

A General Framework for Prediction in Time Series Models

Citation for published version (APA):

Beutner, E., Heinemann, A., & Smeekes, S. (2019). *A General Framework for Prediction in Time Series Models*. Cornell University - arXiv. arXiv.org No. 1902.01622

Document status and date:

Published: 05/02/2019

Document Version:

Early version submitted to journal OR preprint in preprint archive

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

A General Framework for Prediction in Time Series Models

Eric Beutner[†]

Alexander Heinemann[†]

Stephan Smeekes[†]

[†]Department of Quantitative Economics

Maastricht University

February 6, 2019

Abstract

In this paper we propose a general framework to analyze prediction in time series models and show how a wide class of popular time series models satisfies this framework. We postulate a set of high-level assumptions, and formally verify these assumptions for the aforementioned time series models. Our framework coincides with that of Beutner et al. (2019) who establish the validity of conditional confidence intervals for predictions made in this framework. The current paper therefore complements the results in Beutner et al. (2019) by providing practically relevant applications of their theory.

1 Introduction

In time series prediction one is frequently interested in objects that do not only depend on parameters but also on the time series' past. Popular examples are conditional means or conditional variances. Analyzing predictions in this context involves a fundamental issue that is well-recognized in the econometric literature. It stems from the fact that on the one hand one must *condition* on the sample as the past informs about the present and future, yet on the other hand one must treat the data up to now as random to take into account parameter uncertainty. Nevertheless the issue is often ignored in standard practice or bypassed by assuming two independent processes with the same stochastic structure, using one for the conditioning and one for the estimation of the parameters. While the latter is a mathematically convenient assumption, it is rarely satisfied in practice. An alternative, more realistic approach is based on sample-splitting, in which one splits the sample into two (asymptotically) independent subsamples.

In this paper we provide a general framework to analyze prediction in time series models. We postulate a set of high-level assumptions under which Beutner et al. (2019) (henceforth BHS) establish the validity of conditional confidence intervals for predictions while demonstrating an asymptotic equivalence of two-independent processes and the sample-split approach. We show how a wide class of popular time series models satisfies this framework. In particular, we consider autoregressive moving-average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) type models and formally verify the postulated high-level assumptions. Therefore the current paper complements the results in BHS by providing practically relevant applicants to their theory.

The rest of this paper is organized as follows. The general framework to analyze

prediction in time series models is proposed in Section 2 and an accompanying set of high-level assumption is postulated. In Sections 3 and 4 we revisit the leading examples of BHS, i.e. the simple case of a conditional mean in an AR(1) and the conditional variance in a GARCH(1, 1) model. In Section 5 we focus on the conditional mean in a slightly more general model: the ARMA(1, 1) with drift. Section 6 studies the conditional volatility in a threshold GARCH (T-GARCH) model. Concluding remarks are presented in Section 7.

2 General Framework

Let $\{X_t\}$ be a univariate stochastic process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote the relevant parameter (vector) by θ_0 , which belongs to some set $\Theta \subseteq \mathbb{R}^r$, $r \in \mathbb{N}$. The general framework involves inference on objects, which are a function not only of the parameter but also the the time series' past. Mathematically, such object can be written as follows:

$$\psi_{T+1} := \psi(X_T, X_{T-1}, \dots; \theta_0) \tag{2.1}$$

for some function $\psi : \mathbb{R}^\infty \times \Theta \rightarrow \mathbb{R}$. Such prediction function can generally not be determined completely given a sample X_1, \dots, X_T . Replacing the unknown presample values by arbitrary starting values $\{s_t\}$, yields the following approximation:

$$\psi_{T+1}^s(\mathbf{X}_{1:T}; \theta_0) := \psi(X_T, X_{T-1}, \dots, X_1, s_0, s_{-1}, \dots; \theta_0), \tag{2.2}$$

where $\mathbf{X}_{t_1:t_2} = (X_{t_1}, \dots, X_{t_2})'$ for any integers $1 \leq t_1 \leq t_2 \leq T$. To estimate the prediction function in practice, the standard approach is to replace the unknown parameter θ_0 by an estimator $\hat{\theta}(\mathbf{X}_{1:T})$. Conditioning on the entire sample for the

evaluation of the prediction function entails that there is no randomness to account for parameter uncertainty, which highlights the severity of the fundamental issue at hand. The issue is frequently bypassed by making the unrealistic assumption of observing two independent processes, where one is used for the evaluation of the prediction function and the other for parameter estimation.

An alternative, more realistic approach is based on splitting the sample into two (asymptotically) independent subsamples. The successive decline of the influence of past observations, which motivated the approximation in (2.2), entails that

$$\psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) = \psi(X_T, X_{T-1}, \dots, X_{t_1}, c_{t_1-1}, \dots, c_1, s_0, s_{-1}, \dots; \theta_0) \quad (2.3)$$

serves as an approximation for (2.2) (and hence for (2.1)) for an appropriate choice of t_1 . Here $\mathbf{X}_{t_1:T}^c = (c_1, \dots, c_{t_1-1}, X_{t_1}, \dots, X_T)'$ is a vector where a subsample is substituted by a sequence of constants $\{c_t\}$, in a similar way as done for the starting values. Denoting the appropriate choice of t_1 by T_P , which indicates the starting point of the prediction sample, the sample-split estimator is obtained by replacing θ_0 in (2.3) by an estimator $\hat{\theta}(\mathbf{X}_{1:T_E})$, where T_E stands for the end of the estimation sample. Choosing T_E to satisfy $1 < T_E < T_P \leq T$ yields an estimation subsample that does not overlap with the subsample used for prediction.

Next, we postulate a set of high-level assumptions under which BHS establish the validity of conditional confidence intervals for predictions while demonstrating an asymptotic equivalence of two-independent processes and the sample-split approach.

Assumption 1. (*General Assumptions*)

- 1.a (*Estimator*) $m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) \xrightarrow{d} G_\infty$ as $T \rightarrow \infty$ for some cdf $G_\infty : \mathbb{R}^r \rightarrow [0, 1]$,
 where m_T is a sequence of normalizing constants with $m_T \rightarrow \infty$;

- 1.b (*Differentiability*) $\psi(\cdot; \theta)$ is continuous on Θ and twice differentiable on $\overset{\circ}{\Theta}$;
- 1.c (*Gradient*) $\left\| \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = O_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm;
- 1.d (*Hessian*) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ for some open neighborhood $\mathcal{V}(\theta_0)$ around θ_0 ;
- 1.e (*Initial Condition*) Given sequences $\{s_t\}$ and $\{c_t\}$, we have

$$\begin{aligned} m_T(\psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) - \psi(X_T, X_{T-1}, \dots; \theta_0)) &= o_p(1), \\ \left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| &= o_p(1), \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| &= o_p(1) \end{aligned}$$

for any $t_1 \geq 1$ such that $(T - t_1)/l_T \rightarrow \infty$ as $T \rightarrow \infty$ and for some model-specific l_T with $l_T \rightarrow \infty$.

Assumption 2. (*Two Independent Processes*)

- 2.a (*Existence*) $\{Y_t\}$ is a process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, distributed as $\{X_t\}$;
- 2.b (*Independence*) $\{Y_t\}$ is independent of $\{X_t\}$.

Assumption 3. (*SPL Estimator*)

- 3.a (*Rates*) The functions $T_P : \mathbb{N} \rightarrow \mathbb{N}$ and $T_E : \mathbb{N} \rightarrow \mathbb{N}$ satisfy $T_E(T) < T_P(T)$ for all T , while $\frac{T - T_P(T)}{l_T} \rightarrow \infty$ and $m_{T_E(T)}/m_T \rightarrow 1$ as $T \rightarrow \infty$;
- 3.b (*Strict Stationarity*) $\{X_t\}$ is a strictly stationary process;
- 3.c (*Weak Dependence*) $\{X_t\}$ satisfies for each bounded, real-valued Lipschitz func-

tion h on \mathbb{R}^r

$$\int h d\left(G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) - G_{T_E}^{SPL}\right) \xrightarrow{p} 0 \quad \text{as} \quad T \rightarrow \infty,$$

where $G_{T_E}^{SPL}$ denotes the unconditional cumulative distribution function (cdf) of $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$ and $G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T})$ the corresponding conditional cdf given the σ -algebra $\mathcal{I}_{T_P:T} = \sigma(X_t : T_P \leq t \leq T)$.

Assumption 4. (*CDF Estimator*) Let $\hat{G}_T(\cdot)$ denote a random (r -dimensional) cdf as a function of $\mathbf{X}_{1:T}$, used to estimate G_∞ . Then $\int h d\hat{G}_T(\cdot) \xrightarrow{p} \int h dG_\infty$ as $T \rightarrow \infty$ for each bounded, real-valued Lipschitz function h on \mathbb{R}^r .

Assumption 5. (*Normality*) Let G_∞ be the cdf of the $N(0, \Upsilon_0)$ distribution with $\Upsilon_0 = \Upsilon(\theta_0, \xi_0)$ and assume there exist $\hat{\Upsilon}(\mathbf{X}_{1:T})$ converging in probability to Υ_0 .

Assumption 1 ensures that the prediction function is well behaved and that one can estimate the parameter it depends on. Whereas Assumption 2 formalizes the unrealistic two-independent-processes assumption, the stationarity and weak dependence condition in Assumption 3 allow to split the sample into (asymptotically) independent and identical subsamples. The consistent estimation of the asymptotic distribution of the parameter estimator, G_∞ , is stated in Assumption 4, which simplifies in the case of asymptotic normality (Assumption 5).

