Electronic Journal of Statistics
Vol. 16 (2022) 116–157
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1950

Inference and model selection in general causal time series with exogenous covariates*

Mamadou Lamine Diop and William Kengne

THEMA, CY Cergy Paris Université,
33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France
e-mail: mamadou-lamine.diop@u-cergy.fr; william.kengne@u-cergy.fr

Abstract: In this paper, we study a general class of causal processes with exogenous covariates, including many classical processes such as the ARMA-GARCH, APARCH, ARMAX, GARCH-X and APARCH-X processes. Under some Lipschitz-type conditions, the existence of a τ-weakly dependent strictly stationary and ergodic solution is established. We provide conditions for the strong consistency and derive the asymptotic distribution of the quasi-maximum likelihood estimator (QMLE), both when the true parameter is an interior point of the parameters space and when it belongs to the boundary. A significance Wald-type test of parameter is developed. This test is quite extensive and includes the test of nullity of the parameter's components, which in particular, allows us to assess the relevance of the exogenous covariates. Relying on the QMLE of the model, we also propose a penalized criterion to address the problem of the model selection for this class. The weak and the strong consistency of the procedure are established. Finally, Monte Carlo simulations are conducted to numerically illustrate the main results.

MSC2020 subject classifications: Primary 60G10, 62M10, 62F05, 62F12; secondary 60G10, 62M10, 62F05, 62F12

Keywords and phrases: Causal processes, exogenous covariates, quasi-maximum likelihood estimator, consistency, boundary, significance test, model selection, penalized criterion.

Received July 2021.

Contents

*M. L. Diop was supported by the MME-DII center of excellence (ANR-11-LABEX-0023-01). W. Kengne was supported by the ANR BREAKRISK: ANR-17-CE26-0001-01 and the CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013.

1. Introduction

Autoregressive time series with exogenous covariates provide effective ways to take into account some available extra information in the models. The well known example that has been widely studied is the ARMAX model, see Hannan (1976) [13], Hannan and Deistler (2012) [14]. The GARCH-type models with exogenous covariates have recently attracted much attention in the literature, see for instance Han and Kristensen (2014) [11] for GARCH-X, Francq and Thieu (2019) [8] for APARCH-X. Guo *et al.* (2014) [10] considered the factor double autoregressive model, whose ARX and ARCH-X are particular cases. We consider a large class of causal time series models, whose ARMAX and GARCH-X type models are specific examples.

Let $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d_x,t}) \in \mathbb{R}^{d_x}$ be a vector of covariates, with $d_x \in \mathbb{N}$. Consider the class of affine causal models with exogenous covariates,

Class $\mathcal{AC}$-$X(M_\theta, f_\theta)$: A process $\{Y_t, t \in \mathbb{Z}\}$ belongs to $\mathcal{AC}$-$X(M_\theta, f_\theta)$ if it satisfies:

$$Y_t = M_\theta(Y_{t-1}, \ldots; X_{t-1}, \ldots)\xi_t + f_\theta(Y_{t-1}, \ldots; X_{t-1}, \ldots), \tag{1.1}$$

where M_θ, $f_\theta : \mathbb{R}^{\mathbb{N}} \times (\mathbb{R}^{d_x})^{\mathbb{N}} \to \mathbb{R}$ are two measurable functions and assumed to be known up to the parameter θ, which belongs in a compact subset $\Theta \subset \mathbb{R}^d$ ($d \in \mathbb{N}$); and $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of zero-mean independent, identically distributed (*i.i.d*) random variable satisfying $\mathbb{E}(\xi_0^r) < \infty$ for some $r \geq 2$ and $\mathbb{E}(\xi_0^2) = 1$. Remark that, if $X_t \equiv C$ for some constant C (absence of covariates), then (1.1) reduces to the classical affine causal models that has already been

considered in the literature (see, for instance, Bardet and Wintenberger (2009) [2], Bardet *et al.* (2012) [1], Bardet *et al.* (2020) [3]). One can see that, the ARMAX, GARCH-X, APARCH-X models belong to the class $\mathcal{AC}$-$X(M_\theta, f_\theta)$.

There exist several important contributions devoted to autoregressive models with covariates; we refer to Hannan and Deistler (2012) [14], Han and Kristensen (2014) [11], Sucarrat *et al.* (2016) [29], Francq and Sucarrat (2017) [7], Pedersen and Rahbek (2018) [25], Francq and Thieu (2019) [8], Grønneberg and Holcblat (2019) [9], Zambom and Gel (2020) [31] and the references therein for some developments on ARMAX and conditional volatility type models with exogenous covariates. The class $\mathcal{AC}$-$X(M_\theta, f_\theta)$ is more general than the models considered in the aforementioned works, as well as the factor double autoregressive model proposed by Guo *et al.* (2014) [10] which is a particular case of the model (1.1). Note as well that, the class $\mathcal{AC}$-$X(M_\theta, f_\theta)$ provides a more general way to take into account covariates in the model, and one can see that the linear covariates regressors considered by Francq and Thieu (2019) [8] and many other works is a specific case. Compared to Bardet and Wintenberger (2009) [2], besides taking into account covariates in the model (1.1), we address the inference when the true parameter belongs to the boundary of the parameter set Θ and the model selection question.

In this new contribution, we consider the class of model (1.1) and address the following issues.

(i) **Existence of a stationary solution**. We provide sufficient conditions that ensure the existence of a τ-weakly dependent stationary and ergodic solution $Z_t = (Y_t, X_t)$ of (1.1). At a first glance, one might think that these conditions are the same as those obtained by Bardet and Wintenberger (2009) [2], but in our case, the existence of the covariates must be taken into account.

(ii) **Inference for the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$**. An inference based on the quasi likelihood of the model is carried out. The consistency of the quasi-maximum likelihood estimator (QMLE) is established and we derived the asymptotic distribution of this estimator (even when θ^* belongs to the boundary of Θ).

(iii) **Significance test of parameter**. A Wald-type significance test of parameter of the model (1.1) is conducted. The proposed test is quite extensive and includes the test of nullity of the parameter's components. An asymptotic study is carried out, which shows in particular that, when the true parameter belongs to the boundary of Θ, the asymptotic distribution of the test statistic under the null hypothesis is quite different from the classical chi-square distribution.

(iv) **Model selection**. A penalized criterion based on the quasi likelihood of the model is proposed for model selection in the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$. We provides conditions that ensure the weak and the strong consistency of the proposed procedure. These conditions shows in particular that, the

Hannan-Quinn information Criterion (HQC) with a regularization parameter $\kappa_n = c \log\log n$ (see (3.2)) is strongly consistent for sufficiently large c.

The article is organized as follows. In Section 2, firstly, we provide conditions for stability properties. Secondly, we give the definition of the QMLE and study its asymptotic properties; a significance test of parameter with an asymptotic study is also addressed. Section 3 focuses on the model selection and the consistency of the proposed procedure. Some classical examples of processes belonging to the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ are detailed in Section 4. Section 5 gives some empirical results, whereas Section 6 is devoted to a summary and conclusion. Section 7 contains the proofs of the main results.

2. Assumptions, inference and test of the parameters

2.1. Assumptions

Throughout the sequel, the following norms will be used:

- $\|x\| \coloneqq \sqrt{\sum_{i=1}^{p} x_i^2}$, *for any* $x \in \mathbb{R}^p$, $p \in \mathbb{N}$;
- $\|V\| \coloneqq \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{q} v_{i,j}^2}$, *for any matrix* $V \in M_{p,q}(\mathbb{R})$, *where* $M_{p,q}(\mathbb{R})$ *denotes the set of matrices of dimension* $p \times q$ *with coefficients in* $\mathbb{R}$, *for* $p, q \in \mathbb{N}$;
- $\|g\|_{\mathcal{K}} \coloneqq \sup_{\theta\in\mathcal{K}} (\|g(\theta)\|)$ *for any compact set* $\mathcal{K} \subseteq \mathbb{R}^d$ *and function* $g : \mathcal{K} \longrightarrow M_{p,q}(\mathbb{R})$;
- $\|Y\|_r \coloneqq \mathbb{E}\left(\|Y\|^r\right)^{1/r}$, *if* Y *is a random vector with finite* r*-order moments, for* $r > 0$.

We will denote by 0 the null vector of any vector space. Let Ψ_θ be the generic symbol for any of the functions f_θ or M_θ. We set the following classical Lipschitz-type conditions for any compact set $\mathcal{K} \subseteq \Theta$.

Assumption $\mathbf{A}_i(\Psi_\theta, \mathcal{K})$ $(i = 0, 1, 2)$: For any $(y, x) \in \mathbb{R}^{\mathbb{N}} \times (\mathbb{R}^{d_x})^{\mathbb{N}}$, the function $\theta \mapsto \Psi_\theta(y)$ is i times continuously differentiable on $\mathcal{K}$ with $\left\|\frac{\partial^i \Psi_\theta(0)}{\partial\theta^i}\right\|_{\mathcal{K}} < \infty$; and there exists two sequences of non-negative real numbers $(\alpha_{k,Y}^{(i)}(\Psi_\theta, \mathcal{K}))_{k\geq 1}$ and $(\alpha_{k,X}^{(i)}(\Psi_\theta, \mathcal{K}))_{k\geq 1}$ satisfying: $\sum_{k=1}^{\infty} \alpha_{k,Y}^{(i)}(\Psi_\theta, \mathcal{K}) < \infty$, $\sum_{k=1}^{\infty} \alpha_{k,X}^{(i)}(\Psi_\theta, \mathcal{K}) < \infty$ for $i = 0, 1, 2$; such that for any $(y, x), (y', x') \in \mathbb{R}^{\mathbb{N}} \times (\mathbb{R}^{d_x})^{\mathbb{N}}$,

$$\left\|\frac{\partial^i \Psi_\theta(y, x)}{\partial\theta^i} - \frac{\partial^i \Psi_\theta(y', x')}{\partial\theta^i}\right\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_{k,Y}^{(i)}(\Psi_\theta, \mathcal{K})|y_k - y'_k| + \sum_{k=1}^{\infty} \alpha_{k,X}^{(i)}(\Psi_\theta, \mathcal{K})\|x_k - x'_k\|,$$

where $\|\cdot\|$ denotes any vector, matrix norm.

The following assumption is considered on the function $H_\theta = M_\theta^2$ in the cases of ARCH-X type process.

Assumption $\mathbf{A}_i(H_\theta, \mathcal{K})$ $(i = 0, 1, 2)$: Assume that $f_\theta = 0$. There exists two sequences of non-negative real numbers $(\alpha_{k,Y}^{(i)}(H_\theta, \mathcal{K}))_{k\geq 1}$ and $(\alpha_{k,X}^{(i)}(H_\theta, \mathcal{K}))_{k\geq 1}$ satisfying: $\sum_{k=1}^{\infty} \alpha_{k,Y}^{(i)}(H_\theta, \mathcal{K}) < \infty$, $\sum_{k=1}^{\infty} \alpha_{k,X}^{(i)}(H_\theta, \mathcal{K}) < \infty$ for $i = 0, 1, 2$; such that for any $(y, x), (y', x') \in \mathbb{R}^\infty \times (\mathbb{R}^{d_x})^\infty$,

$$\left\| \frac{\partial^i H_\theta(y, x)}{\partial \theta^i} - \frac{\partial^i H_\theta(y', x')}{\partial \theta^i} \right\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_{k,Y}^{(i)}(H_\theta, \mathcal{K}) |y_k^2 - {y'}_k^2| + \sum_{k=1}^{\infty} \alpha_{k,X}^{(i)}(H_\theta, \mathcal{K}) \|x_k - x'_k\|.$$

In the whole paper, we impose an autoregressive-type structure on the covariates:

$$X_t = g(X_{t-1}, X_{t-2}, \ldots; \eta_t), \tag{2.1}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of zero-mean random variables such as $(\eta_t, \xi_t)_{t\in\mathbb{Z}}$ is *i.i.d* and g is a function with values in $\mathbb{R}^{d_x}$ satisfying for all $x, x' \in (\mathbb{R}^{d_x})^{\mathbb{N}}$,

$$\mathbb{E}\left[\|g(0, \eta_0)\|^r\right] < \infty \quad \text{and} \quad \|g(x; \eta_0) - g(x'; \eta_0)\|_r \leq \sum_{k=1}^{\infty} \alpha_k(g) \|x_k - x'_k\| \tag{2.2}$$

for some $r \geq 1$ and non-negative sequence $(\alpha_k(g))_{k\geq 1}$ such that $\sum_{k=1}^{\infty} \alpha_k(g) < 1$.

For $r \geq 1$, when (2.2) holds, we define the set

$$\begin{aligned}\Theta(r) = \Big\{ \theta \in \mathbb{R}^d \,/\, & \mathbf{A}_0(f_\theta, \{\theta\}) \text{ and } \mathbf{A}_0(M_\theta, \{\theta\}) \text{ hold with} \\ & \sum_{k=1}^{\infty} \max\left\{ \alpha_k(g),\, \alpha_{k,Y}^{(0)}(f_\theta, \{\theta\}) + \|\xi_0\|_r \alpha_{k,Y}^{(0)}(M_\theta, \{\theta\} \right\} < 1 \Big\} \\ \bigcup \Big\{ \theta \in \mathbb{R}^d \,/\, & f_\theta = 0 \text{ and } \mathbf{A}_0(H_\theta, \{\theta\}) \text{ holds with} \\ & \|\xi_0\|_r^2 \sum_{k=1}^{\infty} \max\left\{ \alpha_k(g),\, \alpha_{k,Y}^{(0)}(H_\theta, \{\theta\}) \right\} < 1 \Big\}.\end{aligned}$$

In the sequel, we make the convention that if $\mathbf{A}_i(M_\theta, \Theta)$ holds then $\alpha_{k,Y}^{(i)}(H_\theta, \Theta) = \alpha_{k,X}^{(i)}(H_\theta, \Theta) = 0$ for all $k \in \mathbb{N}$ and if $\mathbf{A}_i(H_\theta, \Theta)$ holds then $\alpha_{k,Y}^{(i)}(M_\theta, \Theta) = \alpha_{k,X}^{(i)}(M_\theta, \Theta) = 0$ for all $k \in \mathbb{N}$.

The condition (2.2) allows to assure the stability of the process X_t. Together the aforementioned assumptions assure the existence of a stationary and weakly dependent solution of order r to the model (1.1), as shown in the following proposition.

Proposition 1. *Assume that* $\boldsymbol{A}_0(f_\theta, \Theta)$, $\boldsymbol{A}_0(M_\theta, \Theta)$ *(or* $\boldsymbol{A}_0(H_\theta, \Theta)$*) and (2.2) hold. If* $\theta^* \in \Theta \cap \Theta(r)$ *with* $r \geq 1$*, then there exists a* τ*-weakly dependent stationary, ergodic and non anticipative solution* $(Z_t)_{t\in\mathbb{Z}}$ $Z_t = (Y_t, X_t)$*, to (1.1), satisfying* $\mathbb{E}[\|Z_0\|^r] < \infty$.

2.2. Inference and significance test of parameter

In this paragraph, we describe the use of the Gaussian quasi-maximum likelihood to obtain an estimator of the parameters of the model (1.1). The main asymptotic properties of this estimator are also established. Assume that the observations $(Y_1, X_1), \ldots, (Y_n, X_n)$ are generated from (1.1) and (2.1) according to the true parameter $\theta^* \in \Theta$ which is unknown. For all $t \in \mathbb{Z}$, denote by $\mathcal{F}_t = \sigma((Y_s, X_s),\, s \leq t)$ the σ-field generated by the whole past at time t. The mean and the variance of $Y_t|\mathcal{F}_{t-1}$ and $f_{\theta^*}(Y_{t-1}, \ldots; X_{t-1}, \ldots)$ and variance $M^2_{\theta^*}(Y_{t-1}, \ldots; X_{t-1}, \ldots)$ respectively. For any $\theta \in \Theta$, the conditional Gaussian quasi-log-likelihood is given by (up to an additional constant)

$$L_n(\theta) := -\frac{1}{2}\sum_{t=1}^{n} q_t(\theta) \quad \text{with} \quad q_t(\theta) = \frac{(Y_t - f^t_\theta)^2}{H^t_\theta} + \log H^t_\theta,$$

where $f^t_\theta := f_\theta(Y_{t-1}, Y_{t-2} \ldots; X_{t-1}, X_{t-2}, \ldots)$, $M^t_\theta := M_\theta(Y_{t-1}, Y_{t-2} \ldots; X_{t-1}, X_{t-2}, \ldots)$ and $H^t_\theta := (M^t_\theta)^2$.
Since $(Y_0, X_0), (Y_{-1}, X_{-1}), \ldots$ are not observed, $L_n(\theta)$ is approximated by

$$\widehat{L}_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n} \widehat{q}_t(\theta) \quad \text{with} \quad \widehat{q}_t(\theta) = \frac{(Y_t - \widehat{f}^t_\theta)^2}{\widehat{H}^t_\theta} + \log \widehat{H}^t_\theta,$$

where $\widehat{f}^t_\theta := f_\theta(Y_{t-1}, \ldots, Y_1, 0; X_{t-1}, \ldots, X_1, 0)$, $\widehat{M}^t_\theta := M_\theta(Y_{t-1}, \ldots, Y_1, 0; X_{t-1},$
$\ldots, X_1, 0)$ and $\widehat{H}^t_\theta := (\widehat{M}^t_\theta)^2$. Thus, the QMLE of θ^* is defined by

$$\widehat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}}\big(\widehat{L}_n(\theta)\big).$$

We set the following regularity conditions to assure the identifiability of the model and to derive the asymptotic behavior of the QMLE.

