WORKING PAPER SERIES

Noncausal AR processes driven by causal GARCH volatility

Daniel Velasquez-Gaviria, Jean-Michel Zakoïan

N°2/February 2026

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

Noncausal AR processes driven by causal GARCH volatility

Daniel Velásquez-Gaviria* and Jean-Michel Zakoïan†

January 2026

Abstract

This paper studies the introduction of causal conditional heteroskedasticity in noncausal autoregressive (AR) models. We demonstrate that, in this framework, large shocks to the independent innovation that drives the GARCH error term of a noncausal AR(1) model result in heightened volatility following a bubble crash. The non-coincidence of the σ-fields generated by past observations and past values of the GARCH process makes estimation nonstandard. In particular, the full quasi-maximum likelihood estimator (QMLE) is generally inconsistent. We investigate the asymptotic properties of three-step weighted least squares estimators of the AR coefficient and the QMLE of the volatility parameters. Our findings are illustrated via Monte Carlo experiments and real financial data.

JEL Classification: C13, C22 and C58

Keywords: Bubbles, Conditional Heteroskedasticity, Noncausal Autoregression, Quasi-Maximum Likelihood Estimation, Weighted Least Squares.

*School of Business and Economics, Department of Quantitative Economics, Maastricht University, Maastricht, The Netherlands. E-mail: d.velasquezgaviria@maastrichtuniversity.nl

†Corresponding author: Jean-Michel Zakoïan, CREST-ENSAE, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. E-mail: Zakoïan@ensae.fr.

1 Introduction

Noncausal autoregressive (AR) models have recently attracted increasing attention for their ability to capture nonlinear features commonly observed in economic and financial time series (see, among others, Lanne and Saikkonen, 2013; Gouriéroux and Jasiak, 2016, 2017; Gouriéroux and Zakoïan, 2017; Lof and Nyberg, 2017; Velasco and Lobato, 2018; Cavaliere, Nielsen and Rahbek, 2018; Fries and Zakoïan, 2019; Fries, 2020; Davis and Song, 2020; Hecq, Issler and Telg, 2020; Blasques, Koopman and Mingoli, 2023; Hecq and Velásquez-Gaviria, 2025). In particular, Gouriéroux and Zakoïan (2017) show that simple stationary and noncausal AR(1) dynamics can generate local explosive episodes—so-called financial bubbles—that frequently arise in speculative markets.

Other approaches to modelling locally explosive behaviours stem from literature on rational expectations models, see for instance Blanchard and Watson (1982), in which asset prices are described as the sum of two components: fundamental value and bubble component. A recent contribution has been made by Blasques, Koopman and Nientker (2022), who propose an observation-driven, time-varying parameter approach. In contrast, noncausal specifications do not incorporate any latent variables and do not rely on a decomposition of asset prices into two components.

One disadvantage of existing contributions to the non-causal literature is that they predominantly rely on models with independent and homoskedastic error terms. For financial applications, however, this assumption is difficult to justify given that volatility in asset returns is well known to be time-varying and strongly persistent. Two notable exceptions are Zhan, Ling, Liu and Wang (2025) and Catania and Mingoli (2026) who studied noncausal ARCH specifications in which conditional heteroskedasticity depends on leads of the process.

In contrast, the present paper studies noncausal AR processes driven by a *causal* conditional variance. This specification preserves the forward-looking structure underlying noncausal mean dynamics while maintaining the economically intuitive assumption that volatility responds only to past shocks, rather than anticipating future ones. This feature aligns with classical interpretations of volatility clustering and leverage effects, whereby markets adjust to unexpected news arriving over time. We study the effects on the observed process of shocks on the innovation process. As in the homoskedastic case, the noncausal AR part of the model is able to produce bubbles when large shocks arise. In addition, the noncausal GARCH specification of the error term induces increased volatility after the bubble crash.

Introducing a causal GARCH-type volatility into a noncausal AR model leads to several nonstandard inferential challenges. In particular, the σ-fields generated by past observations and past innovations no longer coincide, so that the usual quasi-maximum likelihood (QML) framework becomes misspecified. We show that the full QML estimator is generally inconsistent. To circumvent this issue, we propose a family of simple weighted least-squares (WLS) estimators for the AR parameter that explicitly account for the heteroskedastic structure without requiring full knowledge of the distribution of the innovation term. We derive their asymptotic properties and demonstrate that the resulting estimators can substantially improve efficiency relative to ordinary least squares.

The paper also investigates the asymptotic properties of a QML estimator for the volatility parameters in the standard GARCH setting, as well as the behaviour of the misspecified full QMLE for the joint parameter vector. Monte Carlo experiments illustrate the theoretical results and quantify the accuracy gains associated with the WLS approach. Finally, an empirical application highlights the impact of causal conditional heteroskedasticity combined with noncausal mean dynamics on bubble formation and on post-crash volatility patterns.

The remainder of the paper is organized as follows. In Section 2, we present the model, discuss the σ-fields generated by past observations or past innovations, and analyse the effect of shocks on bubble behaviour. Section 3 studies WLS estimators of the AR coefficient and derives their asymptotic properties under conditional heteroskedasticity of general form. We also propose a portmanteau test for the null assumption of correct specification. In Section 4, we derive a consistent and asymptotically normal estimator of the volatility parameter of a standard GARCH model. Section 5 examines the asymptotic properties of a misspecified full QMLE. Section 6 provides Monte Carlo experiments and Section 7 discusses an empirical application. Section 8 concludes. Proofs and complements on the empirical study are relegated to an appendix.

2 Model and the effect of shocks on bubble formation

We consider the semi-parametric noncausal AR(1) model driven by a causal GARCH-type volatility process,

$$X_t = \varphi_0 X_{t+1} + \epsilon_t, \quad |\varphi_0| < 1, \quad \epsilon_t = \sigma_t \eta_t, \quad \forall t \in \mathbb{Z}, \tag{2.1}$$

where $\{\eta_t\}$ is a sequence of independent and identically distributed random variables (iid) with $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$. The volatility σ_t^2 of the error term ϵ_t is strictly positive and depends only on past values of ϵ_t:

$$\sigma_t^2 = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots), \tag{2.2}$$

for some measurable function $h : \mathbb{R}^\infty \to (0, \infty)$. A standard GARCH(1,1) specification corresponds to

$$\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \qquad \omega_0 > 0,\ \alpha_0, \beta_0 \geq 0. \tag{2.3}$$

As is well known, the presence of a root inside the unit disk in the AR polynomial generates a noncausal solution with locally explosive dynamics. Bubble-like behaviour arises in sample paths, especially when the error term ϵ_t occasionally attains large values. The introduction of a GARCH specification makes such extreme realizations substantially more likely, even when the innovation distribution itself has light or moderately heavy tails. In addition, the volatility dynamics influences the trajectory of the process before and after the bubble collapses, a feature that will play a central role in the analysis below.

2.1 Conditional moments for different information sets

Under appropriate assumptions, the causal strictly stationary solution of the GARCH model has the form

$$\epsilon_t = \sigma_t \eta_t, \qquad \sigma_t = \sigma(\eta_{t-1}, \eta_{t-2}, \ldots), \tag{2.4}$$

for a measurable function $\sigma : \mathbb{R}^\infty \to (0, \infty)$. For most volatility specifications, the strictly stationary solution admits a small moment, $E|\epsilon_t|^s < \infty$ for some $s > 0$. It follows that the process $\{X_t\}$ is strictly stationary and admits the noncausal MA(∞) representation

$$X_t = \sum_{i=0}^{\infty} \varphi_0^i \epsilon_{t+i}. \tag{2.5}$$

If $\epsilon_t \in L^2$, the linear prediction of X_t based on its past values is given by

$$EL(X_t \mid \mathcal{F}_{t-1}^X) = \varphi_0 X_{t-1},$$

where, for any process (Y_t), $\mathcal{F}_{t-1}^Y$ denotes the σ-field generated by the past of $\{Y_t\}$, and $EL()$ denotes the linear projection. However, except in the Gaussian case, the true conditional expectation satisfies

$$E(X_t \mid \mathcal{F}_{t-1}^X) \neq \varphi_0 X_{t-1}.$$

For iid errors, Gouriéroux and Zakoïan (2017) established that $E(X_t|\mathcal{F}_{t-1}^X) = \text{sign}(\varphi_0)X_{t-1}$ when the process (ϵ_t) is Cauchy distributed. In general, $E(X_t|\mathcal{F}_{t-1}^X)$ depends on the law of η_t and cannot be obtained explicitly. The same property holds for the conditional variance $\text{Var}(X_t \mid \mathcal{F}_{t-1}^X)$.

A key implication of (2.5) is that the information sets generated by past observations and past innovations differ:

$$\mathcal{F}_{t-1}^X \neq \mathcal{F}_{t-1}^\eta = \mathcal{F}_{t-1}^\epsilon.$$

In particular, the conditional variance of X_t given its past does not coincide with the GARCH variance σ_t^2:

$$\sigma_t^2 \notin \mathcal{F}_{t-1}^X,$$

and the volatility of $\{X_t\}$ does not follow a GARCH recursion.

Let $\mathcal{F}_t = \mathcal{F}_t^\epsilon$ and denote conditional expectation w.r.t. this σ-field by $E_t(\cdot)$. Using (2.4)-(2.5), we obtain

$$E_{t-1}(X_t) = \sum_{i=0}^{\infty} \varphi_0^i E_{t-1}(\epsilon_{t+i}) = 0, \qquad \text{Var}_{t-1}(X_t) = \sum_{i=0}^{\infty} \varphi_0^{2i} E_{t-1}(\sigma_{t+i}^2),$$

noting that $E_{t-1}(\epsilon_{t+i}\epsilon_{t+j}) = 0$ for $i \neq j$, and using the independence between η_t and $\mathcal{F}_{t-1}$. Thus, unlike in causal AR–GARCH models, the conditional variances of X_t and ϵ_t differ, and the volatility of X_t inherently incorporates forward-looking effects. This pattern is consistent with empirical evidence that volatility remains low during bubble expansion and rises sharply during and after the crash.

2.2 Effect of shocks on bubble formation and collapse

We now examine the effect of a large shock at time t_0, represented by a large innovation η_{t_0}. From the MA(∞) representation (2.5), for any $t \leq t_0$ we may write

$$X_t = \sum_{i=0}^{t_0-t-1} \varphi_0^i \epsilon_{t+i} + \varphi_0^{t_0-t} \sigma_{t_0} \eta_{t_0} + \sum_{i=t_0-t+1}^{\infty} \varphi_0^i \epsilon_{t+i}. \tag{2.6}$$

The decomposition in (2.6) highlights three components:

(i) The first sum is unaffected by the shock at time t_0 as it is independent of η_{t_0}.

(ii) The second term, $\varphi_0^{t_0-t}\sigma_{t_0}\eta_{t_0}$, grows in magnitude as t approaches t_0, generating the characteristic locally explosive behaviour or "bubble". However, note that the scaling variable σ_{t_0}

is independent of the shock; thus the resulting "volatility" will not increase along the sample path.

(iii) The last sum depends on future volatility: $\epsilon_{t+i} = \sigma(\eta_{t+i-1}, \ldots, \eta_{t_0}, \ldots)\, \eta_{t+i}$, so the shock also propagates through its effect on $\sigma(\cdot)$.

For $t > t_0$, we have

$$X_t = \sum_{i=0}^{\infty} \varphi_0^i \sigma(\eta_{t+i-1}, \eta_{t+i-2}, \ldots \eta_{t_0}, \ldots)\eta_{t+i}, \tag{2.7}$$

showing that the effect of the shock persists in the conditional variance after the bubble bursts. This last feature distinguishes the causal volatility specification from a homoskedastic noncausal AR(1): in the latter case, the post-crash behaviour would be independent of η_{t_0}, while here the volatility transmits the shock beyond the crash. In most GARCH-type models, the most recent innovations carry the largest weight, so the volatility typically jumps at the crash and decays gradually afterwards. This pattern is consistent with empirical evidence that volatility remains low during bubble expansion and rises sharply during and after the crash. In particular, Brunnermeier and Oehmke (2013) argue that bubble run-up phases are typically characterized by low volatility, whereas volatility is higher during and following the collapse.

To illustrate the underlying mechanism, we simulate sample paths under both causal and noncausal volatility specifications. In Figure 1, under a causal GARCH volatility specification the bubble builds up under low conditional volatility and collapses with a sharp spike in volatility, consistent with well-documented stylized facts in financial markets. On the same graph, we also report a simulated trajectory of a homoskedastic noncausal AR(1) process with the same autoregressive coefficient and marginal variance. In that benchmark, the conditional variance is constant, so there is no post-crash volatility spike. By contrast, under the causal GARCH specification the volatility jumps at the crash and remains elevated for several periods before gradually reverting.

For the sake of comparison, the right panel of Figure 1 displays a trajectory generated by a noncausal GARCH specification with identical parameters and innovations. In this model, studied by Zhan et al. (2025), both the mean and variance have a noncausal specification. In contrast to Figure 1, volatility in this case anticipates the crash, increasing before the bubble bursts and declining immediately afterward.

Finally, the magnitude of volatility feedback plays a central role. Increasing the GARCH impact

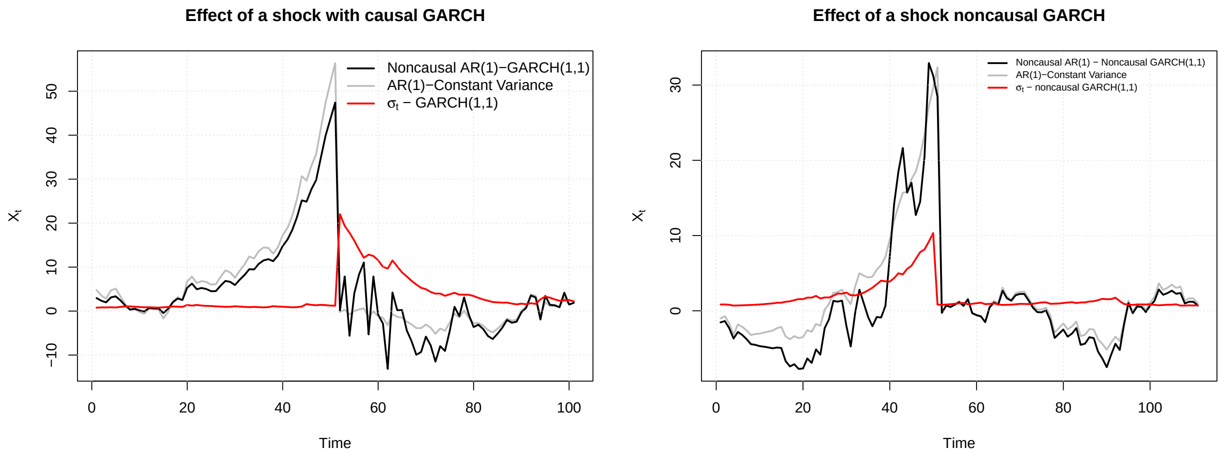

Figure 1: Effect of a large shock at $t_0 = 30$ in the noncausal AR(1) with $\varphi_0 = 0.9$ and (i) left panel: causal GARCH(1,1) volatility Model (2.3) with $\omega_0 = 0.1$, $\alpha_0 = 0.15$, $\beta_0 = 0.78$, and $\eta_t \sim t_{10}$, (ii) right panel: as in the left panel but with the noncausal GARCH(1,1) volatility $\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t+1}^2 + \beta_0 \sigma_{t+1}^2$.

parameter α amplifies the transmission of shocks to future volatility, producing more pronounced clustering, larger post-crash swings, and more frequent extreme realizations (see Figure 2).

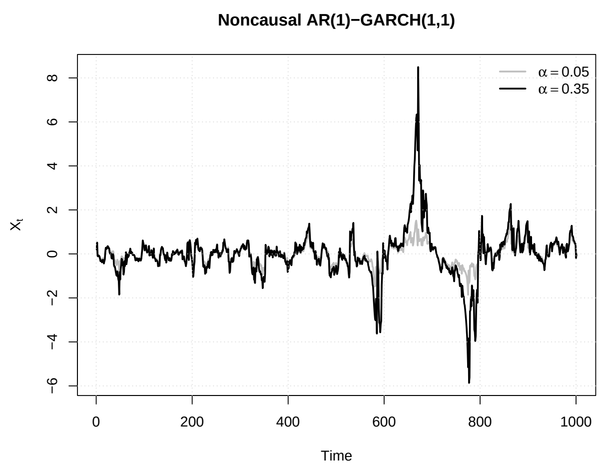

Figure 2: Noncausal AR(1)-GARCH(1,1) process $X_t = 0.9\, X_{t+1} + \eta_t\, \sigma_t$, $\sigma_t^2 = 0.01 + \sigma_{t-1}^2\, (\alpha\, \eta_{t-1}^2 + 0.6)$, with $\alpha \in \{0.05, 0.35\}$, and $\eta_t \sim t_{\nu=2.5}$.

3 Estimation of the AR parameter

We now turn to the estimation of the autoregressive parameter φ_0 in Model (2.1)-(2.2), based on the observed sample $X_1, \dots, X_n$. Even in the case where the volatility process were observable and the distribution of $\{\eta_t\}$ known, exact likelihood-based inference would not be feasible because the conditional moments $E(X_t|\mathcal{F}_{t-1}^X)$ and $\mathrm{V}ar(X_t|\mathcal{F}_{t-1}^X)$ are unknown. This motivates the use of weighted least squares (WLS) methods as a tractable alternative. Self-weighted estimators, using weights that are functions of the past observations, have been studied for a long time in the time series literature (see for instance Klimko and Nelson (1978), Francq, Roy and Saidi (2011), Zhu and Ling (2011), Aknouche, (2013)).

The asymptotic properties of WLS estimators (WLSE) for a broad class of conditional mean models have recently been studied by Aknouche and Francq (2023). Their results, however, cannot be applied here because our AR specification is noncausal. As a consequence, the difference between the WLSE and the true parameter does not admit a martingale difference representation, in contrast with the framework studied in their paper.

3.1 A theoretical one-step WLSE

We start by considering a theoretical WLSE based on a causal weighting function depending on the past values of the iid process. Let $\tau_t = \tau(\eta_{t-1}, \eta_{t-2}, \dots)$ where τ is a measurable function from $\mathbb{R}^\infty$ to $[\underline{\tau}, \infty)$ with $\underline{\tau} > 0$. A WLS estimator (WLSE) of φ_0 is defined as

$$\widehat{\varphi}_{WLS} = \frac{\sum_{t=1}^n \frac{1}{\tau_t} X_t X_{t+1}}{\sum_{t=1}^n \frac{1}{\tau_t} X_{t+1}^2}.$$

The almost sure (a.s.) convergence and asymptotic normality of this estimator are established in the following result.

Proposition 3.1. *Let (X_t) be the noncausal solution of Model (2.1) where (ϵ_t) is the strictly stationary, ergodic, causal solution of (2.4). Then, if $E\epsilon_t^2 < \infty$ we have $\widehat{\varphi}_{WLS} \to \varphi_0$, a.s. Moreover, if $E\epsilon_t^4 < \infty$, we have*

$$\sqrt{n}(\widehat{\varphi}_{WLS} - \varphi_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{WLS}^2 := \mu_\tau^{-2} \sum_{h=-\infty}^{\infty} \gamma(h)\right), \tag{3.1}$$

where $\mu_\tau = E\left(\frac{X_{t+1}^2}{\tau_t}\right)$ *and, for* $h \geq 0$,

$$\gamma(h) = \gamma(-h) \quad = \quad Cov\left(\frac{\epsilon_t X_{t+1}}{\tau_t}, \frac{\epsilon_{t+h} X_{t+1+h}}{\tau_{t+h}}\right) = \varphi_0^h \sum_{j=0}^{\infty} \varphi_0^{2j} E\left(\frac{\epsilon_t \epsilon_{t+h} \sigma_{t+1+h+j}^2}{\tau_t \tau_{t+h}}\right).$$

In particular, for the least-squares estimator (LSE) of φ_0 defined by $\widehat{\varphi}_{LS} = \frac{\sum_{t=1}^n X_t X_{t+1}}{\sum_{t=1}^n X_{t+1}^2}$, we have

$$\sqrt{n}(\widehat{\varphi}_{LS} - \varphi_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{LS}^2 := \left\{E\left(X_{t+1}^2\right)\right\}^{-2} \sum_{h=-\infty}^{\infty} \delta(h)\right), \tag{3.2}$$

where for $h \geq 0$, $\delta(h) = \delta(-h) = \mathrm{Cov}\left(\epsilon_t X_{t+1}, \epsilon_{t+h} X_{t+1+h}\right) = \varphi_0^h \sum_{j=0}^{\infty} \varphi_0^{2j} E\left(\epsilon_t \epsilon_{t+h} \sigma_{t+1+h+j}^2\right)$.