In the following sections we formally verify the high-level assumptions stated above for a wide class of popular time series models satisfying this framework. Since the subsequently considered ARMA and GARCH models exhibit an exponential decay in memory we henceforth set $l_T = \log T$. Further, we constrain ourselves to \sqrt{T} -consistent estimators of the parameters such that $m_T = \sqrt{T}$ throughout the paper.

3 Conditional Mean in an AR(1)

3.1 Model Description

An autoregressive model represents a process in terms of its lagged value(s) and some stochastic innovation process. The first order autoregressive process without drift is defined by the following recursion

$$X_t = \beta_0 X_{t-1} + \varepsilon_t, \quad (3.1)$$

for $t \in \mathbb{Z}$, where the parameter $\beta_0 \in \Theta$ satisfies $|\beta_0| < 1$ and $\{\varepsilon_t\}$ is a sequence of innovations. Subsequently, we make the following assumptions.

Assumption 6. (*AR(1)-Model*)

6.1 (*Compactness*) Θ is compact;

6.2 (*Interior*) $\beta_0 \in \mathring{\Theta}$, where $\mathring{\Theta}$ denotes the interior of Θ ;

6.3 (*Causality*) $|\beta| < 1$ for all $\beta \in \Theta$;

6.4 (*Innovations*) ε_t are *i.i.d.* from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive on \mathbb{R} ;

Θ is assumed to be compact in Assumption 6.1, which holds true, for instance, if it is of the form $\Theta = \{\beta' \in \mathbb{R} : |\beta| \leq 1 - \delta\}$, where $\delta > 0$ is a sufficiently small constant. Assumption 6.2 states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The causality condition is stated in 6.3. Assumption 6.4

imposes further restrictions on the distribution of the innovation process. Next, we turn to the estimation of the model.

3.2 Estimation

To estimate the model in equation (3.1), we employ the OLS estimator given by

$$\hat{\beta}(\mathbf{X}_{1:T}) = \frac{\sum_{t=2}^T X_t X_{t-1}}{\sum_{t=2}^T X_{t-1}^2} \quad (3.2)$$

As the sample size grows large, the OLS estimator approaches a normal distribution under regulatory conditions.

Theorem 1. *(Hamilton, 1994) Under Assumption 6*

$$\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T}) - \beta_0) \xrightarrow{d} N(0, \sigma_\beta^2) \quad (3.3)$$

with $\sigma_\beta^2 = 1 - \beta_0^2$.

3.3 Mapping

The mapping of the AR(1) process into the general framework is straightforward: β_0 corresponds to θ_0 and the conditional mean of X_{T+1} is equal to

$$\psi_{T+1} = \psi(X_T, X_{T-1}, \dots; \theta_0) = \beta_0 X_T. \quad (3.4)$$

3.4 Verification of Assumptions

Assumption 1

For Assumption 1.a to be met, we consider the OLS estimator in (3.2), whose asymptotic distribution is specified in Theorem 1.

As the function $\psi(\dots; \theta)$ given in (3.4) is continuous on Θ and twice differentiable on $\mathring{\Theta}$, Assumption 1.b is met.

Consider Assumption 1.c and notice that the gradient simplifies to

$$\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} = X_T.$$

Clearly, X_T is $O_p(1)$ since the process $\{X_t\}$ is strictly stationary; see also Assumption 3.c, which is verified below.

The condition in Assumption 1.d is met as

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = 0.$$

Regarding Assumption 1.e, we obtain for $t_1 < T$

$$m_T \left(\psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) - \psi(X_T, X_{T-1}, \dots; \theta_0) \right) = \sqrt{T}(\beta_0 X_T - \beta_0 X_T) = 0$$

and

$$\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = |X_T - X_T| = 0$$

as well as

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = |0 - 0| = 0,$$

which completes the verification of Assumption 1.

Assumption 3

The condition in Assumption 3.a is satisfied for instance by $T_E(T) \sim T - \lfloor T^b \rfloor$ and $T_P(T) \sim T - \lfloor T^a \rfloor$ with $0 < a < b < 1$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x .

The process $\{X_t\}$ is strictly stationary since $|\beta_0| < 1$ and $\mathbb{E} \log^+ |\varepsilon_t| \leq \mathbb{E} |\varepsilon_t| < \infty$, where $\log^+ x = \max\{\log x, 0\}$ (Bougerol and Picard, 1992, Thm. 4.1).

The process $\{X_t\}$ is β -mixing with exponential decay (Mokkadem, 1988, Thm. 1'). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 3.c is met with regard to remark 3 of BHS and noting that $T_P(T) - T_E(T) \sim \lfloor T^b \rfloor - \lfloor T^a \rfloor \rightarrow \infty$ as $T \rightarrow \infty$. For alternative mixing results we refer to Davidson (1994, Thm. 14.9) or Andrews (1983, Thm. 1).

Assumptions 4 and 5

Assumption 4 is implied by Assumption 5, which, in turn, is verified by Theorem 1 and $\hat{\sigma}_\beta^2(\mathbf{X}_{1:T}) = 1 - \hat{\beta}(\mathbf{X}_{1:T})^2 \xrightarrow{p} \sigma_\beta^2$.

Assumptions within Corollary 1 of BHS

We show $1/\hat{v}_T^{2IP} = O_p(1)$. By independence of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, the law of $X_T = \sum_{k=0}^{\infty} \beta_0^k \varepsilon_{T-k}$ is equal to $\mathcal{L}(X_T) = \mathcal{L}(\varepsilon_T) * \mathcal{L}(\beta_0 \varepsilon_{T-1}) * \mathcal{L}(\beta_0^2 \varepsilon_{T-2}) * \dots$. As $\mathcal{L}(\varepsilon_t)$ is continuous and non-degenerate, so is $\mathcal{L}(X_T)$, which does not depend of T as $\{X_t\}_{t \in \mathbb{Z}}$ is strictly stationary. It follows that X_T is bounded away from zero. Further, write $\hat{v}_T^{2IP} = X_T^2 \hat{\sigma}_\beta^2(\mathbf{X}_{1:T}) = X_T^2 \sigma_\beta^2 + S_T$ and note that $S_T = X_T^2 (\hat{\sigma}_\beta^2(\mathbf{X}_{1:T}) - \sigma_\beta^2) = o_p(1)$. For every $\epsilon > 0$, we have

$$\begin{aligned} \mathbb{P}[\hat{v}_T^{2IP} \geq \epsilon] &\geq \mathbb{P}[X_T^2 \sigma_\beta^2 + S_T \geq \epsilon \cap |S_T| \leq \epsilon] \\ &\geq \mathbb{P}[X_T^2 \sigma_\beta^2 \geq 2\epsilon \cap |S_T| \leq \epsilon] \\ &\geq \mathbb{P}[X_T^2 \sigma_\beta^2 \geq 2\epsilon] - \mathbb{P}[|S_T| > \epsilon], \end{aligned}$$

where the last inequality follows from $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] - \mathbb{P}[B^c]$. Fix $\delta > 0$; since X_T and hence X_T^2 are bounded away from zero and $\sigma_\beta^2 > 0$, there exists an $\epsilon = \epsilon(\delta)$ such that $\mathbb{P}[X_T^2 \geq 2\epsilon/\sigma_\beta^2] \geq 1 - \delta/2$. For such ϵ , there exists an $\bar{T} = \bar{T}(\epsilon(\delta), \delta) = \bar{T}(\delta)$ such that $\mathbb{P}[|S_T| > \epsilon] < \delta/2$ for all $T \geq \bar{T}$ since $S_T = o_p(1)$. It follows that $\mathbb{P}[\hat{v}_T^{2IP} \geq \epsilon] \geq 1 - \delta$ for all $T \geq \bar{T}$. As $\delta > 0$ was arbitrarily chosen, this completes the proof of \hat{v}_T^{2IP} being bounded away from zero. The proof of \hat{v}_T^{SPL} being bounded away from zero is analogous and hence omitted.

4 Conditional Variance in a GARCH(1,1)

4.1 Model Description

Autoregressive conditional heteroscedasticity models were originally introduced by Engle (1982) and extended to GARCH models by Bollerslev (1986). The model reflects the predominant characteristics of financial returns justifying its popularity among practitioners. The model's temporal dependence structure captures the slow decaying autocorrelations of absolute financial returns, also known as volatility clustering. The GARCH(1,1) process $\{X_t\}$ is defined by

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega_0 + \alpha_0 X_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \end{aligned} \tag{4.1}$$

for all $t \in \mathbb{Z}$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ are non-negative parameters in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. In the traditional GARCH model, Bollerslev (1986) assumed the innovations $\{\varepsilon_t\}$ to be independent following a standard normal distribution. The normality assumption is commonly relaxed to account for stylized statistical properties of financial returns such as *skewness* due to leverage effects and *kurtosis*, also known as fat tails. We denote by $\theta = (\omega, \alpha, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 7. (*GARCH(1,1)-Model*)

7.1 (*Compactness*) Θ is compact;

7.2 (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;

7.3 (*Non-negativity*) $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ for all $\theta \in \Theta$;

7.4 (*Strict Stationarity*) $\mathbb{E}[\ln(\alpha_0 \varepsilon_t^2 + \beta_0)] < 0$ and $\beta < 1$ for all $\theta \in \Theta$;

7.5 (*Roots*) $\alpha_0 z > 0$ and $1 - \beta_0 z > 0$ have no common root, and $\alpha_0 > 0$;

7.6 (*Innovations*) ε_t are *i.i.d.* from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = 1$ and $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive in a neighborhood of zero;

Θ is assumed to be compact in Assumption 7.1, which holds true, for instance, if it is of the form $\Theta = [\delta, 1/\delta] \times [0, 1/\delta] \times [0, 1 - \delta]$, where $\delta \in (0, 1)$ is a sufficiently small constant. Assumption 7.2 states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The non-negativity constraints in 7.3 are standard ensuring the conditional variance to be strictly positive. Assumption 7.4 is necessary and sufficient for $\{X_t\}$ being strictly stationary (cf. Francq and Zakoïan, 2011, Thm. 2.1). The root condition in 7.5 guarantees that the GARCH model is irreducible. Assumption 7.6 imposes further restrictions on the moments and density of the innovation process. Next, we turn to the estimation of the model in (4.1).