(**A0**): for all $\theta \in \Theta$ and some $t \in \mathbb{Z}$, $\big(f^t_{\theta^*} = f^t_\theta$ and $H^t_{\theta^*} = H^t_\theta \;\; a.s.\big) \Rightarrow \; \theta = \theta^*$;
(**A1**): $\exists \underline{h} > 0$ such that $\inf_{\theta\in\Theta} H_\theta(y, x) \geq \underline{h}$, for all $(y, x) \in \mathbb{R}^{\mathbb{N}} \times (\mathbb{R}^{d_x})^{\mathbb{N}}$;
(**A2**): for all $\theta \in \Theta$, $c \in \mathbb{R}^d$, $\left(c' \frac{\partial}{\partial\theta} f^0_{\theta^*} = 0 \text{ or } c' \frac{\partial}{\partial\theta} H^0_{\theta^*} = 0 \right) a.s. \implies c = 0$,
where $'$ denotes the transpose.

Assumption (**A0**) is an identifiability condition and it will be discussed in detail for each of the examples of processes studied in the paper. From (**A1**), the quasi likelihood is well defined, whereas (**A2**), which is classical (see for instance [2]) allows to derive the asymptotic distribution of the QMLE. The following theorem addresses the strong consistency of the QMLE.

Theorem 2.1. *Assume that (**A0**), (**A1**), $\boldsymbol{A}_0(f_\theta, \Theta)$, $\boldsymbol{A}_0(M_\theta, \Theta)$ and (2.2) (with $r \geq 2$) hold with*

$$\alpha_{k,Y}^{(0)}(f_\theta, \Theta) + \alpha_{k,X}^{(0)}(f_\theta, \Theta) + \alpha_{k,Y}^{(0)}(M_\theta, \Theta) + \alpha_{k,X}^{(0)}(M_\theta, \Theta) + \alpha_{k,Y}^{(0)}(H_\theta, \Theta) + \alpha_{k,X}^{(0)}(H_\theta, \Theta) = O(k^{-\gamma}), \quad (2.3)$$

for some $\gamma > 3/2$.
If $\theta^* \in \Theta \cap \Theta(r)$ with $r \geq 2$, then

$$\widehat{\theta}_n \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} \theta^*.$$

To derive the asymptotic distribution of the QMLE, it is necessary to take into account the constraints in the parameter space Θ corresponding to the model. For example, in some processes belonging to (1.1), such as the ARCH-X models (see below), the components of θ^* are constrained to be positive or equal to zero. In order to propose a parsimonious representation, it is often required to test whether or not the exogenous covariates are relevant. For example, in an ARCH-X(1) model defined by $Y_t = \xi_t \sigma_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* Y_{t-k}^2 + \gamma^{*\prime} X_{t-1}$, the true parameter vector is $\theta^* = (\alpha_0^*, \alpha_1^*, \gamma^*) \in \Theta \subset]0, \infty[\times[0, \infty[^{dx+1}$. The significance test of the covariate X_t consists to verify the nullity of the parameter γ^*; that is, if the true parameter vector can be of the form $\theta_0 = (\alpha_0, \alpha_1, 0)$ which is not an interior point of Θ. In this situation, it is impossible to apply the asymptotic normality results based on the classical assumption of "interior point" to derive the asymptotic behavior of the test statistic used. To take into account such a scenario in the general class (1.1), we will consider that the component i of θ^* is constrained if the i-th section of Θ is of the form $[\underline{\theta}_i, \overline{\theta}_i]$ with $\underline{\theta}_i < \overline{\theta}_i$. Assume that the d_2 (with $d_2 \in \{0, \ldots, d\}$) last components of θ^* are constrained, and let $d_1 = d - d_2$. Therefore, if $d_2 \geq 1$ and $\theta_i^* \in \{\underline{\theta}_i, \overline{\theta}_i\}$ with $i > d_1$, then θ^* is not an interior point of Θ. For instance, in a scenario where $\theta_i^* = \underline{\theta}_i$, with the QMLE $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$, it holds that $\sqrt{n}(\widehat{\theta}_{i,n} - \theta_i^*) \in [0, \infty)$ which cannot tend to a Gaussian distribution with mean 0. By convention, it is assumed that $\theta^* \in \overset{\circ}{\Theta}$ if $d_2 = 0$. When $d_2 \geq 1$ and the set Θ is assumed to be large enough, then the following relation holds:

$$\bigcup_{n\geq 1} \left\{ \sqrt{n}\,(\theta - \theta^*)\,,\ \theta \in \Theta \right\} = \mathcal{C} \quad \text{with} \quad \mathcal{C} = \prod_{i=1}^{d} \mathcal{C}_i, \quad (2.4)$$

where $\mathcal{C}_i = [0, \infty[$ when $i > d_1$ and $\theta_i^* = \underline{\theta}_i$, $\mathcal{C}_i =]\infty, 0]$ when $i > d_1$ and $\theta_i^* = \overline{\theta}_i$, and $\mathcal{C}_i = \mathbb{R}$ otherwise. The set $\mathcal{C}$ is a convex cone which is equal to $\mathbb{R}^d$ if $\theta^* \in \overset{\circ}{\Theta}$.

Let us define the following matrices

$$F = \mathbb{E}\Big[\frac{\partial^2 q_0(\theta^*)}{\partial\theta\partial\theta'}\Big] \quad \text{and} \quad G = \mathbb{E}\Big[\frac{\partial q_0(\theta^*)}{\partial\theta}\frac{\partial q_0(\theta^*)}{\partial\theta'}\Big]. \tag{2.5}$$

Under the assumptions $\mathbf{A}_i(f_\theta,\Theta)$, $\mathbf{A}_i(M_\theta,\Theta)$ (with $i=0,1,2$), one can show the existence of F and G. In addition, in view to (**A2**), the same arguments as in [2] allow to establish that matrix F is positive definite. Consider then the F-scalar product $\langle x,y\rangle_F = x'Fy$ and the norm $\|x\|_F^2 = x'Fx$ for $x,y\in\mathbb{R}^d$. Let us define the F-orthogonal projection of a vector $Z\in\mathbb{R}^d$ on the cone $\mathcal{C}$ as follows:

$$Z^{\mathcal{C}} = \arg\inf_{C\in\mathcal{C}} \|C-Z\|_F .$$

This definition is equivalent to

$$Z^{\mathcal{C}}\in\mathcal{C} \quad \text{with} \quad \big\langle Z-Z^{\mathcal{C}}, C-Z^{\mathcal{C}}\big\rangle \le 0, \quad \forall C\in\mathcal{C}. \tag{2.6}$$

Note that, when $\theta^*\in\overset{\circ}{\Theta}$, we have $Z^{\mathcal{C}}=Z$. Combining all the regularity conditions and definitions given above, we obtain the following main result.

Theorem 2.2. *Assume that (**A0**)-(**A2**), ($\mathbf{A}_i(f_\theta,\Theta)$), ($\mathbf{A}_i(M_\theta,\Theta)$) (for $i=0,1,2$) and (2.2) (with $r\ge 4$) hold with*

$$\alpha^{(i)}_{k,Y}(f_\theta,\Theta)+\alpha^{(i)}_{k,X}(f_\theta,\Theta)+\alpha^{(i)}_{k_\theta,Y}(M_\theta,\Theta)+\alpha^{(i)}_{k,X}(M_\theta,\Theta) + \alpha^{(i)}_{k_\theta,Y}(H_\theta,\Theta)+\alpha^{(i)}_{k,X}(H_\theta,\Theta) = O(k^{-\gamma}), \tag{2.7}$$

for $i=0,1,2$ and some $\gamma>3/2$.

- *If $\theta^*\in\Theta\cap\Theta(r)$ with $r\ge 4$, then*

  $$\sqrt{n}\left(\widehat{\theta}_n-\theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}} Z^{\mathcal{C}} \quad \text{with} \quad Z\sim\mathcal{N}_d(0,\Sigma), \text{ where } \Sigma := F^{-1}GF^{-1}.$$

- *If $\theta^*\in\overset{\circ}{\Theta}\cap\Theta(r)$ with $r\ge 4$, then*

  $$\widehat{\theta}_n-\theta^* = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$$

The matrix Σ can be consistently estimated by $\widehat{\Sigma}_n = F_n(\widehat{\theta}_n)^{-1}G_n(\widehat{\theta}_n)F_n(\widehat{\theta}_n)^{-1}$, where

$$F_n(\widehat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 q_t(\widehat{\theta}_n)}{\partial\theta\partial\theta'} \quad \text{and} \quad G_n(\widehat{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial q_t(\widehat{\theta}_n)}{\partial\theta}\frac{\partial q_t(\widehat{\theta}_n)}{\partial\theta'}.$$

Now, we are interested to investigate whether or not a given subset of components of θ^* are equal to some fixed vector. To do so, consider the following hypothesis testing:

$$H_0 : \Gamma\theta^* = \vartheta_0 \qquad \text{against} \qquad H_1 : \Gamma\theta^*\neq\vartheta_0, \tag{2.8}$$

where Γ is a $d_0 \times d$ full-rank matrix and ϑ_0 is a vector of dimension d_0. Define the Wald-type test statistic given by

$$W_n = n(\Gamma\widehat{\theta}_n - \vartheta_0)'(\Gamma\widehat{\Sigma}_n\Gamma')^{-1}(\Gamma\widehat{\theta}_n - \vartheta_0). \tag{2.9}$$

Under H_0, the asymptotic behavior of W_n is given by the following theorem.

Theorem 2.3. *Under H_0, assume that the assumptions of Theorem 2.2 hold. Then*

$$W_n \xrightarrow[n\to+\infty]{\mathcal{L}} (\Gamma Z^{\mathcal{C}})'(\Gamma\Sigma\Gamma')^{-1}\Gamma Z^{\mathcal{C}} \quad \text{with} \quad Z \sim \mathcal{N}_d(0,\Sigma).$$

By the above theorem, at a nominal level $\alpha \in (0,1)$, the critical region of the test is $(W_n > q_\alpha)$, where q_α is the $(1-\alpha)$-quantile of the distribution of $(\Gamma Z^{\mathcal{C}})'(\Gamma\Sigma\Gamma')^{-1}\Gamma Z^{\mathcal{C}}$. The critical value q_α can be computed through Monte-Carlo simulations. The following corollary follows immediately when θ^* belongs to the interior of the parameter space.

Corollary 1. *Assume that the conditions of Theorem 2.3 hold. If $\theta^* \in \overset{\circ}{\Theta}$, then W_n converges to a chi-square distribution with d_0 degrees of freedom.*

Under H_1, one can easily see that $W_n \xrightarrow[n\to\infty]{a.s} +\infty$; which shows that the test is consistent in power. In the empirical studies, we will restrict our attention to test the relevance of the exogenous covariates by using the hypothesis (2.8) with $\vartheta_0 = 0$ and an appropriate matrix Γ.

3. Model selection

3.1. Model selection framework

Assume that $(Y_1, \ldots, Y_n)$ is a trajectory of the process $Y = \{Y_t,\ t \in \mathbb{Z}\}$ satisfying $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ (defined as in (1.1)), where the true parameter θ^* is unknown. Let $\mathcal{M}$ be a finite collection of models belonging to $\mathcal{AC}$-$X(M_\theta, f_\theta)$ with $\theta \in \Theta$. Assume that $\mathcal{M}$ contains at least the true model m^* corresponding to the parameter θ^*. Our objective is to develop a procedure that allows to select the "best model" (that we denote by $\widehat{m}_n$) among the collection $\mathcal{M}$ such that it is "close" to m^* for n large enough. To this end, we consider the following definitions and notations in the sequel:

- a model $m \in \mathcal{M}$ is considered as a subset of $\{1, \ldots, d\}$ and denote by $|m|$ the dimension of m (i.e, $|m| = \#(m)$);
- for $m \in \mathcal{M}$, $\Theta_m = \{(\theta_i)_{1\le i\le d} \in \Theta$ with $\theta_i = 0$ if $i \notin m\}$ is a compact set containing $\theta(m)$, where $\theta(m)$ denotes the parameter vector associated to the model m;
- $\mathcal{M}$ is considered as a subset of the power set of $\{1, \ldots, d\}$; that is, $\mathcal{M} \subset \mathcal{P}(\{1, \ldots, d\})$.

For instance, when the observations $Y_1, \ldots, Y_n$ are generated from a ARMAX(p^*, q^*, s^*) model (defined below), the collection $\mathcal{M}$ of the competing models could

be considered as a family of ARMAX(p,q,s) with $(p,q,s) \in \{0,1,\ldots,p_{max}\} \times \{0,1,\ldots,q_{max}\} \times \{0,1,\ldots,s_{max}\}$, where $p_{max}, q_{max}, s_{max}$ are the fixed upper bounds of the orders satisfying $p_{max} \geq p^*$, $q_{max} \geq q^*$, $s_{max} \geq s^*$. The parameter space Θ is a compact subset of $\mathbb{R}^{p_{max}+q_{max}+s_{max}}$, and thus a model m is a subset of $\{1,2,\ldots,p_{max}+q_{max}+s_{max}\}$.

3.2. Model selection criterion and asymptotic results

Note that, under the identifiability assumption (**A0**), one can show that, for all $m \in \mathcal{M}$, the function $\theta \mapsto -\mathbb{E}[q_0(\theta)]$ has a unique maximum in Θ_m (see proof of Theorem 2.1). Let us thus define the "best" parameter associated to the model m as

$$\theta^*(m) := \underset{\theta \in \Theta_m}{\operatorname{argmin}} \left(\mathbb{E}[q_0(\theta)]\right).$$

When $m \supseteq m^*$, we have $\theta^*(m) = \theta^*(m^*) = \theta^*$; that is, $\theta^*(m)$ will play the role of the true parameter θ^* in cases of "true" or overfitted model. For $m \in \mathcal{M}$, we define the QMLE of $\theta^*(m)$ as

$$\widehat{\theta}(m) := \underset{\theta \in \Theta_m}{\operatorname{argmax}} \left(\widehat{L}_n(\theta)\right). \tag{3.1}$$

Now, define the penalized criteria by

$$\widehat{C}_n(m) := -2\widehat{L}_n(\widehat{\theta}(m)) + \kappa_n|m|, \text{ for all } m \in \mathcal{M}, \tag{3.2}$$

where $(\kappa_n)_{n\in\mathbb{N}}$ is an increasing sequence of the regularization parameter (possibly data-dependent) that will be used to calibrate the penalty term, and $|m|$ is the number of non-zero components of $\theta^*(m) \in \Theta_m$ that will be called the dimension of the model m. The selection of the "best" model $\widehat{m}_n$ is then obtained by minimizing the penalized contrast; that is,

$$\widehat{m}_n := \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left(\widehat{C}_n(m)\right). \tag{3.3}$$

Using the results of Theorems 2.1 and 2.2, we establish the asymptotic behavior of the model selection procedure, as shown in the following theorem.

Theorem 3.1. *Let $(Y_1,\ldots,Y_n)$ be a trajectory of a process belonging to $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$, where $\theta^* \in \Theta \cap \Theta(r)$ with $r > 4$. Assume that (**A0**)-(**A2**), ($\boldsymbol{A}_i(f_\theta, \Theta)$), ($\boldsymbol{A}_i(M_\theta, \Theta)$) (or ($\boldsymbol{A}_i(H_\theta, \Theta)$)) (for $i = 0,1,2$) and (2.2) (with $r > 4$) hold with $\kappa_n/n \underset{n\to\infty}{\longrightarrow} 0$. Suppose that when $\theta^* \in \overset{\circ}{\Theta}$,*

$$\sum_{k\geq 1} \frac{1}{\sqrt{k \log\log k}} \sum_{j\geq k} \sum_{i=0}^{2} \left\{\alpha_{j,Y}^{(i)}(f_\theta,\Theta) + \alpha_{j,X}^{(i)}(f_\theta,\Theta) + \alpha_{j,Y}^{(i)}(M_\theta,\Theta) + \alpha_{j,X}^{(i)}(M_\theta,\Theta) + \alpha_{j,Y}^{(i)}(H_\theta,\Theta) + \alpha_{j,X}^{(i)}(H_\theta,\Theta)\right\} < \infty. \tag{3.4}$$

(i.) If $\kappa_n/\sqrt{\log\log n} \underset{n\to\infty}{\longrightarrow} \infty$, *then*

$$\widehat{m}_n \underset{n\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} m^*.$$

(ii.) When $\theta^* \in \overset{\circ}{\Theta}$, *there exists a constant* c *such that if* $\liminf_{n\to\infty}(\kappa_n/\log\log n) > c$, *then*

$$\widehat{m}_n \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} m^*.$$

(iii.) If $\theta^* \in \overset{\circ}{\Theta}$ *and (3.4) holds, then*

$$\widehat{\theta}(\widehat{m}_n) - \theta^* = O\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Remark that, if $\sum_{i=0}^{2}\left\{\alpha_{j,Y}^{(i)}(f_\theta,\Theta)+\alpha_{j,X}^{(i)}(f_\theta,\Theta)+\alpha_{j,Y}^{(i)}(M_\theta,\Theta)+\alpha_{j,X}^{(i)}(M_\theta,\Theta)+\alpha_{j,Y}^{(i)}(H_\theta,\Theta)+\alpha_{j,X}^{(i)}(H_\theta,\Theta)\right\} = O(j^{-\gamma})$ for some $\gamma > 3/2$, then (3.4) is satisfied. The first and second parts of Theorem 3.1 show the consistency of the selection procedure; in particular, the second part provides sufficient conditions for the consistency of the HQC procedure. The last part establishes that the estimator of the parameter of the selected model $\widehat{\theta}(\widehat{m}_n)$ obeys the law of iterated logarithm.