A substantial simplification arises under symmetry of volatility and of the innovations distribution.

Corollary 3.1. *Under the assumptions of Proposition 3.1, when the distribution of* η_t *is symmetric, and when the volatility function* $\sigma(\cdot)$ *in* (2.4) *is even in each variable (i.e.* $\sigma(\ldots, x_{i-1}, x_i, x_{i+1}, \ldots) = \sigma(\ldots, x_{i-1}, -x_i, x_{i+1}, \ldots)$ *for any* $x \in \mathbb{R}^\infty$*), we have* $\gamma(h) = 0$ *for any* $h > 0$ *and thus*

$$\sigma_{WLS}^2 = \mu_\tau^{-2} \gamma(0) = \mu_\tau^{-2} E\left(\epsilon_t^2 \frac{X_{t+1}^2}{\tau_t^2}\right).$$

Note that the symmetry of the volatility function is satisfied by the standard GARCH model, and by many of its extensions, though not by those (like the TGARCH or EGARCH) that take the leverage effect into account.

When $\tau_t = \sigma_t^2$ and (ϵ_t) follows a standard GARCH(1,1), explicit expressions may be obtained for the asymptotic variances of both the WLSE and the LSE.

Corollary 3.2. *Let* $\tau_t = \sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ *with* $\omega > 0, \alpha, \beta \geq 0$*. Then, if* $(\alpha+\beta)^2 + (\kappa-1)\alpha^2 < 1$ *where* $\kappa = E(\eta_t^4)$*, and if the distribution of* η_t *is symmetric, we have*

$$\sigma_{WLS}^2 = \frac{\{1 - (\alpha+\beta)\varphi_0^2\}\{\xi + (\alpha\kappa + \beta)(1 - \varphi_0^2)\}\{1 - \varphi_0^2\}}{\{\xi + (\alpha+\beta)(1 - \varphi_0^2)\}^2}, \quad \textit{where} \quad \xi = E\left(\frac{\omega}{\sigma_t^2}\right)$$

and

$$\sigma_{LS}^2 = \frac{\{1 - (\alpha+\beta)^2 - (\kappa-1)\alpha^2 + (\alpha\kappa+\beta)(1+\alpha+\beta)(1-\varphi_0^2)\}\{1-\alpha-\beta\}\{1-\varphi_0^2\}}{\{1 - (\alpha+\beta)^2 - (\kappa-1)\alpha^2\}\{1 - (\alpha+\beta)\varphi_0^2\}}.$$

Moreover, $\sigma_{WLS}^2 \leq \sigma_{LS}^2$ *with strict inequality unless* σ_t^2 *is constant (that is,* $\alpha = 0$ *or* $\eta_t^2 = 1$ *a.s.).*

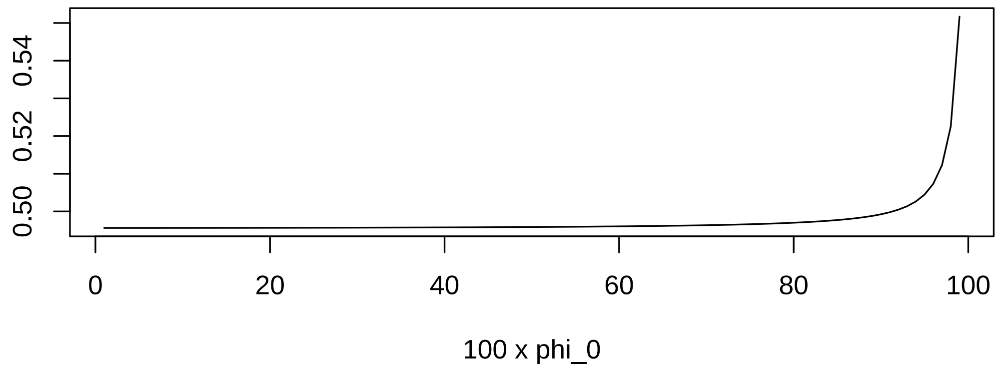

Figure 3: Ratio of asymptotic variances $\sigma^2_{WLS}/\sigma^2_{LS}$ for $\varphi_0 \in \{0.01, \dots, 0.99\}$. The volatility model is as in Figure 1, but with $\alpha_0 = 0.10$ and $\beta_0 = 0.88$ and $\eta_t \sim \mathcal{N}(0, 1)$. The fourth moment condition is satisfied: $(\alpha + \beta)^2 + (\kappa - 1)\alpha^2 \approx 0.98$.

Figure 3 below shows that for a GARCH(1,1) model with Gaussian innovations and coefficients close to those typically obtained in empirical studies, the WLS estimator is almost twice as efficient as the LS estimator.

The ratio of asymptotic variances $\sigma^2_{WLS}/\sigma^2_{LS}$ also depends on the tails of the law of η_t. In Figure 4, for a Student distribution we observe that as the degrees of freedom decrease, the ratio also decreases. This indicates that σ^2_{LS} increases faster than σ^2_{WLS} in the presence of heavy tails.

3.2 Three-step WLSE of the noncausal AR parameter

In practice, the WLSE of φ_0 is infeasible as the weights depend on the unobservable variables η_t. From now on, we assume that the volatility is parameterised. We propose a three-step strategy for consistently estimating φ_0 while taking into account the volatility: *i)* estimate the AR parameter φ_0 by LS, or any other method providing a consistent estimator; *ii)* estimate the volatility parameter $\boldsymbol{\theta}_0$ using the residuals of the first step; *iii)* estimate by WLS the AR parameter using the estimated volatilities as weights.

Step *ii)* can be achieved using the QML estimation method which is the standard procedure for GARCH-type models. This approach will be developed in the next section for the standard GARCH model. For now, we only assume that a consistent estimator $\widehat{\boldsymbol{\theta}}_n$ of the volatility parameter $\boldsymbol{\theta}_0$ is available, without relying on a specific estimation approach. Unexpectedly, we find that the

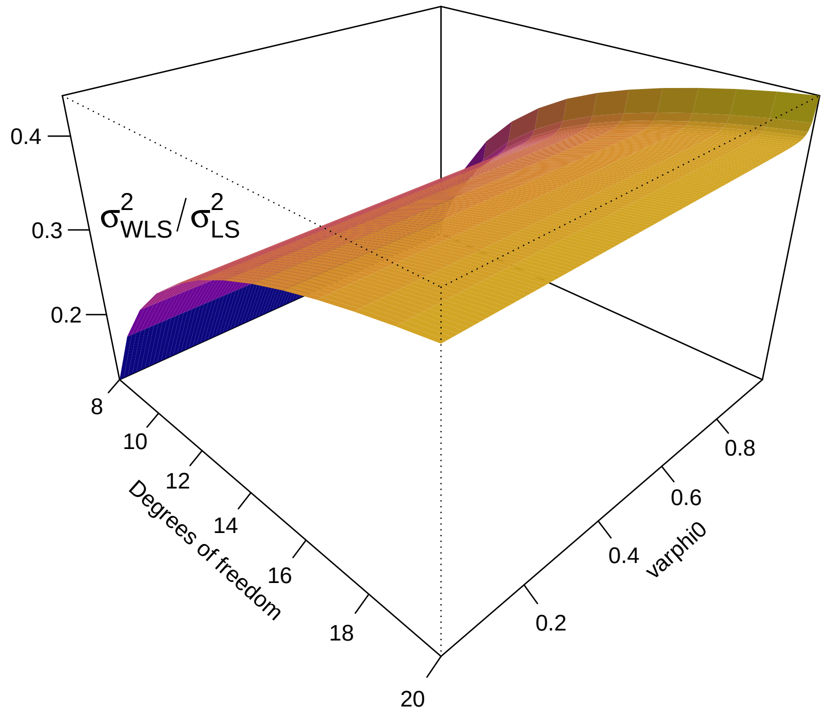

Figure 4: As Figure 3 but with $\eta_t \sim t(\nu)$ for $\nu \in \{8, 9, \ldots, 20\}$.

asymptotic distribution of the estimator of the AR parameter is independent of the estimation method used for the volatility.

Without purporting a specific parametric form for the volatility, we assume that

$$\epsilon_t = \sigma_t \eta_t, \qquad \sigma_t^2 = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \boldsymbol{\theta}_0), \tag{3.3}$$

where $h(\cdot) > \underline{\omega}$ for some $\underline{\omega} > 0$, and $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_\sigma$ where $\boldsymbol{\Theta}_\sigma$ a compact subset of $\mathbb{R}^d$. Let $\boldsymbol{\vartheta}_0 = (\varphi_0, \boldsymbol{\theta}_0')'$. For $\boldsymbol{\vartheta} \in \boldsymbol{\Theta} := (-1,1) \times \boldsymbol{\Theta}_\sigma$, define $\tilde{\sigma}_t^2(\boldsymbol{\vartheta})$ for $t \geq 1$, by

$$\tilde{\sigma}_t^2(\boldsymbol{\vartheta}) = h(\epsilon_{t-1}(\varphi), \ldots, \epsilon_1(\varphi), \widetilde{\epsilon}_0, \widetilde{\epsilon}_{-1}, \ldots; \boldsymbol{\theta}), \tag{3.4}$$

where $\epsilon_t(\varphi) = X_t - \varphi X_{t+1}$ for $\varphi \in (-1,1)$, with fixed initial values $\widetilde{\epsilon}_{-i}$ for $i \geq 0$. Let $\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) = \tilde{\sigma}_t^2(\widehat{\varphi}_{LS}, \boldsymbol{\theta})$. This leads to define the three-stage WLSE of φ_0 as

$$\widehat{\varphi}_{3WLS} = \frac{\sum_{t=1}^n \frac{1}{\tilde{\sigma}_{t,n}^2} X_t X_{t+1}}{\sum_{t=1}^n \frac{1}{\tilde{\sigma}_{t,n}^2} X_{t+1}^2} \quad \text{with} \quad \tilde{\sigma}_{t,n} := \tilde{\sigma}_{t,n}(\widehat{\boldsymbol{\theta}}_n). \tag{3.5}$$

We also need to define the stationary version of (3.4), that is

$$\sigma_t^2(\boldsymbol{\vartheta}) = h(\epsilon_{t-1}(\varphi), \epsilon_{t-2}(\varphi), \ldots; \boldsymbol{\theta}), \quad \forall t. \tag{3.6}$$

We make the following assumptions.

A1: (ϵ_t) is a strictly stationary and nonanticipative solution of (3.3) such that $E\epsilon_t^4 < \infty$.

A2: Let $\tilde{\sigma}_t(\boldsymbol{\vartheta}) > 0$ and $\sigma_t(\boldsymbol{\vartheta}) > 0$ satisfying (3.4) and (3.6), respectively. We have, a.s.

$$\max\left\{\sup_{\boldsymbol{\vartheta}\in\Theta} |\sigma_t(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t(\boldsymbol{\vartheta})|, \sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\{\sigma_t(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t(\boldsymbol{\vartheta})\right\}\right\|, \sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\left\{\sigma_t(\boldsymbol{\vartheta}) - \widetilde{\sigma}_t(\boldsymbol{\vartheta})\right\}\right\|\right\} \leq \rho_t,$$

for some deterministic sequence (ρ_t) such that $\sum_{t=0}^{\infty} \rho_t^2 < \infty$.

A3: $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1)$.

A4: For any real sequence $(x_i)_{i\in\mathbb{N}}$, the function $\boldsymbol{\vartheta} \mapsto h(x_1 - \varphi x_2, x_2 - \varphi x_3, \ldots; \boldsymbol{\theta})$ is continuously differentiable on Θ. Moreover, there exists a neighbourhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that $E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\vartheta})}\frac{\partial}{\partial\boldsymbol{\vartheta}}\sigma_t^2(\boldsymbol{\vartheta})\right\|^4 < \infty$ and $E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\vartheta})}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\sigma_t^2(\boldsymbol{\vartheta})\right\|^2 < \infty$.

A1 is a standard assumption that can generally be displayed more explicitly for specific GARCH-type models. Assumption **A2** is introduced to control the effects of the initial values in (3.4). Note that the convergence of ρ_t to zero does not need to be at an exponential rate, thus allowing for possible long memory in the volatility (see Royer (2023)). Assumption **A3** is obviously satisfied when the estimator of $\boldsymbol{\theta}_0$ is asymptotically Gaussian, but we do not require this assumption. Finally, Assumption **A4** is satisfied for many GARCH-type models.

We can now show the strong consistency and derive the asymptotic distribution of the three-stage WLSE of φ_0.

Theorem 3.1. *Let (X_t) be the noncausal solution of Model (2.1). Under* **A1-A4***, we have $\widehat{\varphi}_{3WLS} \to \varphi_0$ a.s. as $n \to \infty$. Moreover,*

$$\sqrt{n}(\widehat{\varphi}_{3WLS} - \varphi_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{3WLS}^2 := \mu_{\sigma^2}^{-2}\sum_{h=-\infty}^{\infty}\zeta(h)\right). \tag{3.7}$$

for $h \geq 0$, where $\mu_{\sigma^2} = E\left(\frac{X_{t+1}^2}{\sigma_t^2}\right)$

$$\zeta(h) = Cov(Y_t, Y_{t+h}), \quad Y_t = \left(\frac{1}{\sigma_t^2} + \frac{\Phi}{E(X_t^2)}\right)\epsilon_t X_{t+1}, \quad \Phi = E\left(-\frac{\eta_t X_{t+1}}{\sigma_t^3}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right), \tag{3.8}$$

and $\zeta(h) = \zeta(-h)$. *Furthermore, under the symmetry assumptions of Corollary 3.1, we have* $\sigma^2_{3WLS} = \mu^{-2}_{\sigma^2}\zeta(0)$.

A notable feature arising from Theorem 3.1 is that the asymptotic distribution of the third-step estimator is *independent* of the volatility parameter estimator used. estimator used for the volatility parameter θ_0. The effect on (3.7) of the volatility estimation is materialized through the number Φ which depends on the volatility dynamics but does not depend on the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$. In general, $\Phi \neq 0$ because $\frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \varphi}$ is not $\mathcal{F}_{t-1}$-measurable, contrary to $\frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta}}$.

3.3 Portmanteau goodness-of-fit test

Since the pioneering works of Box and Pierce (1970) and Ljung and Box (1978), portmanteau tests based on residual autocorrelations have become a common tool in time series analysis for assessing model suitability. While the seminal papers relied on independence assumptions regarding the error terms, Romano and Thombs (1996), Francq, Roy and Zakoïan (2005), Mainassara (2011) and Zhu (2016), among others, developed diagnostic tests based on weak white noise assumptions in univariate or multivariate ARMA models.

Under correct specification of the noncausal AR(1) model driven by a volatility of general form, corresponding to the null hypothesis

$$H_0: \text{ the process } (X_t) \text{ satisfies Model (2.1) and (3.3) with } \{\eta_t\} \text{ iid},$$

the standardized residuals should resemble a realization of an iid process. For $|h| < n$, define the empirical autocovariances of the standardized residuals

$$\widehat{r}_h = \frac{1}{n} \sum_{t=|h|+1}^{n} \widehat{\eta}_t \widehat{\eta}_{t-|h|},$$

where

$$\widehat{\epsilon}_t = X_t - \widehat{\varphi}_{3WLS} X_{t+1}, \qquad \widehat{\sigma}_t = \widetilde{\sigma}_t(\widehat{\varphi}_{3WLS}, \widehat{\boldsymbol{\theta}}_n), \qquad \widehat{\eta}_t = \frac{\widehat{\epsilon}_t}{\widehat{\sigma}_t}.$$

For any fixed m with $1 \leq m < n$, set $\widehat{\boldsymbol{r}}_m = (\widehat{r}_1, \ldots, \widehat{r}_m)'$. Let also $\boldsymbol{z}_m = (z_1, \ldots, z_m)'$, $\boldsymbol{\iota}_m = (\iota_1, \ldots \iota_m)'$ where for $h = 1, \ldots, m$,

$$z_h = -\varphi_0^{h-1} E\left(\frac{\sigma_t}{\sigma_{t-h}}\right) - \frac{1}{2} E\left[\eta_{t-h}\eta_t \left(\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi} + \frac{1}{\sigma_{t-h}^2}\frac{\partial \sigma_{t-h}^2}{\partial \varphi}\right)\right], \tag{3.9}$$

$$\iota_h = \mu_{\sigma^2}^{-1} \sum_{j=1}^{\infty} \varphi_0^{j-1} E\left(\nu_t \epsilon_t \sigma_{t+j} \eta_{t+j-h}\right), \quad \nu_t = \sigma_t^{-2} + \Phi\left\{E(X_t^2)\right\}^{-1}.$$

Note that in a standard causal setting, the second expectation in (3.9) would be equal to zero. The following result demonstrates that the standard Box–Pierce statistic must be modified to account for causal conditional heteroskedasticity in a noncausal AR(1).

Theorem 3.2. *Under the assumptions of Theorem 3.1, for any* $m \geq 1$, *let* $\hat{\sigma}^2_{3WLS}$, $\widehat{\boldsymbol{\iota}}_m$ *and* $\widehat{\boldsymbol{z}}_m$ *be consistent estimators of* σ^2_{3WLS}, $\boldsymbol{\iota}_m$ *and* $\boldsymbol{z}_m$, *respectively. Let* $\widehat{\boldsymbol{Z}} = [\widehat{\boldsymbol{z}}_m \quad \widehat{\boldsymbol{\iota}}_m]$. *Then, if* $\widehat{\boldsymbol{M}} = \begin{pmatrix} \|\widehat{\boldsymbol{z}}_m\|^2 & 1 + \widehat{\boldsymbol{z}}_m' \widehat{\boldsymbol{\iota}}_m \\ 1 + \widehat{\boldsymbol{z}}_m' \widehat{\boldsymbol{\iota}}_m & \|\widehat{\boldsymbol{\iota}}_m\|^2 - \hat{\sigma}^2_{3WLS} \end{pmatrix}$ *is nonsingular, we have*

$$Q_m := n \widehat{\boldsymbol{r}}_m' [\boldsymbol{I}_m - \widehat{\boldsymbol{Z}} \widehat{\boldsymbol{M}}^{-1} \widehat{\boldsymbol{Z}}'] \widehat{\boldsymbol{r}}_m \xrightarrow{d} \chi^2_m, \tag{3.10}$$

where χ^2_m *denotes the chi-square distribution with* m *degrees of freedom.*

As with all portmanteau tests, the choice of m affects the power of the test. The number of autocorrelations taken must be balanced, as taking too few can affect the power of the tests, while taking too many can result in poor estimation. The impact of m will be examined through numerical experiments in Section 6.

4 Noncausal AR(1) with standard GARCH volatility

We assume in this section that the volatility has the standard GARCH(p, q) parametric form, but it is clear that alternative volatility models could be considered as well. We start by considering the existence and properties of stationary solutions.

4.1 Stationary solution and ARMA representation

Let

$$X_t = \varphi_0 X_{t+1} + \epsilon_t, \quad \epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} \sigma_{t-j}^2, \quad \forall t \in \mathbb{Z}, \tag{4.1}$$

where $|\varphi_0| < 1, \omega_0 > 0, \alpha_{0i}, \beta_{0j} \geq 0$ for $i = 1, \ldots, q$ and $j = 1, \ldots, p$, and $\boldsymbol{\vartheta}_0 = (\varphi_0, \boldsymbol{\theta}_0')' = (\varphi_0, \omega_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})' \in \mathbb{R}^{p+q+2}$ is an unknown parameter, under the same assumptions on (η_t) as in (2.1). For this model, the strict stationarity condition is explicit.