4.2 Estimation

We consider the quasi maximum likelihood (QML) estimator proposed by Francq and Zakoïan (2004) to estimate the GARCH(1, 1) model. For a generic $\theta \in \Theta$ we set

$$\sigma_{t+1}^2(\theta) = \sum_{k=0}^{\infty} \beta^k (\omega + \alpha X_{t-k}^2) \quad (4.2)$$

and note that $\sigma_{t+1}^2 = \sigma_{t+1}^2(\theta_0)$. Replacing the unknown presample observations by arbitrary values, say s_t , $t \leq 0$, we denote the modified version of (4.2) by $\tilde{\sigma}_{t+1}^2(\theta)$.

Then the QML estimator of θ_0 is defined as any measurable solution $\hat{\theta}(\mathbf{X}_{1:T})$ of

$$\hat{\theta}(\mathbf{X}_{1:T}) = \arg \max_{\theta \in \Theta} \tilde{L}_T(\theta; \mathbf{X}_{1:T}) \quad (4.3)$$

with

$$\tilde{L}_T(\theta; \mathbf{X}_{1:T}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\theta)}} \exp\left(-\frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)}\right).$$

Assumption 7 implies that the estimator follows asymptotically a normal distribution.

Theorem 2. (*Francq and Zakoïan, 2004*) *Under Assumption 7*

$$\sqrt{T}(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0), \quad (4.4)$$

where $\Upsilon_0 = (\mathbb{E}[\varepsilon_t^4] - 1) \mathbb{E} \left[\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} \right]^{-1}$ and $\sigma_t^2(\theta)$ is given in (4.2).

It is worth stressing that Υ_0 does not only depend on θ_0 but also on some nuisance parameters such as $\mathbb{E}[\varepsilon_t^4]$.

4.3 Mapping

Having described the model and its estimation, we turn to map the model into the general setup. The conditional variance σ_{T+1}^2 is equal to

$$\psi_{T+1} = \psi(X_T, X_{T-1}, \dots; \theta_0) = \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0 X_{T-k}^2). \quad (4.5)$$

To verify Assumption 1 the first and second derivatives of $\psi(X_T, X_{T-1}, \dots; \theta)$ w.r.t. θ are needed. The first order derivatives are

$$\begin{aligned}\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega} &= \frac{1}{1 - \beta}, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha} &= \sum_{k=0}^{\infty} \beta^k X_{T-k}^2, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} (\omega + \alpha X_{T-k}^2),\end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned}\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} &= 0, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha} &= 0, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} &= 0, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \beta} &= -\frac{1}{(1 - \beta)^2}, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{T-k}^2, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} &= \sum_{k=2}^{\infty} k(k-1) \beta^{k-2} (\omega + \alpha X_{T-k}^2).\end{aligned}$$

4.4 Verification of Assumptions

Before turning to the verification of the high-level assumptions, note that the strict stationarity condition implies the existence of fractional moments: there exists an $s \in (0, 1)$ such that $\mathbb{E}X_t^{2s} < \infty$ (Nelson, 1990, Thm. 2). For such $s \in (0, 1)$ the following elementary inequalities hold: $(a + b)^s \leq a^s + b^s$ for all $a, b \geq 0$ and $c^s \leq c$ for all $c \geq 1$.

Assumption 1

For Assumption 1.a to be met, we consider the QML estimator of Francq and Zakoïan (2004), whose asymptotic distribution is specified in Theorem 2.

As the function $\psi(\dots; \theta)$, given in (4.5), is continuous on Θ and twice differentiable on $\mathring{\Theta}$, Assumption 1.b is satisfied.

Consider Assumption 1.c and note that

$$\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\omega} = \frac{1}{1 - \beta_0}$$

is trivially $O(1)$. For showing $\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha}\right| \geq M\right] < \epsilon$ for T sufficiently large. Employing the Markov inequality, we obtain

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{T-k}^2\right)^s\right] \\ &\leq \frac{1}{M^s} \sum_{k=0}^{\infty} \beta_0^{sk} \mathbb{E}X_t^{2s} = \frac{\mathbb{E}X_t^{2s}}{(1 - \beta_0^s)M^s} \end{aligned}$$

such that $M > \left(\frac{\mathbb{E}X_t^{2s}}{(1 - \beta_0^s)\epsilon}\right)^{1/s}$ gives the desired result. Similarly, we get

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\beta}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=1}^{\infty} k\beta_0^{k-1}(\omega_0 + \alpha_0 X_{T-k}^2)\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\sum_{k=1}^{\infty} k\beta_0^{s(k-1)}(\omega_0^s + \alpha_0^s X_{T-k}^{2s})\right] = \frac{\omega_0^s + \alpha_0^s \mathbb{E}X_t^{2s}}{M^s(1 - \beta_0^s)^2} \end{aligned}$$

such that $M > \left(\frac{\omega_0^s + \alpha_0^s \mathbb{E}X_t^{2s}}{\epsilon(1 - \beta_0^s)^2}\right)^{1/s}$ establishes $\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\beta} = O_p(1)$, which completes

the verification of Assumption 1.c.

Focusing on Assumption 1.d we notice that

$$\begin{aligned} \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} \right| &= 0, \\ \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha} \right| &= 0, \\ \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \beta} \right| = \frac{1}{(1 - \beta_{\text{sup}})^2} = O(1),$$

where $\beta_{\text{sup}} = \sup_{\theta \in \mathcal{Y}(\theta_0)} \beta$. To show $\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| = O_p(1)$, we need to find an M for every $\epsilon > 0$ such that $\mathbb{P} \left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] < \epsilon$ holds.

We find

$$\begin{aligned} \mathbb{P} \left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M \right] &\leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^2 \right)^s \right] \\ &\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\text{sup}}^{s(k-1)} X_{T-k}^{2s} \right] \leq \frac{1}{M^s} \sum_{k=1}^{\infty} k \beta_{\text{sup}}^{s(k-1)} \mathbb{E} X_t^{2s} = \frac{\mathbb{E} X_t^{2s}}{M^s (1 - \beta_{\text{sup}}^s)^2}. \end{aligned}$$

Taking $M > \left(\frac{\mathbb{E}X_t^{2s}}{\epsilon(1-\beta_{\text{sup}}^s)^2} \right)^{1/s}$ leads to the desired result. Similarly, we have

$$\begin{aligned}
& \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} \right| \geq M \right] \\
&= \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=2}^{\infty} k(k-1) \beta_{\text{sup}}^{k-2} (\omega_{\text{sup}} + \alpha_{\text{sup}} X_{T-k}^2) \right)^s \right] \\
&\leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=2}^{\infty} k^s (k-1)^s \beta_{\text{sup}}^{s(k-2)} (\omega_{\text{sup}}^s + \alpha_{\text{sup}}^s X_{T-k}^{2s}) \right] \\
&\leq \frac{1}{M^s} \sum_{k=2}^{\infty} k(k-1) \beta_{\text{sup}}^{s(k-2)} (\omega_{\text{sup}}^s + \alpha_{\text{sup}}^s \mathbb{E}X_t^{2s}) = \frac{2(\omega_{\text{sup}}^s + \alpha_{\text{sup}}^s \mathbb{E}X_t^{2s})}{M^s(1-\beta_{\text{sup}}^s)^3},
\end{aligned}$$

where $\omega_{\text{sup}} = \sup_{\theta \in \mathcal{V}(\theta_0)} \omega$ and $\alpha_{\text{sup}} = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha$. Taking $M > \left(\frac{2(\omega_{\text{sup}}^s + \alpha_{\text{sup}}^s \mathbb{E}X_t^{2s})}{\epsilon(1-\beta_{\text{sup}}^s)^3} \right)^{1/s}$ completes the verification of Assumption 1.d.