4. Some examples

In this section, we detail some particular processes satisfying the class (1.1). We show that the regularity conditions required for the main results are satisfied for these processes, with a particular emphasis on the identifiability assumption. For each example discussed, we consider that $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d_x,t}) \in \mathbb{R}^{d_x}$ ($d_x \in \mathbb{N}$) represents a vector of covariates; and $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of zero-mean *i.i.d.* random variable satisfying $\mathbb{E}(\xi_0^r) < \infty$ for some $r \geq 2$ and $\mathbb{E}(\xi_0^2) = 1$.

4.1. Threshold ARX(∞) models

Consider the threshold autoregressive model with exogenous covariates (TARX (∞)) defined by,

$$Y_t = \psi_0(\theta^*) + \sum_{k\geq 1}\Big(\psi_k^+(\theta^*)\max(Y_{t-k},0) + \psi_k^-(\theta^*)\min(Y_{t-k},0)\Big) + \sum_{k\geq 1}\gamma_k'(\theta^*)X_{t-k} + \xi_t,\ \forall t\in\mathbb{Z}, \quad (4.1)$$

where θ^* is the true parameter and $\psi_0(\cdot)$, $\psi_k^+(\cdot)$, $\psi_k^-(\cdot)$, $\gamma_k(\cdot)$ (for $k \geq 1$) are assumed to be twice continuously differentiable functions on Θ. This model is a generalization of the threshold AR process of Tong (1990) [30]. Also, the ARMAX process (see Hannan and Deistler (2012) [14]) is a specific example of the

model (4.1). Set $f_\theta^t = \psi_0(\theta) + \sum_{k\geq 1}\left(\psi_k^+(\theta)\max(Y_{t-k},0) + \psi_k^-(\theta)\min(Y_{t-k},0)\right)$ $+\sum_{k\geq 1}\gamma_k'(\theta)X_{t-k}$ for all $\theta\in\Theta$. If $\sum_{k\geq 1}\|\gamma_k'(\theta^*)\| < \infty$ and $\sum_{k\geq 1}\max\left\{\alpha_k(g), |\psi_k^+(\theta^*)|, |\psi_k^-(\theta^*)|\right\} < 1$, then, assumption $\mathbf{A}_0(f_\theta, \{\theta^*\})$ holds and there exists a stationary and ergodic solution with r-order moment. The assumption (**A1**) holds with $\underline{h} = 1$. Denote for $t, i \in \mathbb{Z}$,

$$\mathcal{F}_{t,i} = \sigma(\xi_{t-j}, j > i,\ X_{t-k}, k > 0), \tag{4.2}$$

the σ-field generated by $\{\xi_{t-j}, j > i,\ X_{t-k}, k > 0\}$. Let us set following additional conditions:

(**B0**): $\mathbb{E}\left[\xi_t X_{t'}\right] = 0$ for all $(t,t') \in \mathbb{Z}^2$;
(**B1**): for $c^+, c^- \in \mathbb{R}$ such that $c^+ \neq 0$ or $c^- \neq 0$, $c^+\max(Y_{t-i},0) + c^-\min(Y_{t-i},0)$ given $\mathcal{F}_{t,i}$ is non-degenerate;
(**B2**): if $(c_k)_{k\in\mathbb{N}}$ is a sequence of vector of $\mathbb{R}^{d_x}$ such as $\exists c_{k_0} \neq 0$ (with $k_0 \in \mathbb{N}$), then $\sum_{k\geq 1} c_k' X_{t-k}$ is non-degenerate;
(**B3**): the function $\theta \mapsto \psi_{k_0}^+(\theta)$ or $\theta \mapsto \psi_{k_0}^-(\theta)$, for some $k_0 \geq 1$ is injective;
(**B4**): The function $\theta \mapsto \psi_0(\theta)$ (or $\theta \mapsto \gamma_{k_0}(\theta)$ for some $k_0 > 0$) is is injective and holds (**B2**).

To ensure the identifiability, both the assumptions (**B3**) and (**B4**) are not necessary; that is, the model is identifiable if (**B1**) and ((**B3**) or (**B4**)) hold. Indeed, let $\theta \in \Theta$ such that $f_\theta^t = f_{\theta^*}^t$. Then,

$$\sum_{k\geq 1}\Big(\big(\psi_k^+(\theta^*) - \psi_k^+(\theta)\big)\max(Y_{t-k},0) + \big(\psi_k^-(\theta^*) - \psi_k^-(\theta)\big)\min(Y_{t-k},0)\Big)$$
$$= \psi_0(\theta) - \psi_0(\theta^*) + \sum_{k\geq 1}\big(\gamma_k'(\theta) - \gamma_k'(\theta^*)\big)X_{t-k}. \tag{4.3}$$

By contradiction, assume that $\psi_k^+(\theta^*) - \psi_k^+(\theta) \neq 0$ or $\psi_k^-(\theta^*) - \psi_k^-(\theta) \neq 0$ for some $k > 0$ and let $m > 0$ be the smallest integer satisfying $\psi_m^+(\theta^*) - \psi_m^+(\theta) \neq 0$ or $\psi_m^-(\theta^*) - \psi_m^-(\theta) \neq 0$. It holds from (4.3) that,

$$\begin{aligned}
&\big(\psi_m^+(\theta^*) - \psi_m^+(\theta)\big)\max(Y_{t-m},0) + \big(\psi_m^-(\theta^*) - \psi_m^-(\theta)\big)\min(Y_{t-m},0)\\
&= \psi_0(\theta) - \psi_0(\theta^*)\\
&\quad - \sum_{k>m}\Big(\big(\psi_k^+(\theta^*) - \psi_k^+(\theta)\big)\max(Y_{t-k},0) + \big(\psi_k^-(\theta^*) - \psi_k^-(\theta)\big)\min(Y_{t-k},0)\Big)\\
&\quad + \sum_{k\geq 1}\big(\gamma_k'(\theta) - \gamma_k'(\theta^*)\big)X_{t-k}.
\end{aligned} \tag{4.4}$$

Since the right-hand side of (4.4) is $\mathcal{F}_{t,m}$-measurable (thanks to the non anticipative property of the process (Y_t)), $\big(\psi_m^+(\theta^*) - \psi_m^+(\theta)\big)\max(Y_{t-m},0) + \big(\psi_m^-(\theta^*) - \psi_m^-(\theta)\big)\min(Y_{t-m},0)$ given $\mathcal{F}_{t,m}$ is degenerate, which contradicts the assumption (**B1**). Thus, $\psi_k^+(\theta^*) - \psi_k^+(\theta) = 0$ and $\psi_k^-(\theta^*) - \psi_k^-(\theta) = 0$ for all $k \geq 1$. Therefore,

- if (**B3**) holds, then $\theta = \theta^*$;
- else, if (**B4**) holds, then we have from (4.4),

$$\sum_{k=1}^{\infty}(\gamma_k'(\theta^*) - \gamma_k'(\theta))X_{t-k} = \psi_0(\theta) - \psi_0(\theta^*),$$

which implies that $\gamma_k(\theta^*) = \gamma_k(\theta)\ \forall k \geq 1$ (by the assumption (**B2**)); and consequently, $\psi_0(\theta) = \psi_0(\theta^*)$. Hence, $\theta = \theta^*$.

In the case of the ARX(∞) models (obtained when $\psi_k^+(\theta^*) = \psi_k^-(\theta^*)$, for all $k \geq 1$), the condition (**B1**) is not necessary; the identifiability holds with (**B0**) and ((**B3**) or (**B4**)).

4.2. Asymmetric Power ARCH-X(δ, ∞) models

Consider the Asymmetric Power ARCH with with exogenous covariates (APA RCH-X(δ, ∞)) defined by,

$$Y_t = \xi_t\sigma_t((\theta^*) \text{ with}$$
$$\sigma_t^{\delta}(\theta^*) = \phi_0(\theta^*) + \sum_{k=1}^{\infty}\Big(\phi_k^+(\theta^*)\big(\max(Y_{t-k},0)\big)^{\delta} + \phi_k^-(\theta^*)\big(\max(-Y_{t-k},0)\big)^{\delta}\Big) + \sum_{k=1}^{\infty}\gamma_k'(\theta^*)X_{t-k},\ \forall t \in \mathbb{Z}, \qquad (4.5)$$

where θ^* is the true parameter, $\phi_0(\cdot)$, $\phi_k^+(\cdot)$, $\phi_k^-(\cdot)$, $\gamma_k(\cdot)$, $k \geq 1$ are non-negative (componentwise for $\gamma_k(\theta^*)$) functions assumed to be twice continuously differentiable on Θ, with $(\phi_0(\theta)) > 0$ for all $\theta \in \Theta$, X_t is a vector of non-negative (componentwise) covariates, and $\delta > 0$. This process is an example of the class (1.1) with $f_\theta^t = 0$ and $M_\theta^t = \sigma_t(\theta)$. Numerous classical ARCH-type parametrizations, for instance, GARCH-X (obtained with $\delta = 2$ and $\phi_k^+(\theta^*) = \phi_k^-(\theta^*)$), TARCH-X (obtained with $\delta = 1$) are specific example of (4.5). There are several works in the literature based on the GARCH-X model, see for instance Han and Kristensen (2014) [11], Nana *et al.* (2013) [22], Han (2015) [12]. Model (4.5) is a generalization of the class of APARCH-X(δ, p, q) studied by Francq and Thieu (2019) [8]. Assume that $\theta^* \in \Theta(r)$, where $\Theta(r)$ is given by

$$\Theta(r) = \Big\{\theta \in \mathbb{R}^d \ / \ \sum_{k\geq 1}\max\Big\{\alpha_k(g), \|\xi_0\|_r|\phi_k^+(\theta)|^{1/\delta}, \|\xi_0\|_r|\phi_k^-(\theta)|^{1/\delta}\Big\} < 1 \text{ and } \sum_{k\geq 1}\|\gamma_k(\theta)\|^{1/\delta} < \gamma_U \text{ for some } \gamma_U > 0\Big\}.$$

Therefore, $\mathbf{A}_0(H_\theta, \{\theta^*\})$ holds; and a stationary and ergodic solution with r-order moment exists. If $\inf_{\theta\in\Theta}\phi_0(\theta) > 0$, then the assumption (**A1**) is satisfied. The following assumptions are needed to ensure the identifiability.

(**B5**): for all $i \geq 1$ and $t \in \mathbb{Z}$, the support of the distribution of ξ_{t-i} given $\mathcal{F}_{t,i}$ is not included in $[0, \infty)$ or in $(-\infty, 0]$ and contains at least three points.
(**B6**): the function $\theta \mapsto \phi^+_{k_0}(\theta)$ or $\theta \mapsto \phi^-_{k_0}(\theta)$ for some $k_0 > 0$, is injective.
(**B7**): The function $\theta \mapsto \phi_0(\theta)$ (or $\theta \mapsto \gamma_{k_0}(\theta^*)$ for some $k_0 > 0$) is injective, and the condition (**B2**) holds for this model.

By going along similar lines as in the Subsection 4.1 and by using the Lemma 4 in [8], one get that, if (**B5**) and ((**B6**) or (**B7**)) hold, then the model (4.5) is identifiable.

Let us stress that, the *i.i.d.* assumption for $(\xi_t)_{t\in\mathbb{Z}}$ is a bit strong for the model (4.5). This assumption, which is needed to the large class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ can be relaxed to $(\xi_t, \mathcal{F}_{t,0})_{t\in\mathbb{Z}}$ is a martingale difference sequence (see, for instance, [8] in the case of APARCH-X(δ, p, q) model) when checking the identifiability. Also, as pointed out by Francq and Thieu (2019) [8], in the absence of covariates and when (ξ_t) is *i.i.d* (i.e., the case of the APARCH(∞) model), the assumption (**B5**) can be automatically reduced to: $P(\xi_0 > 0) \in (0, 1)$ and the support of the distribution of ξ_0 contains at least three points. Moreover, Assumption (**B5**) prevents taking redundant covariate; for instance, it excludes the situation where $X_{t-1} = \max(Y_{t-i}, 0)^\delta$ or $\max(-Y_{t-i}, 0)^\delta$ for some $i \geq 1$.

4.3. ARX(∞)-ARCH(∞) models

Consider the ARX(∞)-ARCH(∞) model given by,

$$\begin{cases} Y_t = \psi_0(\theta^*) + \sum\limits_{k\geq 1} \psi_k(\theta^*) Y_{t-k} + \sum\limits_{k\geq 1} \gamma'_k(\theta^*) X_{t-k} + \varepsilon_t \\ \varepsilon_t = \xi_t \sigma_t(\theta^*) \quad \text{with} \quad \sigma^2_t(\theta^*) = \phi_0(\theta^*) + \sum\limits_{k\geq 1} \phi_k(\theta^*) \varepsilon^2_{t-k}, \end{cases} \tag{4.6}$$

where θ^* is the true parameter, and $\psi_0(\cdot)$, $\psi_k(\cdot)$, $\gamma_k(\cdot)$, $\phi_0(\cdot)$, $\phi_k(\cdot)$, $k \geq 1$ are assumed to be twice continuously differentiable on Θ, and satisfying $\phi_0(\theta) > 0$ for all $\theta \in \Theta$. Model (4.6) is an extension of the ARMA-GARCH, ARMAX-GARCH processes. This model belongs to the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ with

$$f^t_\theta = \psi_0(\theta) + \sum_{k\geq 1} \psi_k(\theta) Y_{t-k} + \sum_{k\geq 1} \gamma'_k(\theta) X_{t-k} \quad \text{and}$$

$$M^t_\theta = \sqrt{\phi_0(\theta) + \sum_{k\geq 1} \phi_k(\theta) \Big(Y_{t-k} - \psi_0(\theta) - \sum_{j\geq 1} \psi_j(\theta) Y_{t-k-j} - \sum_{j\geq 1} \gamma'_j(\theta) X_{t-k-j} \Big)^2},$$

for all $\theta \in \Theta$. Hence, the assumption $\mathbf{A}_0(f_\theta, \{\theta\})$ holds with $\alpha^{(0)}_{k,Y}(f_\theta, \{\theta\}) = |\psi_k(\theta)|$ and $\alpha^{(0)}_{k,X}(f_\theta, \{\theta\}) = \|\gamma_k(\theta)\|$. From an expansion of M^t_θ, one can easily

get that, $\mathbf{A}_0(M_\theta, \{\theta\})$ holds with

$$\begin{cases} \alpha_{1,Y}^{(0)}(M_\theta, \{\theta\}) = \sqrt{\phi_k(\theta)} \ \text{ and} \\ \qquad \alpha_{k,Y}^{(0)}(M_\theta, \{\theta\}) = \sqrt{\phi_k(\theta)} + \sum_{i=1}^{k-1} \sqrt{\phi_i(\theta)}|\psi_{k-i}(\theta)| \quad \text{for } k \geq 2; \\ \alpha_{1,X}^{(0)}(M_\theta, \{\theta\}) = 0 \ \text{ and } \ \alpha_{k,X}^{(0)}(M_\theta, \{\theta\}) = \sum_{i=1}^{k-1} \sqrt{\phi_i(\theta)}\|\gamma_{k-i}(\theta)\| \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } k \geq 2. \end{cases}$$

Therefore, the stationarity set $\Theta(r)$ is defined by

$$\Theta(r) = \Big\{\theta \in \mathbb{R}^d \ / \ \sum_{k\geq 1} \max\Big\{\alpha_k(g), |\psi_k(\theta)| + \|\xi_0\|_r\Big(\sqrt{\phi_k(\theta)} + \sum_{i=1}^{k-1}\sqrt{\phi_i(\theta)}|\psi_{k-i}(\theta)|\Big)\Big\} < 1 \ \text{ and } \sum_{k\geq 1}\|\gamma_k(\theta)\| < \gamma_U \text{ for some } \gamma_U > 0\Big\}.$$

Assumption (**A1**) is satisfied if $\inf_{\theta\in\Theta}\phi_0(\theta) > 0$. The identifiability conditions can be obtained as in Subsection 4.1 and 4.2.