Let $\gamma(\boldsymbol{A}_0)$ denote the top-Lyapunov exponent of the sequence of matrices driving the GARCH(p,q) equation, when the latter is written in vector form (see for instance Francq and Zakoïan, 2019, Section 2.2.2). The following result gives a condition of existence of the strictly stationary solution to the noncausal AR(1)-GARCH(p,q) model.

Proposition 4.1. *Assume $\gamma(\boldsymbol{A}_0) < 0$ and $|\varphi_0| < 1$. Then, there exists a unique strictly stationary and ergodic solution admitting a small-order moment to Model* (4.1)*. This solution is of the form*

$$X_t = \sum_{i=0}^{\infty} \varphi_0^i \eta_{t+i} H(\eta_{t+i-1}^2, \eta_{t+i-2}^2, \ldots),$$

for some measurable function H. When $p = q = 1$, the solution takes the explicit form

$$X_t = \omega_0^{1/2} \sum_{i=0}^{\infty} \varphi_0^i \eta_{t+i} \left\{ 1 + \sum_{j=1}^{\infty} a(\eta_{t+i-1}) \ldots a(\eta_{t+i-j}) \right\}^{1/2},$$

where $a(z) = \alpha_0 z^2 + \beta_0$.

Note that the solution is neither nonanticipative nor anticipative, as X_t is a function of the infinite past and future of η_t. If (X_t) has finite variance, we know that its autocorrelation structure $\rho_X(\cdot)$ of is that of a causal AR(1): $\rho_X(h) = \varphi_0^h$ for $h \geq 0$. The next results provides the autocorrelation structure of (X_t^2) under a symmetry assumption.

Proposition 4.2. *Suppose that $E\epsilon_t^4 < \infty$ and that the law of η_t is symmetric. Then $EX_t^4 < \infty$ and the autocorrelation function $\rho_{X^2}(\cdot)$ of X_t^2 satisfies*

$$(1 - \varphi_0^2 L)\{1 - \alpha(L) - \beta(L)\}\rho_{X^2}(h) = 0, \quad \text{for } \; h > p,$$

where L is the lag operator, $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$ and $\beta(L) = \sum_{i=1}^{p} \beta_i L^i$. Thus, (X_t^2) admits an ARMA$(p \vee q + 1, p)$ representation.

4.2 QML estimation of the volatility parameter

We turn to the estimation of the GARCH parameters using a first-step consistent estimator $\widehat{\varphi}_n$ of the noncausal AR(1) coefficient. We rely on the QML approach which does not require any distributional assumption on the iid process (η_t). Let $\boldsymbol{\Theta}_\sigma$ a compact subset of $(0,\infty) \times [0,\infty)^{p+q}$. For $\boldsymbol{\vartheta} \in (-1,1) \times \boldsymbol{\Theta}_\sigma$, define $\tilde{\sigma}_t^2(\boldsymbol{\vartheta})$ recursively, for $t \geq 1$, by

$$\tilde{\sigma}_t^2(\boldsymbol{\vartheta}) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2(\varphi) + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2(\boldsymbol{\vartheta}), \tag{4.2}$$

where $\epsilon_t(\varphi) = X_t - \varphi X_{t+1}$ for $\varphi \in (-1,1)$, with fixed initial values $\epsilon_0^2, \dots, \epsilon_{1-q}^2$ and $\tilde{\sigma}_0^2(\boldsymbol{\vartheta}), \dots, \tilde{\sigma}_{1-p}^2(\boldsymbol{\vartheta})$. Let $\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) = \tilde{\sigma}_t^2(\widehat{\varphi}_n, \boldsymbol{\theta})$. Define the QML estimator of $\boldsymbol{\theta}_0$ as

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_\sigma} \tilde{\mathrm{l}}_n(\boldsymbol{\theta}), \quad \tilde{\mathrm{l}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{t,n}(\boldsymbol{\theta}), \quad \tilde{\ell}_{t,n}(\boldsymbol{\theta}) = \tilde{\ell}_t(\widehat{\varphi}_n, \boldsymbol{\theta}) = \frac{\epsilon_t^2(\widehat{\varphi}_n)}{\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta})} + \log \tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}).$$

To show the strong consistency and the asymptotic normality of the QML estimator of $\boldsymbol{\theta}_0$, the following assumptions will be made:

B1: $\boldsymbol{\theta}_0$ belongs to the compact set $\boldsymbol{\Theta}_\sigma \subset \mathbb{R}^d$, $d = p+q+1$. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma$, we have $\sum_{i=1}^p \beta_i < 1$.

B2: η_t^2 has a non-degenerate distribution.

B3: $\gamma(\boldsymbol{A}_0) < 0$ and $E\epsilon_t^4 < \infty$.

B4: If $p > 0$, the polynomials $\sum_{i=1}^q \alpha_{0i} z^i$ and $1 - \sum_{i=1}^p \beta_{0i} z^i$ have no common roots, $\sum_{i=1}^q \alpha_{0i} \neq 0$ and $\alpha_{0q} + \beta_{0p} \neq 0$.

B5: $\widehat{\varphi}_n \to \varphi_0$ a.s. as $n \to \infty$.

These assumptions are the same as those ensuring the consistency of the QMLE of the volatility parameter when ϵ_t is observed, except the fourth-order moment condition in **B3**.

Theorem 4.1. *Under* **B1-B5***, let* (X_t) *be the strictly stationary and ergodic solution of Model* (4.1)*. We have* $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ *a.s. as* $n \to \infty$.

Now we turn to the asymptotic normality. The steps of the proof are similar to those used in the standard case of causal GARCH models. However, the presence of a first-step estimator and the fact that $\sigma_t^2(\boldsymbol{\vartheta})$ and its derivatives do not belong to $\mathcal{F}_{t-1}$ entails additional difficulties. To obtain an explicit asymptotic covariance matrix, we need to specify a first step estimator of φ_0 and we consider $\widehat{\varphi}_{WLS}$ studied in Section 3. Note that Assumptions **A1-A4** are satisfied under **B1-B4**. Thus, by Theorem 3.1, **B5** holds for this estimator. The following assumption is standard for asymptotic normality results.

B6: $\boldsymbol{\theta}_0$ belongs to the interior the compact set $\boldsymbol{\Theta}_\sigma$.

Theorem 4.2. *Under* **B1-B4***, assuming* $E\epsilon_t^8 < \infty$ *and* **B6***, let* (X_t) *be the strictly stationary and ergodic solution of Model* (4.1)*. Then we have, using* $\widehat{\varphi}_n = \widehat{\varphi}_{WLS}$*,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma} := \sigma_{WLS}^2 \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{K}'\boldsymbol{J}^{-1} + (\kappa - 1)\boldsymbol{J}^{-1}\right),$$

where

$$\boldsymbol{J} = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}'}(\boldsymbol{\vartheta}_0)\right)_{d\times d}, \boldsymbol{K} = E\left(\left\{1-\eta_t^2\right\}\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\boldsymbol{\theta}\partial\varphi}(\boldsymbol{\vartheta}_0) + \left\{2\eta_t^2-1\right\}\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\varphi}(\boldsymbol{\vartheta}_0)\right)_{d\times 1}.$$

The asymptotic variance of the volatility parameter estimator is thus the sum of two matrices: $(\kappa-1)\boldsymbol{J}^{-1}$ is the asymptotic variance of the QML when the ϵ_t's are observed; the other matrix represents the price paid for the first-step estimation of φ_0. Notice that in the matrix $\boldsymbol{K}$, terms involving η_t^2 cannot be simplified as in causal AR-GARCH models, because the derivatives of σ_t^2 with respect to φ are not $\mathcal{F}_{t-1}$-measurable.

5 Properties of the misspecified full QMLE

A naive approach would be to consider an estimation criterion that mimics the usual QML. This approach consists in minimizing with respect to $\boldsymbol{\vartheta}$ the criterion

$$L_n(\boldsymbol{\vartheta}) = \sum_{t=1}^{n}\frac{(X_t - m_t(\boldsymbol{\vartheta}))^2}{\sigma_t^2(\boldsymbol{\vartheta})} + \log\sigma_t^2(\boldsymbol{\vartheta}), \tag{5.1}$$

where $m_t(\boldsymbol{\vartheta})$ is the *noncausal* conditional mean of the process (X_t) and $\sigma_t^2(\boldsymbol{\vartheta})$ is the *causal* conditional variance. Let $\widehat{\boldsymbol{\vartheta}}$ be the estimator of $\boldsymbol{\vartheta}_0$ obtained by minimizing (5.1) over a compact subset $\boldsymbol{\Theta}$ of $(-1,1)\times(0,\infty)\times[0,\infty)^{p+q}$. Notice that this QMLE is misspecified since $m_t(\boldsymbol{\vartheta}_0) = \varphi_0 X_{t+1}$ is not the conditional mean of X_t given $\mathcal{F}_t$. Consider the limit criterion

$$L_\infty(\boldsymbol{\vartheta}) = E\left\{\frac{\epsilon_t^2(\varphi)}{\sigma_t^2(\boldsymbol{\vartheta})} + \log\sigma_t^2(\boldsymbol{\vartheta})\right\}, \quad \epsilon_t(\varphi) = X_t - \varphi X_{t+1},$$

assuming the expectation exists. We have, for any $\boldsymbol{\vartheta}\in\boldsymbol{\Theta}$,

$$\begin{aligned} L_\infty(\boldsymbol{\vartheta}) - L_\infty(\boldsymbol{\vartheta}_0) &= E\log\frac{\sigma_t^2(\boldsymbol{\vartheta})}{\sigma_t^2(\boldsymbol{\vartheta}_0)} + E\left[\frac{\epsilon_t^2(\varphi)}{\sigma_t^2(\boldsymbol{\vartheta})} - \frac{\epsilon_t^2(\varphi_0)}{\sigma_t^2(\boldsymbol{\vartheta}_0)}\right] \\ &= E\left\{\log\frac{\sigma_t^2(\boldsymbol{\vartheta})}{\sigma_t^2(\boldsymbol{\vartheta}_0)} + \frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})} - 1\right\} + E\frac{\{\epsilon_t(\varphi)-\epsilon_t(\varphi_0)\}^2}{\sigma_t^2(\boldsymbol{\vartheta})} \\ &\quad + (\varphi_0-\varphi)E\left[\frac{2\eta_t\sigma_t(\boldsymbol{\vartheta}_0)X_{t+1}}{\sigma_t^2(\boldsymbol{\vartheta})}\right] + E\left\{(\eta_t^2-1)\frac{\sigma_t^2(\boldsymbol{\vartheta}_0)}{\sigma_t^2(\boldsymbol{\vartheta})}\right\}. \end{aligned}$$

The first two expectations in the r.h.s. are nonnegative, using the elementary inequality $\log(x)\leq x-1$ for $x>0$, but the last two expectations do not cancel as in the usual causal framework. This is due to the fact that $\sigma_t^2(\boldsymbol{\vartheta})\notin\mathcal{F}_{t-1}^{\eta}$ (except at $\boldsymbol{\vartheta}_0$). Of course, this does not prove that the QML estimator is inconsistent without additional assumptions on the iid process, but it makes it plausible. Simulations not reported here support this claim, by showing that the limit criterion may not be minimized at the true parameter value.

However, under a symmetry assumption on the distribution of the innovation process (η_t) we show that the derivative of the limiting criterion $L_\infty(\boldsymbol{\vartheta})$ cancels at $\boldsymbol{\vartheta}_0$. We also show that, in a particular case, the criterion reaches its minimum at the true parameter value.

Proposition 5.1. *Under* **B1-B4** *and* **B6**, *let* (X_t) *be the strictly stationary and ergodic solution of Model* (4.1)*. Then, we have*

$$E\eta_t^3 = 0 \quad \Rightarrow \quad E\left(\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right) = 0. \tag{5.2}$$

Moreover, the second-order derivative of the limiting criterion, $\boldsymbol{J} = E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right)$, *satisfies*

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \varphi \partial \boldsymbol{\theta}}\right) = \mathbf{0}, \quad E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) > \mathbf{0}.$$

in the sense of positive-definite matrices. In the ARCH(1) case, under the previous assumptions, we also have $E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \varphi^2}\right) > 0$ *whenever* $\alpha_0(\kappa - 1 + (2-\kappa)\varphi_0^2) < 1$, *thus* $L_\infty(\boldsymbol{\vartheta})$ *is minimized at* $\boldsymbol{\vartheta}_0$.

6 Monte Carlo simulation

We assess the finite sample behaviour of the LSE and three-step WLSE by means of a Monte Carlo study for the noncausal AR(1) model with GARCH(1,1) volatility

$$X_t = 0.5\, X_{t+1} + \epsilon_t, \qquad \epsilon_t = \sigma_t \eta_t, \qquad \sigma_t^2 = 0.01 + 0.08\,\epsilon_{t-1}^2 + 0.90\,\sigma_{t-1}^2.$$

The innovations η_t follow either a standardised Student-t distribution $\eta_t \sim t_\nu$ or a skew-t distribution $\eta_t \sim st_{\nu,2}$, with $\nu \in \{8, 9, 10\}$, scaled so that $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$; these choices ensure finite fourth moments for ϵ_t and satisfy the usual regularity conditions for the GARCH(1,1) specification. For each combination of innovation distribution and sample size $n \in \{1{,}000,\ 2{,}000,\ 4{,}000,\ 8{,}000\}$, we generate $M = 10{,}000$ Monte Carlo replications and apply the three step estimation procedure.

6.1 Bias and RMSE

Table 1 summarises the finite sample bias and RMSE of all parameters in the three steps. For the first step, the LS estimator $\hat{\varphi}_{\text{LS}}$ exhibits a small negative bias that decreases in magnitude as n

grows, while the RMSE shrinks monotonically with the sample size. The effect of heavier tails is visible when comparing $\nu = 8$ with $\nu = 10$: for fixed n, both bias and RMSE are slightly larger at $\nu = 8$, but the differences are modest and vanish as n increases. The second–step QML estimates of the GARCH parameters (ω, α, β) show the same pattern. Relative to the symmetric t cases, the skew $t_{8,2}$ distribution produces marginally larger RMSEs, especially for α and β, reflecting the additional variability induced by conditional asymmetry. In the third step, the weighted estimator $\hat{\varphi}_{3\text{WLS}}$ improves systematically on LS. For all distributions and sample sizes, both the bias and RMSE of $\hat{\varphi}_{3\text{WLS}}$ are slightly smaller than those of $\hat{\varphi}_{\text{LS}}$, with the gain more pronounced at the smallest sample size and under skew $t_{8,2}$ innovations. Overall, the results confirm that the three–step procedure is effectively unbiased in large samples, and that exploiting conditional heteroskedasticity through 3WLS yields a tangible efficiency gain relative to the simple LS approach.

Table 1: Monte Carlo bias (RMSE) based on $M = 10{,}000$ replications.

Distribution	n	First step		Second step						Third step	
		$\hat{\varphi}_{\text{LS}}$		$\hat{\omega}$		$\hat{\alpha}$		$\hat{\beta}$		$\hat{\varphi}_{3\text{WLS}}$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
t_8	1000	-0.003	0.035	0.003	0.009	0.000	0.023	-0.009	0.035	-0.002	0.031
	2000	-0.002	0.027	0.002	0.005	0.001	0.016	-0.005	0.021	-0.002	0.022
	4000	-0.001	0.019	0.001	0.003	0.000	0.011	-0.002	0.015	-0.001	0.016
	8000	-0.001	0.014	0.000	0.002	0.001	0.008	-0.002	0.010	-0.001	0.011
t_9	1000	-0.003	0.036	0.003	0.010	0.001	0.022	-0.009	0.037	-0.003	0.031
	2000	-0.002	0.027	0.002	0.005	0.000	0.015	-0.005	0.021	-0.001	0.022
	4000	0.000	0.019	0.001	0.003	0.000	0.011	-0.002	0.014	-0.001	0.015
	8000	-0.001	0.014	0.000	0.002	0.000	0.008	-0.001	0.010	-0.001	0.011
t_{10}	1000	-0.003	0.035	0.004	0.010	0.000	0.022	-0.010	0.036	-0.003	0.030
	2000	-0.002	0.025	0.002	0.005	0.000	0.015	-0.004	0.021	-0.001	0.022
	4000	-0.001	0.019	0.001	0.003	0.000	0.010	-0.002	0.014	-0.001	0.015
	8000	0.000	0.014	0.000	0.002	0.000	0.007	-0.001	0.009	0.000	0.011
skew-$t_{8,2}$	1000	-0.005	0.036	0.005	0.012	0.002	0.028	-0.014	0.046	-0.003	0.031
	2000	-0.003	0.027	0.002	0.005	0.001	0.019	-0.005	0.024	-0.002	0.022
	4000	-0.001	0.021	0.001	0.003	0.001	0.013	-0.003	0.016	0.000	0.016
	8000	0.000	0.016	0.001	0.002	0.000	0.009	-0.002	0.011	0.000	0.011

6.2 Asymptotic variance

Table 2 compares the finite sample (Monte Carlo) variances of the LS and 3WLS estimators with the analytical expressions derived in Section 3. For LS, the table reports the empirical quantities $n\widehat{\text{Var}}(\hat{\varphi}_{LS})$ together with the two analytical variances from Corollary 3.1 and Equation (3.2), the latter retaining the infinite sum $\sum_{h=-\infty}^{\infty} \delta(h)$. For 3WLS, we report the empirical $n\widehat{\text{Var}}(\hat{\varphi}_{3WLS})$ and the analytical variance σ^2_{3WLS} that uses the sum $\sum_{h=-\infty}^{\infty} \zeta(h)$.

Across all sample sizes and distributions, the analytical formulas reproduce the Monte Carlo variances very accurately. In the symmetric Student t cases with $\nu \geq 8$, the closed form expressions based on Corollary 3.1 are virtually indistinguishable from the empirical $n\widehat{\text{Var}}(\hat{\varphi})$ values, indicating that the diagonal terms dominate the variance when both the innovations and the conditional variance are symmetric. The versions that keep the infinite sums over $\delta(h)$ and $\zeta(h)$ perform equally well; the small discrepancies are due only to truncation of the sums and are slightly more noticeable for LS, since $\delta(h)$ involves an additional inner summation. As expected, for any estimator the variance increases when ν decreases from 10 to 8, reflecting heavier tails, while for fixed ν the scaling by n stabilizes the empirical variances already at $n = 4{,}000$, where further increases in n change the entries by less than two percent.

The efficiency ordering predicted by Corollary 3.2 is clearly visible. The 3WLS estimator is uniformly more efficient than LS. The additional term in the feasible 3WLS variance, captured by the ζ sum and the derivative of σ_t^2 with respect to φ, is numerically negligible for the GARCH parameters used here. In the skew $t_{8,2}$ case, conditional asymmetry activates the off diagonal covariances, so that truncation of the infinite sums introduces a small positive bias in the analytical variances. This bias is modest, decreases with n, and confirms that the symmetric shortcut in Corollary 3.1 should not be used when skewness or leverage is present, whereas the full formulas remain accurate in practice.

Note: The quantities $n\widehat{\text{Var}}(\hat{\varphi}_{LS})$ and $n\widehat{\text{Var}}(\hat{\varphi}_{3WLS})$ denote empirical variances of the corresponding estimators, computed over the $M = 10{,}000$ Monte Carlo replications. The values based on Corollary 3.1 are reported only for symmetric innovation laws and are omitted for skewed distributions.

Table 2: Monte Carlo asymptotic variances σ^2_{LS} and σ^2_{3WLS} based on $M = 10{,}000$ replications.

