarrow{P} 0$, ii) $\sup_{\theta \in V(\theta_0)} \sqrt{N} \left\| \frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta} - \frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta} \right\| \xrightarrow{P} 0$

for some neighborhood $V(\theta_0)$ of $\theta_0$.

Proof Following the same lines of Francq and Zakoian (2019, Section 7.4), it is easily seen that under A1,

$$E_{\theta_0} \left\| \frac{1}{\sigma^2_1(\theta)} \frac{\partial \psi_1(\theta)}{\partial \theta} \right\| < 1, \quad E_{\theta_0} \left\| \frac{1}{\sigma^2_2(\theta)} \frac{\partial^2 \psi_1(\theta)}{\partial \theta \partial \theta} \right\| < 1, \quad E_{\theta_0} \left\| \frac{1}{\sigma^2_1(\theta)} \frac{\partial \psi_1(\theta)}{\partial \theta} \frac{\partial \psi_1(\theta)}{\partial \theta} \right\| < 1. \quad (A.19)$$

By (A.17), the compactness of $\Theta$ (cf. A2) and the fact that $\rho \left( \prod_{v=0}^{S-1} \beta_{S-v} \right) < 1$ (cf. A1) we have

$$\left| \frac{1}{\psi_1(\theta)} - \frac{1}{\psi_1(\theta)} \right| \leq \frac{M\rho^l}{\psi_1(\theta)} \left( 1 + M \right) \rho^l, \quad \sup_{\theta \in \Theta} \left| \frac{\partial \psi_1(\theta)}{\partial \theta} - \frac{\partial \psi_1(\theta)}{\partial \theta} \right| \leq M\rho^l, \quad \sup_{\theta \in \Theta} \left| \frac{\partial^2 \psi_1(\theta)}{\partial \theta \partial \theta} - \frac{\partial^2 \psi_1(\theta)}{\partial \theta \partial \theta} \right| \leq M\rho^l.$$
from which the proof of the lemma follows.

Lemma 6 Under A1-A6,

\[
\frac{\partial^2 L_T(\theta^*)}{\partial \theta \partial \theta^*} \xrightarrow{\mathcal{P}} \frac{1}{S} J \left( \theta_0, \sigma^2 \right)
\]

for any \( \theta^* \) between \( \hat{\theta}_G \) and \( \theta_0 \).

Proof Let \( V_k(\theta_0) \) \((k \in \mathbb{N}^*)\) be the open ball with center \( \theta_0 \) and radius \( 1/k \). Assume that \( n \) is large enough so that \( \theta^* \) belongs to \( V_k(\theta_0) \). By periodic stationarity and periodic ergodicity of \( \left\{ \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^*} \right| \right. \) we have

\[

\left| \frac{\partial^2 L_T(\theta^*)}{\partial \theta \partial \theta^*} - J \left( \theta_0, \sigma^2 \right) \right|_{i,j} = \left| \frac{\partial^2 L_T(\theta^*)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right| = \frac{1}{T} \left| \sum_{t=1}^{T} \frac{\partial^2 l_i(\theta^*)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right|

\leq \frac{1}{NS} \sum_{t=1}^{T} \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right| \leq \frac{1}{S} \sum_{v=1}^{S} E \left( \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right| \right).

The Lebesgue dominated convergence theorem yields

\[
\lim_{k \to \infty} E \left( \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right| \right) = E \left( \lim_{k \to \infty} \sup_{\theta \in V_k(\theta_0)} \left| \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^*} - E \left( \frac{\partial^2 l_i(\theta_0)}{\partial \theta \partial \theta^*} \right) \right| \right) = 0,
\]

from which the proof of the lemma follows. □
Lemma 7 Under A1-A6

\[ \sqrt{N} \frac{\partial L_T(\theta_0)}{\partial \theta} \xrightarrow{L^2} N \left( 0, \frac{1}{N^2} I \left( \theta_0, \sigma^2 \right) \right). \]

Proof It is clear that \( \sqrt{N} \frac{\partial L_T(\theta_0)}{\partial \theta} = \sum_{t=1}^{T} \frac{1}{\sqrt{N}} \frac{\partial l_t(\theta)}{\partial \theta} \) is a term of a square integrable periodically ergodic Martingale. Since by the periodic ergodic theorem (cf. Supplementary material)

\[ \frac{1}{N} \sum_{t=1}^{T} \frac{\partial l_t(\theta_0)}{\partial \theta} = \frac{1}{N} \sum_{t=1}^{T} (1 - \xi_t)^2 \frac{1}{\sigma_t^2 \psi_t(\theta_0)} \frac{\partial \psi_t(\theta)}{\partial \theta} \xrightarrow{a.s.} I \left( \theta_0, \sigma^2 \right), \]

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

Proof of Theorem 4.3 Set \( U_{v,n}(\theta) = \frac{(Y_{v+n,S} - \psi_{v+n,S}(\theta))^2}{\psi_{v+n,S}^2(\theta)} \), and denote by \( o_a.s. \) (1) a term converging almost surely to 0 as \( N \to \infty \). If we show

\[ \hat{\sigma}_v^2 = \frac{1}{N} \sum_{n=0}^{N-1} U_{v,n}(\theta_0) + o_a.s. \ (1) \quad (A.21) \]

then the result (4.9a) would follow from standard arguments.