5. Simulation study

In this section, we consider a double autoregressive model with exogenous covariates, defined by

$$Y_t = \phi_0 + \sum_{i=1}^{p_1}\phi_i Y_{t-i} + \sum_{i=1}^{q_1}\psi_i' X_{t-i} + \xi_t\sqrt{\alpha_0 + \sum_{i=1}^{p_2}\alpha_i Y_{t-i}^2 + \sum_{i=1}^{q_2}\beta_i'(X_{t-i}\odot X_{t-i})}, \tag{5.1}$$

where $(X_t)_{t\in\mathbb{Z}}$ is an exogenous multivariate covariate process with values in $\mathbb{R}^{d_x}$ ($d_x \in \mathbb{N}$), $p_1, p_2, q_1, q_2 \in \mathbb{N}$, $\phi_0, \phi_1, \cdots, \phi_{p_1} \in \mathbb{R}$, $\psi_1, \cdots, \psi_{q_1} \in \mathbb{R}^{dx}$, $\alpha_0 > 0$, $\alpha_1, \cdots, \alpha_{p_2} \geq 0$, $\beta_1, \cdots, \beta_{q_2} \in [0, \infty)^{d_x}$, $\odot$ denotes the Hadamard product (componentwise multiplication), ξ_t is a white noise with $\mathbb{E}\xi_0^2 = 1$. This model is a generalization of the factor double autoregressive (FDAR) process introduced by Guo *et al.* (2014) [10] to extend the double AR(p) model proposed by Ling (2007) [21]. The ARX(p) and ARCH-X(p) are particular cases of the model (5.1). We assume that $(X_t)_{t\in\mathbb{Z}}$ is a VAR(1) (vector autoregressive) process:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \eta_t \ \text{ for all } t \in \mathbb{Z}, \tag{5.2}$$

where $\varphi_0 \in \mathbb{R}^{d_x}$, φ_1 is a real coefficients $(d_x \times d_x)$-matrix and η_t is a white noise with $\mathbb{E}(\eta_0\eta_0') = \Sigma_\eta$. If $\|\varphi_1\| < 1$, then the stability condition (2.2) holds with $\alpha_1(g) = \|\varphi_1\|$ and $\alpha_k(g) = 0$ for $k \geq 2$. The stationarity set $\Theta(r)$ is

$$\Theta(r) = \Big\{\theta = (\phi_0, \phi_1, \cdots, \phi_{p_1}, \alpha_0, \alpha_1, \cdots, \alpha_{p_2}, \psi_1', \cdots, \psi_{q_1}', \beta_1', \cdots, \beta_{q_2}') \in \Theta,$$
$$\text{where } \Theta \subset \mathbb{R}^{p_1+1} \times (0,\infty) \times [0,\infty)^{p_2} \times \mathbb{R}^{d_x q_1} \times [0,\infty)^{d_x q_2} \text{ and}$$
$$\sum_{i=1}^{\max(1,p_1,p_2)} \max\{\alpha_i(g), |\phi_i| + \|\xi_0\|_r \alpha_i\} < 1\Big\},$$

with $\phi_i = 0$ if $i > p_1$ and $\alpha_i = 0$ if $i > p_2$. Based on the examples discussed in Section 4, if the conditions (**B0**) and (**B3**) hold for (5.1), then to satisfy the identifiability condition, it suffices to impose the following assumption on the covariate:

(**B7**): if $(c_{k,1})_{1\le k\le q_1}$ and $(c_{k,2})_{1\le k\le q_2}$ are sequences of vector of $\mathbb{R}^{d_x}$ such as $\exists c_{k_1,1} \neq 0$ and $\exists c_{k_2,2} \neq 0$ (with $1 \le k_1 \le q_1$, $1 \le k_2 \le q_2$), then $\sum_{i=1}^{q_1} c_{k,1}' X_{t-k}$ and $\sum_{i=1}^{q_2} c_{k,2}' X_{t-k} \odot X_{t-k}$ are not degenerated.

Set $\psi = (\psi_1, \ldots, \psi_q)$ and $\beta = (\beta_1, \ldots, \beta_q)$; the true parameter is $\theta^* = (\phi_0, \phi_1, \alpha_0, \alpha_1, \psi, \beta)$. In the sequel, we focus on the following two cases.

Case 1. We consider an example of model (5.1) with univariate covariates where $p_1 = p_2 = 1$ and $q_1 = q_2 = q$; the AR parameter is set to $(\varphi_0, \varphi_1) = (0.5, 0.5)$ and $\Sigma_\eta = 1$.

Case 2. In this second example, model (5.1) is considered with univariate/multivariate covariates where $p_1 = p_2 = 1$, $q_1 = 1$ and $q_2 = 0$; thus, the true parameter is $\theta^* = (\phi_0, \phi_1, \alpha_0, \alpha_1, \psi_1')$. This second example is related to the real data application (see Section 6).

5.1. Estimation and significance test

Some results from Monte Carlo simulations are displayed to assess the asymptotic properties of the QMLE, as well as to investigate the empirical size and power of the proposed procedure on testing the significance of the covariate X_t. We will consider samples where the innovation $(\xi_t)_{t\in\mathbb{Z}}$ is generated from Gaussian and Student distributions (with 5 with degrees of freedom). The model (5.1) is considered in the **Case 1** with $p_1 = p_2 = q_1 = q_2 = 1$ (scenario $\mathbf{S}_0$, $\mathbf{S}_1$, $\mathbf{S'}_0$ and $\mathbf{S'}_1$ below). In the **Case 2** ($p_1 = p_2 = q_1 = 1$, $q_2 = 0$), we consider scenarios with univariate covariate (scenario $\mathbf{S''}_0$ and $\mathbf{S''}_1$ below), the AR parameter are $(\varphi_0, \varphi_1, \Sigma_\eta) = (23.61, 0.7, 21.56)$.

- **scenario** $\mathbf{S}_0$: $\theta^* = (0.15, -0.2, 0.4, 0.3, 0, 0)$;
- **scenario** $\mathbf{S}_1$: $\theta^* = (0.15, -0.2, 0.4, 0.3, 0.08, 0)$;
- **scenario** $\mathbf{S'}_0$: $\theta^* = (1, 0.4, 0.5, 0.2, 0, 0)$;
- **scenario** $\mathbf{S'}_1$: $\theta^* = (1, 0.4, 0.5, 0.2, 0.07, 0.07)$.
- **scenario** $\mathbf{S''}_0$: $\theta^* = (37.95, 0.33, 32.11, 0.02, 0)$;
- **scenario** $\mathbf{S''}_1$: $\theta^* = (37.95, 0.33, 32.11, 0.02, -0.21)$.

The scenarios $\mathbf{S}_0$, $\mathbf{S'}_0$ and $\mathbf{S''}_0$ correspond to cases where the covariate is absent; $\mathbf{S''}_0$ and $\mathbf{S''}_1$ are related to the real data application. We consider the following significance tests:

$$H_0 : \theta^* = (0.15, -0.2, 0.4, 0.3, 0, 0) \quad (S_0) \quad \text{against} \quad H_1 : \theta^* \neq (0.15, -0.2, 0.4, 0.3, 0, 0);$$

$$H_0 : \theta^* = (1, 0.4, 0.5, 0.2, 0, 0) \quad (S_0') \quad \text{against} \quad H_1 : \theta^* \neq (1, 0.4, 0.5, 0.2, 0, 0).$$

$$H_0 : \theta^* = (37.95, 0.33, 32.11, 0.02, 0) \quad (S_0'') \quad \text{against} \quad H_1 : \theta^* \neq (37.95, 0.33, 32.11, 0.02, 0).$$

In each of the scenarios $\mathbf{S}_0$, $\mathbf{S}_1$, $\mathbf{S'}_0$, $\mathbf{S'}_1$, $\mathbf{S''}_0$ and $\mathbf{S''}_1$, we simulate 200 replications with the sample size $n = 500$, 1000 and test the nullity of the vector (ψ, β) after estimating the parameters of interest. Table 1 contains the empirical mean and root mean square error (RMSE) of each component of the estimator. The last column of Table 1 indicates the empirical levels and powers of the above tests at the nominal level $\alpha = 0.05$, where the empirical powers are computed under the alternative H_1 respectively in the scenario $\mathbf{S}_1$, $\mathbf{S'}_1$ and $\mathbf{S''}_1$. For the scenario $\mathbf{S'}_1$, the histograms and estimated densities of the estimates are plotted in Figure 1.

From these findings, one can see that, in all scenarios, the performance of the QMLE is satisfactory in terms of the mean and that, the RMSE of the estimators decreases when n increases. This is consistent with the results of Theorem 2.1. Also remark that, the fact of computing the QMLE with $(\widehat{\psi}, \widehat{\beta})$ for trajectories generated without covariates (see the scenarios $\mathbf{S}_0$, $\mathbf{S'}_0$ and $\mathbf{S''}_0$) does not affect the performance of the QMLE, which again confirms its good theoretical properties. As seen in Figure 1, for each component of $\widehat{\theta}_n$, the estimated density is very close to that of the normal distribution; which is in accordance with the asymptotic results obtained from Theorem 2.2 when $\theta^* \in \overset{\circ}{\Theta}$. The results of the test (see Table 1) show that, the statistic W_n is slightly oversized for $n = 500$ in cases where the innovation is generated from Student distributions, but the empirical levels are reasonable when $n = 1000$ in the sense that, they are very close to the nominal one. Further, the empirical powers displayed increases with the sample size and are quite accurate.

5.2. Model selection

Now, we are going to carry out other simulation experiments aimed at evaluating the effectiveness of the proposed model selection procedure in the model (5.1) for choosing the order $q_1 = q_2 = q$ in the **Case 1**. To this end, $q = 2$ is set as the "true" model m^* and that the following scenarios are considered:

- **scenario** $\mathbf{S}_1^*$: $\theta^* = (0.6, 0.45, 0.5, 0.15, 1, 0.7, 0.6, 0.35)$;
- **scenario** $\mathbf{S}_2^*$: $\theta^* = (0.15, 0.4, 0.5, 0.2, 0.1, 0.1, 0.03, 0.3)$.

The competing models used are all process satisfying (5.1) with $q \in \{0, 1, \ldots, 9\}$, which leads us to a collection of 10 models.

In the **Case 2**, consider the multivariate covariate $X_t = (X_{t,1}, X_{t,2}, \cdots, X_{t,5})$, a VAR(1) (see (5.2)) with parameter $\varphi_0 = (23.61, 4.95, 12.12, 716.70, 13.01)'$, $\varphi_1 = diag(0.7, 0.57, 0.51, 0.26, 0.56)$ and

$$\Sigma_\eta = \begin{pmatrix} 42.62 & -2.59 & 11.14 & 281.05 & -14.65 \\ -2.59 & 7.50 & -1.29 & 200.64 & -8.24 \\ 11.14 & -1.29 & 41.13 & 1019.57 & -1.10 \\ 281.05 & 200.64 & 1019.57 & 64948.60 & -745.27 \\ -14.64 & -8.24 & -1.10 & -745.27 & 59.00 \end{pmatrix}$$

Consider the scenario $\mathbf{S"}_1$ with the covariate $X_{t,1}$ as the true model. The **Case 2** with all the combination of the covariates is performed on the data; that is, there are 32 competing models. This example is related and close to the real data application.

For $n = 100, 125, 150, \ldots, 1000$, we simulate 100 independent replications in each of the three scenarios $\mathbf{S}_1^*$, $\mathbf{S}_2^*$ and $\mathbf{S"}_1$. We compare the performances of the procedure with $\kappa_n = \log n$ (see (3.2)) linked to the Bayesian Information Criteria (BIC) and the procedure with $\kappa_n = c \log\log n$ ($c \in \{2, 3.5, 5\}$) linked to the Hannan-Quinn information Criterion (HQC). For the scenarios $\mathbf{S}_1^*$ and $\mathbf{S}_2^*$, Figures 2 displays the points $(n, \widehat{q}_n)$, where $\widehat{q}_n$ denotes the average of the orders selected with trajectories of length n, as well as the curve of the proportions of number of replications (frequencies) where the associated criterion selects the true order. For the scenario of the **Case 2** (i.e, $\mathbf{S"}_1$), the probabilities of choosing the true covariate are displayed in Figure 3.

From these figures, the first remark is that, for all the penalties, the performances of the procedure increase with n in each scenario. Further, the probability of selecting the true order is very close to 1 when $n = 1000$. This shows that, these procedures are in accordance with the results of Theorem 3.1. One can notice that, in the scenario $\mathbf{S}_1^*$, the $\log n$ penalty is more interesting for selecting the true order than the others penalties for a small sample size (see Figure 2 ((a) and (b)) for $n \leq 250$), while in the scenarios $\mathbf{S}_2^*$ and $\mathbf{S"}_1$, the HQC with $c = 2$ slightly outperforms the BIC penalization when $n \leq 350$ (see Figure 2 ((c) and (d)) and Figure 3). However, the larger the sample size, the $c \log\log n$ penalty (except for the case where $c = 2$) provides the same accuracies in comparison with the $\log n$ penalty, and displays satisfactory results. The results also show that, as c increases, the performances of the $c \log\log n$ penalty increase, which reveals that the common use of the classical HQC penalization (i.e, the $c \log\log n$ penalty with $c = 2$) is not always the optimal choice to select the best model with this information criterion.

TABLE 1

Sample mean and RMSE of the QMLE for the model (5.1) following the scenarios $\boldsymbol{S}_0$, $\boldsymbol{S}_1$, $\boldsymbol{S'}_0$, $\boldsymbol{S'}_1$, $\boldsymbol{S''}_0$ *and* $\boldsymbol{S''}_1$, *where G and St refers to the Gaussian and Student innovation. The last column shows the empirical levels (scenarios* $\boldsymbol{S}_0$, $\boldsymbol{S'}_0$, $\boldsymbol{S''}_0$*) and powers (scenarios* $\boldsymbol{S}_1$, $\boldsymbol{S'}_1$, $\boldsymbol{S''}_1$*) at the nominal level* 0.05 *for the test of the relevance of the exogenous covariates.*

			QMLE						
Scenario	n		$\widehat{\phi}_0$	$\widehat{\phi}_1$	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\psi}$	$\widehat{\beta}$	W_n
$\mathbf{S}_0$ G	500	Mean	*0.1488*	−0.2001	0.3887	0.2987	0.0019	0.0033	0.035
		Rmse	0.0435	0.0607	0.0421	0.0924	0.0276	0.0063	
	1000	Mean	0.1502	−0.2023	0.3911	0.2979	−0.0006	0.0032	0.045
		Rmse	0.0326	0.0370	0.0288	0.0704	0.0832	0.0055	
$\mathbf{S}_0$ St	500	Mean	0.1511	−0.1956	0.3820	0.3332	−0.0013	0.00585	0.070
		Rmse	0.0445	0.0711	0.0630	0.1934	0.0286	0.0115	
	1000	Mean	0.1475	−0.1998	0.3902	0.3071	0.0004	0.0049	0.055
		Rmse	0.0302	0.0400	0.0545	0.1394	0.0180	0.0092	
$\mathbf{S}_1$ G	500	Mean	0.1520	−0.2012	0.3857	0.2997	0.0806	0.0049	0.750
		Rmse	0.0488	0.0547	0.0439	0.0902	0.0316	0.0095	
	1000	Mean	0.1511	−0.1996	0.3889	0.3006	0.0793	0.0031	0.970
		Rmse	0.0327	0.0365	0.0295	0.0672	0.0212	0.0055	
$\mathbf{S}_1$ St	500	Mean	0.1518	−0.2004	0.3790	0.3149	0.0803	0.0089	0.825
		Rmse	0.0396	0.0616	0.0695	0.1734	0.0275	0.0178	
	1000	Mean	0.1497	−0.1977	0.3862	0.3076	0.0805	0.0047	0.990
		Rmse	0.0295	0.0452	0.0512	0.1334	0.0198	0.0090	
$\mathbf{S'}_0$ G	500	Mean	1.0003	0.4009	0.4793	0.2004	0.0020	0.0065	0.030
		Rmse	0.0815	0.0447	0.0672	0.0285	0.0383	0.0126	
	1000	Mean	1.0015	0.4005	0.4854	0.1986	−0.0012	0.0050	0.055
		Rmse	0.0481	0.0277	0.0485	0.0200	0.0295	0.0095	
$\mathbf{S'}_0$ St	500	Mean	1.0061	0.3957	0.4734	0.2011	0.0005	0.0101	0.075
		Rmse	0.0830	0.0465	0.1230	0.0583	0.0393	0.0197	
	1000	Mean	1.0025	0.3967	0.4852	0.1976	0.0003	0.0088	0.065
		Rmse	0.0554	0.0296	0.0880	0.0349	0.0245	0.0181	
$\mathbf{S'}_1$ G	500	Mean	1.0022	0.3993	0.4943	0.2005	0.0704	0.0702	0.795
		Rmse	0.0821	0.0443	0.0803	0.0293	0.0449	0.0315	
	1000	Mean	1.0019	0.3970	0.5052	0.1979	0.0680	0.0672	0.975
		Rmse	0.0587	0.0310	0.0548	0.0217	0.0322	0.0224	
$\mathbf{S'}_1$ St	500	Mean	1.0098	0.3979	0.5015	0.1956	0.0711	0.0759	0.665
		Rmse	0.0837	0.0461	0.1575	0.0576	0.0431	0.0522	
	1000	Mean	0.9996	0.3990	0.4949	0.2034	0.0733	0.0692	0.930
		Rmse	0.0576	0.0318	0.1119	0.0420	0.0295	0.0399	
$\mathbf{S''}_0$ G	500	Mean	37.9776	0.3259	32.2873	0.0198	0.0025		0.085
		Rmse	2.4524	0.0373	2.8635	0.0021	0.0350		
	1000	Mean	37.7283	0.3274	31.9529	0.0200	0.0042		0.060
		Rmse	2.2560	0.0302	2.7856	0.0018	0.0325		
$\mathbf{S''}_0$ St	500	Mean	38.0380	0.3254	32.0056	0.0194	0.0027		0.065
		Rmse	2.3269	0.04313	2.8796	0.0036	0.0370		
	1000	Mean	37.8905	0.3279	31.7781	0.0199	0.0019		0.040
		Rmse	2.1403	0.0273	2.8731	0.0026	0.0291		
$\mathbf{S''}_1$ G	500	Mean	37.9698	0.3265	33.0176	0.0189	−0.2092		0.985
		Rmse	2.4005	0.0401	2.6480	0.0040	0.0310		
	1000	Mean	38.1928	0.3255	32.7826	0.0193	−0.2113		1.000
		Rmse	2.1828	0.0289	2.5628	0.0031	0.0263		
$\mathbf{S''}_1$ St	500	Mean	38.2794	0.3232	32.5255	0.0194	−0.2117		0.995
		Rmse	2.2897	0.0397	2.8555	0.0065	0.0282		
	1000	Mean	38.3885	0.3288	32.1560	0.0198	−0.2147		1.000
		Rmse	2.1398	0.0258	2.7629	0.0046	0.02653		