		σ^2_{LS}			σ^2_{3WLS}	
Dist.	n	$n\widehat{\mathrm{Var}}(\hat{\varphi}_{LS})$	$\frac{E(\epsilon_t^2 X_{t+1}^2)}{\{E(X_{t+1}^2)\}^2}$	$\frac{\sum_{h=-\infty}^{\infty}\delta(h)}{\{E(X_{t+1}^2)\}^2}$	$n\widehat{\mathrm{Var}}(\hat{\varphi}_{3WLS})$	$\frac{\sum_{h=-\infty}^{\infty}\zeta(h)}{\{E(X_{t+1}^2/\sigma_t^2)\}^2}$
t_8	1000	1.214	1.249	1.323	0.927	0.917
	2000	1.400	1.404	1.463	0.991	0.920
	4000	1.475	1.509	1.571	0.976	0.943
	8000	1.630	1.661	1.710	0.978	0.949
t_9	1000	1.266	1.245	1.300	0.966	0.933
	2000	1.480	1.349	1.408	0.922	0.919
	4000	1.486	1.486	1.521	0.944	0.929
	8000	1.613	1.608	1.632	0.936	0.931
t_{10}	1000	1.167	1.188	1.255	0.889	0.883
	2000	1.332	1.324	1.370	0.887	0.893
	4000	1.356	1.405	1.443	0.870	0.907
	8000	1.557	1.512	1.544	0.919	0.912
skew-$t_{8,2}$	1000	1.270	-	1.440	0.940	0.970
	2000	1.440	-	1.630	0.980	0.990
	4000	1.680	-	1.840	0.980	1.010
	8000	2.060	-	2.040	0.990	0.990

6.3 Goodness-of-fit

Table 3 reports the empirical rejection frequencies (finite sample size) of the portmanteau statistic Q_m of Theorem 3.2 based on standardized residuals for the noncausal AR(1) - GARCH(1,1) model. For each design, we generate artificial series from the noncausal AR(1) process with GARCH(1,1) volatility, estimate the same model by the three step procedure, and compute Q_m for $m \in \{8, 10, 20\}$. The entries are the proportions of Monte Carlo replications in which the null hypothesis is rejected at the nominal levels 5% and 10%, for Student-t innovations with $\nu \in \{8, 10, 15\}$ and sample sizes $n \in \{1000, 2000, 4000\}$.

Overall, the empirical sizes are close to the nominal values. For all combinations of ν, m and N, the rejection frequencies at the 5% level remain in a narrow band around 0.05, and the same holds for the 10% level around 0.10, with only mild deviations at $n = 1000$ and the shortest lag truncation $m = 8$. As n increases from 1000 to 4000, these deviations become smaller, and there is no systematic drift with respect to m or the degrees of freedom.

Figure 5 complements the table by showing the empirical densities of Q_m under the same designs together with the theoretical χ^2_m density. The simulated distributions of Q_m align very closely with the χ^2_m benchmark already at $N = 2000$, and the match improves further at $N = 4000$. Taken together, Table 3 and Figure 5 indicate that, under the correctly specified noncausal AR(1) - GARCH(1,1) model, the portmanteau statistic Q_m is well calibrated and its χ^2_m approximation provides an accurate basis for residual diagnostics in practice. We also provide the empirical distribution of Q_m under a misspecified noncausal AR(1) with noncausal GARCH(1,1) which differs significantly from the chi-squared distribution, particularly when $\nu = 10$. This demonstrates that the test should be effective in detecting this type of misspecification.

Table 3: Empirical rejection frequencies (size) of the portmanteau test Q_m at nominal levels 5% and 10% for the noncausal AR(1) - GARCH(1,1) model.

Distribution	m	$n = 1000$		$n = 2000$		$n = 4000$	
		5%	10%	5%	10%	5%	10%
	8	0.055	0.110	0.062	0.112	0.060	0.121
t_8	10	0.062	0.118	0.061	0.112	0.067	0.122
	20	0.045	0.090	0.055	0.102	0.052	0.115
	8	0.053	0.110	0.053	0.113	0.061	0.123
t_{10}	10	0.056	0.104	0.068	0.110	0.069	0.125
	20	0.039	0.085	0.050	0.094	0.055	0.111
	8	0.062	0.124	0.067	0.124	0.052	0.105
t_{15}	10	0.072	0.115	0.079	0.135	0.055	0.115
	20	0.048	0.104	0.054	0.101	0.049	0.102

7 Empirical Application

In our application we consider two cryptocurrencies: XRP and Solana (SOL), using daily USD prices from April 10, 2020, to July 1, 2024 ($n = 1,544$). Both series exhibit bubble-like dynamics, characterized by a rapid run-up starting in 2021, followed by a sharp correction in 2022 that brought prices back to levels seen during the COVID-19 period. Figure 9 in the appendix displays this boom and bust pattern. Consistent with cryptocurrency markets, both assets exhibit pronounced volatility and volatility clustering, characterized by large swings and abrupt drawdowns; SOL also displays a stronger rebound in 2023–2024.

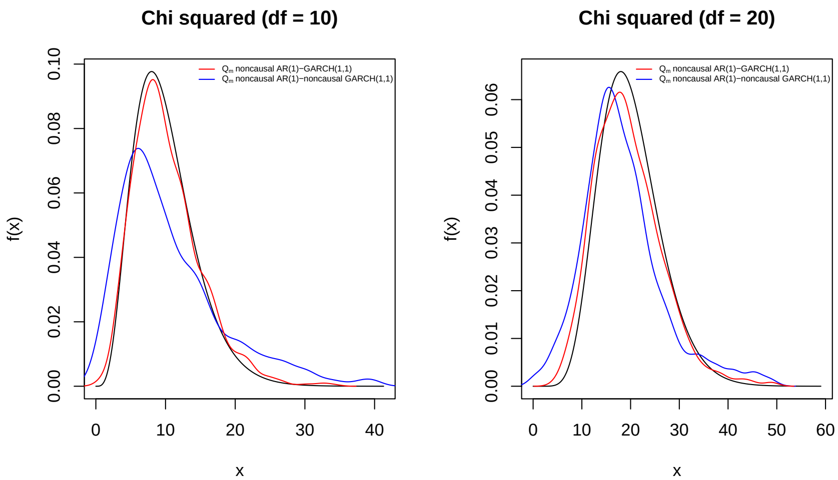

Figure 5: Probability density function of the χ^2_m distribution and empirical density of the test statistic Q_m for $m \in \{10, 20\}$ and $T = 2{,}000$. The DGP is the noncausal AR(1) with causal GARCH(1,1), $X_t = 0.9\, X_{t+1} + \epsilon_t$, $\epsilon_t = \sigma_t \eta_t$, $\sigma_t^2 = 0.1 + 0.15\, \epsilon_{t-1}^2 + 0.84\, \sigma_{t-1}^2$. The black curve corresponds to the χ^2_m density, the red curve to the empirical density of Q_m under the correctly specified noncausal AR(1) with causal GARCH(1,1), and the blue curve to the empirical density of Q_m under the misspecified noncausal AR(1) with noncausal GARCH(1,1): $\sigma_t^2 = 0.1 + 0.15\, \epsilon_{t+1}^2 + 0.84\, \sigma_{t+1}^2$.

Both price series display clear signatures of nonstationarity: in levels, their ACFs decay slowly, indicating a persistent trend component. To separate this low-frequency component from short-run dynamics without imposing ad hoc polynomial trends or introducing look-ahead bias, we extract a one-sided stochastic trend using a local-level state-space specification estimated by the Kalman filter (see Blasques, Koopman, and Mingoli, 2023). In this framework the trend is stochastic, and adapts gradually over time, so sharp run-ups associated with bubble-like behavior remain largely in the detrended component. We tune the filter's signal-to-noise ratio to yield a sufficiently smooth trend and then verify that the resulting residuals are stationary (the detrending method is detailed in the Appendix)

Figure 6 displays XRP and SOL prices (left panels) with the estimated trend in red, and the corresponding detrended series (right panels), obtained by subtracting the trend from the price level. The detrended series fluctuate around zero and appear approximately stationary, while preserving the salient short-run dynamics, including explosive run-ups associated with bubble formation.

For the detrended series, denoted X_t, we start by fitting the noncausal AR(1) model via LS:

$$X_t \;=\; \widehat{\varphi}\, X_{t+1} \;+\; \epsilon_t. \tag{7.1}$$

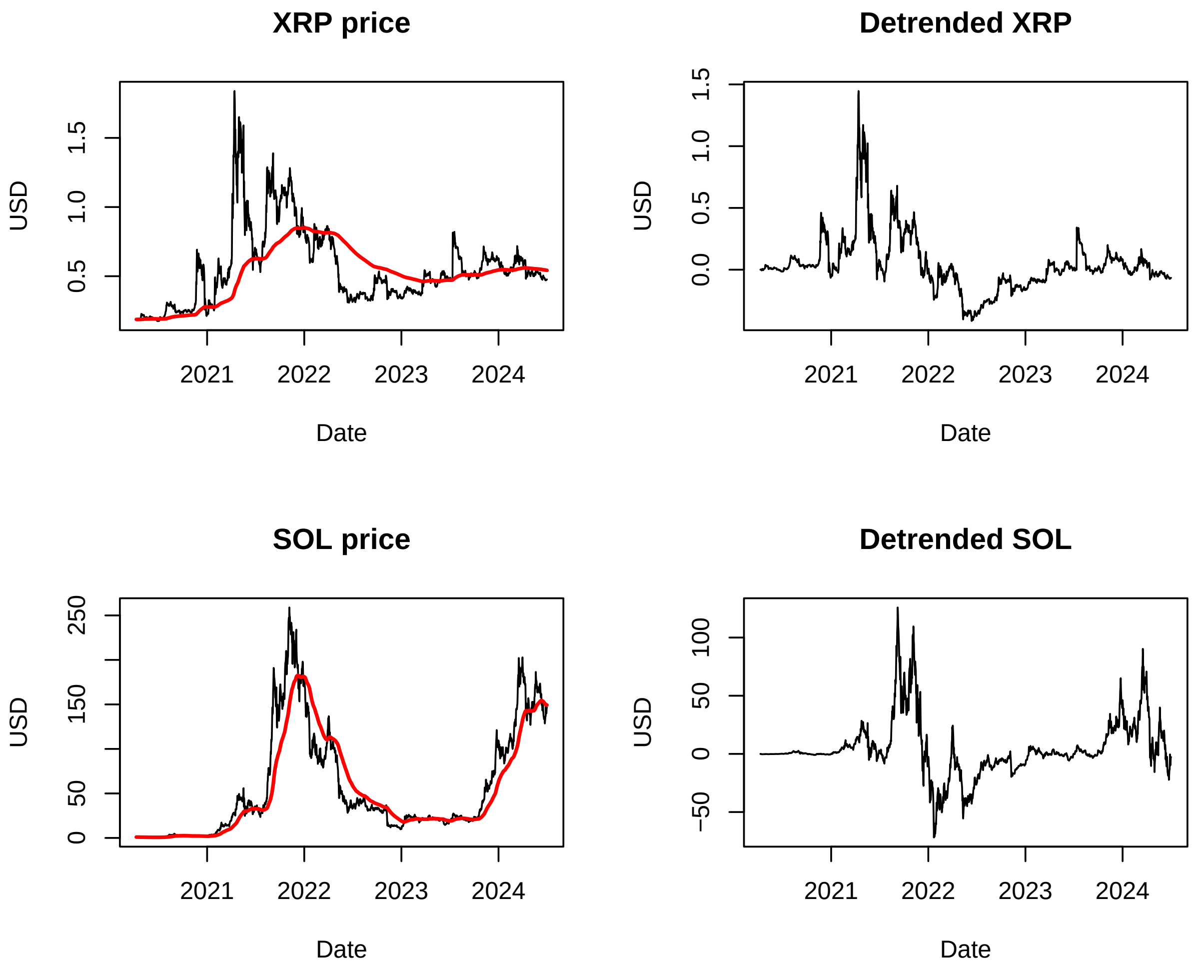

Figure 6: Detrended XRP and SOL prices.

Table 4: Parameter estimation - first, second, and third step.

XRP					SOL				
First step	Second step			Third step	First step	Second step			Third step
$\hat{\varphi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\varphi}_{3\text{WLS}}$	$\hat{\varphi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\varphi}_{3\text{WLS}}$
0.9840	0.0000	0.1407	0.8541	0.9822	0.9832	0.0012	0.1902	0.8088	0.9890
(0.0044)	(0.0000)	(0.0001)	(0.0004)	(0.0044)	(0.0045)	(0.0005)	(0.0120)	(0.0113)	(0.0045)

Note: Standard errors are reported in parentheses. The standard error for the third-step are obtained using the asymptotic variance.

Table 4 reports the first-step estimates of the noncausal AR(1) parameter for both assets. The coefficients are close to unity indicating very persistent forward dependence.

The estimated autocorrelations of the residuals do not allow to reject the white noise assumption, with minimal significant lags. In contrast, the correlation in the squared-residuals present strong dynamics, confirming the presence of conditional heteroskedasticity.

In the second step, we estimate the GARCH(1,1) parameters by QML. Table 4 reports the estimates. The variance intercept $\hat{\omega}$ is small in both assets and highly significant. The short-run sensitivity of volatility to new shocks is higher for SOL than for XRP: $\hat{\alpha}_{\text{SOL}} = 0.1902\,(0.0120)$ versus $\hat{\alpha}_{\text{XRP}} = 0.1407\,(0.0001)$. By contrast, volatility clustering is stronger in XRP than in SOL: $\hat{\beta}_{\text{XRP}} = 0.8541\,(0.0004)$ versus $\hat{\beta}_{\text{SOL}} = 0.8088\,(0.0113)$. Overall persistence, measured by $\hat{\alpha} + \hat{\beta}$, is very high in both series: 0.9948 for XRP and 0.9990 for SOL.

We proceed to the third step to obtain $\hat{\varphi}_{3\text{WLS}}$. The estimates are similar to those from the first step and remain highly significant (see Table 4). We then construct standardized residuals using the estimates from the second and third steps, $\hat{\eta}_t = \hat{\epsilon}_t/\hat{\sigma}_t$.

Table 5 reports the Q_m test statistics and p-values for standardized residuals. We fail to reject the null hypothesis for all the lags, which supports the noncausal AR(1)–GARCH(1,1) specification.

Figure 7 displays the conditional volatility estimated in the second step. The series reacts strongly to new information and exhibits marked persistence, evidenced by volatility clustering once it rises above its unconditional level.

Table 5: Portmanteau statistic Q_m and p-values for standardized residuals of the noncausal AR(1) - GARCH(1,1) model.

	XRP		SOL	
Lag m	Q_m	p-value	Q_m	p-value
1	0.143	0.705	0.021	0.885
5	4.383	0.496	5.329	0.377
10	7.570	0.671	10.811	0.372
15	10.429	0.792	18.017	0.262
20	11.534	0.931	18.724	0.540

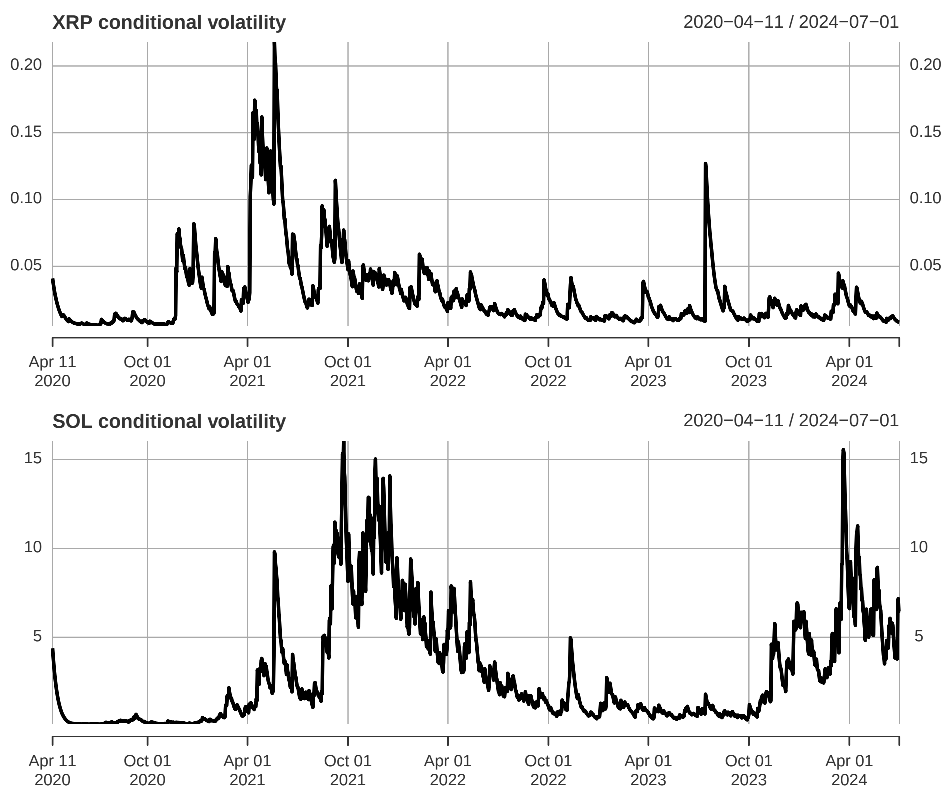

Figure 7: Estimated conditional volatility $\hat{\sigma}_t$ for XRP and SOL

Figure 8 plots the SOL detrended price from 3 February 2021 to 1 August 2021. From early 2021 through mid-May, SOL rose steadily, supported by a broad post-pandemic risk-on environment and by ecosystem momentum: the rollout of native, on-chain central-limit-order-book infrastructure like

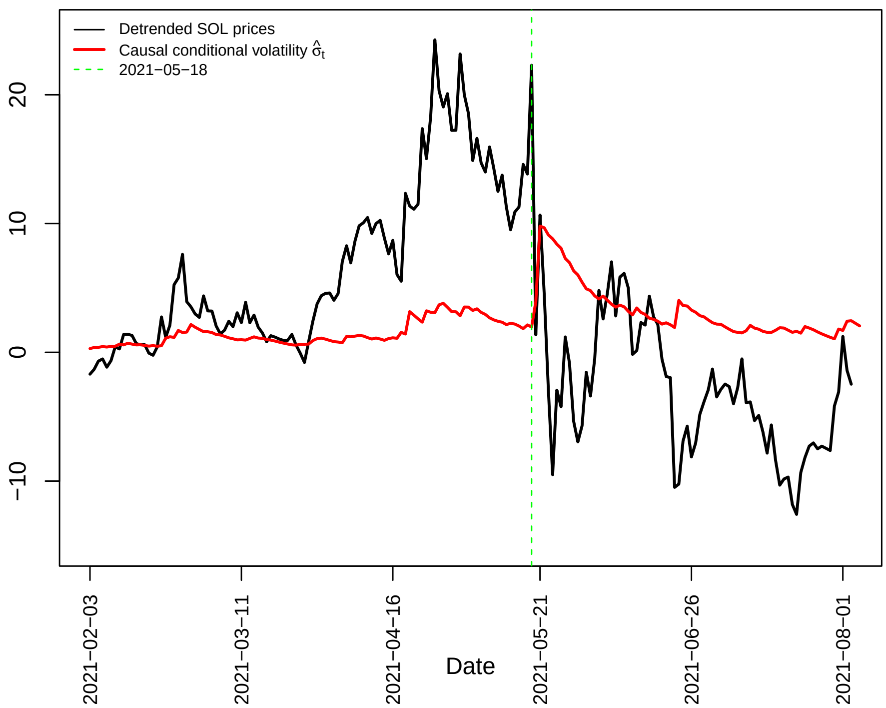

Figure 8: Estimated causal conditional variance σ_t^2 for SOL from 2021-02-03 to 2021-08-01.

Serum and its integrations as Raydium deepened liquidity, lowered effective trading frictions, and attracted substantial speculative participation within decentralized finance (DeFi).

This momentum culminated around 18 May 2021, when SOL briefly approached 56 USD. A sector-wide sell-off followed regulatory announcements in China restricting crypto-related financial services (18 May) and, shortly after, signaling a crackdown on Bitcoin mining (21 May). On 19 May, SOL fell from about 55.9 to 35.1 USD and continued to decline, reaching roughly 23–25 USD by 23 May. Thereafter, the price stabilized, but at a lower plateau relative to its previous levels. Figure 8 suggests a bubble-like pattern: an organic, sustained appreciation followed by an abrupt collapse. The estimated conditional volatility from the residuals of the noncausal AR(1) model is shown for the causal specification (red). On the crash day, 19 May 2021, the causal conditional variance spikes contemporaneously and then decays gradually as prices stabilize after the drop, matching the observed dynamics of a higher volatility after the crash.