Regarding Assumption 1.e we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e. $c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_T, X_{T-1}, \dots; \theta) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta) = \sum_{k=T-t_1+1}^{\infty} \beta^k \alpha X_{T-k}^2.$$

We have

$$m_T \left(\psi(X_T, X_{T-1}, \dots; \theta_0) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) \right) = \sqrt{T} \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^k \alpha_0 X_{t_1-k}^2. \quad (4.6)$$

Clearly, the sum is of order $O_p(1)$. Further, for any $t_1 \geq 1$ such that $(T-t_1)/l_T \rightarrow \infty$ we get $\sqrt{T} \beta_0^{T-t_1} \rightarrow 0$. Hence, (4.6) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

and

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \alpha} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha} \right| = \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^2$$

being $o_p(1)$ since the sum is $O_p(1)$ and $\beta_0^{T-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned} & \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta} \right| = \sum_{k=T-t_1+1}^{\infty} k \beta_0^{k-1} \alpha_0 X_{T-k}^2 \\ & = (T-t_1) \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^{k-1} \alpha_0 X_{t_1-k}^2 + \beta_0^{T-t_1} \sum_{k=1}^{\infty} k \beta_0^{k-1} \alpha_0 X_{t_1-k}^2 \end{aligned}$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| = (0, 0, 0), \\ & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^2} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} \right| = 0 \end{aligned}$$

and

$$\begin{aligned} & \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha \partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| = \sum_{k=T-t_1+1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^2 \\ & = (T-t_1) \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} \beta_{\text{sup}}^{k-1} X_{t_1-k}^2 + \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{t_1-k}^2 \end{aligned}$$

is $o_p(1)$ by previous arguments noting that $\beta_{\text{sup}} \in (0, 1)$. Similarly, it can be shown

that

$$\begin{aligned} & \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \beta \partial \beta'} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta \partial \beta'} \right| \\ &= \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \sum_{k=T-t_1+1}^{\infty} k(k-1) \beta^{k-2} \alpha X_{T-k}^2 \right| \leq \sum_{k=T-t_1+1}^{T-1} k(k-1) \beta_{\text{sup}}^{k-2} \alpha_{\text{sup}} X_{T-k}^2 \end{aligned}$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{Y}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 3

The condition in Assumption 3.a is satisfied for instance by $T_E(T) \sim T - \lfloor T^b \rfloor$ and $T_P(T) \sim T - \lfloor T^a \rfloor$ with $0 < a < b < 1$.

With regard to Assumption 7.4, $\{X_t\}$ is a strictly stationary process such that Assumption 3.b is satisfied.

The process $\{X_t\}$ is β -mixing with exponential decay (Francq and Zakoïan, 2011, Thm. 3.4). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 3.c is met with regard to remark 3 of BHS noting that $T_P(T) - T_E(T) \rightarrow \infty$.

Assumptions 4 and 5

Assumption 4 is implied by Assumption 5, which, in turn, is verified by Theorem 2 and the consistent¹ estimator

$$\hat{\Upsilon}(\mathbf{X}_{1:T}) = \left(\frac{1}{T} \sum_{t=1}^T \frac{X_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:T}))} - 1 \right) \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:T}))} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta'} \right)^{-1}.$$

Assumptions within Corollary 1 of BHS

To show $1/\hat{v}_T^{2IP} = O_p(1)$, recall that $\hat{v}_T^{2IP} = v_T^{2IP} + o_p(1)$ (see proof of Corollary 2 of BHS) and define $\kappa = \text{eig}_{\min} \Upsilon_0$, the minimum eigenvalue of Υ_0 . Since Υ_0 is positive definite, we have $\kappa > 0$ such that $1/\hat{v}_T^{2IP} = O_p(1)$ is implied by

$$\hat{v}_T^{2IP} + o_p(1) = v_T^{2IP} \geq \kappa \left\| \frac{\partial \psi_{T+1}}{\partial \theta} \right\|^2 \geq \kappa \left| \frac{\partial \psi_{T+1}}{\partial \omega} \right|^2 \geq \kappa.$$

Similarly, we obtain $\hat{v}_T^{SPL} + o_p(1) \geq \kappa$ such that $1/\hat{v}_T^{SPL} = O_p(1)$.

5 Conditional Mean in an ARMA(1, 1)

5.1 Model Description

The ARMA model was popularized by the classical book of Box and Jenkins (1971). It represents a stationary stochastic process in terms of an autoregressive and a moving-average part. The ARMA(1, 1) process with drift is given by

$$X_t - \omega_0 = \alpha_0 \varepsilon_{t-1} + \beta_0 (X_{t-1} - \omega_0) + \varepsilon_t \quad (5.1)$$

¹A formal proof of consistency under Assumption 7 is along the lines of the intermediary results (iv) and (vi) included in Francq and Zakoïan (2004, Thm. 2.2).

for $t \in \mathbb{Z}$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ is a parameter vector in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. We denote by $\theta = (\omega, \alpha, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 8. (*ARMA(1,1)-Model*)

8.1 (*Compactness*) Θ is compact;

8.2 (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;

8.3 (*Invertibility*) $|\alpha| < 1$ for all $\theta \in \Theta$;

8.4 (*Causality*) $|\beta| < 1$ for all $\theta \in \Theta$;

8.5 (*Roots*) $1 - \beta z$ and $1 + \alpha z$ have no common root, and $\alpha, \beta \neq 0$ for all $\theta \in \Theta$;

8.6 (*Innovations*) ε_t are *i.i.d.* from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = \sigma_\varepsilon^2 < \infty$ and having a Lebesgue density strictly positive on \mathbb{R} ;

Θ is assumed to be compact in Assumption 8.1, which holds true, for instance, if it is of the form $\Theta = \{(\omega, \alpha, \beta)' \in \mathbb{R}^3 : |\omega| \leq \delta^{-1}, \delta \leq |\alpha| \leq 1 - \delta \text{ and } \delta \leq |\beta| \leq 1 - \delta\}$, where $\delta > 0$ is a sufficiently small constant. Assumption 8.2 states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The invertibility and causality conditions are stated in 8.3 and 8.4. Assumption 8.5 ensures that the ARMA model is irreducible. Assumption 8.6 imposes further restrictions on the distribution of the innovation process. Next, we turn to the estimation of the model.

5.2 Estimation

To estimate the model in equation (5.1), we consider a least squares estimator in the spirit of Brockwell and Davis (1991).² Other estimators such as the QML estimator based on the Gaussian likelihood can alternatively be considered. Let $G_T(\alpha_0, \beta_0)$ be the correlation matrix of $(X_1, \dots, X_T)'$ with elements given by

$$\text{Cor}(X_t, X_{t-k}) = \frac{(\alpha_0 + \beta_0)(1 + \alpha_0\beta_0)}{1 + 2\alpha_0\beta_0 + \alpha_0^2} \beta_0^{k-1}$$

for $k \geq 1$. The (weighted) least squares estimator of θ_0 is given by

$$\hat{\theta}(\mathbf{X}_{1:T}) = \arg \min_{\theta \in \Theta} (\mathbf{X}_{1:T} - \omega \iota_T)' G_T^{-1}(\alpha, \beta) (\mathbf{X}_{1:T} - \omega \iota_T). \quad (5.2)$$

with $\iota_T = (1, \dots, 1)' \in \mathbb{R}^T$. As the sample size grows large, the estimator approaches a normal distribution under regulatory conditions.

Theorem 3. (Brockwell and Davis, 1991; Bao, 2018) Under Assumption 8

$$\sqrt{T}(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0) \quad (5.3)$$

with

$$\Upsilon_0 = \begin{pmatrix} \frac{\sigma_\varepsilon^2(1-\alpha_0)^2}{(1-\beta_0)^2} & 0 & 0 \\ 0 & \frac{(1-\alpha_0\beta_0)^2(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} & \frac{(1-\alpha_0^2)(1-\alpha_0\beta_0)(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} \\ 0 & \frac{(1-\alpha_0^2)(1-\alpha_0\beta_0)(1-\beta_0^2)}{(\alpha_0-\beta_0)^2} & \frac{(1-\alpha_0\beta_0)^2(1-\alpha_0^2)}{(\alpha_0-\beta_0)^2} \end{pmatrix}. \quad (5.4)$$

It is worth highlighting that Υ_0 does not only depend on $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$, but also on the nuisance parameter σ_ε^2 .

²Brockwell and Davis (1991) consider $\omega = 0$ for simplicity. The extension to $\omega \neq 0$ is straightforward.

5.3 Mapping

Having described the model and its estimation, we write the model in terms of the general framework. The conditional mean of X_{T+1} is equal to

$$\psi_{T+1} = \psi(X_T, X_{T-1}, \dots; \theta_0) = \omega_0 + \sum_{k=0}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) (X_{T-k} - \omega_0). \quad (5.5)$$

To verify Assumption 1 requires the first and second derivatives of $\psi(X_T, X_{T-1}, \dots; \theta)$ w.r.t. θ . The first order derivatives are

$$\begin{aligned} \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega} &= \frac{1 - \beta}{1 + \alpha}, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=0}^{\infty} (-\alpha)^k (X_{T-k} - \omega), \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha} &= \sum_{k=0}^{\infty} (k + 1) (-\alpha)^k \left((X_{T-k} - \omega) - \beta (X_{T-k-1} - \omega) \right), \end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned} \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} &= 0, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} &= 0, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \beta} &= -\frac{1}{1 + \alpha}, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha} &= -\frac{1 - \beta}{(1 + \alpha)^2}, \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} &= \sum_{k=0}^{\infty} (k + 1) (-\alpha)^k (\omega - X_{T-k-1}), \\ \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} &= -\sum_{k=1}^{\infty} (k + 1) k (-\alpha)^{k-1} \left((X_{T-k} - \omega) - \beta (X_{T-k-1} - \omega) \right). \end{aligned}$$

5.4 Verification of Assumptions

Before turning to the verification of the high-level assumptions, note that $\mathbb{E}|X_t| < \infty$ as the process $\{X_t\}$ is assumed to be causal and $\mathbb{E}|\varepsilon_t| < \infty$.³

Assumption 1

For Assumption 1.a to be met, we consider the least squares estimator in equation (5.2), whose asymptotic distribution is specified in Theorem 3.

As the function $\psi_{T+1}(\dots; \theta)$ given in (5.5) is continuous on Θ and twice differentiable on $\mathring{\Theta}$, Assumption 1.b is met.