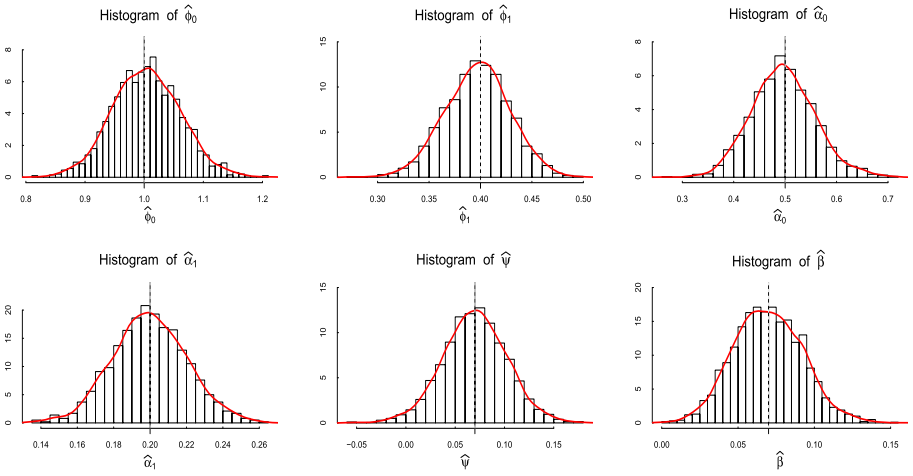

FIG 1. *Histograms of the components of* $\widehat{\theta}_n$ *in the scenario* $\boldsymbol{S'}_1$ *with sample size* $n = 1000$. *The overlaying curves are the density estimates and the dotted vertical lines represent the true values of the parameters.*

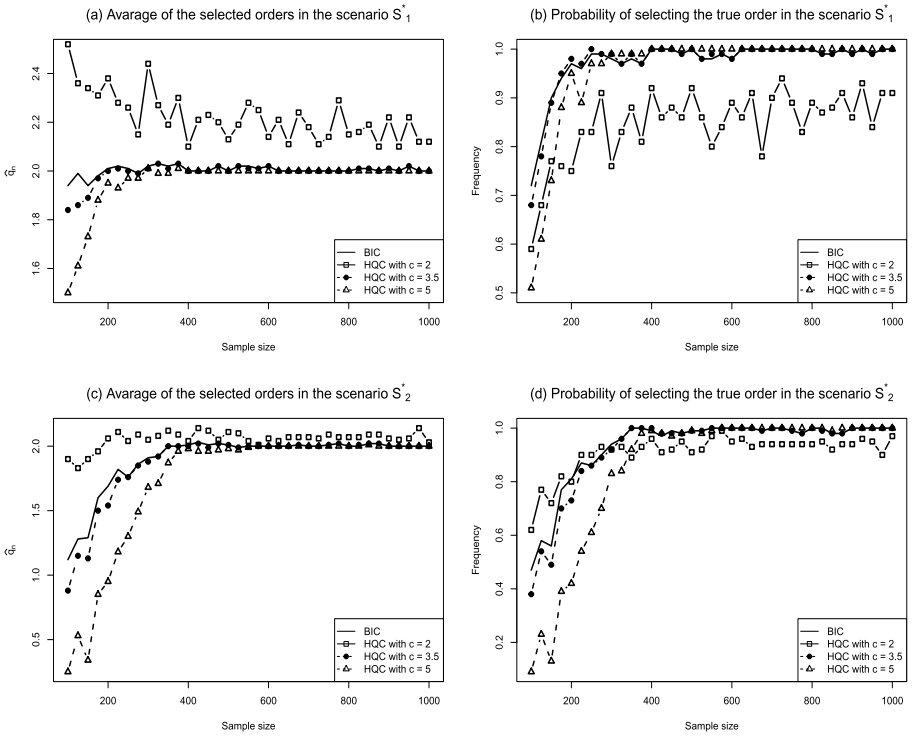

FIG 2. *The averages of the orders selected and the frequencies of selecting the true order based on 100 independent replications depending on sample's length in the scenarios* $\mathbf{S}_1^*$ *and* $\mathbf{S}_2^*$.

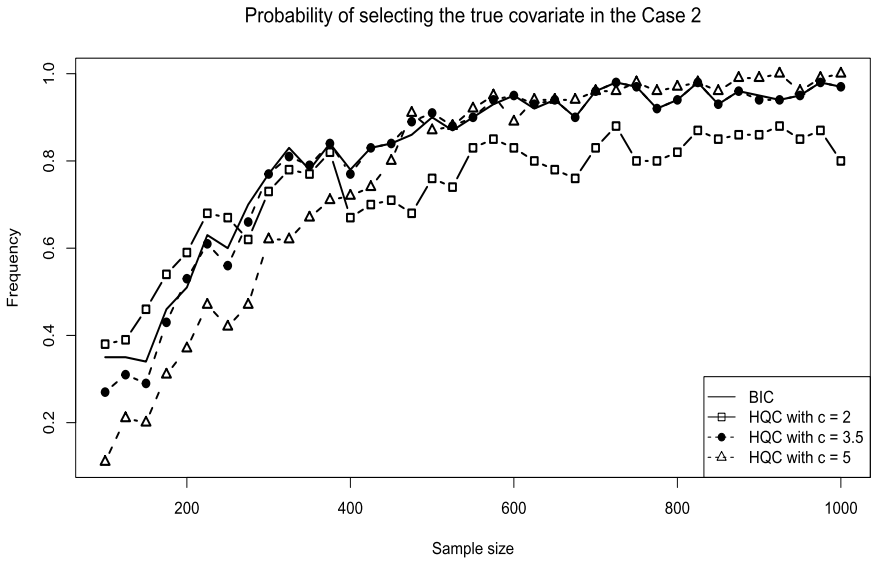

FIG 3. *The frequencies of selecting the true covariate based on 100 independent replications depending on sample's length in the scenario* $\boldsymbol{S}$"$_1$.

6. Real data example

We consider the daily concentrations of the PM_{10} (particulate matter with a diameter less than 10 μm) in the Vitória metropolitan area; see Figure 4(a). These data as well as those of the other meteorological variables (see below) are obtained from the State Environment and Water Resources Institute, and were collected at eight monitoring stations. We focus on the data from January 21st, 2005 to March 04th 2006, these observations on 408 days are a part of a large dataset (available at https://rss.onlinelibrary.wiley.com/pb-assets/hub-assets/rss/Datasets/RSSC%2067.2/C1239deSouza-1531120585220.zip) which were analyzed by Souza *et al.* (2018) [26] to quantify the association between respiratory disease and air pollution concentrations.

The variables considered are: the average concentration for the particulate matter (PM_{10}, μgm^{-3}), sulphur dioxide (SO_2, μgm^{-3}), nitrogen dioxide (NO_2, μgm^{-3}), carbon monoxide (CO, μgm^{-3}), ozone (O_3, μgm^{-3}); and Air relative humidity (RH, %). Table 2 displays some elementary statistics of these variables.

As pointed out by Ng and Awang (2018) [24], PM_{10} is a notorious air pollutant associated in particular with detrimental health impacts; it affects the respiratory and cardiopulmonary functions and increases the morbidity and mortality rate of related diseases. Therefore, forecasting the PM_{10} concentration and understanding its relation with other factors is an important issue. Several

TABLE 2
Some elementary statistics of the variables PM_{10}, SO_2, NO_2, CO, O_3, RH, for the period from January 21st, 2005 to March 04th 2006.

Variable	Mean	SD	Min	Q_1	Med	Q_3	Max
PM_{10} (μgm^{-3})	32.04	8.62	11.16	26.19	31.87	36.92	66.60
SO_2 (μgm^{-3})	11.64	2.74	4.89	9.75	11.63	13.48	19.29
NO_2 (μgm^{-3})	24.57	6.41	10.47	19.93	23.79	28.84	46.84
CO (μgm^{-3})	969.70	254.85	456.00	785.10	951.00	1129.70	2141.50
O_3 (μgm^{-3})	29.96	7.68	16.76	24.79	28.46	33.96	66.52
RH (%)	79.21	6.52	62.45	74.36	78.67	83.62	95.39

models, including among others models, ARIMA, MLR (multiple linear regression), RTSE (Regression with time series error), were considered; we refer to Ng and Awang (2018) [24] and Ng (2017) [23] and the references therein for an overview of this issue.

In this section, we focus on the forecast of the PM_{10} concentration from some meteorological variables of the previous day. One know that (see for instance, Ng and Awang (2018) [24] and the references therein), the RTSE can be used for such data. We apply the model (5.1) (whose the RTSE is a specific case when $p_2 = q_2 = 0$) with $p_1, p_2, q_1, q_2 \in \{0, 1\}$, and the covariate $X_t = (SO_{2,t}, NO_{2,t}, CO_t, O_{3,t}, RH_t)'$ (the value of the corresponding variable at day t). The following issues are addressed.

1. **Model selection**. The aim is to select the orders p_1, p_2, q_1, q_2 and the "best" subset of the covariates that are the major factor related to the next-day PM_{10} concentration. For this purpose, we consider all the combination of the covariates with $p_1, p_2, q_1, q_2 \in \{0, 1\}$; which represents a collection of 376 models. The procedure based on the penalized criteria $\widehat{C}_n(m)$ (see (3.2)) is applied with the regularization parameter $\kappa_n = \log n,\ 2\log\log n,\ 3.5\log\log n,\ 5\log\log n$. These criteria (BIC and HQC) select the model with $p_1 = p_2 = q_1 = 1$, $q_2 = 0$ and the covariate RH. This result shows that, compared to the RTSE, the selected model is preferred. This result is in accordance with some existing works (see, for instance, [24] and [23]) which have found that the air humidity of the previous day is an important factor related to PM_{10} concentration.
2. **Estimation and significance test**. The estimated model is:

$$PM_{10,t} = \underset{(2.671)}{37.946} + \underset{(0.024)}{0.330}\, PM_{10,t-1} - \underset{(0.028)}{0.210}\, RH_{t-1} + \xi_t \sqrt{\underset{(3.362)}{32.108} + \underset{(0.003)}{0.023}\, PM_{10,t-1}^2},$$

where in parentheses are the standard errors of the estimators obtained from the robust sandwich matrix. The test (2.8) with $\vartheta_0 = 0$ is now applied

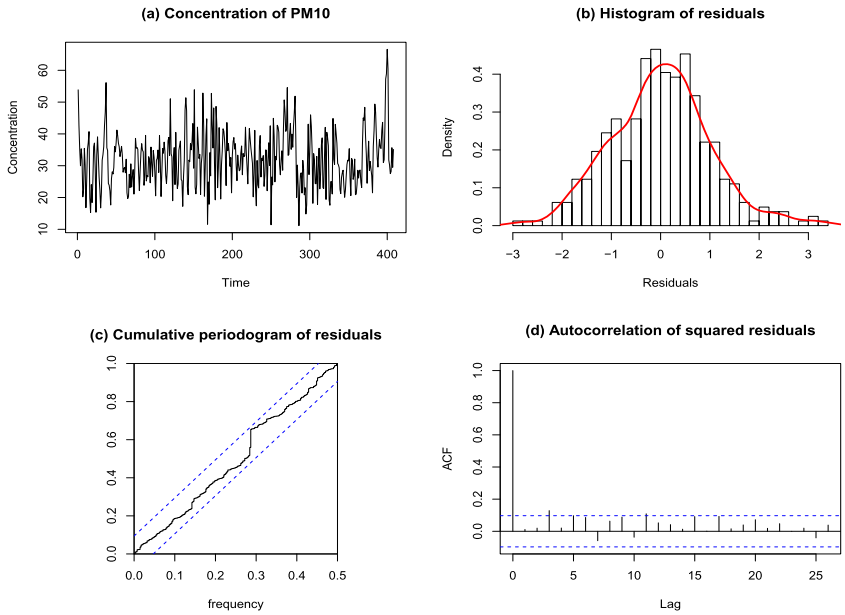

FIG 4. *(a) The daily concentrations of the PM_{10} from January 21st, 2005 to March 04th 2006; (b), (c) the histogram and the cumulative periodogram of the residuals; (d) the autocorrelation functions of the squared residuals.*

for testing the significance of the covariate. At the nominal level $\alpha = 0.05$, the critical value of the test, computed from $(\Gamma Z^{\mathcal{C}})'(\Gamma\Sigma\Gamma')^{-1}\Gamma Z^{\mathcal{C}}$ is 2.68 and the statistic, computed from W_n (see (2.9)) is 10.86. Thus, the null hypothesis is rejected. Figure 4 displays the histogram and cumulative periodogram of the residuals as well as the autocorrelation functions of the squared residuals. From these findings, the residuals do not show any signs of correlation.

In conclusion of this section, let us stress that, several authors have applied the MLR and the RTSE on other data and have found that NO_2, CO, O_3 were also important factors associated to the PM_{10} concentration (see [24] and [23]) and the references therein). For the data considered here, and by applying the model (5.1), it appears that these variables were less important in for forecasting the PM_{10} concentrations.

7. Summary and conclusion

This paper considers a general class of causal processes with exogenous covariates in a semiparametric framework. This class is quite extensive and many classical processes such as ARMA-GARCH, ARMAX-GARCH-X, APARCH-X,$\cdots$ are particular cases. Sufficient conditions for the existence of a stationary

and ergodic solution are provided.
A quasi likelihood estimator is performed for inference; the consistency of this estimator is established and the asymptotic distribution is derived. This distribution coincides with the Gaussian one, when the true parameter is an interior point of the parameter's space.
A Wald-type statistic is proposed for testing the significance test of parameter. The asymptotic studies show that, this test has correct size asymptotically and is consistent in power. In certain cases, this test can be used in particular to test the relevance of the exogenous covariates.
The model selection question for the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ is carried out by a penalized quasi likelihood contrast. The weak and the strong consistency of the proposed procedure is established. These results provide sufficient conditions for the consistency of the BIC and the HQC procedures. Simulation study shows that, the empirical and the theoretical results are overall in accordance.

An extension of this works is to address the inference, the significance test of parameter, the model selection problem for the class $\mathcal{AC}$-$X(M_{\theta^*}, f_{\theta^*})$ with a non Gaussian quasi likelihood. For instance, as pointed out by Kengne (2021) [18], the use of the Laplacian quasi likelihood will allow to reduce the order of moments imposed to the process. Other topics of a research project, are the change-point detection and the prediction question (see for instance Ing (2003) [15], Ing and Wei (2003, 2005) [16, 17]) for this class of models.

8. Proofs of the main results

To simplify the expressions, in the proofs of Theorems 2.1, 2.2 and 3.1, we will use the conditional Gaussian quasi-log-likelihood given by $L_n(\theta) = -\sum_{t=1}^{n} q_t(\theta)$ and $\widehat{L}_n(\theta) = -\sum_{t=1}^{n} \widehat{q}_t(\theta)$. Throughout the sequel, C denotes a positive constant whom value may differ from an inequality to another.