8 Conclusion

This article has considered introducing conditional heteroskedasticity to noncausal AR models. By specifying causal volatility dynamics, the model can handle time series displaying increased volatility after bubble crashes. However, the coexistence of causal and noncausal components in the model, contrary to the standard causal framework, results in a discrepancy between the information sets generated by observations and innovations. Consequently, the conditional moments in calendar time of the observed process are unknown. This makes estimation considerably more difficult, as the standard QML approach generally fails. We propose alternative weighted three-step least-squares methods for estimating the AR coefficient, and establish the asymptotic properties of the resulting estimators. In the standard GARCH case, we also derive a consistent and asymptotically normal two-step estimator of the conditional variance parameters.

A more realistic description of financial time series may require a hybrid specification that blends causal and noncausal volatility dynamics or introduces explicit asymmetric terms in the conditional variance. Other potential areas of exploration include the detection of bubbles in their inflationary phase and the prediction of future bubbles. Fries (2021) established expressions of the crash odds in the case of noncausal MA(∞) processes based on extreme value clustering. For noncausal processes, Gouriéroux and Jasiak (2016) proposed computational methods such as sampling importance resampling in order to generate future paths and prediction intervals. De Truchis, Fries and Thomas (2025) propose a method for forecasting extreme trajectories when the innovations follow a stable distribution. Extension of these approaches to noncausal AR models driven by conditionally heteroskedastic errors is left for future research.

Appendix

This Appendix provides the proofs of the asymptotic results and complements on the empirical study. We use throughout the multiplicative norm defined by $\|A\| = \sum |a_{ij}|$ with obvious notation. For convenience, the norm will be denoted identically whatever the dimension of the matrix/vector A.

A Proofs of the results of Section 3

Proof of Proposition 3.1. We have

$$\widehat{\varphi}_{WLS} = \varphi_0 + \frac{\frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tau_t}\epsilon_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tau_t} X_{t+1}^2}. \tag{A.1}$$

The summands, in the numerator and the denominator, are measurable functions of past and future values of η_t by (2.4)-(2.5). Moreover, $E\left(\frac{1}{\tau_t}X_{t+1}^2\right) \leq \frac{1}{\underline{\tau}}E\left(X_{t+1}^2\right) < \infty$ by (2.5). Thus, by the Cauchy-Schwarz inequality, $E\left|\frac{1}{\tau_t}\epsilon_t X_{t+1}\right| \leq \frac{1}{\underline{\tau}}\{E(\epsilon_t^2)E(X_{t+1}^2)\}^{1/2} < \infty$. Hence, the ergodic theorem applies to the empirical means in (A.1). The strong consistency of $\widehat{\varphi}_{WLS}$ follows because

$$E\left(\frac{\epsilon_t X_{t+1}}{\tau_t}\right) = \sum_{i=0}^{\infty} \varphi_0^i E\left(\frac{\epsilon_t \epsilon_{t+1+i}}{\tau_t}\right) = \sum_{i=0}^{\infty} \varphi_0^i E\left(\frac{\epsilon_t \eta_{t+1+i}\sigma_{t+1+i}}{\tau_t}\right) = 0,$$

using the independence between η_{t+1+i} and all other terms in the expectation, since $\tau_t, \epsilon_t, \sigma_{t+1+i} \in \mathcal{F}_{t+i}$.

Now we turn to the asymptotic distribution of $\widehat{\varphi}_{WLS}$. We have

$$\sqrt{n}(\widehat{\varphi}_{WLS} - \varphi_0) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{1}{\tau_t}\epsilon_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tau_t} X_{t+1}^2}. \tag{A.2}$$

Notice that the summands in the numerator do not constitute a martingale difference sequence w.r.t. the sigma-field $\mathcal{F}_t$. Indeed, $\frac{1}{\tau_t}\epsilon_t X_{t+1} \notin \mathcal{F}_{t+k}$, whatever k. However, we will show they constitute a mixingale.

The concept of mixingale was introduced by McLeish (1975). Mixing and martingale difference sequences are particular cases of mixingales (see Hall and Heyde (1980) Section 2.3 for other examples). For the reader's convenience we recall the definition. Let $\{\mathcal{F}_i; i \in \mathbf{Z}\}$ be a nondecreasing sequence of σ-fields. Let $\|\cdot\|_2$ denote the L^2 norm.

The sequence $\{(Z_t, \mathcal{F}_t)\}$ is called a mixingale if there exist sequences of nonnegative constants (c_t) and (ψ_m), with $\psi_m \to 0$ as $m \to \infty$, such that for all $t \geq 1$, $m \geq 0$:

a) $\|E(Z_t|\mathcal{F}_{t-m})\|_2 \leq \psi_m c_t$ and

b) $\|Z_t - E(Z_t|\mathcal{F}_{t+m})\|_2 \leq \psi_{m+1} c_t$.

Let us check that $\{(Z_t, \mathcal{F}_t)\}$ is a mixingale, where $Z_t = \frac{1}{\tau_t}\epsilon_t X_{t+1}$ and $\mathcal{F}_t = \mathcal{F}_t^{\epsilon}$. We have, for $m \geq 0$,

$$E(Z_t|\mathcal{F}_{t-m}) = \sum_{i=0}^{\infty} \varphi_0^i E\left(\frac{\epsilon_t \epsilon_{t+1+i}}{\tau_t}\middle|\mathcal{F}_{t-m}\right) = 0,$$

by arguments already used. Hence a) is satisfied. Moreover, for $m > 0$,

$$E(Z_t|\mathcal{F}_{t+m}) = \sum_{i=0}^{m-1} \varphi_0^i \frac{1}{\tau_t}\epsilon_t\epsilon_{t+1+i} + \sum_{i=m}^{\infty} \varphi_0^i E\left(\frac{1}{\tau_t}\epsilon_t\epsilon_{t+1+i}\middle|\mathcal{F}_{t+m}\right) = \sum_{i=0}^{m-1} \varphi_0^i \frac{1}{\tau_t}\epsilon_t\epsilon_{t+1+i}.$$

Hence, using Hölder's inequality,

$$\begin{aligned}
\|Z_t - E(Z_t|\mathcal{F}_{t+m})\|_2 &\leq |\varphi_0|^m \sum_{i=m}^{\infty} |\varphi_0^{i-m}| \left\| \frac{1}{\tau_t}\epsilon_t\epsilon_{t+1+i}\right\|_2 \\
&\leq \underline{\tau}^{-1}|\varphi_0|^m \sum_{i=m}^{\infty} |\varphi_0|^{i-m} \|\epsilon_t\|_4 \|\epsilon_{t+1+i}\|_4 \leq K|\varphi_0|^m,
\end{aligned}$$

which proves that b) is satisfied with $\psi_m = |\varphi_0|^m$ and $c_t = K = \underline{\tau}^{-1}(1-|\varphi_0|)^{-1}\|\epsilon_t\|_4^2$. We thus have shown that $\{(Z_t, \mathcal{F}_t)\}$ is a mixingale. The asymptotic normality of the numerator in the r.h.s. of (A.2) follows by McLeish (1975), Theorem 5.4 (see also White (2014), Section 5.3), with asymptotic variance $\sum_{h=-\infty}^{\infty}\gamma(h)$, where for any $h \geq 0$, $\gamma(-h) = \gamma(h)$ and

$$\begin{aligned}
\gamma(h) = \mathrm{Cov}\left(\frac{\epsilon_t X_{t+1}}{\tau_t}, \frac{\epsilon_{t+h}X_{t+1+h}}{\tau_{t+h}}\right) &= \sum_{i,j=0}^{\infty} \varphi_0^{i+j} E\left(\frac{\epsilon_t\epsilon_{t+1+i}\epsilon_{t+h}\epsilon_{t+1+h+j}}{\tau_t\tau_{t+h}}\right) \\
&= \varphi_0^h \sum_{j=0}^{\infty} \varphi_0^{2j} E\left(\frac{\epsilon_t\epsilon_{t+h}\sigma_{t+1+h+j}^2\eta_{t+1+h+j}^2}{\tau_t\tau_{t+h}}\right) \\
&= \varphi_0^h \sum_{j=0}^{\infty} \varphi_0^{2j} E\left(\frac{\epsilon_t\epsilon_{t+h}\sigma_{t+1+h+j}^2}{\tau_t\tau_{t+h}}\right), \qquad \text{(A.3)}
\end{aligned}$$

where the last equality follows from the independence between $\eta_{t+1+h+j}^2$ and all other terms involved in the expectation. The conclusion follows, by applying the ergodic theorem to the denominator of the r.h.s. of (A.2). □

Proof of Corollary 3.1. Denote by $\mathcal{F}_{\neq u}$ the σ-field generated by $\{\eta_v, v \neq u\}$. For any $h > 0$ and $j \geq 0$ we have

$$E\left(\frac{\epsilon_t\epsilon_{t+h}\sigma_{t+1+h+j}^2}{\tau_t\tau_{t+h}}\right) = E\left\{\frac{\epsilon_t\sigma_{t+h}}{\tau_t\tau_{t+h}}E\left(\eta_{t+h}\sigma_{t+1+h+j}^2|\mathcal{F}_{\neq t+h}\right)\right\} = 0,$$

using the facts that (i) $\epsilon_t, \sigma_{t+h}, \tau_t, \tau_{t+h}$ are measurable functions of the η_u with $u < t+h$, and (ii) $\eta_{t+h}\sigma^2_{t+1+h+j}$ is an odd function of η_{t+h} given $\mathcal{F}_{\neq t+h}$. In view of (A.3), this entails that $\gamma(h) = 0$ for $h > 0$. □

Proof of Corollary 3.2. For the GARCH(1,1) model, we have

$$\sigma_t^2 = \left\{1 + \sum_{i=1}^{\infty} a(\eta_{t-1}) \dots a(\eta_{t-i})\right\}\omega, \quad a(\eta) = a\eta^2 + \beta,$$

(see for instance Francq and Zakoïan (2019), Theorem 2.1). Hence (2.4) is satisfied, as well as the symmetry assumption made on the volatility function in Corollary 3.1. By arguments already used, we have

$$S := E\left(\epsilon_t^2 \frac{X_{t+1}^2}{\sigma_t^4}\right) = \sum_{i,j\geq 0} \varphi_0^{i+j} E\left(\eta_t^2 \frac{\epsilon_{t+1+i}\epsilon_{t+1+j}}{\sigma_t^2}\right) = \sum_{i\geq 0} \varphi_0^{2i} m_i, \quad m_i = E\left(\eta_t^2 \frac{\sigma_{t+1+i}^2}{\sigma_t^2}\right).$$

It follows that

$$m_i = \xi + (\alpha+\beta)m_{i-1}, \quad \text{for } i > 0, \quad \text{and } m_0 = E\left(\frac{\eta_t^2}{\sigma_t^2}(\omega + \alpha\epsilon_t^2 + \beta\sigma_t^2)\right) = \xi + \alpha\kappa + \beta.$$

Thus

$$S = m_0 + \frac{\xi\varphi_0^2}{1-\varphi_0^2} + (\alpha+\beta)\varphi_0^2 S = \frac{\xi + (\alpha\kappa+\beta)(1-\varphi_0^2)}{\{1-(\alpha+\beta)\varphi_0^2\}(1-\varphi_0^2)}.$$

Similarly,

$$E\left(\frac{X_{t+1}^2}{\sigma_t^2}\right) = \frac{\xi + (\alpha+\beta)(1-\varphi_0^2)}{\{1-(\alpha+\beta)\varphi_0^2\}(1-\varphi_0^2)}.$$

The formula for σ^2_{WLS} follows.

Now, similar calculations yield

$$S^* := E\left(\epsilon_t^2 X_{t+1}^2\right) = \sum_{i\geq 0} \varphi_0^{2i} m_i^*, \quad m_i^* = E\left(\epsilon_t^2 \sigma_{t+1+i}^2\right),$$

$$m_0^* = \omega E\epsilon_t^2 + (\alpha\kappa+\beta)E\sigma_t^4, \quad m_i^* = \omega E\epsilon_t^2 + (\alpha+\beta)m_{i-1}^*, \quad \text{for } i > 0.$$

Thus

$$S^* = m_0^* + \frac{\omega\varphi_0^2}{1-\varphi_0^2}E\epsilon_t^2 + (\alpha+\beta)\varphi_0^2 S^* = \frac{\omega E\epsilon_t^2 + (\alpha\kappa+\beta)E\sigma_t^4(1-\varphi_0^2)}{\{1-(\alpha+\beta)\varphi_0^2\}(1-\varphi_0^2)}.$$

We also have

$$E\sigma_t^2 = \frac{\omega}{1-(\alpha+\beta)}, \quad EX_t^2 = \frac{E\sigma_t^2}{1-\varphi_0^2}, \quad E\sigma_t^4 = \frac{\omega^2(1+\alpha+\beta)}{\{1-(\alpha+\beta)^2-(\kappa-1)\alpha^2\}\{1-(\alpha+\beta)\}}.$$

The formula for σ^2_{LS} follows. Now we prove that $\sigma^2_{WLS} < \sigma^2_{LS}$. A lower bound for ξ is obtained from the Cauchy–Schwarz inequality:

$$1 = E(\sigma_t \cdot \sigma_t^{-1}) \leq \sqrt{E\sigma_t^2 \, E\sigma_t^{-2}} = \sqrt{\frac{\omega}{1-\alpha-\beta} \cdot \frac{\xi}{\omega}} = \sqrt{\frac{\xi}{1-\alpha-\beta}},$$

which implies

$$\xi \geq 1-\alpha-\beta, \tag{A.4}$$

with equality if and only if σ_t^2 is a.s. constant. Write $\sigma^2_{WLS} = \sigma^2_{WLS}(\xi)$. It is easily seen that $\sigma^2_{WLS}(\cdot)$ is a strictly decreasing function. By (A.4) we obtain the bound

$$\sigma^2_{WLS} \leq \sigma^2_{WLS}(1-\alpha-\beta) = \frac{\{\,1-\alpha-\beta+(\alpha\kappa+\beta)(1-\varphi_0^2)\,\}\{1-\varphi_0^2\}}{1-(\alpha+\beta)\varphi_0^2}. \tag{A.5}$$

The inequality $\sigma^2_{LS} \geq \sigma^2_{WLS}(1-\alpha-\beta)$ is equivalent to

$$\frac{\{1-(\alpha+\beta)^2-(\kappa-1)\alpha^2+(\alpha\kappa+\beta)(1+\alpha+\beta)(1-\varphi_0^2)\}\{1-\alpha-\beta\}}{1-(\alpha+\beta)^2-(\kappa-1)\alpha^2} \;\geq\; \{\,1-\alpha-\beta+(\alpha\kappa+\beta)(1-\varphi_0^2)\,\}.$$

which reduces to $(\kappa-1)\alpha^2 \;\geq\; 0$. The latter inequality holds true for all admissible parameters and is strict whenever $\alpha > 0$ and $\kappa > 1$. Combining with (A.5) and the monotonicity of $\sigma^2_{WLS}(\cdot)$ yields $\sigma^2_{WLS} \leq \sigma^2_{LS}$, with strict inequality unless $\xi = 1-\alpha-\beta$ and $(\kappa-1)\alpha^2 = 0$. By (A.4), $\xi = 1-\alpha-\beta$ if and only if σ_t^2 is constant, which occurs when $\alpha = 0$ or $\eta_t^2 = 1$ a.s. In particular, when $\alpha = 0$ the two formulas reduce to $\sigma^2_{WLS} = \sigma^2_{LS} = 1-\varphi_0^2$. □

Proof of Theorem 3.1. We start by showing the consistency of $\widehat{\varphi}_{3WLS}$. We have

$$\widehat{\varphi}_{3WLS} = \varphi_0 + \frac{\frac{1}{n}\sum_{t=1}^n \frac{1}{\tilde{\sigma}^2_{t,n}}\epsilon_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^n \frac{1}{\tilde{\sigma}^2_{t,n}} X^2_{t+1}}. \tag{A.6}$$

Moreover,

$$\left|\frac{1}{n}\sum_{t=1}^n \frac{1}{\tilde{\sigma}^2_{t,n}}\epsilon_t X_{t+1}\right| \leq \left|\frac{1}{n}\sum_{t=1}^n \frac{1}{\sigma^2_t}\epsilon_t X_{t+1}\right| + \frac{1}{n}\sum_{t=1}^n \left|\frac{1}{\tilde{\sigma}^2_{t,n}} - \frac{1}{\sigma^2_t}\right| |\epsilon_t X_{t+1}|, \tag{A.7}$$

where the first term in the r.h.s. converges a.s. to $E(\frac{1}{\sigma_t^2}\epsilon_t X_{t+1}) = 0$ by arguments already given. Now

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n \left|\frac{1}{\tilde{\sigma}^2_{t,n}} - \frac{1}{\sigma^2_t}\right| |\epsilon_t X_{t+1}| \leq& \frac{1}{n}\sum_{t=1}^n \left|\frac{1}{\tilde{\sigma}^2_{t,n}} - \frac{1}{\sigma^2_t(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n)}\right| |\epsilon_t X_{t+1}| + \frac{1}{n}\sum_{t=1}^n \left|\frac{1}{\sigma^2_t(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n)} - \frac{1}{\sigma^2_t}\right| |\epsilon_t X_{t+1}| \\ \leq& \frac{1}{\underline{\omega}^2}\frac{1}{n}\sum_{t=1}^n \rho_t |\epsilon_t X_{t+1}| + \frac{1}{\underline{\omega}}|\widehat{\varphi}_{LS} - \varphi_0| \frac{1}{n}\sum_{t=1}^n \left|\frac{1}{\sigma^2_t(\boldsymbol{\vartheta}^*_n)}\frac{\partial}{\partial\varphi}\sigma^2_t(\boldsymbol{\vartheta}^*_n)\right| |\epsilon_t X_{t+1}| \\ &+ \frac{1}{\underline{\omega}}\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \frac{1}{n}\sum_{t=1}^n \left\|\frac{1}{\sigma^2_t(\boldsymbol{\vartheta}^*_n)}\frac{\partial}{\partial\boldsymbol{\theta}}\sigma^2_t(\boldsymbol{\vartheta}^*_n)\right\| |\epsilon_t X_{t+1}|, \end{aligned} \tag{A.8}$$

using Assumption **A2** for the first term in the right-hand side and, for the second term, a Taylor expansion of $\sigma_t^2(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n)$ around $\boldsymbol{\vartheta}_0$, where $\boldsymbol{\vartheta}_n^*$ is between $(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n')'$ and $\boldsymbol{\vartheta}_0$. For any $\delta > 0$, we have by **A2**, Markov's and Hölder's inequalities

$$\sum_{t=0}^{\infty} P(\rho_t|\epsilon_t X_{t+1}| > \delta) \leq \delta^{-2}\left(\sum_{t=0}^{\infty}\rho_t^2\right) E(|\epsilon_t X_{t+1}|)^2 \leq KE(\epsilon_t^4) < \infty.$$

Hence $\rho_t|\epsilon_t X_{t+1}| \to 0, a.s.$, from which it follows by Cesàro's lemma that the first term in the r.h.s. of (A.8) vanishes a.s. as $n \to \infty$. Moreover, the second term is bounded by

$$\frac{1}{\underline{\omega}}|\widehat{\varphi}_{LS} - \varphi_0|\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{1}{\sigma_t^2(\boldsymbol{\vartheta})}\frac{\partial}{\partial\varphi}\sigma_t^2(\boldsymbol{\vartheta})\right||\epsilon_t X_{t+1}|,$$

which, using again Cesàro's lemma, the strong consistency of the LSE of φ_0, Assumption **A4** and Hölder's inequality, tends to zero a.s. as $n \to \infty$. The third term in the r.h.s. of (A.8) is treated similarly using **A3**. Therefore, the second term in the r.h.s. of (A.7) converges to 0 a.s., and so does the left-hand side. By arguments already used, the denominator of the second term in the r.h.s. of (A.6) converges a.s. to $E(\frac{1}{\sigma_t^2}\epsilon_t X_{t+1}^2) < \infty$. The strong consistency of $\widehat{\varphi}_{3WLS}$ is thus established.