Regarding Assumption 1.c we note that

$$\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \omega} = \frac{1 - \beta_0}{1 + \alpha_0}$$

is trivially $O(1)$. To show $\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta}\right| \geq M\right] < \epsilon$ for sufficiently large T . Employing the Markov inequality, we obtain

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta}\right| \geq M\right] &\leq \frac{1}{M} \mathbb{E}\left[\left|\sum_{k=0}^{\infty} (-\alpha_0)^k (X_{T-k} - \omega_0)\right|\right] \\ &\leq \frac{1}{M} \sum_{k=0}^{\infty} |\alpha_0|^k (\mathbb{E}|X_t| + |\omega_0|) = \frac{\mathbb{E}|X_t| + |\omega_0|}{(1 - |\alpha_0|)M} \end{aligned}$$

³As $\{X_t\}$ is causal, we can write it in the MA(∞) representation: $X_t - \omega_0 = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} |\vartheta_j| < \infty$ such that $\mathbb{E}|X_t| \leq |\omega_0| + \mathbb{E}|\varepsilon_t| \sum_{j=0}^{\infty} |\vartheta_j| < \infty$.

such that $M > \frac{\mathbb{E}|X_t| + |\omega_0|}{(1 - |\alpha_0|)\epsilon}$ gives the desired result. Similarly, we find

$$\begin{aligned}
& \mathbb{P} \left[\left| \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha} \right| \geq M \right] \\
& \leq \frac{1}{M} \mathbb{E} \left[\left| \sum_{k=0}^{\infty} (k+1)(-\alpha_0)^k \left((X_{T-k} - \omega_0) - \beta_0(X_{T-k-1} - \omega_0) \right) \right| \right] \\
& \leq \frac{1}{M} \mathbb{E} \left[\sum_{k=0}^{\infty} (k+1)|\alpha_0|^k \left(|X_{T-k}| + |\omega_0| + |\beta_0| (|X_{T-k-1}| + |\omega_0|) \right) \right] \\
& = \frac{1}{M} \sum_{k=0}^{\infty} (k+1)|\alpha_0|^k (\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|) = \frac{(\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|)}{M(1 - |\alpha_0|)^2}
\end{aligned}$$

such that $M > \frac{(\mathbb{E}|X_t| + |\omega_0|)(1 + |\beta_0|)}{\epsilon(1 - |\alpha_0|)^2}$ establishes $\frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha} = O_p(1)$, which completes the verification of Assumption 1.c.

Consider Assumption 1.d and note that

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} \right| = 0, \\
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} \right| = 0, \\
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \beta} \right| = \frac{1}{1 - \alpha_{\text{sup}}} = O(1)
\end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha} \right| \leq \frac{1 + \beta_{\text{sup}}}{1 - \alpha_{\text{sup}}},$$

where $\alpha_{\text{sup}} = \sup_{\theta \in \mathcal{V}(\theta_0)} |\alpha|$ as well as $\beta_{\text{sup}} = \sup_{\theta \in \mathcal{V}(\theta_0)} |\beta|$. To show that the term $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right|$ is $O_p(1)$, we need to find an M for every $\epsilon > 0$ such that

$\mathbb{P}\left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M\right] < \epsilon$ holds for sufficiently large T . We obtain

$$\begin{aligned} & \mathbb{P}\left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \geq M\right] \\ & \leq \frac{1}{M} \mathbb{E}\left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \sum_{k=0}^{\infty} (k+1)(-\alpha)^k (\omega - X_{T-k-1}) \right|\right] \\ & \leq \frac{1}{M} \mathbb{E}\left[\sum_{k=0}^{\infty} (k+1)\alpha_{\text{sup}}^k (|X_{T-k-1}| + \omega_{\text{sup}})\right] \\ & \leq \frac{1}{M} \sum_{k=0}^{\infty} (k+1)\alpha_{\text{sup}}^k (\mathbb{E}|X_t| + \omega_{\text{sup}}) = \frac{\mathbb{E}|X_t| + \omega_{\text{sup}}}{M(1 - \alpha_{\text{sup}})^2}, \end{aligned}$$

where $\omega_{\text{sup}} = \sup_{\theta \in \mathcal{Y}(\theta_0)} |\omega|$. Taking $M > \frac{\mathbb{E}|X_t| + \omega_{\text{sup}}}{\epsilon(1 - \alpha_{\text{sup}})^2}$ leads to the desired result. Similarly, we find

$$\begin{aligned} & \mathbb{P}\left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} \right| \geq M\right] \\ & \leq \frac{1}{M} \mathbb{E}\left[\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \sum_{k=1}^{\infty} (k+1)k(-\alpha)^{k-1} \left((X_{T-k} - \omega) - \beta(X_{T-k-1} - \omega) \right) \right|\right] \\ & \leq \frac{1}{M} \mathbb{E}\left[\sum_{k=1}^{\infty} (k+1)k\alpha_{\text{sup}}^{k-1} \left(|X_{T-k}| + \omega_{\text{sup}} + \beta_{\text{sup}}(|X_{T-k-1}| + \omega_{\text{sup}}) \right)\right] \\ & \leq \frac{1}{M} \sum_{k=1}^{\infty} (k+1)k\alpha_{\text{sup}}^{k-1} (\mathbb{E}|X_t| + \omega_{\text{sup}})(1 + \beta_{\text{sup}}) = \frac{2(\mathbb{E}|X_t| + \omega_{\text{sup}})(1 + \beta_{\text{sup}})}{M(1 - \alpha_{\text{sup}})^3}. \end{aligned}$$

Taking $M > \frac{2(\mathbb{E}|X_t| + \omega_{\text{sup}})(1 + \beta_{\text{sup}})}{\epsilon(1 - \alpha_{\text{sup}})^3}$ establishes that $\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} \right| = O_p(1)$.

Regarding Assumption 1.e we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e. $c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_T, X_{T-1}, \dots; \theta) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta) = \sum_{k=T-t_1+1}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) X_{T-k}.$$

We have

$$\begin{aligned}
& m_T \left(\psi(X_T, X_{T-1}, \dots; \theta_0) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) \right) \\
&= \sqrt{T} (-\alpha_0)^{T-t_1} \sum_{k=1}^{\infty} (-\alpha_0)^k (\alpha_0 + \beta_0) X_{t_1-k}.
\end{aligned} \tag{5.6}$$

Clearly, the sum is of order $O_p(1)$ as $|\alpha_0| < 1$ and $\{X_t\}$ is strictly stationary. Further, for any $t_1 \geq 1$ such that $(T - t_1)/l_T \rightarrow \infty$ we get $\sqrt{T}(-\alpha_0)^{T-t_1} \rightarrow 0$. Hence, (5.6) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

and

$$\begin{aligned}
& \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta} \right| \\
&= \left| \sum_{k=T-t_1+1}^{\infty} (-\alpha_0)^k X_{T-k} \right| \leq |\alpha_0|^{T-t_1} \sum_{k=1}^{\infty} |\alpha_0|^k |X_{t_1-k}|
\end{aligned}$$

being $o_p(1)$ since the sum is $O_p(1)$ and $|\alpha_0|^{T-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned}
& \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \alpha} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha} \right| \\
&= \left| \sum_{k=T-t_1+1}^{\infty} ((k+1)\alpha_0 + k\beta_0)(-\alpha_0)^{k-1} X_{T-k} \right| \\
&\leq \sum_{k=T-t_1+1}^{\infty} ((k+1)|\alpha_0| + k|\beta_0|)|\alpha_0|^{k-1}|X_{T-k}| \\
&\leq 2(|\alpha_0| + |\beta_0|) \sum_{k=T-t_1+1}^{\infty} k|\alpha_0|^{k-1}|X_{T-k}| \\
&= (T-t_1)|\alpha_0|^{T-t_1} 2(|\alpha_0| + |\beta_0|) \sum_{k=1}^{\infty} |\alpha_0|^{k-1}|X_{t_1-k}| \\
&\quad + |\alpha_0|^{T-t_1} 2(|\alpha_0| + |\beta_0|) \sum_{k=1}^{\infty} k|\alpha_0|^{k-1}|X_{t_1-k}|
\end{aligned}$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| = (0, 0, 0), \\
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \beta^2} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} \right| = 0
\end{aligned}$$

and

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha \partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha \partial \beta} \right| \\
&= \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=T-t_1}^{\infty} (k+1)(-\alpha)^k X_{T-k-1} \right| \leq \sum_{k=T-t_1}^{\infty} (k+1) \alpha_{\text{sup}}^k |X_{T-k-1}| \\
&= (T-t_1) \alpha_{\text{sup}}^{T-t_1} \sum_{k=0}^{\infty} \alpha_{\text{sup}}^k |X_{t_1-k-1}| + \alpha_{\text{sup}}^{T-t_1} \sum_{k=0}^{\infty} (k+1) \alpha_{\text{sup}}^k |X_{t_1-k-1}|
\end{aligned}$$

is $o_p(1)$ by previous arguments as $\alpha_{\text{sup}} \in (0, 1)$. Similarly, it can be shown that

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^2} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^2} \right| \\
&= \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \sum_{k=T-t_1+1}^{\infty} (k+1)k(-\alpha)^{k-1} X_{T-k} - \beta \sum_{k=T-t_1}^{\infty} (k+1)k(-\alpha)^{k-1} X_{T-k-1} \right| \\
&\leq \sum_{k=T-t_1+1}^{\infty} (k+1)k \alpha_{\text{sup}}^{k-1} |X_{T-k}| + \beta_{\text{sup}} \sum_{k=T-t_1}^{\infty} (k+1)k \alpha_{\text{sup}}^{k-1} |X_{T-k-1}|
\end{aligned}$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 3

The condition in Assumption 3.a is satisfied for instance by $T_E(T) \sim T - \lfloor T^b \rfloor$ and $T_P(T) \sim T - \lfloor T^a \rfloor$ with $0 < a < b < 1$.