8.1. Proof of Proposition 1

We verify that the process $Z_t := (Y_t; X_t)$ satisfies the conditions required for the Theorem 3.1 in Doukhan and Wintenberger [6]. According to (1.1), for all $t \in \mathbb{Z}$,

$$\begin{aligned} Z_t &= \big(M_{\theta^*}(Y_{t-1}, \ldots; X_{t-1}, \ldots)\xi_t + f_{\theta^*}(Y_{t-1}, \ldots; X_{t-1}, \ldots);\, g(X_{t-1}, \ldots; \eta_t)\big) \\ &= F(Z_{t-1}, Z_{t-2}, \ldots; U_t), \end{aligned}$$

with $U_t = (\xi_t, \eta_t)$ and $F(\boldsymbol{z}; U_t) = \big(M_{\theta^*}(y_1, \ldots; x_1, \ldots)\xi_t + f_{\theta^*}(y_1, \ldots; x_1, \ldots);\, g(x_1, \ldots; \eta_t)\big)$ for all $\boldsymbol{z} = \big((y_k, x_k)\big)_{k\in\mathbb{N}} \in (\mathbb{R}^{d_x+1})^{\mathbb{N}}$. Thus, the equation (1.1) of [6] holds for $(Z_t)_{t\in\mathbb{Z}}$. For a vector $z = (y, x) \in \mathbb{R}^{d_x+1}$, define the norm $\|z\|_w = |y| + w_x\|x\|$ for some $w_x > 0$. According to Doukhan and Wintenberger (2008) [6], it suffices to show that:

(i) $\mathbb{E}\|F(\boldsymbol{z}; U_0)\|_w^r < \infty$ for some $\boldsymbol{z} \in (\mathbb{R}^{d_x+1})^{\mathbb{N}}$;

(ii) there exists a non-negative sequence $(\alpha_k(F))_{k\geq 1}$ satisfying $\sum_{k\geq 1}\alpha_k(F) < 1$ such that, for all $\boldsymbol{z}, \tilde{\boldsymbol{z}} \in (\mathbb{R}^{d_x+1})^{\mathbb{N}}$,

$$\mathbb{E}\|F(\boldsymbol{z};U_0) - F(\tilde{\boldsymbol{z}};U_0)\|_w^r \leq \sum_{k\geq 1}\alpha_k(F)\|z_k - \tilde{z}_k\|_w.$$

Using the condition (2.2), the part (i) is directly obtained from the assumptions $\mathbf{A}_0(f_\theta,\Theta)$, $\mathbf{A}_0(M_\theta,\Theta)$.
To prove (ii), let $\boldsymbol{z} = (z_1,\ldots)$, $\tilde{\boldsymbol{z}} = (\tilde{z}_1,\ldots) \in (\mathbb{R}^{d_x+1})^{\mathbb{N}}$ such that $z_k = (y_k, x_k)$ and $\tilde{z}_k = (\tilde{y}_k, \tilde{x}_k)$ for all $k \geq 1$. From $\mathbf{A}_0(f_\theta,\Theta)$, $\mathbf{A}_0(M_\theta,\Theta)$ and (2.2), we get

$$\begin{aligned}
&\|\|F(\boldsymbol{z};U_0) - F(\tilde{\boldsymbol{z}};U_0)\|_w\|_r \\
&\leq \Big\| \|(M_\theta(y_1,\ldots;x_1,\ldots) - M_\theta(\tilde{y}_1,\ldots;\tilde{x}_1,\ldots))\,\xi_0\|_\Theta \\
&\qquad\qquad + \|f_\theta(y_1,\ldots;x_1,\ldots) - f_\theta(\tilde{y}_1,\ldots;\tilde{x}_1,\ldots)\|_\Theta \Big\|_r \\
&\qquad\qquad + w_x \|g(x_1,\ldots;\eta_0) - g(\tilde{x}_1,\ldots;\eta_0)\|_r \\
&\leq \sum_{k=1}^{\infty}\Big(\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,Y}^{(0)}(M_\theta,\Theta)\Big)\,|y_k - \tilde{y}_k| \\
&\qquad\qquad + \sum_{k=1}^{\infty}\Big(\alpha_{k,X}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,X}^{(0)}(M_\theta,\Theta)\Big)\,\|x_k - \tilde{x}_k\| \\
&\qquad\qquad + w_x\sum_{k=1}^{\infty}\alpha_k(g)\,\|x_k - \tilde{x}_k\| \\
&\leq \sum_{k=1}^{\infty}\Big(\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,Y}^{(0)}(M_\theta,\Theta)\Big)\,|y_k - \tilde{y}_k| \\
&\qquad + w_x\sum_{k=1}^{\infty}\Big(\frac{1}{w_x}\big(\alpha_{k,X}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,X}^{(0)}(M_\theta,\Theta)\big) + \alpha_k(g)\Big)\|x_k - \tilde{x}_k\| \\
&\leq \sum_{k=1}^{\infty}\alpha_k(F)\|z_k - \tilde{z}_k\|_w
\end{aligned}$$

with

$$\begin{aligned}
\alpha_k(F) = \max\Big\{&\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,Y}^{(0)}(M_\theta,\Theta), \\
&\frac{1}{w_x}\big(\alpha_{k,X}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,X}^{(0)}(M_\theta,\Theta)\big) + \alpha_k(g)\Big\}.
\end{aligned}$$

Thus, to get $\sum_{k=1}^{\infty}\alpha_k(F) < 1$, it suffices to choose w_x sufficiently large, such that

$$w_x > \frac{\sum_{k\geq 1}\Big\{\alpha_{k,X}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,X}^{(0)}(M_\theta,\Theta)\Big\}}{1 - \sum_{k\geq 1}\max\Big\{\alpha_k(g),\, \alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \|\xi_0\|_r\alpha_{k,Y}^{(0)}(M_\theta,\Theta)\Big\}}.$$

This completes the proof of the proposition. ■

8.2. Proof of Theorem 2.1

We consider the following lemma.

Lemma 1. *Assume that the assumptions of Theorem 2.1 hold. Then*

$$\frac{1}{n}\big\|\widehat{L}_n(\theta) - L_n(\theta)\big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof of Lemma 1
Remark that

$$\frac{1}{n}\|\widehat{L}_n(\theta) - L_n(\theta)\|_\Theta \le \frac{1}{n}\sum_{t=1}^{n}\|\widehat{q}_t(\theta) - q_t(\theta)\|_\Theta.$$

Hence, by Corollary 1 of Kounias and Weng (1969) [19]., with $2 \le \tilde{r} \le \min\{3, r\}$ (without loss of generality), it suffices to show that

$$\sum_{\ell\ge 1}\frac{1}{\ell^{\tilde{r}/3}}\mathbb{E}\left(\|\widehat{q}_\ell(\theta) - q_\ell(\theta)\|_\Theta^{\tilde{r}/3}\right) < \infty. \tag{8.1}$$

For any $\theta \in \Theta$, by applying the mean value theorem at the functions $x \mapsto \frac{1}{x^2}$ and $x \mapsto \log x$, we have

$$\begin{aligned}
&|\widehat{q}_t(\theta) - q_t(\theta)| \\
&\le \Big|\frac{(Y_t - \widehat{f}_\theta^t)^2}{\widehat{H}_\theta^t} - \frac{(Y_t - f_\theta^t)^2}{H_\theta^t}\Big| + |\log \widehat{H}_\theta^t - \log H_\theta^t| \\
&\le \Big|\frac{(Y_t - \widehat{f}_\theta^t)^2}{(\widehat{M}_\theta^t)^2} - \frac{(Y_t - f_\theta^t)^2}{(M_\theta^t)^2}\Big| + 2\big|\log|\widehat{M}_\theta^t| - \log|M_\theta^t|\big| \\
&\le \Big|(Y_t - \widehat{f}_\theta^t)^2\Big(\frac{1}{(\widehat{M}_\theta^t)^2} - \frac{1}{(M_\theta^t)^2}\Big) + \frac{1}{(M_\theta^t)^2}\left((Y_t - \widehat{f}_\theta^t)^2 - (Y_t - f_\theta^t)^2\right)\Big| \\
&\qquad\qquad + \frac{2}{\underline{h}^{1/2}}|\widehat{M}_\theta^t - M_\theta^t| \\
&\le \frac{2}{\underline{h}^{3/2}}|Y_t - \widehat{f}_\theta^t|^2|\widehat{M}_\theta^t - M_\theta^t| + \frac{1}{\underline{h}}|\widehat{f}_\theta^t - f_\theta^t||\widehat{f}_\theta^t + f_\theta^t - 2Y_t| + \frac{2}{\underline{h}^{1/2}}|\widehat{M}_\theta^t - M_\theta^t| \\
&\le C\big((|Y_t - f_\theta^t|^2 + 1)|\widehat{M}_\theta^t - M_\theta^t| + |\widehat{f}_\theta^t - f_\theta^t||\widehat{f}_\theta^t + f_\theta^t - 2Y_t|\big).
\end{aligned}$$

This implies

$$\begin{aligned}
\mathbb{E}\big[\|\widehat{q}_t(\theta) - q_t(\theta)\|_\Theta^{\tilde{r}/3}\big] \le C\Big(&\mathbb{E}\Big[\big(\|Y_t - f_\theta^t\|_\Theta^2 + 1\big)^{\tilde{r}/3}\|\widehat{M}_\theta^t - M_\theta^t\|_\Theta^{\tilde{r}/3}\Big] \\
&+ \mathbb{E}\Big[\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^{\tilde{r}/3}\big(\|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta + 2|Y_t|\big)^{\tilde{r}/3}\Big]\Big).
\end{aligned}$$

Moreover, since $\theta^* \in \Theta(r)$ for some $r \geq 2$, by Assumption $\mathbf{A}_0(\Psi_\theta, \Theta)$, one can easily show that:

- $\mathbb{E}\left[|Y_t|^r + \|f_\theta^t\|_\Theta^r + \|\widehat{f}_\theta^t\|_\Theta^r + \|M_\theta^t\|_\Theta^r + \|\widehat{M}_\theta^t\|_\Theta^r + \|H_\theta^t\|_\Theta^{r/2} + \|\widehat{H}_\theta^t\|_\Theta^{r/2}\right] < \infty;$ (8.2)

- $$\begin{cases} \mathbb{E}(\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^r) \leq C\Big(\sum\limits_{k\geq t}\{\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \alpha_{k,X}^{(0)}(f_\theta,\Theta)\}\Big)^r; \\ \mathbb{E}(\|\widehat{M}_\theta^t - M_\theta^t\|_\Theta^r) \leq C\Big(\sum\limits_{k\geq t}\{\alpha_{k,Y}^{(0)}(M_\theta,\Theta) + \alpha_{k,X}^{(0)}(M_\theta,\Theta)\}\Big)^r. \end{cases} \quad (8.3)$$

Then, by the Hölder's inequality, we have

$$\begin{aligned} &\mathbb{E}\Big[\|\widehat{f}_\theta^\ell - f_\theta^\ell\|_\Theta^{\tilde{r}/3}\big(\|\widehat{f}_\theta^\ell\|_\Theta + \|f_\theta^\ell\|_\Theta + 2|Y_\ell|\big)^{\tilde{r}/3}\Big] \\ &\qquad \leq \Big(\mathbb{E}\big[\|\widehat{f}_\theta^\ell - f_\theta^\ell\|_\Theta^{\tilde{r}}\big]\Big)^{1/3}\Big(\mathbb{E}\big[\|\widehat{f}_\theta^\ell\|_\Theta + \|f_\theta^\ell\|_\Theta + 2|Y_\ell|\big]^{\tilde{r}/2}\Big)^{2/3} \\ &\qquad \leq C\Big(\sum_{k\geq t}\{\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \alpha_{k,X}^{(0)}(f_\theta,\Theta)\}\Big)^{\tilde{r}/3}. \end{aligned}$$

Again, by the Hölder's inequality, from (8.2) and (8.3), we obtain

$$\begin{aligned} &\mathbb{E}\Big[\big(\|Y_t - f_\theta^\ell\|_\Theta^2 + 1\big)^{\tilde{r}/3}\|\widehat{M}_\theta^\ell - M_\theta^\ell\|_\Theta^{\tilde{r}/3}\Big] \\ &\leq \mathbb{E}\Big[\|Y_t + f_\theta^\ell + 1\|_\Theta^{2\tilde{r}/3}\|\widehat{M}_\theta^\ell - M_\theta^\ell\|_\Theta^{\tilde{r}/3}\Big] \\ &\leq \big(\mathbb{E}\big[\|Y_t + f_\theta^\ell + 1\|_\Theta^{\tilde{r}}\big]\big)^{2/3}\big(\mathbb{E}\big[\|\widehat{M}_\theta^\ell - M_\theta^\ell\|_\Theta^{\tilde{r}}\big]\big)^{1/3} \\ &\leq C\Big(\sum_{k\geq t}\{\alpha_{k,Y}^{(0)}(M_\theta,\Theta) + \alpha_{k,X}^{(0)}(M_\theta,\Theta)\}\Big)^{\tilde{r}/3}. \end{aligned}$$

Hence, from (2.3), we deduce

$$\begin{aligned} &\sum_{\ell\geq 1}\frac{1}{\ell^{\tilde{r}/3}}\mathbb{E}\left(\|\widehat{q}_\ell(\theta) - q_\ell(\theta)\|_\Theta^{\tilde{r}/3}\right) \leq \\ &C\sum_{\ell\geq 1}\frac{1}{\ell^{\tilde{r}/3}}\Big(\sum_{k\geq \ell}\{\alpha_{k,Y}^{(0)}(f_\theta,\Theta) + \alpha_{k,X}^{(0)}(f_\theta,\Theta) + \alpha_{k,Y}^{(0)}(M_\theta,\Theta) + \alpha_{k,X}^{(0)}(M_\theta,\Theta)\}\Big)^{\tilde{r}/3} \\ &\leq C\sum_{\ell\geq 1}\frac{1}{\ell^{\tilde{r}/3}}\Big(\frac{1}{\ell^{\gamma-1}}\Big)^{\tilde{r}/3} \leq C\sum_{\ell\geq 1}\frac{1}{\ell^{\tilde{r}\gamma/3}} < \infty, \end{aligned}$$

where the last inequality holds since $\gamma > 3/2$. Thus, the condition (8.1) is satisfied. This completes the proof Lemma 1. ■

To complete the proof of Theorem 2.1, we will show that: (1.) $\mathbb{E}\left[\|q_t(\theta)\|_\Theta\right] < \infty$ and (2.) the function $\theta \mapsto -\mathbb{E}[q_0(\theta)]$ has a unique maximum at θ^*.

(1.) For all $\theta \in \Theta$, using the inequality $|\log(x)| \leq |x-1|$ for all $x > 1$, we have

$$\begin{aligned}|q_t(\theta)| &\leq \frac{1}{H_\theta^t}|Y_t - f_\theta^t)|^2 + \big|\log\big(\frac{H_\theta^t}{\underline{h}}\big) + \log(\underline{h})\big| \\ &\leq \frac{1}{\underline{h}}(Y_t^2 + (f_\theta^t)^2 + 2Y_t f_\theta^t) + \big|\frac{H_\theta^t}{\underline{h}} - 1\big| + |\log(\underline{h})| \\ &\leq C\left(Y_t^2 + (f_\theta^t)^2 + 2Y_t f_\theta^t + |M_\theta^t|^2\right) + C.\end{aligned}$$

Hence, from (8.2), we deduce

$$\begin{aligned}\mathbb{E}\left[\|q_t\|_\Theta\right] \leq C\Big(\mathbb{E}[Y_t^2] + \mathbb{E}\|f_\theta^t\|_\Theta^2 + 2\big((\mathbb{E}[Y_t^2])^{1/2}(\mathbb{E}[\|f_\theta^t\|_\Theta^2])^{1/2}\big) \\ + \mathbb{E}[\|M_\theta^t\|_\Theta^2]\Big) + C < \infty,\end{aligned}$$

which shows that (1.) holds.

(2.) Let $\theta \in \Theta$ with $\theta \neq \theta^*$. We have

$$\mathbb{E}[q_0(\theta)] - \mathbb{E}[q_0(\theta^*)] = \mathbb{E}\big[\mathbb{E}\left[(q_0(\theta) - q_0(\theta^*))\,|\mathcal{F}_{-1}\right]\big]. \tag{8.4}$$

Moreover,

$$\begin{aligned}&\mathbb{E}\left[(q_0(\theta) - q_0(\theta^*))\,|\mathcal{F}_{-1}\right] \\ &= \mathbb{E}\big[\frac{(Y_0 - f_\theta^0)^2}{H_\theta^0} + \log H_\theta^0 - \frac{(Y_0 - f_{\theta^*}^0)^2}{H_{\theta^*}^0} - \log H_{\theta^*}^0\,\big|\mathcal{F}_{-1}\big] \\ &= -\log\big(\frac{H_{\theta^*}^0}{H_\theta^0}\big) + \frac{\mathbb{E}\left[(Y_0 - f_\theta^0)^2\big|\mathcal{F}_{-1}\right]}{H_\theta^0} - \frac{\mathbb{E}\left[(Y_0 - f_{\theta^*}^0)^2\big|\mathcal{F}_{-1}\right]}{H_{\theta^*}^0} \\ &= -\log\big(\frac{H_{\theta^*}^0}{H_\theta^0}\big) - 1 + \frac{\mathbb{E}\left[(Y_0 - f_{\theta^*}^0 + f_{\theta^*}^0 - f_\theta^0)^2\big|\mathcal{F}_{-1}\right]}{H_\theta^0} \\ &= \frac{H_{\theta^*}^0}{H_\theta^0} - \log\big(\frac{H_{\theta^*}^0}{H_\theta^0}\big) - 1 + \frac{(f_{\theta^*}^0 - f_\theta^0)^2}{H_\theta^0}.\end{aligned}$$

Therefore, using (8.4) and by applying the Jensen's inequality, we get

$$\begin{aligned}\mathbb{E}[q_0(\theta)] - \mathbb{E}[q_0(\theta^*)] &= \mathbb{E}\big[\frac{H_{\theta^*}^0}{H_\theta^0} - \log\big(\frac{H_{\theta^*}^0}{H_\theta^0}\big) - 1 + \frac{(f_{\theta^*}^0 - f_\theta^0)^2}{H_\theta^0}\big] \\ &\geq \mathbb{E}\big[\frac{H_{\theta^*}^0}{H_\theta^0}\big] - \log\big(\mathbb{E}\big[\frac{H_{\theta^*}^0}{H_\theta^0}\big]\big) - 1 + \mathbb{E}\big[\frac{(f_{\theta^*}^0 - f_\theta^0)^2}{H_\theta^0}\big].\end{aligned}$$

Since $x - \log(x) - 1 > 0$ for any $x > 0,\ x \neq 1$; and $x - \log(x) - 1 = 0$ for $x = 1$, we deduce:

- if $f_{\theta^*}^0 \neq f_\theta^0$ *a.s.*, then $\mathbb{E}\big[\frac{(f_{\theta^*}^0 - f_\theta^0)^2}{H_\theta^0}\big] > 0$ and $\mathbb{E}[q_0(\theta)] - \mathbb{E}[q_0(\theta^*)] > 0$,
- if $f_{\theta^*}^0 = f_\theta^0$ *a.s.*, then

$$\mathbb{E}[q_0(\theta)] - \mathbb{E}[q_0(\theta^*)] = \mathbb{E}\big[\frac{H_{\theta^*}^0}{H_\theta^0} - \log\big(\frac{H_{\theta^*}^0}{H_\theta^0}\big) - 1\big]$$

From the identifiability condition (**A0**), when $\theta^* \neq \theta$ and $f^0_{\theta^*} = f^0_\theta$ *a.s.*, we necessarily have $H^0_{\theta^*} \neq H^0_\theta$ *a.s.*. This implies $\frac{H^0_{\theta^*}}{H^0_\theta} \neq 1$ *a.s.*, and thus $\mathbb{E}[q_0(\theta)] - \mathbb{E}[q_0(\theta^*)] > 0$.