Turning to the asymptotic normality, we note that

$$\sqrt{n}(\widehat{\varphi}_{3WLS} - \varphi_0) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\tilde{\sigma}_{t,n}^2}\epsilon_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\tilde{\sigma}_{t,n}^2}X_{t+1}^2},$$

with

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\tilde{\sigma}_{t,n}^2}\epsilon_t X_{t+1} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t^2}\epsilon_t X_{t+1} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{1}{\sigma_t^2(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n)} - \frac{1}{\sigma_t^2}\right)\epsilon_t X_{t+1} \\ &+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{1}{\tilde{\sigma}_{t,n}^2} - \frac{1}{\sigma_t^2(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n)}\right)\epsilon_t X_{t+1} := A_n + B_n + C_n.\end{aligned}$$

Using again a Taylor expansion, with $\widehat{\boldsymbol{\vartheta}}_n^*$ between $(\widehat{\varphi}_{LS}, \widehat{\boldsymbol{\theta}}_n')'$ and $\boldsymbol{\vartheta}_0$, we have

$$\begin{aligned}B_n &= \left(\frac{-1}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^4(\widehat{\boldsymbol{\vartheta}}_n^*)}\frac{\partial\sigma_t^2(\widehat{\boldsymbol{\vartheta}}_n^*)}{\partial\varphi}\right)\sqrt{n}(\widehat{\varphi}_{LS} - \varphi_0) + \left(\frac{-1}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^4(\widehat{\boldsymbol{\vartheta}}_n^*)}\frac{\partial\sigma_t^2(\widehat{\boldsymbol{\vartheta}}_n^*)}{\partial\boldsymbol{\theta}'}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &:= \Phi_n\sqrt{n}(\widehat{\varphi}_{LS} - \varphi_0) + \boldsymbol{M}_n'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).\end{aligned}$$

Let us show that $\boldsymbol{M}_n \to \mathbf{0}$ in probability when $n \to \infty$. Another Taylor expansion yields

$$\begin{aligned}\boldsymbol{M}_n =& \frac{-1}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^4}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}} \\ &+ \left(\frac{2}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^6(\widetilde{\boldsymbol{\vartheta}}_n)}\frac{\partial\sigma_t^2(\widetilde{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\widetilde{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}'} - \frac{1}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^4(\widetilde{\boldsymbol{\vartheta}}_n)}\frac{\partial^2\sigma_t^2(\widetilde{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right)(\widehat{\boldsymbol{\vartheta}}_n^* - \boldsymbol{\vartheta}_0), \qquad \text{(A.9)}\end{aligned}$$

where $\widetilde{\boldsymbol{\vartheta}}_n$ is between $\widehat{\boldsymbol{\vartheta}}_n^*$ and $\boldsymbol{\vartheta}_0$. The first sum converges a.s. by the ergodic theorem to

$$E\left(\frac{\epsilon_t X_{t+1}}{\sigma_t^4}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}}\right)=\sum_{i=0}^{\infty}\varphi_0^i E\left(\frac{\epsilon_t\sigma_{t+1+i}\eta_{t+1+i}}{\sigma_t^4}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}}\right)=0,$$

where the last equality holds because for any $i\geq 0$, the variable η_{t+1+i} is independent from all the other variables involved in the expectation. In particular, it is important to note that $\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}}$ belongs to $\mathcal{F}_{t-1}$. Moreover,

$$\begin{aligned}\limsup_{n\to\infty}\left\|\frac{2}{n}\sum_{t=1}^{n}\frac{\epsilon_t X_{t+1}}{\sigma_t^6(\widetilde{\boldsymbol{\vartheta}}_n)}\frac{\partial\sigma_t^2(\widetilde{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\widetilde{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}'}\right\| &\leq \limsup_{n\to\infty}\frac{2}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{\epsilon_t X_{t+1}}{\sigma_t^6(\boldsymbol{\vartheta})}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}'}\right\|\\ &\leq \limsup_{n\to\infty}\frac{2}{n}\frac{1}{\underline{\omega}}\sum_{t=1}^{n}|\epsilon_t X_{t+1}|\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^4(\boldsymbol{\vartheta})}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}'}\right\|\\ &=\frac{2}{\underline{\omega}}E\left(|\epsilon_t X_{t+1}|\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|\frac{1}{\sigma_t^4(\boldsymbol{\vartheta})}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}'}\right\|\right)<\infty,\end{aligned}$$

using the ergodic theorem and Assumptions **A1** and **A4**. The third sum in the r.h.s. of (A.9) can be handled similarly. It follows from the consistency of $\widehat{\boldsymbol{\vartheta}}_n^*$ to $\boldsymbol{\vartheta}_0$ that $\boldsymbol{M}_n\to\mathbf{0}$ in probability.

We can similarly show that

$$\Phi_n\to\Phi=E\left(-\frac{\eta_t X_{t+1}}{\sigma_t^3}\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right),\quad C_n\to 0,\quad \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\tilde{\sigma}_{t,n}^2}X_{t+1}^2\to E(\sigma_t^{-2}X_{t+1}^2)$$

in probability when $n\to\infty$.

It follows from the previous results, together with the equality (A.2) written with $\tau_t=1$ for the LS estimator, that

$$\sqrt{n}(\widehat{\varphi}_{3WLS}-\varphi_0)=\frac{1}{\mu_{\sigma^2}\sqrt{n}}\sum_{t=1}^{n}\left(\sigma_t^{-2}+\Phi\left\{E(X_t^2)\right\}^{-1}\right)\epsilon_t X_{t+1}+o_P(1)=\frac{1}{\mu_{\sigma^2}\sqrt{n}}\sum_{t=1}^{n}Y_t+o_P(1). \tag{A.10}$$

We can now deduce the asymptotic normality in (3.7) by the same arguments as in the proof of Proposition 3.1, using the fact that $\{(Y_t,\mathcal{F}_t)\}$ is a mixingale. The last statement of the theorem is proven in a similar way to the proof of Corollary 3.1. □

Proof of Theorem 3.2. Let

$$r_h=\frac{1}{n}\sum_{t=h+1}^{n}\eta_t\eta_{t-h},\qquad \boldsymbol{r}_m=(r_1,\ldots,r_m)'.$$

For $\boldsymbol{\vartheta} = (\varphi, \boldsymbol{\theta}')'$, let $\eta_t(\boldsymbol{\vartheta}) = \epsilon_t(\varphi)/\sigma_t(\boldsymbol{\vartheta})$ and $\widetilde{\eta}_t(\boldsymbol{\vartheta}) = \epsilon_t(\varphi)/\widetilde{\sigma}_t(\boldsymbol{\vartheta})$, and let $r_h(\boldsymbol{\vartheta})$ and $\boldsymbol{r}_m(\boldsymbol{\vartheta})$ (resp. $\widetilde{r}_h(\boldsymbol{\vartheta})$ and $\widetilde{\boldsymbol{r}}_m(\boldsymbol{\vartheta})$) be obtained by replacing η_t with $\eta_t(\boldsymbol{\vartheta})$ (resp. $\widetilde{\eta}_t(\boldsymbol{\vartheta})$) in r_h and $\boldsymbol{r}_m$. Then $\boldsymbol{r}_m = \boldsymbol{r}_m(\boldsymbol{\vartheta}_0)$ and $\widehat{\boldsymbol{r}}_m = \widetilde{\boldsymbol{r}}_m(\widehat{\boldsymbol{\vartheta}}_n)$ with $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\varphi}_n, \widehat{\boldsymbol{\theta}}_n')'$. We similarly define $s_t(\boldsymbol{\vartheta})$ and $\widetilde{s}_t(\boldsymbol{\vartheta})$.

Let $\nabla Q = (\nabla_1 Q, \dots, \nabla_{d+1} Q)'$ be the vector of the first-order derivatives of a function $Q : \boldsymbol{\Theta} \mapsto \mathbb{R}$. By the same arguments as in Francq, Wintenberger and Zakoïan (2018), it can be shown that

$$\sqrt{n}\, \|\boldsymbol{r}_m - \widetilde{\boldsymbol{r}}_m(\boldsymbol{\vartheta}_0)\| = o_P(1), \qquad \sup_{\boldsymbol{\vartheta} \in \mathcal{V}} \|\nabla \boldsymbol{r}_m(\boldsymbol{\vartheta}) - \nabla \widetilde{\boldsymbol{r}}_m(\boldsymbol{\vartheta})\| = o_P(1), \tag{A.11}$$

for any neighborhood $\mathcal{V}$ of $\boldsymbol{\vartheta}_0$. Arguing as in Francq, Wintenberger and Zakoïan (2018), we thus find that

$$\begin{aligned}\sqrt{n}\, \widehat{\boldsymbol{r}}_m &= \sqrt{n}\, \widetilde{\boldsymbol{r}}_m(\boldsymbol{\vartheta}_0) + [\nabla \widetilde{\boldsymbol{r}}_m'(\boldsymbol{\vartheta}^*)]' \sqrt{n}\, (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \\ &= \sqrt{n}\, \boldsymbol{r}_m + \boldsymbol{K}_m \sqrt{n}\, (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) + o_P(1),\end{aligned} \tag{A.12}$$

where the h-th row of the matrix $[\nabla \widetilde{\boldsymbol{r}}_m'(\boldsymbol{\vartheta}^*)]'$ is the transpose of $\nabla \widetilde{r}_h(\boldsymbol{\vartheta}_h^*)$ for some $\boldsymbol{\vartheta}_h^*$ between $\widehat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$, while the h-th row of the $m \times (d+1)$ matrix $\boldsymbol{K}_m$ is c_h' with

$$c_h = E\{\eta_{t-h} \nabla \eta_t(\boldsymbol{\vartheta}_0) + \eta_t \nabla \eta_{t-h}(\boldsymbol{\vartheta}_0)\}.$$

Because $\eta_t = \epsilon_t(\varphi)/\sigma_t(\boldsymbol{\vartheta})$, we have

$$\nabla_\varphi \eta_t(\boldsymbol{\vartheta}_0) = -\frac{X_{t+1}}{\sigma_t} - \frac{\eta_t}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \varphi}(\boldsymbol{\vartheta}_0), \quad \nabla_{\boldsymbol{\theta}} \eta_t(\boldsymbol{\vartheta}_0) = -\frac{\eta_t}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}(\boldsymbol{\vartheta}_0),$$

where the latter derivative belongs to $\mathcal{F}_{t-1}$. This entails $c_h = (z_h, \mathbf{0}_{1\times d})'$ where

$$\begin{aligned} z_h &= E\{\eta_{t-h} \nabla_\varphi \eta_t(\boldsymbol{\vartheta}_0) + \eta_t \nabla_\varphi \eta_{t-h}(\boldsymbol{\vartheta}_0)\} \\ &= -\varphi_0^{h-1}\, E\left(\frac{\sigma_t}{\sigma_{t-h}}\right) - \frac{1}{2}\, E\left[\eta_{t-h}\eta_t \left(\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi} + \frac{1}{\sigma_{t-h}^2}\frac{\partial \sigma_{t-h}^2}{\partial \varphi}\right)\right]. \end{aligned}$$

It follows from (A.12) that

$$\sqrt{n}\, \widehat{\boldsymbol{r}}_m = \sqrt{n}\, \boldsymbol{r}_m + \boldsymbol{z}_m \sqrt{n}\, (\widehat{\varphi}_{3WLS} - \varphi_0) + o_P(1). \tag{A.13}$$

By (A.10) we have, with $\boldsymbol{\eta}_{t-1:t-m} = (\eta_{t-1}, \dots, \eta_{t-m})'$,

$$\sqrt{n} \begin{pmatrix} \widehat{\varphi}_{3WLS} - \varphi_0 \\ \boldsymbol{r}_m \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t \begin{pmatrix} \mu_{\sigma^2}^{-1} \sigma_t \nu_t X_{t+1} \\ \boldsymbol{\eta}_{t-1:t-m} \end{pmatrix} + o_P(1),$$

where $\nu_t = \sigma_t^{-2} + \Phi \left\{ E(X_t^2) \right\}^{-1}$.

To derive the off-diagonal term of the asymptotic variance of the left-hand side, we compute for any $h > 0$ and $i = 1, \ldots, m$,

$$\begin{aligned}
\gamma_i(h) = \mathrm{Cov}\left(\mu_{\sigma^2}^{-1} \nu_t \epsilon_t X_{t+1}, \eta_{t+h}\eta_{t+h-i}\right) &= \mu_{\sigma^2}^{-1} \sum_{j=0}^{\infty} \varphi_0^j E\left(\nu_t \epsilon_t \epsilon_{t+1+j} \eta_{t+h}\eta_{t+h-i}\right) \\
&= \mu_{\sigma^2}^{-1} \varphi_0^{h-1} E\left(\nu_t \epsilon_t \sigma_{t+h}\eta_{t+h-i}\right).
\end{aligned}$$

We also have $\gamma_i(h) = 0$ for $h \leq 0$. The asymptotic covariance is thus $\boldsymbol{\iota}_m$ where $\iota_i = \sum_{h=-\infty}^{\infty} \gamma_i(h)$. We thus have, using Theorem 3.1,

$$\sqrt{n} \begin{pmatrix} \widehat{\varphi}_{3WLS} - \varphi_0 \\ \boldsymbol{r}_m \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{ \boldsymbol{0}, \begin{pmatrix} \sigma_{3WLS}^2 & \boldsymbol{\iota}_m' \\ \boldsymbol{\iota}_m & \boldsymbol{I}_m \end{pmatrix} \right\}, \tag{A.14}$$

where the convergence in distribution follows from arguments already used in the proof of Proposition 3.1. Combining (A.14) with (A.13) yields $\sqrt{n}\,\widehat{\boldsymbol{r}}_m \Rightarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ with

$$\boldsymbol{D} = \boldsymbol{I}_m + \sigma_{WLS}^2\, \boldsymbol{z}_m \boldsymbol{z}_m' + \boldsymbol{z}_m \boldsymbol{\iota}' + \boldsymbol{\iota} \boldsymbol{z}_m' = I_m + \boldsymbol{Z}_m \boldsymbol{M}_1 \boldsymbol{Z}_m', \quad \boldsymbol{Z}_m = [\boldsymbol{z}_m \;\; \boldsymbol{\iota}],\; \boldsymbol{M}_1 = \begin{pmatrix} \sigma_{3WLS}^2 & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that,

$$\boldsymbol{D}^{-1} = \boldsymbol{I}_m - \boldsymbol{Z}\boldsymbol{M}_2^{-1}\boldsymbol{Z}', \qquad \boldsymbol{M}_2 = \boldsymbol{M}_1^{-1} + \boldsymbol{Z}'\boldsymbol{Z}$$

whenever $\boldsymbol{M}_2$ is nonsingular. The convergence in distribution in (3.10) straightforwardly follows from $n\widehat{\boldsymbol{r}}_m^{(1')} \boldsymbol{D}^{-1} \widehat{\boldsymbol{r}}_m^{(1)} \Rightarrow \chi_m^2$, by plugging consistent estimators of the entries of $\boldsymbol{D}^{-1}$. □

B Proofs of the results of Section 4

Proof of Proposition 4.1. The existence and uniqueness of a strictly stationary and ergodic solution (ϵ_t) to the GARCH(p, q) model is ensured by the condition $\gamma(\boldsymbol{A}_0) < 0$ (see Bougerol and Picard, 1992). Moreover, we have $E|\epsilon_t|^s < \infty$ for some $s \in (0, 1)$ (see for instance, Francq and Zakoïan, 2019, Corollary 2.3). Therefore, $X_t = \sum_{i=0}^{\infty} \varphi_0^i \epsilon_{t+i}$ defines a strictly stationary and ergodic process such that $E|X_t|^s < \infty$. If (X_t^*) is another strictly stationary solution to Model (4.1) with $E|X_t^*|^s < \infty$, we have $X_t - X_t^* = \varphi_0^N (X_{t+N} - X_{t+N}^*)$ for any $N \geq 0$, hence $E|X_t - X_t^*|^s \leq |\varphi_0|^{Ns}(|X_{t+N}|^s + |X_{t+N}^*|^s) \to 0$ as $N \to \infty$. We thus have $E|X_t - X_t^*|^s = 0$, from which we deduce that $X_t = X_t^*$ for any t a.s. Hence the existence and uniqueness of a strictly stationary solution to Model (4.1). □

Proof of Proposition 4.2. For $h \geq 0$ we have

$$\gamma_{X^2}(h) := \mathrm{Cov}(X_t^2, X_{t+h}^2) = \varphi_0^2 \gamma_{X^2}(h-1) + \mathrm{Cov}(\epsilon_t^2, X_{t+h}^2) + 2\varphi_0 \mathrm{Cov}(\epsilon_t X_{t+1}, X_{t+h}^2).$$

For $h = 0$ the last covariance is equal to 0 by arguments already used. For $h > 0$, we have

$$\begin{aligned}
\mathrm{Cov}(\epsilon_t X_{t+1}, X_{t+h}^2) &= \sum_{i,j,k=0}^{\infty} \varphi_0^{i+j+k} \mathrm{Cov}(\epsilon_t \epsilon_{t+1+i}, \epsilon_{t+h+j} \epsilon_{t+h+k}) \\
&= \sum_{i,j,k=0}^{\infty} \varphi_0^{i+j+k} E(\eta_t \sigma_t \epsilon_{t+1+i} \epsilon_{t+h+j} \epsilon_{t+h+k}) \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{h+j-2} \varphi_0^{i+2j} E(\eta_t \sigma_t \epsilon_{t+1+i} \epsilon_{t+h+j}^2) + \sum_{j=0}^{\infty} \varphi_0^{3j+h-1} E(\eta_t \sigma_t \epsilon_{t+h+j}^3) \\
&+ 2 \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \varphi_0^{2j+k+h-1} E(\eta_t \sigma_t \epsilon_{t+h+k} \epsilon_{t+h+j}^2) = 0
\end{aligned}$$

where the last equality follows from the assumption of a symmetric law for η_t. Indeed, the expectations involved in the sums are all of the form, for some

$$E\{\eta_t Z_1(\eta_{t-1}, \ldots) Z_2(\eta_{t+h+k}, \eta_{t+h+k-1}, \ldots, \eta_t, \eta_{t-1}, \ldots)\}$$

for some measurable functions Z_1 and Z_2, where Z_2 is an even function of η_t. We thus have shown that for any $h \geq 0$,

$$\gamma_{X^2}(h) = \varphi_0^2 \gamma_{X^2}(h-1) + \mathrm{Cov}(\epsilon_t^2, X_{t+h}^2).$$

Now

$$\mathrm{Cov}(\epsilon_t^2, X_{t+h}^2) = \mathrm{Cov}(\epsilon_t^2, \sum_{i,j=0}^{\infty} \varphi_0^{i+j} \epsilon_{t+h+i} \epsilon_{t+h+j}) = \mathrm{Cov}(\epsilon_t^2, \sum_{i=0}^{\infty} \varphi_0^{2i} \epsilon_{t+h+i}^2) = \sum_{i=0}^{\infty} \varphi_0^{2i} \gamma_{\epsilon^2}(h+i).$$

Therefore, for any $h \geq 0$,

$$(1 - \varphi_0^2 L) \gamma_{X^2}(h) = \sum_{i=0}^{\infty} \varphi_0^{2i} \gamma_{\epsilon^2}(h+i).$$

Finally, we know that (ϵ_t^2) admits an ARMA$(p \vee q, p)$ representation characterized by

$$\{1 - \alpha(L) - \beta(L)\} \gamma_{\epsilon^2}(h) = 0, \quad \text{for} \ \ h > p,$$

see for instance Francq and Zakoïan (2019), Equation (2.4). The conclusion follows. □

Proof of Theorem 4.1. Define

$$\mathbf{l}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \ell_t(\varphi_0, \boldsymbol{\theta}), \qquad \ell_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2(\varphi)}{\sigma_t^2(\boldsymbol{\vartheta})} + \log \sigma_t^2(\boldsymbol{\vartheta}),$$

where $\sigma_t^2(\boldsymbol{\vartheta})$ is the strictly stationary solution of the stochastic recurrence equation $\sigma_t^2(\boldsymbol{\vartheta}) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2(\varphi) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2(\boldsymbol{\vartheta})$ for $t \in \mathbb{Z}$ under, in particular, the condition $\sum_{j=1}^{p} \beta_j < 1$.