Now a Taylor expansion of \( U_{v,n}(\hat{\theta}_G) \) around \( \theta_0 \) yields

\[ \hat{\sigma}_v^2 = \frac{1}{N} \sum_{n=0}^{N-1} U_{v,n}(\hat{\theta}_G) = \frac{1}{N} \sum_{n=0}^{N-1} U_{v,n}(\theta_0) + \left( \hat{\theta}_G - \theta_0 \right) \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial U_{v,n}(\theta^*)}{\partial \theta} \quad (A.22) \]

where \( \theta^* \) is between \( \hat{\theta}_G \) and \( \theta_0 \). Note that

\[ \frac{\partial U_{v,n}(\theta)}{\partial \theta} = 2 \left( \frac{Y_{v+n,S}^2}{\psi_{v+n,S}^2(\theta)} - \frac{Y_{v+n,S}}{\psi_{v+n,S}(\theta)} \right) \frac{1}{\psi_{v+n,S}(\theta)} \frac{\partial \psi_{v+n,S}(\theta)}{\partial \theta}. \]

Following the same lines of Francq and Zakoian (2019, p. 197) it can be easily seen that

\[ E \left( \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{Y_{v+n,S}^2}{\psi_{v+n,S}(\theta)} \right) < \infty, \quad E \left( \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{Y_{v+n,S}}{\psi_{v+n,S}(\theta)} \right) < \infty, \quad E \left( \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\psi_{v+n,S}(\theta)} \frac{\partial \psi_{v+n,S}(\theta)}{\partial \theta} \right\| \right) < \infty \]

for some neighborhood \( \mathcal{V}(\theta_0) \) of \( \theta_0 \). Hence, by the ergodic theorem and the consistency of \( \hat{\theta}_G \)
we get

\[
\limsup_{N \to \infty} \mathbb{E} \left( \sup_{\theta \in \mathcal{Y}(\theta_0)} \left\| \frac{\partial U_{\nu,n}(\theta^*)}{\partial \theta} \right\| \right) < \infty.
\]

Thus

\[
\left( \hat{\theta}_G - \theta_0 \right) \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial U_{\nu,n}(\theta^*)}{\partial \theta} = o_{a.s.} (1),
\]

and in view of (A.22) we obtain (A.21). Result (4.9b) follows by a similar argument. □

References


Supplementary material for: "Periodic Autoregressive Conditional Duration"

Abdelhakim Aknouche, Bader Almohaimeed and Stefanos Dimitrakopoulos

1 Definitions of periodic stationarity and periodic ergodicity

A positive real-valued stochastic process \( fY_t; t \in \mathbb{Z}g \) defined on a probability space \((\Omega, \mathcal{F}, P)\) is said to be strictly periodically stationary with period \( S \in \mathbb{N}^* \) (henceforth \( sps_S \)) iff each one of its \( S \) corresponding "sub-processes" \( \{Y_{nS+v}, n \in \mathbb{Z}\} \) \((1 \leq v \leq S)\) is strictly stationary in the standard sense. The simplest \( sps_S \) process is an \( ipd_S \) sequence. The periodic analog of the ergodic theorem for \( sps_S \) processes (e.g. Boyles and Gardener, 1983) can be stated as follows. If \( \{Y_t, t \in \mathbb{Z}\} \) is \( sps_S \) with \( E(Y_v) < \infty \) for all \( 1 \leq v \leq S \) then

\[
\frac{1}{n} \sum_{t=1}^{n} Y_t \xrightarrow{a.s.} \frac{1}{S} \sum_{v=1}^{S} Y_v^*, \tag{S.1}
\]

for some random variables \( (Y_v^*)_{1 \leq v \leq S} \) defined on \((\Omega, \mathcal{F}, P)\) and satisfying

\[
Y_v^* = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} Y_{kS+v}, \text{ a.s.}
\]