The equality $\mathbb{E}[q_0(\theta)] = \mathbb{E}[q_0(\theta^*)]$ holds a.s. if and only if $\theta^* = \theta$. This achieves the proof of (2.).

Since $\{(Y_t, X_t),\, t \in \mathbb{Z}\}$ is stationary and ergodic, the process $\{q_t(\theta),\ t \in \mathbb{Z}\}$ is also a stationary and ergodic sequence. Then, according to (1.), by the uniform strong law of large number applied on the process $\{q_t(\theta),\ t \in \mathbb{Z}\}$, it holds that

$$\big\|\frac{1}{n}L_n(\theta) + \mathbb{E}(q_0(\theta))\big\|_\Theta = \big\|\frac{1}{n}\sum_{t=1}^{n} q_t(\theta) - \mathbb{E}(q_0(\theta))\big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

Then, by Lemma 1, we obtain

$$\big\|\frac{1}{n}\widehat{L}_n(\theta) + \mathbb{E}(q_0(\theta))\big\|_\Theta \leq \frac{1}{n}\big\|\widehat{L}_n(\theta) - L_n(\theta)\big\|_\Theta + \big\|\frac{1}{n}L_n(\theta) + \mathbb{E}(q_0(\theta))\big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0. \tag{8.5}$$

The part (2.) and (8.5) lead to conclude the proof of the theorem. ■

8.3. Proof of Theorem 2.2

The following lemma is needed.

Lemma 2. *Assume that the conditions of Theorem 2.2 hold. Then*

(i.) $\mathbb{E}\big[\frac{1}{\sqrt{n}}\big\|\frac{\partial \widehat{L}_n(\theta)}{\partial\theta} - \frac{\partial L_n(\theta)}{\partial\theta}\big\|_\Theta\big] \xrightarrow[n\to\infty]{} 0$;

(ii.) $\frac{1}{n}\big\|\frac{\partial^2 \widehat{L}_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}\big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0$;

(iii.) $\big\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} - \mathbb{E}\big(\frac{\partial^2 q_0(\theta)}{\partial\theta\partial\theta'}\big)\big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$

Proof of Lemma 2

(i.) Remark that

$$\big\|\frac{\partial \widehat{L}_n(\theta)}{\partial\theta} - \frac{\partial L_n(\theta)}{\partial\theta}\big\|_\Theta \leq \sum_{t=1}^{n}\big\|\frac{\partial \widehat{q}_t(\theta)}{\partial\theta} - \frac{\partial q_t(\theta)}{\partial\theta}\big\|_\Theta. \tag{8.6}$$

Moreover, for all $\theta \in \Theta$,

$$\begin{aligned}\frac{\partial q_t(\theta)}{\partial\theta} &= -(H^t_\theta)^{-2}\big(2H^t_\theta(Y_t - f^t_\theta)\frac{\partial f^t_\theta}{\partial\theta} + (Y_t - f^t_\theta)^2\frac{\partial H^t_\theta}{\partial\theta}\big) + (H^t_\theta)^{-1}\frac{H^t_\theta}{\partial\theta}\\ &= -2(H^t_\theta)^{-1}(Y_t - f^t_\theta)\frac{\partial f^t_\theta}{\partial\theta} + (Y_t - f^t_\theta)^2\frac{\partial (H^t_\theta)^{-1}}{\partial\theta} + (H^t_\theta)^{-1}\frac{\partial H^t_\theta}{\partial\theta},\end{aligned} \tag{8.7}$$

which implies

$$\Big|\frac{\partial \widehat{q}_t(\theta)}{\partial\theta}-\frac{\partial q_t(\theta)}{\partial\theta}\Big| \le 2\Big|(\widehat{H}_\theta^t)^{-1}(Y_t-\widehat{f}_\theta^t)\frac{\partial \widehat{f}_\theta^t}{\partial\theta}-(H_\theta^t)^{-1}(Y_t-f_\theta^t)\frac{\partial f_\theta^t}{\partial\theta}\Big| + \Big|(Y_t-\widehat{f}_\theta^t)^2\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-(Y_t-f_\theta^t)^2\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big| + \Big|(\widehat{H}_\theta^t)^{-1}\frac{\partial \widehat{H}_\theta^t}{\partial\theta}-(H_\theta^t)^{-1}\frac{\partial H_\theta^t}{\partial\theta}\Big|.$$

Using the relation $|a_1b_1c_1-a_2b_2c_2| \le |a_1-a_2||b_2||c_2|+|a_1||b_1-b_2||c_2|+|a_1||b_1||c_1-c_2|,\ \forall a_1,a_2,b_1,b_2,c_1,c_2,\in\mathbb{R}$, we get

$$\begin{aligned}
&\Big\|\frac{\partial \widehat{q}_t(\theta)}{\partial\theta}-\frac{\partial q_t(\theta)}{\partial\theta}\Big\|_\Theta \\
&\le 2\Big(\big\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\big\|_\Theta\|Y_t-\widehat{f}_\theta^t\|_\Theta\Big\|\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta \\
&+\big\|(\widehat{H}_\theta^t)^{-1}\big\|_\Theta\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}\Big\|_\Theta \\
&+\big\|(\widehat{H}_\theta^t)^{-1}\big\|_\Theta\|Y_t-f_\theta^t\|_\Theta\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}-\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta\Big) \\
&+\|(Y_t-\widehat{f}_\theta^t)\|_\Theta^2\Big\|\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta+2|Y_t|\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta\Big\|\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta \\
&+\big\|(\widehat{H}_\theta^t)^{-1}\big\|_\Theta\Big\|\frac{\partial \widehat{H}_\theta^t}{\partial\theta}-\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta+\big\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\big\|_\Theta\Big\|\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta \\
&\le 2(\underline{h})^{-1}\Big(\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}\Big\|_\Theta \\
&+\|Y_t-f_\theta^t\|_\Theta\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}-\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta+\frac{1}{2}\Big\|\frac{\partial \widehat{H}_\theta^t}{\partial\theta}-\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta\Big) \\
&+2\big\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\big\|_\Theta\|Y_t-\widehat{f}_\theta^t\|_\Theta\Big\|\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta \\
&+\|(Y_t-\widehat{f}_\theta^t)\|_\Theta^2\Big\|\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta \\
&+2|Y_t|\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta\Big\|\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta+\big\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\big\|_\Theta\Big\|\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta. \quad (8.8)
\end{aligned}$$

By applying the Hölder's inequality to the terms of the right hand side of (8.8), we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big\|\frac{\partial \widehat{q}_t(\theta)}{\partial\theta}-\frac{\partial q_t(\theta)}{\partial\theta}\Big\|_\Theta\Big] \\
&\le C\Bigg[\Big(\mathbb{E}[\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta^4]\Big)^{1/4}\Big(\mathbb{E}\Big[\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}\Big\|_\Theta^{4/3}\Big]\Big)^{3/4} \\
&+\big(\mathbb{E}[\|Y_t-f_\theta^t\|_\Theta^{4/3}]\big)^{3/4}\Big(\mathbb{E}\Big[\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial\theta}-\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta^4\Big]\Big)^{1/4}
\end{aligned}$$

$$
\begin{aligned}
&+\Big(\mathbb{E}\Big[\Big\|\frac{\partial\widehat{H}_\theta^t}{\partial\theta}-\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big[\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\|_\Theta^4\Big]\Big)^{1/4}\Big(\mathbb{E}[\|Y_t-\widehat{f}_\theta^t\|_\Theta^4]\Big)^{1/4}\Big(\mathbb{E}\Big[\Big\|\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}\\
&+\Big(\mathbb{E}[\|(Y_t-\widehat{f}_\theta^t)\|_\Theta^4]\Big)^{1/2}\Big(\mathbb{E}\Big[\Big\|\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}\\
&+\big(\mathbb{E}[|Y_t|^4]\big)^{1/4}\Big(\mathbb{E}[\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta^4]\Big)^{1/4}\times\Big(\mathbb{E}\Big[\Big\|\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}\\
&+\Big(\mathbb{E}\big[\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\|_\Theta^4\big]\Big)^{1/4}\Big(\mathbb{E}\Big[\Big\|\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta^{4/3}\Big]\Big)^{3/4}\Big].
\end{aligned}
$$

Moreover, since $\theta^*\in\Theta(r)$, using $\mathbf{A}_i(f_\theta,\Theta)$ and $\mathbf{A}_i(M_\theta,\Theta)$ (with $i=0,1$), one can go along similar lines as in [2] to establish the following results:

- $$\mathbb{E}\Big[\Big\|\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta^r+\Big\|\frac{\partial\widehat{f}_\theta^t}{\partial\theta}\Big\|_\Theta^r+\Big\|\frac{\partial M_\theta^t}{\partial\theta}\Big\|_\Theta^r+\Big\|\frac{\partial\widehat{M}_\theta^t}{\partial\theta}\Big\|_\Theta^r+\Big\|\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta^{r/2}+\Big\|\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta^{r/2}\Big]<\infty, \tag{8.9}$$

- $$\begin{cases}
\mathbb{E}\Big[\Big\|\frac{\partial\widehat{f}_\theta^t}{\partial\theta}-\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta^r\Big]\le C\Big(\sum\limits_{k\ge t}\{\alpha_{k,Y}^{(1)}(f_\theta,\Theta)+\alpha_{k,X}^{(1)}(f_\theta,\Theta)\}\Big)^r,\\
\mathbb{E}\Big[\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\|_\Theta^r\Big]\le C\Big(\sum\limits_{k\ge t}\{\alpha_{k,Y}^{(0)}(M_\theta,\Theta)+\alpha_{k,X}^{(0)}(M_\theta,\Theta)\}\Big)^r,\\
\mathbb{E}\Big[\Big\|\frac{\partial\widehat{H}_\theta^t}{\partial\theta}-\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta^{r/2}\Big]\le C\Big(\sum\limits_{k\ge t}\{\alpha_{k,Y}^{(0)}(M_\theta,\Theta)+\alpha_{k,X}^{(1)}(M_\theta,\Theta)\}\Big)^{r/2},\\
\mathbb{E}\Big\|\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta^{\frac{r}{2}}\le C\Big(\sum\limits_{k\ge t}\{\alpha_{k,Y}^{(0)}(M_\theta,\Theta)+\alpha_{k,X}^{(1)}(M_\theta,\Theta)\}\Big)^{\frac{r}{2}}.
\end{cases} \tag{8.10}$$

Thus, using (8.2), (8.3), (8.9) and (8.10) with $r=4$, we obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big\|\frac{\partial\widehat{q}_t(\theta)}{\partial\theta}-\frac{\partial q_t(\theta)}{\partial\theta}\Big\|_\Theta\Big]\\
&\le C\Big[\Big(\mathbb{E}[\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta^4]\Big)^{1/4}+\Big(\mathbb{E}\Big[\Big\|\frac{\partial\widehat{f}_\theta^t}{\partial\theta}-\frac{\partial f_\theta^t}{\partial\theta}\Big\|_\Theta^4\Big]\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big[\Big\|\frac{\partial\widehat{H}_\theta^t}{\partial\theta}-\frac{\partial H_\theta^t}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}+\Big(\mathbb{E}\Big[\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\|_\Theta^4\Big]\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big[\Big\|\frac{\partial(\widehat{H}_\theta^t)^{-1}}{\partial\theta}-\frac{\partial(H_\theta^t)^{-1}}{\partial\theta}\Big\|_\Theta^2\Big]\Big)^{1/2}+\Big(\mathbb{E}[\|\widehat{f}_\theta^t-f_\theta^t\|_\Theta^4]\Big)^{1/4}\\
&+\Big(\mathbb{E}\big[\|(\widehat{H}_\theta^t)^{-1}-(H_\theta^t)^{-1}\|_\Theta^4\big]\Big)^{1/4}\Big]
\end{aligned}
$$

$$\leq C\sum_{k\geq t}\Big\{\alpha^{(0)}_{k,Y}(f_\theta,\Theta)+\alpha^{(0)}_{k,X}(f_\theta,\Theta)+\alpha^{(1)}_{k,Y}(f_\theta,\Theta)+\alpha^{(1)}_{k,X}(f_\theta,\Theta)$$
$$+\alpha^{(0)}_{k,Y}(M_\theta,\Theta)+\alpha^{(0)}_{k,X}(M_\theta,\Theta)+\alpha^{(1)}_{k,Y}(M_\theta,\Theta)+\alpha^{(1)}_{k,X}(M_\theta,\Theta)\Big\}.$$

Therefore, in view of the condition (2.7), it holds that

$$\mathbb{E}\Big[\Big\|\frac{\partial\widehat{q}_t(\theta)}{\partial\theta}-\frac{\partial q_t(\theta)}{\partial\theta}\Big\|_\Theta\Big]\leq C\sum_{k\geq t}k^{-\gamma}=C\frac{1}{t^{\gamma-1}}.$$

By the inequality (8.6), we deduce

$$\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial\widehat{L}_n(\theta)}{\partial\theta}-\frac{\partial L_n(\theta)}{\partial\theta}\Big\|_\Theta\Big]\leq C\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{t^{\gamma-1}}=C\frac{1}{\sqrt{n}}(1+n^{2-\gamma})$$
$$\underset{n\to\infty}{\longrightarrow}0.$$

This proves the part (i.) of Lemma 2.

(ii.) This part can be established by using the same arguments as in the proof of Lemma 1.

(iii.) Let us show that $\mathbb{E}\Big[\Big\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta\Big]<\infty$, for all $i,j\in\{1,\ldots,d\}$. From (8.7), for any $i,j\in\{1,\ldots,d\}$, we have

$$\frac{\partial^2 q_t(\theta)}{\partial\theta\partial_i\theta_j}$$
$$=-2(H^t_\theta)^{-1}(Y_t-f^t_\theta)\frac{\partial^2 f^t_\theta}{\partial\theta_i\partial\theta_j}+(Y_t-f^t_\theta)^2\frac{\partial^2(H^t_\theta)^{-1}}{\partial\theta_i\partial\theta_j}$$
$$-2(Y_t-f^t_\theta)\Big(\frac{\partial f^t_\theta}{\partial\theta_i}\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_j}+\frac{\partial f^t_\theta}{\partial\theta_j}\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_i}\Big)$$
$$+2(H^t_\theta)^{-1}\frac{\partial f^t_\theta}{\partial\theta_i}\frac{\partial f^t_\theta}{\partial\theta_j}+\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_j}\frac{\partial H^t_\theta}{\partial\theta_i}+(H^t_\theta)^{-1}\frac{\partial^2 H^t_\theta}{\partial\theta_i\partial\theta_j}.$$

Therefore, according to (**A1**), we get

$$\Big\|\frac{\partial^2 q_t(\theta)}{\partial\theta\partial_i\theta_j}\Big\|_\Theta$$
$$\leq C\|(Y_t-f^t_\theta)\|_\Theta\times$$
$$\Big(\Big\|\frac{\partial^2 f^t_\theta}{\partial\theta_i\partial\theta_j}\Big\|_\Theta+\Big\|\frac{\partial f^t_\theta}{\partial\theta_i}\Big\|_\Theta\Big\|\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_j}\Big\|_\Theta+\Big\|\frac{\partial f^t_\theta}{\partial\theta_j}\Big\|_\Theta\Big\|\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_i}\Big\|_\Theta\Big)$$
$$+C\Big(\Big\|\frac{\partial f^t_\theta}{\partial\theta_i}\Big\|_\Theta\Big\|\frac{\partial f^t_\theta}{\partial\theta_j}\Big\|_\Theta+\Big\|\frac{\partial^2 H^t_\theta}{\partial\theta_i\partial\theta_j}\Big\|_\Theta\Big)$$
$$+\|(Y_t-f^t_\theta)\|^2_\Theta\Big\|\frac{\partial^2(H^t_\theta)^{-1}}{\partial\theta_i\partial\theta_j}\Big\|_\Theta+\Big\|\frac{\partial(H^t_\theta)^{-1}}{\partial\theta_j}\Big\|_\Theta\Big\|\frac{\partial H^t_\theta}{\partial\theta_i}\Big\|_\Theta. \tag{8.11}$$

Moreover, by $\mathbf{A}_2(f_\theta,\Theta)$ and $\mathbf{A}_2(M_\theta,\Theta)$, one can show that

$$\mathbb{E}\Big[\Big\|\frac{\partial^2 f_\theta^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^4+\Big\|\frac{\partial^2 H_\theta^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^2+\Big\|\frac{\partial^2 (H_\theta^t)^{-1}}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^2\Big]<\infty.$$

Thus, by applying the Hölder's inequality to the terms of the right hand side of (8.11), it suffices to use (8.2) and (8.9) to obtain $\mathbb{E}\Big[\Big\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta\Big]<\infty$.
Since $\mathbb{E}\Big[\Big\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta\Big]<\infty$ for all $i,j\in\{1,\ldots,d\}$, from the stationarity and ergodicity properties of $\big\{\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'},\ t\in\mathbb{Z}\big\}$ and the uniform strong law of large numbers, it holds that

$$\Big\|\frac{1}{n}\sum_{t=1}^n\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'}-\mathbb{E}\Big(\frac{\partial^2 q_0(\theta)}{\partial\theta\partial\theta'}\Big)\Big\|_\Theta\xrightarrow[n\to\infty]{a.s.}0.$$

This completes the proof of Lemma 2. ■

The following lemma is also needed.