The strong consistency of $\widehat{\boldsymbol{\theta}}_n$ will be a consequence of the following intermediate steps (see Francq and Zakoïan, Section 7.4, 2019). A major difference with the standard QML estimation of GARCH models is that the summands in $\tilde{\ell}_{t,n}(\boldsymbol{\theta})$ depend on n, through the estimator of φ_0.

i) $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_\sigma} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})| = 0, \quad a.s.$

ii) if $\sigma_t^2(\varphi_0, \boldsymbol{\theta}) = \sigma_t^2(\varphi_0, \boldsymbol{\theta}_0) \quad a.s.,$ then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

iii) if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then $E\ell_t(\varphi_0, \boldsymbol{\theta}) > E\ell_t(\varphi_0, \boldsymbol{\theta}_0)$,

iv) any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ has a neighborhood $V(\boldsymbol{\theta})$ such that $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) > E\ell_1(\varphi_0, \boldsymbol{\theta}_0) \quad a.s.$

We first show *i)*. It will be convenient to rewrite (4.2) in matrix form when $\varphi = \widehat{\varphi}_n$, as

$$\tilde{\boldsymbol{v}}_{t,n}(\boldsymbol{\theta}) = \tilde{\boldsymbol{c}}_{t,n}(\boldsymbol{\theta}) + \boldsymbol{B}\tilde{\boldsymbol{v}}_{t-1,n}(\boldsymbol{\theta}), \tag{B.1}$$

where

$$\tilde{\boldsymbol{v}}_{t,n}(\boldsymbol{\theta}) = \begin{pmatrix} \tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) \\ \tilde{\sigma}_{t-1,n}^2(\boldsymbol{\theta}) \\ \vdots \\ \tilde{\sigma}_{t-p+1,n}^2(\boldsymbol{\theta}) \end{pmatrix}, \quad \tilde{\boldsymbol{c}}_{t,n}(\boldsymbol{\theta}) = \begin{pmatrix} \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2(\widehat{\varphi}_n) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_p \\ 1 & 0 & \dots & 0 \\ \vdots & & & \\ 0 & \dots & 1 & 0 \end{pmatrix}.$$

It follows that

$$\tilde{\boldsymbol{v}}_{t,n}(\boldsymbol{\theta}) = \sum_{k=0}^{t-1} \boldsymbol{B}^k \tilde{\boldsymbol{c}}_{t-k,n}(\boldsymbol{\theta}) + \boldsymbol{B}^t \tilde{\boldsymbol{v}}_{0,n}(\boldsymbol{\theta}).$$

Similarly, we have

$$\boldsymbol{v}_t(\boldsymbol{\vartheta}) = \boldsymbol{c}_t(\boldsymbol{\vartheta}) + \boldsymbol{B}\boldsymbol{v}_{t-1}(\boldsymbol{\vartheta}) = \sum_{k=0}^{t-1} \boldsymbol{B}^k \boldsymbol{c}_{t-k}(\boldsymbol{\vartheta}) + \boldsymbol{B}^t \boldsymbol{v}_0(\boldsymbol{\vartheta}) = \sum_{k=0}^{\infty} \boldsymbol{B}^k \boldsymbol{c}_{t-k}(\boldsymbol{\vartheta}), \tag{B.2}$$

where $\boldsymbol{v}_t(\boldsymbol{\vartheta}) = (\sigma_t^2(\boldsymbol{\vartheta}), \dots, \sigma_{t-p+1}^2(\boldsymbol{\vartheta}))'$ and $\boldsymbol{c}_t(\boldsymbol{\vartheta}) = \left(\omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2(\varphi), 0, \dots, 0\right)'$. Note that

$$\begin{aligned}
\tilde{\boldsymbol{c}}_{t,n}(\boldsymbol{\theta}) - \boldsymbol{c}_t(\varphi_0, \boldsymbol{\theta}) &= \left(\sum_{i=1}^q \alpha_i \{\epsilon_{t-i}^2(\widehat{\varphi}_n) - \epsilon_{t-i}^2(\varphi_0)\}, 0, \dots, 0\right)' \\
&= (\varphi_0 - \widehat{\varphi}_n) \left(\sum_{i=1}^q \alpha_i X_{t-i+1}\{2X_{t-i} - (\widehat{\varphi}_n + \varphi_0) X_{t-i+1}\}, 0, \dots, 0\right)'.
\end{aligned}$$

Hence

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \|\tilde{\boldsymbol{c}}_{t,n}(\boldsymbol{\theta}) - \boldsymbol{c}_t(\varphi_0, \boldsymbol{\theta})\| \leq K|\widehat{\varphi}_n - \varphi_0| Y_t, \quad Y_t = \sum_{i=1}^q |X_{t-i+1}|(|X_{t-i}| + |X_{t-i+1}|),$$

where $K > 0$ denotes a generic variable belonging to $\mathcal{F}_{\eta,0}$ whose value can vary from line to line.

Recall that the condition $\sum_{i=1}^p \beta_i < 1$ of **B1** entails $\rho(\boldsymbol{B}) < 1$, where $\rho(\boldsymbol{B})$ is the spectral radius of $\boldsymbol{B}$. The compactness of $\boldsymbol{\Theta}_\sigma$ entails that $\rho := \sup_{\theta \in \boldsymbol{\Theta}_\sigma} \rho(\boldsymbol{B}) < 1$. We thus have

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \|\tilde{\boldsymbol{v}}_{t,n}(\boldsymbol{\theta}) - \boldsymbol{v}_t(\varphi_0, \boldsymbol{\theta})\| &= \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \left\| \sum_{k=0}^{t-1} \boldsymbol{B}^k \{\tilde{\boldsymbol{c}}_{t-k,n}(\boldsymbol{\theta}) - \boldsymbol{c}_{t-k}(\varphi_0, \boldsymbol{\theta})\} + \boldsymbol{B}^t \{\tilde{\boldsymbol{v}}_{0,n}(\boldsymbol{\theta}) - \boldsymbol{v}_0(\varphi_0, \boldsymbol{\theta})\} \right\| \\
&\leq K|\widehat{\varphi}_n - \varphi_0| \sum_{k=0}^{t-1} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \left\| \boldsymbol{B}^k \right\| Y_{t-k} + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \left\| \boldsymbol{B}^t \right\| \|\tilde{\boldsymbol{v}}_{0,n}(\boldsymbol{\theta}) - \boldsymbol{v}_0(\varphi_0, \boldsymbol{\theta})\| \\
&\leq K|\widehat{\varphi}_n - \varphi_0| Y_t^\infty + K\rho^t, \qquad Y_t^\infty = \sum_{k=0}^\infty \rho^k Y_{t-k}. \qquad \text{(B.3)}
\end{aligned}$$

Note that $\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) > \underline{\omega}$ and $\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \sigma_t^2(\varphi_0, \boldsymbol{\theta}) > \underline{\omega}$ for some $\underline{\omega} > 0$. Thus

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} |\mathbf{l}_n(\boldsymbol{\theta}) - \tilde{\mathbf{l}}_n(\boldsymbol{\theta})| \\
\leq\ & \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\sigma} \left\{ |\epsilon_t^2 - \epsilon_t^2(\widehat{\varphi}_n)| \frac{1}{\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta})} + \epsilon_t^2 \frac{|\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) - \sigma_t^2(\varphi_0, \boldsymbol{\theta})|}{\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta}) \sigma_t^2(\varphi_0, \boldsymbol{\theta})} + \left| \log \left(\frac{\sigma_t^2(\varphi_0, \boldsymbol{\theta})}{\tilde{\sigma}_{t,n}^2(\boldsymbol{\theta})} \right) \right| \right\} \\
\leq\ & K|\widehat{\varphi}_n - \varphi_0| \frac{1}{n} \sum_{t=1}^n \{|X_{t+1}|(|X_t| + |X_{t+1}|) + (\epsilon_t^2 + 1) Y_t^\infty\} + \frac{K}{n} \sum_{t=1}^n \rho^t (\epsilon_t^2 + 1)
\end{aligned}$$

using (B.3) and the elementary inequality $|\log(x/y)| \leq |x - y|/(x \wedge y)$ for $x, y > 0$. The first term converges to zero by the a.s. convergence of $\widehat{\varphi}_n$ to φ_0 and because the term into accolades has a finite expectation by **B3**. The second term goes to zero a.s. using Cesàro's lemma and $\rho^t(\epsilon_t^2 + 1) \to 0$ a.s. Thus *i)* is established.

The proof of *ii)* and *iii)* is the same as for standard GARCH models. The proof of iv) follows the lines of the proof of Theorem 2.1 in Francq, Horváth and Zakoïan (2011): letting $V_k(\boldsymbol{\theta})$ denote the open ball with center $\boldsymbol{\theta}$ and radius $1/k$, we have, using *i), iii)*, the monotone convergence and ergodic theorems,

$$\begin{aligned}\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}_\sigma} \tilde{\mathbf{l}}_n(\boldsymbol{\theta}^*) &\geq \liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}_\sigma} \mathbf{l}_n(\boldsymbol{\theta}^*) - \limsup_{n\to\infty} \sup_{\boldsymbol{\Theta}_\sigma} |\tilde{\mathbf{l}}_n(\boldsymbol{\theta}) - \mathbf{l}_n(\boldsymbol{\theta})| \\ &\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}_\sigma} \ell_t(\varphi_0,\boldsymbol{\theta}^*) > E\ell_1(\varphi_0,\boldsymbol{\theta}_0),\end{aligned}$$

for k large enough and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. The consistency of $\widehat{\boldsymbol{\theta}}_n$ is thus established. □

Proof of Theorem 4.2. We have

$$\begin{aligned}\mathbf{0}_d &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_{t,n}(\widehat{\boldsymbol{\theta}}_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t(\widehat{\varphi}_{WLS},\widehat{\boldsymbol{\theta}}_n) \\ &= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{J}_n\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \boldsymbol{K}_n\sqrt{n}(\widehat{\varphi}_{WLS} - \varphi_0),\end{aligned} \tag{B.4}$$

where, for some vectors $\boldsymbol{\vartheta}_i^*$ between $\boldsymbol{\vartheta}_0$ and $(\widehat{\varphi}_{WLS}, \widehat{\boldsymbol{\theta}}_n')'$,

$$\boldsymbol{J}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\ell}_t(\boldsymbol{\vartheta}_i^*)\right), \quad \boldsymbol{K}_n = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\varphi}\tilde{\ell}_t(\boldsymbol{\vartheta}_i^*)\right).$$

The asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ will follow from the following results

i) $E\left\|\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'}\right\| < \infty, \quad E\left\|\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\| < \infty,$

ii) $\boldsymbol{J}$ is non-singular,

iii) there exists a neighborhood $\mathcal{V}(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$ such that, for all $i, j \in \{1,\dots,d\}$,

$$E\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left|\frac{\partial^3\ell_t(\boldsymbol{\vartheta})}{\partial\theta_i\partial\theta_j\partial\varphi}\right| < \infty,$$

iv) $\left\|n^{-1/2}\sum_{t=1}^{n}\left\{\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{\ell}_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}}\right\}\right\| \to 0$ and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}(\boldsymbol{\vartheta}_0)}\left\|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi} - \frac{\partial^2\tilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi}\right\}\right\| \to 0$

in probability when $n \to \infty$,

v) $n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\boldsymbol{\vartheta}_k^*) \to \boldsymbol{J}(i,j)$ *a.s.*, $\quad n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\varphi}\ell_t(\boldsymbol{\vartheta}_k^*) \to \boldsymbol{K}(i)$ *a.s.*,

vi) $n^{-1/2}\sum_{t=1}^{n}\boldsymbol{W}_t \Rightarrow \mathcal{N}\left\{\mathbf{0}, \sum_{h=-\infty}^{\infty}\boldsymbol{\Gamma}(h)\right\}.$

The derivatives of $\ell_t(\boldsymbol{\vartheta}) = \epsilon_t^2(\varphi)/\sigma_t^2(\boldsymbol{\vartheta}) + \log\sigma_t^2(\boldsymbol{\vartheta})$ are given by

$$\frac{\partial \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = 2\frac{\epsilon_t}{\sigma_t^2}\frac{\partial \epsilon_t}{\partial \boldsymbol{\vartheta}} + \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}(\boldsymbol{\vartheta}), \tag{B.5}$$

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} &= 2\frac{\epsilon_t}{\sigma_t^2}\frac{\partial \epsilon_t}{\partial \boldsymbol{\vartheta}}\frac{\partial \epsilon_t}{\partial \boldsymbol{\vartheta}'} - 2\frac{\epsilon_t}{\sigma_t^2}\left(\frac{\partial \epsilon_t}{\partial \boldsymbol{\vartheta}}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'} + \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\frac{\partial \epsilon_t}{\partial \boldsymbol{\vartheta}'}\right) \\ &\quad + \left\{1 - \frac{\epsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}) + \left\{2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}).\end{aligned} \tag{B.6}$$

Similar formulas hold for the derivatives of $\tilde{\ell}_t$, with σ_t^2 replaced by $\tilde{\sigma}_t^2$. We thus have, at the true parameter value,

$$\frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} = \frac{2\eta_t}{\sigma_t}\boldsymbol{Z}_{t+1} + (1-\eta_t^2)\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}(\boldsymbol{\vartheta}_0), \quad \frac{\partial \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta}} = (1-\eta_t^2)\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\right\}(\boldsymbol{\theta}_0), \tag{B.7}$$

$$\begin{aligned}\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} &= \frac{2\eta_t}{\sigma_t}\boldsymbol{Z}_{t+1}\boldsymbol{Z}'_{t+1} - 2\frac{\eta_t}{\sigma_t}\left(\boldsymbol{Z}_{t+1}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}\right\} + \left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}\boldsymbol{Z}'_{t+1}\right) \\ &\quad + \left\{1-\eta_t^2\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}_0) + \left\{2\eta_t^2 - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\vartheta}'}\right\}(\boldsymbol{\vartheta}_0),\end{aligned} \tag{B.8}$$

where $\boldsymbol{Z}_{t+1} = \frac{\partial \epsilon_t(\varphi)}{\partial \boldsymbol{\vartheta}} = (-X_{t+1}, \mathbf{0}'_d)'$. It follows that

$$\begin{aligned}E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) &= E\left(\left\{1-\eta_t^2\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right\}(\boldsymbol{\vartheta}_0) + \left\{2\eta_t^2 - 1\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'}\right\}(\boldsymbol{\vartheta}_0)\right) \\ &= E\left\{\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'}\right\}(\boldsymbol{\vartheta}_0),\end{aligned}$$

using the fact that the first and second derivatives of σ_t^2 with respect to $\boldsymbol{\theta}$ at the point $\boldsymbol{\vartheta}_0$ belong to $\mathcal{F}_{t-1}$, whereas the existence of the expectation is guaranteed by the moment condition in **B3**. However, this property does not hold for the derivatives w.r.t. φ. We have

$$\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \varphi^2} = \left(1 - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \varphi^2} + \left(2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^4}\left(\frac{\partial \sigma_t^2}{\partial \varphi}\right)^2 + \frac{2}{\sigma_t^2}\left(\frac{\partial \epsilon_t}{\partial \varphi}\right)^2 - \frac{4\epsilon_t}{\sigma_t^4}\frac{\partial \epsilon_t}{\partial \varphi}\frac{\partial \sigma_t^2}{\partial \varphi}, \tag{B.9}$$

$$\frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \varphi\partial \boldsymbol{\theta}} = \left(1 - \frac{\epsilon_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \varphi\partial \boldsymbol{\theta}} + \left(2\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \varphi}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} - \frac{2\epsilon_t}{\sigma_t^2}\frac{\partial \epsilon_t}{\partial \varphi}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}, \tag{B.10}$$

where the first derivatives of σ_t^2 are given by

$$\frac{\partial \sigma_t^2(\boldsymbol{\vartheta})}{\partial \varphi} = -2\sum_{i=1}^q \alpha_i \epsilon_{t-i}(\varphi)X_{t-i+1} + \sum_{j=1}^p \beta_j \frac{\partial \sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial \varphi}, \tag{B.11}$$

$$\frac{\partial \sigma_t^2(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} = \frac{\partial \omega}{\partial \boldsymbol{\theta}} + \sum_{i=1}^q \frac{\partial \alpha_i}{\partial \boldsymbol{\theta}}\epsilon_{t-i}^2(\varphi) + \sum_{j=1}^p \frac{\partial \beta_j}{\partial \boldsymbol{\theta}}\sigma_{t-j}^2(\boldsymbol{\vartheta}) + \sum_{j=1}^p \beta_j \frac{\partial \sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}}. \tag{B.12}$$

and the second-order derivatives are

$$\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta})}{\partial\varphi^2} = 2\sum_{i=1}^{q}\alpha_i X_{t-i+1}^2 + \sum_{j=1}^{p}\beta_j\frac{\partial^2\sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial\varphi^2}, \tag{B.13}$$

$$\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi} = -2\sum_{i=1}^{q}\frac{\partial\alpha_i}{\partial\boldsymbol{\theta}}X_{t-i+1}\epsilon_{t-i}(\varphi) + \sum_{j=1}^{p}\frac{\partial\beta_j}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial\varphi} + \sum_{j=1}^{p}\beta_j\frac{\partial^2\sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi}. \tag{B.14}$$

It follows from (B.10) that

$$E\left(\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\varphi}\right) = E\left(\left\{1-\eta_t^2\right\}\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\boldsymbol{\theta}\partial\varphi}(\boldsymbol{\vartheta}_0) + \left\{2\eta_t^2-1\right\}\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2}{\partial\varphi}(\boldsymbol{\vartheta}_0)\right), \tag{B.15}$$

noting that, by arguments already used,

$$E\left(\frac{\epsilon_t}{\sigma_t^2}\frac{\partial\epsilon_t}{\partial\varphi}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right) = E\left(-\frac{\eta_t}{\sigma_t}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}X_{t+1}\right) = 0.$$

To prove *i)*, it suffices to show

$$E\left\{(1+\eta_t^2)\left(\frac{\partial\epsilon_t(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right)^2\right\} < \infty, \quad E\left\{\frac{(1+\eta_t^2)^2}{\sigma_t^4}\left(\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right)^2\right\} < \infty, \tag{B.16}$$

$$E\left(\frac{1+\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi^2}\right) < \infty, \quad E\left\|\frac{1+\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\varphi}\right\| < \infty. \tag{B.17}$$

The first inequality in (B.16) writes $E((1+\eta_t^2)X_{t+1}^2) < \infty$, which follows from Hölder's inequality. In view of (B.11), there exists an absolutely summable sequence of coefficients $c_i(\boldsymbol{\vartheta})$ such that

$$\frac{\partial\sigma_t^2(\boldsymbol{\vartheta})}{\partial\varphi} = \sum_{i=0}^{\infty}c_i(\boldsymbol{\vartheta})\epsilon_{t-i-1}(\varphi)X_{t-i} = \sum_{i,j=0}^{\infty}\varphi_0^j c_i(\boldsymbol{\vartheta})\epsilon_{t-i-1}(\varphi)\epsilon_{t-i+j}(\varphi_0). \tag{B.18}$$

By Hölder's and Minkowski's inequalities it follows that

$$\begin{aligned}\left\|\frac{1+\eta_t^2}{\sigma_t^2}\left(\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right)\right\|_2 &\leq \left\|\frac{1+\eta_t^2}{\underline{\omega}}\right\|_4\left\|\frac{\partial\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right\|_4 \\ &\leq K\sum_{i,j=0}^{\infty}|\varphi_0|^j|c_i(\boldsymbol{\vartheta})|\|\epsilon_{t-i-1}(\varphi)\|_8\|\epsilon_{t-i+j}(\varphi_0)\|_8 < \infty,\end{aligned}$$

since $E|\epsilon_t|^8 < \infty$ entails $E|\eta_t|^8 < \infty$. The second inequality in (B.16) is thus established.