When for a given season \( v_0 \in \{1, ..., S\} \) the sub-process \( \{Y_{nS+v_0}, n \in \mathbb{Z}\} \) is ergodic, the limiting random variable \( Y_{v_0}^* \) is almost surely constant and then \( Y_{v_0}^* = E(Y_{v_0}) \), almost surely. The process \( \{Y_t, t \in \mathbb{Z}\} \) is said to be periodically ergodic with period \( S \) (\( pe_S \)) iff all sub-processes \( \{Y_{nS+v}, n \in \mathbb{Z}\} \) \((v \in \{1, ..., S\})\) are ergodic in the usual sense. The simplest \( pe_S \)
process is an \( ipd_S \) sequence. It follows that the limiting variable in (S.1) simplifies to

\[
\frac{1}{n} \sum_{t=1}^{n} Y_t \xrightarrow{a.s.} \frac{1}{S} \sum_{v=1}^{S} E(Y_v),
\]

the mean of the seasonal means. Like strict stationarity and ergodicity (see e.g. Billingsley, 1995, Theorem 36.4), strict periodic stationarity and periodic ergodicity are preserved under certain periodic transformations. Indeed, if \( \{Y_t, t \in \mathbb{Z}\} \) is \( sps_S \) and periodically ergodic and if \( \{Z_t, t \in \mathbb{Z}\} \) is given by \( Z_t = f_t(\ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots) \), where \( f_t \) is a function from \( \mathbb{R}^\mathbb{Z} \) into \( \mathbb{R} \), which is measurable, \( S \)-periodic over \( t \) (for all \( n \) and \( t \)) then so is \( \{Z_t, t \in \mathbb{Z}\} \).

## 2 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 if

\( \gamma^S(A) < 0 \) then there exists \( \delta > 0 \) and \( n_0 \) such that

\[
E(\|A_{n_0} S A_{n_0} S^{-1} A_1\|^\delta) < 1. \tag{S.2}
\]

Since \( \gamma^S(A) = \inf_{n \in \mathbb{N}^*} \{\frac{1}{n} E(\|A_{n} S A_{n} S^{-1} A_1\|)\} \) is strictly negative, there exists a positive integer \( n_0 \) such that

\[
E(\log \|A_{n_0} S A_{n_0} S^{-1} A_1\|) < 0.
\]

Using a multiplicative norm and by the \( ipd_S \) property of the sequence \( \{A_t, t \in \mathbb{Z}\} \) we have

\[
E(\|A_{n_0} S A_{n_0} S^{-1} A_1\|) = \|E(\|A_{n_0} S A_{n_0} S^{-1} A_1\|)\| \\
\leq \|E(\|A_{n_0} S A_{n_0} S^{-1} A_1\|)^{n_0} < \infty.
\]

Let \( f(x) = E(\|A_{n_0} S A_{n_0} S^{-1} A_1\|^x). \) Since under (3.2) in the main paper \( f'(0) = E(\log \|A_{n_0} S A_{n_0} S^{-1} A_1\|) \) \( \gamma^S < 0 \), the function \( f(x) \) decreases in a neighborhood of 0 and as \( f(0) = 1 \), it follows that there exists \( 0 < \delta < 1 \) such that (S.2) holds. Now from (3.3) in the main paper we have for some \( v \in \{1, \ldots, S\} \)

\[
\|Y_v\| \leq \sum_{k=1}^{\infty} \left\| \prod_{j=0}^{k-1} A_{v-j} \right\| \|B_{v-k}\| + \|B_v\|.
\]
Since $0 < \kappa < 1$, it follows that

$$\|Y_v \|^\kappa \leq \sum_{k=1}^{\infty} \left( \prod_{j=0}^{k-1} |A_{v-j}| \right)^\kappa \|B_{v-k}\|^\kappa + \|B_v\|^\kappa,$$

and then by the independence of $\{\xi_t, t \in \mathbb{Z}\}$

$$E \|Y_v \|^\kappa \leq \sum_{k=1}^{\infty} E \left( \left( \prod_{j=0}^{k-1} |A_{v-j}| \right)^\kappa \right) E (\|B_{v-k}\|^\kappa) + E (\|B_v\|^\kappa)$$

$$\leq B(\kappa) \sum_{k=1}^{\infty} E \left( \left( \prod_{j=0}^{k-1} |A_{v-j}| \right)^\kappa \right) + E (\|B_v\|^\kappa),$$

where $B(\kappa) = \max_{0 \leq v \leq S-1} E (\|B_{v-k}\|^\kappa)$. In view of (S.2) there exists $a_v > 0$ and $0 < b_v < 1$ such that

$$E \left( \left( \prod_{j=0}^{k-1} |A_{v-j}| \right)^\kappa \right) \leq a_v b_v^k \leq c,$$

where $c = \max_{0 \leq v \leq S-1} \{a_v b_v^k\}$. This proves that $E \|Y_v \|^\kappa < \infty$ establishing the results.
3 Kernel densities of the Bitcoin trade volume across days

- Kernel density of BTV over Monday
- Kernel density of BTV over Tuesday
- Kernel density of BTV over Wednesday
- Kernel density of BTV over Thursday
Figure S.1. Kernel densities of Bitcoin Trade Volume across days
4 Kernel densities of the UN realized volatility across days

Figure S.2. Kernel densities of UN Realized Volatility across days.