The process $\{X_t\}$ is strictly stationary since $|\beta_0| < 1$ and $\mathbb{E} \log^+ |\varepsilon_t| \leq \mathbb{E} |\varepsilon_t| < \infty$ (Bougerol and Picard, 1992, Thm. 4.1).

The process $\{(\varepsilon_t, X_t)\}$ is β -mixing with exponential decay (Mokkadem, 1988, Thm.

1'). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 3.c is met with regard to remark 3 of BHS noting that $T_P(T) - T_E(T) \rightarrow \infty$. For an alternative mixing result we refer to Davidson (1994, Thm. 14.9).

Assumptions 4 and 5

Assumption 4 is implied by Assumption 5, which, in turn, is verified by Theorem 3 and the consistent estimator

$$\hat{\Upsilon}(\mathbf{X}_{1:T}) = \begin{pmatrix} \hat{\Upsilon}_{11}(\mathbf{X}_{1:T}) & 0 & 0 \\ 0 & \hat{\Upsilon}_{22}(\mathbf{X}_{1:T}) & \hat{\Upsilon}_{23}(\mathbf{X}_{1:T}) \\ 0 & \hat{\Upsilon}_{23}(\mathbf{X}_{1:T}) & \hat{\Upsilon}_{33}(\mathbf{X}_{1:T}) \end{pmatrix},$$

where

$$\begin{aligned} \hat{\Upsilon}_{11}(\mathbf{X}_{1:T}) &= \frac{\hat{\sigma}_\varepsilon^2(\mathbf{X}_{1:T})(1 - \hat{\alpha}(\mathbf{X}_{1:T}))^2}{(1 - \hat{\beta}(\mathbf{X}_{1:T}))^2} \\ \hat{\Upsilon}_{22}(\mathbf{X}_{1:T}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:T})\hat{\beta}(\mathbf{X}_{1:T}))^2(1 - \hat{\beta}(\mathbf{X}_{1:T}))^2}{(\hat{\alpha}(\mathbf{X}_{1:T}) - \hat{\beta}(\mathbf{X}_{1:T}))^2} \\ \hat{\Upsilon}_{23}(\mathbf{X}_{1:T}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:T})^2)(1 - \hat{\alpha}(\mathbf{X}_{1:T})\hat{\beta}(\mathbf{X}_{1:T}))(1 - \hat{\beta}(\mathbf{X}_{1:T}))^2}{(\hat{\alpha}(\mathbf{X}_{1:T}) - \hat{\beta}(\mathbf{X}_{1:T}))^2} \\ \hat{\Upsilon}_{33}(\mathbf{X}_{1:T}) &= \frac{(1 - \hat{\alpha}(\mathbf{X}_{1:T})\hat{\beta}(\mathbf{X}_{1:T}))^2(1 - \hat{\alpha}(\mathbf{X}_{1:T}))^2}{(\hat{\alpha}(\mathbf{X}_{1:T}) - \hat{\beta}(\mathbf{X}_{1:T}))^2} \end{aligned}$$

and

$$\hat{\sigma}_\varepsilon^2(\mathbf{X}_{1:T}) = \frac{1}{T-3} (\mathbf{X}_{1:T} - \hat{\omega}(\mathbf{X}_{1:T})\nu_T)' G_T^{-1}(\hat{\alpha}(\mathbf{X}_{1:T}), \hat{\beta}(\mathbf{X}_{1:T})) (\mathbf{X}_{1:T} - \hat{\omega}(\mathbf{X}_{1:T})\nu_T).$$

Assumptions within Corollary 1 of BHS

To show $1/\hat{v}_T^{2IP} = O_p(1)$, recall that $\hat{v}_T^{2IP} = v_T^{2IP} + o_p(1)$ (see proof of corollary 2 of BHS) and define $\kappa = \text{eig}_{\min} \Upsilon_0$, the minimum eigenvalue of Υ_0 . Since Υ_0 is positive definite, we have $\kappa > 0$. Together with $|\beta_0| < 1$ and $|\alpha_0| < 1$

$$\hat{v}_n^{2IP} + o_p(1) = v_n^{2IP} \geq \kappa \left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta_0)}{\partial \theta} \right\|^2 \geq \kappa \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta_0)}{\partial \omega} \right|^2 = \kappa \frac{(1 - \beta_0)^2}{(1 + \alpha_0)^2} > 0$$

implies $1/\hat{v}_T^{2IP} = O_p(1)$. Analogously, we have $1/\hat{v}_T^{SPL} = O_p(1)$.

6 Conditional Volatility in a T-GARCH(1,1)

6.1 Model Description

The T-GARCH model was first introduced by Zakoian (1994). It accounts for the stylized fact that past positive and negative innovations appear not to have the same impact on current volatility, which is also known as leverage effect. The T-GARCH(1, 1) process $\{X_t\}$ is defined by

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t &= \omega_0 + \alpha_0^+ X_{t-1}^+ + \alpha_0^- X_{t-1}^- + \beta_0 \sigma_{t-1} \end{aligned} \tag{6.1}$$

for all $t \in \mathbb{Z}$ using the notation $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)'$ are non-negative parameters in a parameter set Θ and $\{\varepsilon_t\}$ is a sequence of innovations. We denote by $\theta = (\omega, \alpha^+, \alpha^-, \beta)'$ a generic parameter vector and subsequently make the following assumptions:

Assumption 9. (*T-GARCH(1,1)-Model*)

9.1 (*Compactness*) Θ is compact;

- 9.2 (*Interior*) θ_0 belongs to $\overset{\circ}{\Theta}$;
- 9.3 (*Non-negativity*) $\omega > 0$, $\alpha^+ \geq 0$, $\alpha^- \geq 0$ and $\beta \geq 0$ for all $\theta \in \Theta$;
- 9.4 (*Strict Stationarity*) $\mathbb{E}[\ln(\alpha_0^+ \varepsilon_t^+ + \alpha_0^- \varepsilon_t^- + \beta_0)] < 1$ and $\beta < 1$ for all $\theta \in \Theta$;
- 9.5 (*Roots*) $1 - \beta_0 z > 0$ has no common root with $\alpha_0^+ z$ and $\alpha_0^- z$, and $\alpha_0^+ + \alpha_0^- \neq 0$;
- 9.6 (*Innovations*) ε_t are *i.i.d.* from an absolutely continuous distribution with respect to the Lebesgue measure on \mathbb{R} satisfying $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}[\varepsilon_t^2] = 1$ and $\mathbb{E}[\varepsilon_t^4] < \infty$ and having a Lebesgue density strictly positive in a neighborhood of zero.

Θ is assumed to be compact in Assumption 9.1, which holds true, for instance, if it is of the form $\Theta = [\delta, 1/\delta] \times [0, 1/\delta]^2 \times [0, 1 - \delta]$, where $\delta \in (0, 1)$ is a sufficiently small constant. Assumption 9.2 states that the true parameter vector lies in the interior of the parameter set and is necessary to obtain asymptotic normality of the parameter estimator. The non-negativity constraints in 9.3 are standard ensuring the conditional standard deviation to be strictly positive. Assumption 9.4 is necessary and sufficient for $\{X_t\}$ being strictly stationary (cf. Hamadeh and Zakoïan, 2011). The root condition in 9.5 guarantees that the T-GARCH model is irreducible. Assumption 9.6 imposes further restrictions on the moments and density of the innovation process. Next, we turn to the estimation of the model in (6.1).

6.2 Estimation

We consider the Gaussian QML estimator proposed by Hamadeh and Zakoïan (2011).

For a generic $\theta \in \Theta$ we set

$$\sigma_{t+1}(\theta) = \sum_{k=0}^{\infty} \beta^k (\omega + \alpha^+ X_{t-k}^+ + \alpha^- X_{t-k}^-) \quad (6.2)$$

and note that $\sigma_{t+1} = \sigma_{t+1}(\theta_0)$. Replacing the unknown presample observations by arbitrary values, say s_t , $t \leq 0$, we denote the modified version of (6.2) by $\tilde{\sigma}_{t+1}^2(\theta)$.

Then the QML estimator of θ_0 is defined as any measurable solution $\hat{\theta}(\mathbf{X}_{1:T})$ of

$$\hat{\theta}(\mathbf{X}_{1:T}) = \arg \max_{\theta \in \Theta} \tilde{L}_T(\theta; \mathbf{X}_{1:T}) \quad (6.3)$$

with

$$\tilde{L}_T(\theta; \mathbf{X}_{1:T}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2(\theta)}} \exp\left(-\frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)}\right).$$

Assumption 9 implies that the estimator follows asymptotically a normal distribution.

Theorem 4. *(Hamadeh and Zakoïan, 2011) Under Assumption 9*

$$\sqrt{T}(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) \xrightarrow{d} N(0, \Upsilon_0), \quad (6.4)$$

where $\Upsilon_0 = \frac{1}{4}(\mathbb{E}[\varepsilon_t^4] - 1) \mathbb{E}\left[\frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'}\right]^{-1}$ and $\sigma_t(\theta)$ is given in (6.2).