Now, by (B.13) we have that $\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta})}{\partial\varphi^2} \in L^2$. The first moment condition in (B.17) follows by Hölder's inequality. By (B.14) we have $\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi} = \sum_{\ell=0}^{\infty}d_\ell(\boldsymbol{\vartheta})Y_{t-\ell}$ where $Y_t = -2\sum_{i=1}^{q}\frac{\partial\alpha_i}{\partial\boldsymbol{\theta}}X_{t-i+1}\epsilon_{t-i} + \sum_{j=1}^{p}\frac{\partial\beta_j}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_{t-j}^2(\boldsymbol{\vartheta})}{\partial\varphi}$ for an absolutely summable sequence of coefficients $d_\ell(\boldsymbol{\vartheta})$. By already given arguments we have $Y_t \in L^2$, whence $\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\varphi} \in L^2$. The second moment condition in (B.17) follows. This completes the proof of *i)*.

Property *ii)* only depends on the causal GARCH model and thus is already known. Properties *iii)-v)* follow by arguments already used.

We thus have shown, in view of (A.2), (B.4) and (B.7), that

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) &= -\boldsymbol{J}_n^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t(\boldsymbol{\vartheta}_0) + \boldsymbol{K}_n\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\tau_t}\epsilon_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\tau_t}X_{t+1}^2}\right)\\ &= -\boldsymbol{J}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(1-\eta_t^2)\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right\}(\boldsymbol{\theta}_0) + \boldsymbol{K}\frac{1}{\mu_\tau}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\tau_t}\epsilon_t X_{t+1}\right) + o_P(1)\\ &= -\boldsymbol{J}^{-1}\boldsymbol{L}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t + o_P(1).\end{aligned}$$

where $\boldsymbol{L} = [\mu_\tau^{-1}\boldsymbol{K}\ \ \boldsymbol{I}_d]_{d\times(d+1)}$ and $\boldsymbol{W}_t = \begin{pmatrix} \frac{\sigma_t}{\tau_t}\eta_t X_{t+1} \\ \{1-\eta_t^2\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}\right\}(\boldsymbol{\vartheta}_0)\end{pmatrix}$. By the arguments used for the WLS estimator it can be shown that $\{(\boldsymbol{W}_t, \mathcal{F}_t)\}$ is a mixingale. We thus have shown that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{J}^{-1}\boldsymbol{L}\boldsymbol{\Gamma}\boldsymbol{L}'\boldsymbol{J}^{-1}\right),$$

where $\boldsymbol{\Gamma} = \sum_{h=-\infty}^{\infty}\boldsymbol{\Gamma}(h)$, and for $h \geq 0$, $\boldsymbol{\Gamma}(h) = \mathrm{Cov}(\boldsymbol{W}_t, \boldsymbol{W}_{t+h}) = \boldsymbol{\Gamma}(-h)'$. Now we note that the bottom-right block of the matrix $\boldsymbol{\Gamma}(h)$, for $h > 0$, is equal to $\boldsymbol{0}_{d\times d}$ because the second component of $(\boldsymbol{W}_t)$ is a martingale difference sequence. We also have that the off-diagonal blocks of $\boldsymbol{\Gamma}(h)$ are equal to the null vector, for any $h \geq 0$. In particular, the covariances between the first and the other components of $\boldsymbol{W}_t$ are equal to zero by already used arguments. Therefore,

$$\boldsymbol{\Gamma} = \begin{pmatrix} \sum_{h=-\infty}^{\infty}\gamma(h) & \boldsymbol{0}_{1\times d} \\ \boldsymbol{0}_{d\times 1} & (\kappa-1)\boldsymbol{J}\end{pmatrix}$$

where $\gamma(h)$ is defined in Theorem 3.1. The conclusion follows. □

C Proofs of the results of Section 5

Proof of Proposition 5.1. In view of (B.7), we have by arguments already used

$$\begin{aligned}E\left(\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}}\right) &= E\left\{(1-\eta_t^2)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_0)\right\} = 0,\\ E\left(\frac{\partial\ell_t(\boldsymbol{\vartheta}_0)}{\partial\varphi}\right) &= E\left\{\frac{-2\eta_t}{\sigma_t}X_{t+1}\right\} + E\left\{(1-\eta_t^2)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\varphi}(\boldsymbol{\vartheta}_0)\right\}.\end{aligned} \tag{C.1}$$

Now,

$$E\left(\frac{\eta_t X_{t+1}}{\sigma_t}\right) = E\left(\frac{\eta_t \sum_{i=0}^{\infty} \varphi_0^i \epsilon_{t+1+i}}{\sigma_t}\right) = 0 \tag{C.2}$$

using the independence between η_{t+1+i} and (η_t, σ_t) for any $i \geq 0$. It follows that the first expectation in the r.h.s. of (C.1) is equal to the zero vector. Turning to the second term we have, in view of (B.18),

$$\begin{aligned} E\left\{\left(1-\eta_t^2\right) \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial \varphi}\right\} &= E\left((1-\eta_t^2)\frac{1}{\sigma_t^2} \sum_{i,j\geq 0} c_i(\boldsymbol{\vartheta}_0)\varphi_0^j \epsilon_{t-1-i}\epsilon_{t-i+j}\right) \\ &= E\left((1-\eta_t^2)\eta_t \frac{1}{\sigma_t} \sum_{i\geq 0} c_i(\boldsymbol{\vartheta}_0)\varphi_0^i \epsilon_{t-1-i}\right) \\ &= -E\eta_t^3 E\left(\frac{1}{\sigma_t} \sum_{i\geq 0} c_i(\boldsymbol{\vartheta}_0)\varphi_0^i \epsilon_{t-1-i}\right), \end{aligned} \tag{C.3}$$

where the second equality follows from the fact that in the double sum all terms such that $i \neq j$ vanish: if $j > i$ this follows from $E\{(1-\eta_t^2)\epsilon_{t-1-i}\epsilon_{t-i+j}\} = E\{(1-\eta_t^2)\epsilon_{t-1-i}\sigma_{t-i+j}\}E(\eta_{t-i+j}) = 0$; if $j < i$, because $\epsilon_{t-1-i}\epsilon_{t-i+j} \in \mathcal{F}_{t-1}^{\eta}$. We thus have shown that (5.2) holds.

By (B.10) we have

$$\begin{aligned} &E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\varphi\partial\boldsymbol{\theta}}\right) \\ &= E\left(\frac{1-\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\varphi\partial\boldsymbol{\theta}}(\boldsymbol{\vartheta}_0)\right) + E\left(\frac{2\eta_t^2-1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\varphi}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}(\boldsymbol{\vartheta}_0)\right) + E\left(\frac{2\eta_t X_{t+1}}{\sigma_t^3}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}(\boldsymbol{\vartheta}_0)\right). \end{aligned} \tag{C.4}$$

For any component θ_k of $\boldsymbol{\theta}$, it can be seen from (B.14) that the second-order partial derivative of σ_t^2 w.r.t. θ_k and φ has the form

$$\frac{\partial^2\sigma_t^2(\boldsymbol{\vartheta}_0)}{\partial\varphi\partial\theta_k} = \sum_{i=1}^{\infty} c_{k,i}X_{t-i+1}\epsilon_{t-i} = \sum_{i=1}^{\infty} c_{k,i}\epsilon_{t-i}\sum_{j=0}^{\infty}\varphi_0^j\epsilon_{t-i+1+j}.$$

Thus

$$\begin{aligned} E\left(\frac{1-\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\varphi\partial\theta_k}(\boldsymbol{\vartheta}_0)\right) &= E\left(\frac{1-\eta_t^2}{\sigma_t^2}\sum_{i=1}^{\infty} c_{k,i}\epsilon_{t-i}\sum_{j=0}^{i-2}\varphi_0^j\epsilon_{t-i+1+j}\right) \\ &\quad + E\left(\frac{1-\eta_t^2}{\sigma_t^2}\sum_{i=1}^{\infty} c_{k,i}\epsilon_{t-i}\sum_{j=i-1}^{\infty}\varphi_0^j\epsilon_{t-i+1+j}\right), \end{aligned}$$

where the first expectation vanishes from the independence between η_t and $\mathcal{F}^{\epsilon}_{t-1}$. We also have

$$E\left(\frac{1-\eta_t^2}{\sigma_t^2}\epsilon_{t-i}\epsilon_{t-i+1+j}\right) = E\left(\frac{1-\eta_t^2}{\sigma_t^2}\epsilon_{t-i}\sigma_{t-i+1+j}\eta_{t-i+1+j}\right) = 0$$

for $j > i-1$, by independence between $\eta_{t-i+1+j}$ and variables belonging to $\mathcal{F}_{\epsilon,t-i+j}$. Finally, for $j = i-1$, $E\left(\frac{1-\eta_t^2}{\sigma_t^2}\epsilon_{t-i}\epsilon_t\right) = 0$ using the symmetry assumption on the law of η_t. We thus have shown that the first expectation in the r.h.s. of (C.4) is equal to the null vector. The second and third expectations can be shown to be zero in a similar fashion. We thus have shown that

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\varphi\partial\boldsymbol{\theta}}\right) = \mathbf{0}.$$

Note that $E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$ is a positive-definite matrix, by results already established for the standard GARCH model (see for instance Francq and Zakoïan (2019), Theorem 7.2).

Turning to the top left block of matrix $\boldsymbol{J}$, by (B.9) we have

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\varphi^2}\right) = E\left(\frac{1-\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\varphi^2}\right) + E\left(\frac{2\eta_t^2-1}{\sigma_t^4}\left(\frac{\partial\sigma_t^2}{\partial\varphi}\right)^2\right) + E\left(\frac{2X_{t+1}^2}{\sigma_t^2}\right) + E\left(\frac{4\eta_t X_{t+1}}{\sigma_t^3}\frac{\partial\sigma_t^2}{\partial\varphi}\right).$$

By the arguments used to prove (C.2), the last expectation cancels and we have

$$E\left(\frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial\varphi^2}\right) = E\left(\frac{1-\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\varphi^2}\right) + E\left(\frac{2\eta_t^2-1}{\sigma_t^4}\left(\frac{\partial\sigma_t^2}{\partial\varphi}\right)^2\right) + E\left(\frac{2X_{t+1}^2}{\sigma_t^2}\right).$$

Note that in the standard case where the AR model for X_t is causal, the first expectation vanishes because $\frac{\partial^2\sigma_t^2}{\partial\varphi^2} \in \mathcal{F}^{\epsilon}_{t-1}$. In the noncausal case, this expectation is non-zero.

In the ARCH(1) case, $\sigma_t^2(\boldsymbol{\theta}_0) = \omega_0 + \alpha_0\epsilon_{t-1}^2$, tedious computations shows that

$$\begin{aligned}
E\left(\frac{1-\eta_t^2}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\varphi^2}\right) &= \frac{2\alpha_0(1-\kappa)}{1-\varphi_0^2\alpha_0} \leq 0,\\
E\left(\frac{2\eta_t^2-1}{\sigma_t^4}\left(\frac{\partial\sigma_t^2}{\partial\varphi}\right)^2\right) &= \frac{4\alpha_0^2(2\kappa-1)}{1-\varphi_0^2\alpha_0}E\left(\frac{\epsilon_{t-1}^2}{\sigma_t^2}\right) + E\left(\frac{\epsilon_{t-1}^2}{\sigma_t^4}\right)\frac{4\alpha_0^2\omega\varphi_0^2}{(1-\varphi_0^2\alpha_0)(1-\varphi_0^2)} \geq 0,\\
E\left(\frac{2X_{t+1}^2}{\sigma_t^2}\right) &= \frac{2}{1-\varphi_0^2\alpha_0}\left(\alpha_0 + E\left(\frac{\omega}{\sigma_t^2}\right)\frac{1}{1-\varphi_0^2}\right) \geq 0.
\end{aligned}$$

By Cauchy-Schwarz inequality, we have $E\left(\frac{\omega}{\sigma_t^2}\right) \geq 1-\alpha_0$. It follows that

$$\frac{2\alpha_0(1-\kappa)}{1-\varphi_0^2\alpha_0} + \frac{2}{1-\varphi_0^2\alpha_0}\left(\alpha_0 + E\left(\frac{\omega}{\sigma_t^2}\right)\frac{1}{1-\varphi_0^2}\right) \geq \frac{2+2\alpha_0(1-\kappa+(\kappa-2)\varphi_0^2)}{(1-\varphi_0^2\alpha_0)(1-\varphi_0^2)} \geq 0$$

under the condition. Hence $\boldsymbol{J}$ is positive definite.

D Detrending method and complements on the empirical section

Let p_t denote the observed price, $p_t > 0$. We set the local-level model

$$p_t = \mu_t + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, H), \tag{D.1}$$

$$\mu_t = \mu_{t-1} + \upsilon_t, \qquad \upsilon_t \sim \mathcal{N}(0, Q), \tag{D.2}$$

with ξ_t and υ_t mutually independent. The parameter $r = Q/H$ governs the smoothness of the trend μ_t; a smaller r implies a slower trend and hence a larger fraction of bubble-like excursions remains in the detrended price $z_t = p_t - \mu_t$. The filtered level obeys the exponential-smoothing recursion

$$\mu_t = \mu_{t-1} + K_\infty(r)\,(p_t - \mu_{t-1}), \tag{D.3}$$

with

$$K_\infty(r) = \frac{\bar{P}(r) + r}{\bar{P}(r) + r + 1}, \qquad \bar{P}(r) = \frac{-r + \sqrt{r^2 + 4r}}{2}. \tag{D.4}$$

For $\{p_s : s \leq t\}$ is used to compute μ_t, so the procedure is strictly one-sided. We choose the smallest r in a grid $\{r_j\}$ that yields stationarity of z_t according to an augmented Dickey–Fuller (ADF) test with drift and selected lag order using the BIC. Denote the ADF regression

$$\Delta z_t = d + \rho z_{t-1} + \sum_{i=1}^{k} \varphi_i \Delta z_{t-i} + u_t, \tag{D.5}$$

For each r_j, we compute the ADF statistic $\tau(r_j)$ and critical value c_α; we select

$$r^\star = \min\{r_j : \tau(r_j) < c_\alpha\}. \tag{D.6}$$

Figure 6 displays the boom and bust pattern of the cryptocurrency prices. Table 6 reports the descriptive statistics for the detrended daily prices. Both series are right-skewed (skewness 1.879 for XRP; 1.114 for SOL) and leptokurtic (kurtosis 9.608 and 6.633), with heavier tails for XRP. Relative volatility, measured by the coefficient of variation sd/mean, is very high and similar across assets, 5.57 for XRP and 5.76 for SOL, indicating only a slightly larger dispersion for SOL. In both cases the mean exceeds the median ($0.042 > 0.009$ for XRP; $4.234 > 0.110$ for SOL), consistent with large positive deviations, mostly during the bubble episodes. The range is wider for SOL (-71.849 to 125.879) than for XRP (-0.415 to 1.447).

Figure 9: XRP-USD and SOL-USD daily prices from 2020-04-10 to 2024-07-01.

Table 6: Descriptive statistics detrended prices

	Mean	Median	Std. Dev.	Skewness	Kurtosis	Min	Max
XRP	0.042	0.009	0.234	1.879	9.608	−0.415	1.447
SOL	4.234	0.110	24.374	1.114	6.633	−71.849	125.879

References

[1] Aknouche, A. (2013). Two-stage weighted least squares estimation of nonstationary random coefficient autoregressions. *Journal of Time Series Econometrics* 5, 25–46.

[2] Aknouche, A. and Francq, C. (2023) Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *Journal of Econometrics* 237, Article 105174.

[3] Berkes, I., Horváth, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201-227.

[4] Blanchard, O.J. and M.W. Watson (1982) Bubbles, rational expectations and financial markets. NBER working paper, w0945.

[5] Blasques, F., Koopman, S. J. and Mingoli, G. (2023) Observation-driven filters for time-series with stochastic trends and mixed causal non-causal dynamics. Tinbergen Institute Discussion Paper.

[6] Blasques, F., Koopman, S. J. and Nientker, M. (2022) A time-varying parameter model for local explosions. *Journal of Econometrics* 227, 65–84.

[7] Bougerol, P., and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of econometrics* 52, 115–127.

[8] Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association* 65, 1509–1526.

[9] Brunnermeier, M. K. and Oehmke, M. (2013) Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance* 2, 1221–1288.

[10] Catania, L. and G. Mingoli (2026) Autoregressive models with non-causal ARCH volatility. Unpublished document.

[11] Cavaliere, G., Nielsen, H. B. and Rahbek, A. (2018) Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. *Journal of Business & Economic Statistics* 38, 55–67.

[12] Davis, R.A. and Song, L. (2020) Noncausal vector AR processes with application to economic time series. *Journal of Econometrics* 216, 246–267.

[13] de Truchis, G., Fries, S., and Thomas, A. (2025). Forecasting extreme trajectories using semi-norm representations. Unpublished document. hal-05007564

[14] Francq, C., Horváth, L. and Zakoïan, J.-M. (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.

[15] Francq, C., Roy, R. and Saidi, A. (2011). Asymptotic properties of weighted least squares estimation in weak PARMA models. *Journal of Time Series Analysis* 32, 699–723.

[16] Francq, C., Roy, R. and Zakoïan, J. M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.

[17] Francq, C., Wintenberger, O. and Zakoïan, J.-M. (2018). Goodness-of-fit tests for Log-GARCH and EGARCH models. *Test* 27, 27–51.

[18] Francq, C. and Zakoïan, J.-M. (2019) *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons, Second Edition.

[19] Fries, S. (2021) Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. *Journal of Business & Economic Statistics* 40, 1596-1616.

[20] Fries, S. and Zakoïan, J.-M. (2019) Mixed causal-noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory* 35, 1234–1270.

[21] Gouriéroux, C. and Jasiak, J. (2016) Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis* 37, 405–430.

[22] Gouriéroux, C. and Jasiak, J. (2017) Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation. *Journal of Econometrics* 200, 118–134.

[23] Gouriéroux, C. and Zakoïan, J.-M. (2017) Local explosion modelling by noncausal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79, 737–756.

[24] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York-London.

[25] Hecq, A., Issler, J. V. and Telg, S. (2020) Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics* 35, 328–343.

[26] Hecq, A. and Velásquez-Gaviria, D. (2025), Spectral estimation for mixed causal-noncausal autoregressive models. *Econometric Reviews* 44, 939–962.

[27] Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics* 6, 629–642.

[28] Lanne, M. and Saikkonen, P. (2013) Noncausal vector autoregression. *Econometric Theory* 29, 447–481.

[29] Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.

[30] Lof, M. and Nyberg, H. (2017) Noncausality and the commodity currency hypothesis. *Energy Economics* 65, 424–433.

[31] Mainassara, Y. B. (2011) Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *Journal of Statistical Planning and Inference* 141, 2961–2975.

[32] McLeish, D. L. (1975). Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32, 165–178.

[33] Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91, 590–600.

[34] Royer, J. (2023). Conditional asymmetry in Power ARCH (∞) models. *Journal of Econometrics* 234, 178–204.

[35] Velasco, C. and Lobato, I.N. (2018) Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models. *The Annals of Statistics* 46, 555–579.

[36] White, H. (2014). *Asymptotic theory for econometricians.* Academic press.

[37] Zhan, Y., Ling, S. Liu, Z. and Wang, S. (2025) Noncausal AR-ARCH Model and Its Applications to Financial Time Series. *International Journal of Finance & Economics*, forthcoming.

[38] Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78, 463–485.

[39] Zhu, K. and Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *The Annals of Statistics* 46, 2131-2163.

CRESTCENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

The Center for Research in Economics and Statistics (CREST) is a leading French scientific institution for advanced research on quantitative methods applied to the social sciences.

CREST is a joint interdisciplinary unit of research and faculty members of CNRS, ENSAE Paris, ENSAI and the Economics Department of Ecole Polytechnique. Its activities are located physically in the ENSAE Paris building on the Palaiseau campus of Institut Polytechnique de Paris and secondarily on the Ker-Lann campus of ENSAI Rennes.