5 Application to the S&P 500 volume

The third dataset is the daily S&P 500 volume over the sample period from January 04, 1999 to December, 31, 2008 with a total of $T = 2382$ observations. The time series plot of
the index series is displayed in Figure S.1.

![Figure S.3. S&P 500 volume series (S&PV).](image)

We followed the same scheme as in the applications of the main paper. The conclusions were similar. The PACD dominates the ACD both in terms of the in-sample and out-of-sample forecasting criteria. So we only reports the results on Tables S.1-S.4.

<table>
<thead>
<tr>
<th>Day</th>
<th>Full series</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>2382</td>
<td>446</td>
<td>488</td>
<td>487</td>
<td>481</td>
<td>480</td>
</tr>
<tr>
<td>Mean</td>
<td>3.7125</td>
<td>2.2519</td>
<td>1.5130</td>
<td>1.5333</td>
<td>1.8267</td>
<td>1.4538</td>
</tr>
<tr>
<td>Std</td>
<td>0.8718</td>
<td>0.8128</td>
<td>0.7704</td>
<td>0.8133</td>
<td>0.8349</td>
<td>1.0651</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.7019</td>
<td>2.2519</td>
<td>1.5130</td>
<td>1.5333</td>
<td>1.8267</td>
<td>1.4538</td>
</tr>
</tbody>
</table>

Table S.1. Day-of-the-week pattern in daily S&PV series.
Figure S.4. Kernel densities of S&P 500 volume across days

<table>
<thead>
<tr>
<th>( \hat{\omega} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\sigma}^2 )</th>
<th>( \hat{\alpha} + \hat{\beta} )</th>
<th>IMSFE</th>
<th>IMAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4822</td>
<td>0.4574</td>
<td>0.4123</td>
<td>0.02515</td>
<td>0.8698</td>
<td>0.3522</td>
<td>0.4004</td>
</tr>
</tbody>
</table>

Table S.2. ACD (1,1) EQML estimates for the S&PV series.
#### Table S.3. PACD(1,1) 2S-GQML estimates for the S&PV series.

<table>
<thead>
<tr>
<th>Day</th>
<th>v</th>
<th>(\hat{\sigma}_v^2)</th>
<th>(\hat{\omega}_v)</th>
<th>(\hat{\alpha}_v)</th>
<th>(\hat{\beta}_v)</th>
<th>(\hat{\alpha}_v + \hat{\beta}_v)</th>
<th>IMSFE</th>
<th>IMAFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon</td>
<td>1</td>
<td>0.02327 (8.8e-06)</td>
<td>0.1633</td>
<td>0.0211</td>
<td>0.8538</td>
<td>0.8748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tue</td>
<td>2</td>
<td>0.0132 (2.3e-06)</td>
<td>0.2731</td>
<td>0.6348</td>
<td>0.3446</td>
<td>0.9794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed</td>
<td>3</td>
<td>0.0156 (4.0e-06)</td>
<td>0.0684</td>
<td>0.5904</td>
<td>0.4081</td>
<td>0.9985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thu</td>
<td>4</td>
<td>0.0178 (2.5e-05)</td>
<td>0.7527</td>
<td>0.4756</td>
<td>0.3290</td>
<td>0.8046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri</td>
<td>5</td>
<td>0.0463 (3.6e-05)</td>
<td>0.5609</td>
<td>0.5713</td>
<td>0.2861</td>
<td>0.8574</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td>(\prod_{v=1}^T (\hat{\alpha}_v + \hat{\beta}_v))</td>
<td></td>
<td></td>
<td></td>
<td>0.5902</td>
<td>0.3141</td>
<td>0.3744</td>
</tr>
</tbody>
</table>

Table S.4. Out-of-sample forecasting performance of PACD and ACD on S&PV series.

<table>
<thead>
<tr>
<th>(T_f)</th>
<th>1000</th>
<th>1200</th>
<th>1400</th>
<th>1600</th>
<th>1800</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSFE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MAFE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MQLI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MAFE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MQLI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S.4. Out-of-sample forecasting performance of PACD and ACD on S&PV series.

**References**