6.3 Mapping

Having described the model and its estimation, we map the model into the general framework. The conditional volatility σ_{T+1} is equal to

$$\psi_{T+1} = \psi(X_T, X_{T-1}, \dots; \theta_0) = \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{T-k}^+ + \alpha_0^- X_{T-k}^-). \quad (6.5)$$

To verify Assumption 1 the first and second derivatives of $\psi(X_T, X_{T-1}, \dots; \theta)$ w.r.t. θ are needed. The first order derivatives are

$$\begin{aligned} \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega} &= \frac{1}{1 - \beta}, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+} &= \sum_{k=0}^{\infty} \beta^k X_{T-k}^+, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^-} &= \sum_{k=0}^{\infty} \beta^k X_{T-k}^-, \\ \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} (\omega + \alpha^+ X_{T-k}^+ + \alpha^- X_{T-k}^-), \end{aligned}$$

whereas the second order derivatives are given by

$$\begin{aligned}
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega^2} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha^+} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \alpha^-} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^{+2}} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+ \partial \alpha^-} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^{-2}} &= 0 \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \beta} &= \frac{1}{(1 - \beta)^2} \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{T-k}^+ \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} &= \sum_{k=1}^{\infty} k \beta^{k-1} X_{T-k}^- \\
\frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} &= \sum_{k=2}^{\infty} k(k-1) \beta^{k-2} (\omega + \alpha^+ X_{T-k}^+ + \alpha^- X_{T-k}^-)
\end{aligned}$$

6.4 Verification of Assumptions

Before turning to the verification of the high-level assumptions, note that the strict stationarity condition implies the existence of fractional moments: there exists an $s \in (0, 1)$ such that $E|X_t|^s < \infty$ (Hamadeh and Zakoïan, 2011, Prop. A.1).

Assumption 1

For Assumption 1.a to be met, we consider the quasi-maximum likelihood estimator by Hamadeh and Zakoïan (2011), whose asymptotic distribution is specified in Theorem 4.

As the function $\psi(\dots; \theta)$, given in (6.5), is continuous on Θ and twice differentiable on $\overset{\circ}{\Theta}$, Assumption 1.b is satisfied.

Consider Assumption 1.c and note that

$$\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\omega} = \frac{1}{1 - \beta_0}$$

is trivially $O(1)$. For showing $\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha^+} = O_p(1)$, we need to find a finite M for every $\epsilon > 0$ such that $\mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha^+}\right| \geq M\right] < \epsilon$ for T sufficiently large. Markov's inequality implies

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha^+}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{T-k}^+\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k |X_{T-k}|\right)^s\right] \leq \frac{1}{M^s} \sum_{k=0}^{\infty} \beta_0^{sk} \mathbb{E}|X_t|^s = \frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s)M^s} \end{aligned}$$

such that $M > \left(\frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s)\epsilon}\right)^{1/s}$ gives the desired result. The same M serves to show $\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha^-} = O_p(1)$:

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\alpha^-}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k X_{T-k}^-\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \beta_0^k |X_{T-k}|\right)^s\right] \leq \frac{\mathbb{E}|X_t|^s}{(1 - \beta_0^s)M^s} < \epsilon. \end{aligned}$$

Similarly, we get

$$\begin{aligned} \mathbb{P}\left[\left|\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\beta}\right| \geq M\right] &\leq \frac{1}{M^s} \mathbb{E}\left[\left(\sum_{k=1}^{\infty} k\beta_0^{k-1}(\omega_0 + \alpha_0^+ X_{T-k}^+ + \alpha_0^- X_{T-k}^-)\right)^s\right] \\ &\leq \frac{1}{M^s} \mathbb{E}\left[\sum_{k=1}^{\infty} k\beta_0^{s(k-1)}(\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s |X_{T-k}|^s)\right] = \frac{\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s \mathbb{E}|X_t|^s}{M^s(1 - \beta_0^s)^2} \end{aligned}$$

such that $M > \left(\frac{\omega_0^s + (\alpha_0^+ + \alpha_0^-)^s \mathbb{E}|X_t|^s}{\epsilon(1 - \beta_0^s)^2}\right)^{1/s}$ establishes $\frac{\partial\psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial\beta} = O_p(1)$.

Concerning Assumption 1.d we notice that

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\omega^2} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\omega\partial\alpha^+} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\omega\partial\alpha^-} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\alpha^{+2}} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\alpha^+\partial\alpha^-} \right| &= 0, \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\alpha^{-2}} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\omega\partial\beta} \right| = \frac{1}{(1 - \beta_{\sup})^2} = O(1),$$

where $\beta_{\sup} = \sup_{\theta \in \mathcal{V}(\theta_0)} \beta$. To show $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\alpha^+\partial\beta} \right| = O_p(1)$, we need to find an M for every $\epsilon > 0$ such that $\mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2\psi(X_T, X_{T-1}, \dots; \theta)}{\partial\alpha^+\partial\beta} \right| \geq M\right] < \epsilon$ holds.

We obtain

$$\begin{aligned} & \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| \geq M \right] \leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^+ \right)^s \right] \\ & \leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\text{sup}}^{s(k-1)} |X_{T-k}|^s \right] \leq \frac{1}{M^s} \sum_{k=1}^{\infty} k \beta_{\text{sup}}^{s(k-1)} \mathbb{E} |X_t|^s = \frac{\mathbb{E} |X_t|^s}{M^s (1 - \beta_{\text{sup}}^s)^2}. \end{aligned}$$

Taking $M > \left(\frac{\mathbb{E} |X_t|^s}{\epsilon (1 - \beta_{\text{sup}}^s)^2} \right)^{1/s}$ leads to the desired result. The same M serves to prove that $\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| = O_p(1)$ since

$$\begin{aligned} & \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| \geq M \right] \leq \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^- \right)^s \right] \\ & \leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=1}^{\infty} k^s \beta_{\text{sup}}^{s(k-1)} |X_{T-k}|^s \right] \leq \frac{\mathbb{E} |X_t|^s}{M^s (1 - \beta_{\text{sup}}^s)^2} < \epsilon. \end{aligned}$$

Similarly, we have

$$\begin{aligned} & \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \beta^2} \right| \geq M \right] \\ & = \frac{1}{M^s} \mathbb{E} \left[\left(\sum_{k=2}^{\infty} k(k-1) \beta^{k-2} (\omega_{\text{sup}} + \alpha_{\text{sup}}^+ X_{T-k}^+ + \alpha_{\text{sup}}^- X_{T-k}^-) \right)^s \right] \\ & \leq \frac{1}{M^s} \mathbb{E} \left[\sum_{k=2}^{\infty} k^s (k-1)^s \beta_{\text{sup}}^{s(k-2)} (\omega_{\text{sup}}^s + (\alpha_{\text{sup}}^+ + \alpha_{\text{sup}}^-)^s |X_{T-k}|^s) \right] \\ & \leq \frac{1}{M^s} \sum_{k=2}^{\infty} k(k-1) \beta_{\text{sup}}^{s(k-2)} (\omega_{\text{sup}}^s + (\alpha_{\text{sup}}^+ + \alpha_{\text{sup}}^-)^s \mathbb{E} |X_t|^s) = \frac{2(\omega_{\text{sup}}^s + (\alpha_{\text{sup}}^+ + \alpha_{\text{sup}}^-)^s \mathbb{E} |X_t|^s)}{M^s (1 - \beta_{\text{sup}}^s)^3}, \end{aligned}$$

where $\omega_{\text{sup}} = \sup_{\theta \in \mathcal{V}(\theta_0)} \omega$, $\alpha_{\text{sup}}^+ = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha^+$ and $\alpha_{\text{sup}}^- = \sup_{\theta \in \mathcal{V}(\theta_0)} \alpha^-$. Taking $M > \left(\frac{2(\omega_{\text{sup}}^s + (\alpha_{\text{sup}}^+ + \alpha_{\text{sup}}^-)^s \mathbb{E} |X_t|^s)}{\epsilon (1 - \beta_{\text{sup}}^s)^3} \right)^{1/s}$ completes the verification of Assumption 1.d.

Regarding Assumption 1.e we choose $\{c_t\}$ and $\{s_t\}$ to be sequences of zeros, i.e.

$c_t = s_t = 0$ for all $t \in \mathbb{Z}$, and note that

$$\psi(X_T, X_{T-1}, \dots; \theta) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta) = \sum_{k=T-t_1+1}^{\infty} \beta^k (\alpha^+ X_{T-k}^+ + \alpha^- X_{T-k}^-).$$

We have

$$\begin{aligned} m_T \left(\psi(X_T, X_{T-1}, \dots; \theta_0) - \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) \right) \\ = \sqrt{T} \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^k (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-). \end{aligned} \quad (6.6)$$

Clearly, the sum is of order $O_p(1)$. Further, for any $t_1 \geq 1$ such that $(T - t_1)/l_T \rightarrow \infty$ we get $\sqrt{T} \beta_0^{T-t_1} \rightarrow 0$. Hence, (6.6) is $o_p(1)$. Moreover, we obtain

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \omega} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \omega} \right| = 0$$

as well as

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \alpha^+} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha^+} \right| = \sum_{k=T-t_1+1}^{\infty} \beta_0^k X_{T-k}^+ \leq \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^+$$

and

$$\left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \alpha^-} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \alpha^-} \right| = \sum_{k=T-t_1+1}^{\infty} \beta_0^k X_{T-k}^- \leq \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^k X_{t_1-k}^-$$

being $o_p(1)$ since the sums are $O_p(1)$ and $\beta_0^{T-t_1} \rightarrow 0$. Similarly, we find

$$\begin{aligned} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta} \right| &= \sum_{k=T-t_1+1}^{\infty} k \beta_0^{k-1} (\alpha_0^+ X_{T-k}^+ + \alpha_0^- X_{T-k}^-) \\ &= (T - t_1) \beta_0^{T-t_1} \sum_{k=1}^{\infty} \beta_0^{k-1} (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-) + \beta_0^{T-t_1} \sum_{k=1}^{\infty} k \beta_0^{k-1} (\alpha_0^+ X_{t_1-k}^+ + \alpha_0^- X_{t_1-k}^-) \end{aligned}$$

being $o_p(1)$ and we conclude that

$$\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = o_p(1).$$

Further, we get

$$\begin{aligned} \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \omega \partial \theta'} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \omega \partial \theta'} \right| &= (0, 0, 0), \\ \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^{+2}} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^{+2}} \right| &= 0, \\ \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^{-2}} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^{-2}} \right| &= 0 \end{aligned}$$

and

$$\sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^+ \partial \alpha^-} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+ \partial \alpha^-} \right| = 0.$$

In addition, we find

$$\begin{aligned} \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^+ \partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^+ \partial \beta} \right| &= \sum_{k=T-t_1+1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^+ \\ &= (T-t_1) \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} \beta_{\text{sup}}^{k-1} X_{t_1-k}^+ + \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{t_1-k}^+ \end{aligned}$$

and

$$\begin{aligned} \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \alpha^- \partial \beta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \alpha^- \partial \beta} \right| &= \sum_{k=T-t_1+1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{T-k}^- \\ &= (T-t_1) \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} \beta_{\text{sup}}^{k-1} X_{t_1-k}^- + \beta_{\text{sup}}^{T-t_1} \sum_{k=1}^{\infty} k \beta_{\text{sup}}^{k-1} X_{t_1-k}^- \end{aligned}$$

being $o_p(1)$ by previous arguments noting that $\beta_{\text{sup}} \in (0, 1)$. Similarly, it can be

shown that

$$\begin{aligned} & \sup_{\theta \in \mathcal{Y}(\theta_0)} \left| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \beta \partial \beta'} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \beta \partial \beta'} \right| \\ & \leq \sum_{k=T-t_1+1}^{\infty} k(k-1) \beta_{\text{sup}}^{k-2} (\alpha_{\text{sup}}^+ X_{T-k}^+ + \alpha_{\text{sup}}^- X_{T-k}^-) \end{aligned}$$

vanishes in probability to zero and we conclude that

$$\sup_{\theta \in \mathcal{Y}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = o_p(1).$$

Assumption 3

The condition in Assumption 3.a is satisfied for instance by $T_E(T) \sim T - \lfloor T^b \rfloor$ and $T_P(T) \sim T - \lfloor T^a \rfloor$ with $0 < a < b < 1$.

With regard to Assumption 9.4, $\{X_t\}$ is a strictly stationary process such that Assumption 3.b is satisfied.

The process $\{X_t\}$ is β -mixing with exponential decay (Francq and Zakoïan, 2006, Thm. 3). As β -mixing implies α -mixing (cf. Bradley, 2005), Assumption 3.c is met with regard to remark 3 of BHS noting that $T_P(T) - T_E(T) \rightarrow \infty$. For an alternative mixing result we refer to Carrasco and Chen (2002).

Assumptions 4 and 5

Assumption 4 is implied by Assumption 5, which, in turn, is verified by Theorem 4 and the consistent estimator

$$\hat{\Upsilon}(\mathbf{X}_{1:T}) = \left(\frac{1}{T} \sum_{t=1}^T \frac{X_t^4}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:T}))} - 1 \right) \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{\tilde{\sigma}_t^4(\hat{\theta}(\mathbf{X}_{1:T}))} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta'} \right)^{-1}.$$

Assumptions within Corollary 1 of BHS

The verification of $1/\hat{v}_T^{2IP} = O_p(1)$ and $1/\hat{v}_T^{SPL} = O_p(1)$ is analogous to the GARCH(1, 1) case and hence omitted.

7 Concluding Remarks

In this paper we establish the mapping of the conditional mean in an AR(1) and ARMA(1, 1) model into the general setup. Further, the conditional variance and the conditional volatility in a GARCH(1, 1) and T-GARCH(1, 1) model, respectively, are shown to be encompassed in that framework. Further, the theoretical results of BHS are validated by verifying the corresponding assumptions for each model. Clearly, the list of nested models is non-exhaustive and can be extended. For instance one could study higher order models such as the ARMA(p, q) or the GARCH(p, q) model with $p, q \in \mathbb{N}$, which come at the cost of a more evolved analysis. Table 1 enlists four other GARCH-type extensions that are frequently encountered in the literature. The family of quadratic GARCH (Q-GARCH) models has been proposed by Sentana (1995). Its Q-GARCH(1, 1) member is very similar to the GARCH(1, 1) model and can be verified in a similar fashion replacing αX_{t-1}^2 by $\alpha X_{t-1}^2 + \phi X_t$. The GJR-GARCH(1, 1) model named after Glosten, Jagannathan and Runkle (1993) is a variant of the T-GARCH(1, 1), which corresponds to squaring the variables involved.

Model	Conditional variance σ_t^2 specification
E-GARCH(1, 1)	$\ln \sigma_t^2 = \omega + \alpha \frac{X_{t-1}}{\sigma_{t-1}} + \phi \left(\left \frac{X_{t-1}}{\sigma_{t-1}} \right - \mathbb{E} \varepsilon_t \right) X_{t-1}^2 + \beta \ln \sigma_{t-1}^2$
N-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha (X_{t-1} - \phi \sigma_{t-1})^2 + \beta \sigma_{t-1}^2$
GJR-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha^+ X_{t-1}^2 I_{\{X_{t-1} \geq 0\}} + \alpha^-, X_{t-1}^2 I_{\{X_{t-1} < 0\}} + \beta \sigma_{t-1}^2$
Q-GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 + \phi X_{t-1}$

Table 1: GARCH extensions. The respective process $\{X_t\}$ is generated by $X_t = \sigma_t \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of innovations and σ_t^2 is the conditional variance at time t .

It can be easily verified along the lines of Section 6. The exponential GARCH (E-GARCH) model suggested by Nelson (1991) and the non-linear GARCH (N-GARCH) introduced by Engle and Ng (1993) can also be embedded into the framework of BHS. For example, the conditional variance in an N-GARCH(1, 1) given by $\sigma_{T+1}^2 = \omega_0 + \alpha_0 (X_T - \phi_0 \sigma_T)^2 + \beta_0 \sigma_T^2$, where $\theta_0 = (\omega_0, \alpha_0, \beta_0, \phi_0)'$ denotes the parameter vector. However, obtaining an explicit expression for the conditional variance in terms of θ_0 and $\{X_t\}_{t \leq T}$ is complicated due to non-linearities in the recursive formula: e.g. σ_{T+1}^2 depends on σ_T^2 and σ_T in the N-GARCH(1, 1).

There are few GARCH extensions such as the fractionally integrated (FI-GARCH) of Baillie et al. (1996) or the fractionally integrated EGARCH (FIE-GARCH) of Bollerslev and Mikkelsen (1996) that cannot be encompassed in the framework at hand. The corresponding processes typically exhibit intermediate or long memory such that standard mixing results do not apply. Establishing the merging results on the basis of verifying Assumption 3.c directly, instead via some mixing result, is an interesting question, which demands further investigation.

Finally, we would like to emphasize that conditional risk measures such as condi-

tional Value-at-Risk (VaR) can be mapped into the general framework. For instance in the T-GARCH(1,1) model of Section 6, the conditional VaR of X_{T+1} given $\{X_t\}_{t \leq T}$ at level $a \in (0, 1)$ reduces to

$$VaR_a(X_{T+1}|X_T, X_{T-1}, \dots) = -\xi_a \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{T-k}^+ + \alpha_0^- X_{T-k}^-) \quad (7.1)$$

with $\xi_a = \inf \{ \tau \in \mathbb{R} : \mathbb{P}[\varepsilon_t \leq \tau] \geq a \}$; see Francq and Zakoïan (2015) for details. Fixing a and treating ξ_a as additional parameter, (7.1) is a function of $\{X_t\}_{t \leq T}$ and $\vartheta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \xi_a)'$ and hence is nested in the setup of Section 2. Similarly, the conditional Expected Shortfall (ES) of X_{T+1} given $\{X_t\}_{t \leq T}$ at level $a \in (0, 1)$

$$ES_a(X_{T+1}|X_T, X_{T-1}, \dots) = -\mu_a \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ X_{T-k}^+ + \alpha_0^- X_{T-k}^-) \quad (7.2)$$

with $\mu_a = -\mathbb{E}[\varepsilon_t | \varepsilon_t < \xi_a]$ can also be mapped into the general framework.

References

- Andrews, D. W. K. (1983). First order autoregressive processes and strong mixing. *Cowles Foundation Discussion Papers*, 664.
- Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.
- Bao, Y. (2018). The asymptotic covariance matrix of the QMLE in ARMA models. *Econometric Reviews*, 37(4):309–324.
- Beutner, E., Heinemann, A., and Smeekes, S. (2019). A justification of conditional confidence intervals. arXiv e-print 1710.00643.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of econometrics*, 73(1):151–184.

- Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52(1-2):115–127.
- Box, G. E. and Jenkins, G. M. (1971). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2(2):107–144.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18(1):17–39.
- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, Oxford.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.
- Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637.
- Francq, C. and Zakoïan, J.-M. (2006). Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory*, 22(5):815–834.
- Francq, C. and Zakoïan, J.-M. (2011). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, Chichester.
- Francq, C. and Zakoïan, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, 184(1):158–173.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.
- Hamadeh, T. and Zakoïan, J.-M. (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference*, 141(1):488–507.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

- Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications*, 29(2):309–315.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3):318–334.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.
- Sentana, E. (1995). Quadratic ARCH models. *The Review of Economic Studies*, 62(4):639–661.
- Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.