Lemma 3. *Assume that the conditions of Theorem 2.2 hold. Then*

(i.) $\big\{\frac{\partial q_t(\theta^*)}{\partial\theta}|\mathcal{F}_{t-1},\ t\in\mathbb{Z}\big\}$ is a stationary ergodic martingale difference sequence with covariance matrix G,

(ii.) $-\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\widehat{L}_n(\tilde{\theta}_n)\xrightarrow[n\to\infty]{a.s.}F$, for any sequence $(\tilde{\theta}_n)_{n\geq1}$ with values in Θ and satisfying $\tilde{\theta}_n\xrightarrow[n\to\infty]{a.s.}\theta^*$,

where G and F are defined in (2.5).

Proof of Lemma 3

(i.) Recall that $G=\mathbb{E}\Big[\frac{\partial q_t(\theta^*)}{\partial\theta}\frac{\partial q_t(\theta^*)}{\partial\theta'}\Big]$ and that for all $\theta\in\Theta$,

$$\frac{\partial q_t(\theta)}{\partial\theta}=-2(H_\theta^t)^{-1}(Y_t-f_\theta^t)\frac{\partial f_\theta^t}{\partial\theta}-\Big(\frac{Y_t-f_\theta^t}{H_\theta^t}\Big)^2\frac{\partial H_\theta^t}{\partial\theta}+(H_\theta^t)^{-1}\frac{\partial H_\theta^t}{\partial\theta}.$$

Since the functions f_θ^t, H_θ^t, $\frac{\partial f_\theta^t}{\partial\theta}$ and $\frac{\partial H_\theta^t}{\partial\theta}$ are $\mathcal{F}_{t-1}$-measurable, we have

$$\mathbb{E}\Big[\frac{\partial q_t(\theta^*)}{\partial\theta}\big|\mathcal{F}_{t-1}\Big]=-(H_{\theta^*}^t)^{-1}\frac{\partial H_{\theta^*}^t}{\partial\theta}\left((H_{\theta^*}^t)^{-1}\mathbb{E}\left[(Y_t-f_{\theta^*}^t)^2|\mathcal{F}_{t-1}\right]-1\right)=0,$$

which shows that (i.) holds.

(ii.) Let $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ be a sequence satisfying $\tilde{\theta}_n\xrightarrow[n\to\infty]{a.s.}\theta^*$. For any $i,j=1,\ldots,d$, we have

$$\Big|\frac{1}{n}\sum_{t=1}^n\frac{\partial}{\partial\theta_j\partial\theta_i}q_t(\tilde{\theta}_n)-\mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\theta^*)\Big)\Big|$$

$$\leq \Big|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_j\partial\theta_i}q_t(\tilde{\theta}_n) - \mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\tilde{\theta}_n)\Big)\Big| + \Big|\mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\tilde{\theta}_n)\Big) - \mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\theta^*)\Big)\Big|$$

$$\leq \Big\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_j\partial\theta_i}q_t(\theta) - \mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\theta)\Big)\Big\|_{\Theta} + \Big|\mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\tilde{\theta}_n)\Big) - \mathbb{E}\Big(\frac{\partial}{\partial\theta_j\partial\theta_i}q_0(\theta^*)\Big)\Big|$$

$\underset{n\to\infty}{\longrightarrow} 0$ (by virtue of Lemma 2 (iii.)).

Thus,

$$-\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}L_n(\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}q_t(\tilde{\theta}_n) \underset{n\to\infty}{\overset{a.s.}{\to}} \mathbb{E}\Big(\frac{\partial^2}{\partial\theta\partial\theta'}q_0(\theta^*)\Big) = F.$$

We conclude the proof of the part (ii.) by using Lemma 2 (ii.). ■

Now, we use the results of Lemma 2 and 3 to prove the first part of Theorem 2.2. The second part can be established by going along similar lines as in Kengne (2021) [18].

By applying a second-order Taylor expansion to the function $\theta \mapsto \widehat{L}_n(\theta)$, for all $\theta \in \Theta$, there exists $\tilde{\theta}$ between θ and θ^* such that

$$\frac{1}{n}\Big\{\widehat{L}_n(\theta) - \widehat{L}_n(\theta^*)\Big\} = \frac{1}{n}\frac{\partial L_n(\theta^*)}{\partial\theta'}(\theta-\theta^*) - \frac{1}{2}(\theta-\theta^*)'F(\theta-\theta^*) + R_n(\theta), \quad (8.12)$$

where

$$R_n(\theta) = \frac{1}{n}\Big\{\frac{\partial\widehat{L}_n(\theta^*)}{\partial\theta'} - \frac{\partial L_n(\theta^*)}{\partial\theta'}\Big\}(\theta-\theta^*) + \frac{1}{2}(\theta-\theta^*)'\Big(\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\widehat{L}_n(\tilde{\theta}) + F\Big)(\theta-\theta^*).$$

Let us define the vector

$$Z_n = F^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta^*)}{\partial\theta}.$$

Then, we can rewrite (8.12) as

$$\frac{1}{n}\Big\{\widehat{L}_n(\theta) - \widehat{L}_n(\theta^*)\Big\} = \frac{1}{2n}\|Z_n\|_F^2 - \frac{1}{2n}\big\|Z_n - \sqrt{n}(\theta-\theta^*)\big\|_F^2 + R_n(\theta). \quad (8.13)$$

Define also

$$\theta_{Z_n} = \arg\inf_{\theta\in\Theta}\big\|Z_n - \sqrt{n}(\theta-\theta^*)\big\|_F.$$

Then, by (2.4), for n large enough, we have

$$\sqrt{n}(\theta_{Z_n} - \theta^*) = Z_n^{\mathcal{C}},$$

where $Z_n^{\mathcal{C}}$ is the F-projection of Z_n on $\mathcal{C}$. Using this relation and the definition of θ_{Z_n}, we have

$$\left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n - Z_n^{\mathcal{C}}\right\|_F^2 = \left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n - \sqrt{n}(\theta_{Z_n} - \theta^*)\right\|_F^2 \geq 0.$$

Furthermore, from (8.13) and the definition of $\widehat{\theta}_n$, it holds that

$$\begin{aligned}&\left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n - \sqrt{n}(\theta_{Z_n} - \theta^*)\right\|_F^2 \\ &= \{\widehat{L}_n(\theta_{Z_n}) - \widehat{L}_n(\widehat{\theta}_n)\} + 2n\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\} \\ &\leq 2n\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\}.\end{aligned}$$

Therefore,

$$\left|\left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n - Z_n^{\mathcal{C}}\right\|_F^2\right| \leq 2n\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\}. \tag{8.14}$$

Let us consider the following Lemma.

Lemma 4. *Assume that the conditions of Theorem 2.2 hold. Then*

$$n\{R_n(\widehat{\theta}_n) - R_n(\theta_{Z_n})\} = o_P(1).$$

By Lemma 4 and (8.14), it follows that

$$\left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n^{\mathcal{C}} - Z_n\right\|_F^2 = o_P(1). \tag{8.15}$$

Moreover, according to the equivalent definition of the F-orthogonal projection in (2.6), we get

$$\begin{aligned}&\left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 \\ &= \left\|Z_n^{\mathcal{C}} - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 + \left\|Z_n^{\mathcal{C}} - Z_n\right\|_F^2 - 2\left\langle Z_n^{\mathcal{C}} - \sqrt{n}(\widehat{\theta}_n - \theta^*), Z_n^{\mathcal{C}} - Z_n\right\rangle \\ &\geq \left\|Z_n^{\mathcal{C}} - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 + \left\|Z_n^{\mathcal{C}} - Z_n\right\|_F^2.\end{aligned}$$

Therefore, from (8.15), we obtain

$$\left\|Z_n^{\mathcal{C}} - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 \leq \left\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\right\|_F^2 - \left\|Z_n^{\mathcal{C}} - Z_n\right\|_F^2 = o_P(1). \tag{8.16}$$

Now, using Lemma 3 (i.), we apply the central limit theorem for the stationary ergodic martingale difference sequence $\left\{\frac{\partial q_t(\theta^*)}{\partial \theta}|\mathcal{F}_{t-1},\, t \in \mathbb{Z}\right\}$. It follows that

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta^*)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial q_t(\theta^*)}{\partial \theta} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_d\left(0, G\right), \tag{8.17}$$

and thus

$$Z_n = F^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta^*)}{\partial \theta} \xrightarrow[n\to+\infty]{\mathcal{L}} Z \sim \mathcal{N}_d\left(0, F^{-1}GF^{-1}\right). \tag{8.18}$$

Hence, $Z_n^{\mathcal{C}} \xrightarrow[n\to+\infty]{\mathcal{L}} Z^{\mathcal{C}}$. From this, it suffices to use (8.16) to conclude the proof of Theorem 2.2. ■

Proof of Lemma 4.
Recall that

$$R_n(\theta) = \frac{1}{n}\Big\{\frac{\partial \widehat{L}_n(\theta^*)}{\partial \theta'} - \frac{\partial L_n(\theta^*)}{\partial \theta'}\Big\}(\theta-\theta^*) + \frac{1}{2}(\theta-\theta^*)'\Big(\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\widehat{L}_n(\tilde{\theta}) + F\Big)(\theta-\theta^*).$$

According to Lemmas 2 (i.) and 3 (ii.), when $\tilde{\theta}_n - \theta^* = o_P(1)$, we have

$$nR_n(\tilde{\theta}_n) = o_P(\sqrt{n}(\tilde{\theta}_n - \theta^*)) + o_P(n\|\tilde{\theta}_n - \theta^*\|^2). \tag{8.19}$$

This implies

$$nR_n(\tilde{\theta}_n) = o_P(1) \quad \text{when} \quad \sqrt{n}(\tilde{\theta}_n - \theta^*) = O_P(1). \tag{8.20}$$

It comes from the definition of θ_{Z_n} that

$$\big\|\sqrt{n}(\theta_{Z_n} - \theta^*)\big\|_F \le \big\|\sqrt{n}(\theta_{Z_n} - \theta^*) - Z_n\big\|_F + \|Z_n\|_F \le 2\,\|Z_n\|_F\,.$$

Moreover, the convergence in (8.18) implies $\|Z_n\|_F = O_P(1)$; and consequently, $\sqrt{n}(\theta_{Z_n} - \theta^*) = O_P(1)$. Thus, $nR_n(\theta_{Z_n}) = o_P(1)$ by virtue (8.20).

We now show that, it also holds $nR_n(\widehat{\theta}_n) = o_P(1)$. From (8.13), we have

$$\|Z_n\|_F^2 - \big\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\big\|_F^2 + 2nR_n(\widehat{\theta}_n) = 2\{\widehat{L}_n(\widehat{\theta}_n) - \widehat{L}_n(\theta^*)\} \ge 0,$$

where the inequality holds since $\widehat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}}(\widehat{L}_n(\theta))$. Thus, it holds that

$$\begin{aligned}\big\|\sqrt{n}(\widehat{\theta}_n - \theta^*)\big\|_F^2 &\le 2\big(\big\|Z_n - \sqrt{n}(\widehat{\theta}_n - \theta^*)\big\|_F^2 + \|Z_n\|_F^2\big)\\ &\le 4\,\|Z_n\|_F^2 + 4nR_n(\widehat{\theta}_n).\end{aligned}$$

Furthermore, since $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^*$, by (8.19), it follows that $nR_n(\widehat{\theta}_n) = o_P\big(\big\|\sqrt{n}(\widehat{\theta}_n - \theta^*)\big\|_F^2\big)$. Consequently, $\sqrt{n}(\widehat{\theta}_n - \theta^*) = O_P(1)$, and $nR_n(\widehat{\theta}_n) = o_P(1)$ holds according to (8.20). This achieves the proof of the lemma. ■

8.4. Proof of Theorem 2.3

Under H_0, we have $\Gamma\widehat{\theta}_n - \vartheta_0 = \Gamma(\widehat{\theta}_n - \theta^*)$. Then, we get

$$\begin{aligned}W_n &= n(\Gamma\widehat{\theta}_n - \vartheta_0)'(\Gamma\widehat{\Sigma}_n\Gamma')^{-1}(\Gamma\widehat{\theta}_n - \vartheta_0)\\ &= n(\widehat{\theta}_n - \theta^*)'\Gamma'(\Gamma\widehat{\Sigma}_n\Gamma')^{-1}\Gamma(\widehat{\theta}_n - \theta^*)\\ &= \sqrt{n}(\widehat{\theta}_n - \theta^*)'\Gamma'(\Gamma\Sigma\Gamma')^{-1}\Gamma\sqrt{n}(\widehat{\theta}_n - \theta^*)\end{aligned}$$

$$+\sqrt{n}(\widehat{\theta}_n-\theta^*)'\Gamma'\left((\Gamma\widehat{\Sigma}_n\Gamma')^{-1}-(\Gamma\Sigma\Gamma')^{-1}\right)\Gamma\sqrt{n}(\widehat{\theta}_n-\theta^*). \tag{8.21}$$

Recall that, by Theorem 2.3, we have $\sqrt{n}(\widehat{\theta}_n-\theta^*) \xrightarrow[n\to+\infty]{\mathcal{L}} Z^{\mathcal{C}}$ with $Z\sim\mathcal{N}_d(0,\Sigma)$.
Furthermore, $(\Gamma\widehat{\Sigma}_n\Gamma')^{-1}-(\Gamma\Sigma\Gamma')^{-1}=o_P(1)$. Thus, from (8.21), it holds that

$$\begin{aligned}W_n&=\sqrt{n}(\widehat{\theta}_n-\theta^*)'\Gamma'(\Gamma\Sigma\Gamma')^{-1}\Gamma\sqrt{n}(\widehat{\theta}_n-\theta^*)+o_P(1)\\&\xrightarrow[n\to+\infty]{\mathcal{L}}(\Gamma Z^{\mathcal{C}})'(\Gamma\Sigma\Gamma')^{-1}\Gamma Z,\end{aligned}$$

which establishes the theorem. ■

Proof of Corollary 1.
When $\theta^*\in\overset{\circ}{\Theta}$, we have $W_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z'\Gamma'(\Gamma\Sigma\Gamma')^{-1}\Gamma Z=\|U\|^2$ with $U=(\Gamma\Sigma\Gamma')^{-1/2}\Gamma Z$ and $Z\sim\mathcal{N}_d(0,\Sigma)$. Since Σ is symmetric, the vector U follows a multivariate Gaussian distribution with mean 0 and covariance matrix I_{d_0}, where I_{d_0} is the identity matrix of size d_0. Therefore, all components of U are independent, standard normal distributed random variables. This leads to the conclusion. ■

8.5. Proof of Theorem 3.1

Consider the following lemma.

Lemma 5. *Assume that the conditions of Theorem 3.1 hold. Then*

$$\frac{1}{\sqrt{n\log\log n}}\left\|\frac{\partial\widehat{L}_n(\theta)}{\partial\theta}-\frac{\partial L_n(\theta)}{\partial\theta}\right\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof of Lemma 5.
Using the inequality (8.6) and Corollary 1 of [19], it suffices to show that

$$\sum_{k\geq 2}\frac{1}{\sqrt{k\log\log k}}\mathbb{E}\left[\left\|\frac{\partial\widehat{q}_k(\theta)}{\partial\theta}-\frac{\partial q_k(\theta)}{\partial\theta}\right\|_\Theta\right]<\infty. \tag{8.22}$$

In the proof of Lemma 3, we have established that

$$\begin{aligned}&\left\|\frac{\partial\widehat{q}_k(\theta)}{\partial\theta}-\frac{\partial q_k(\theta)}{\partial\theta}\right\|_\Theta\\&\leq C\sum_{j\geq k}\Big\{\alpha_{j,Y}^{(0)}(f_\theta,\Theta)+\alpha_{j,X}^{(0)}(f_\theta,\Theta)+\alpha_{j,Y}^{(1)}(f_\theta,\Theta)+\alpha_{k,X}^{(1)}(f_\theta,\Theta)\\&\qquad+\alpha_{j,Y}^{(0)}(M_\theta,\Theta)+\alpha_{j,X}^{(0)}(M_\theta,\Theta)+\alpha_{j,Y}^{(1)}(M_\theta,\Theta)+\alpha_{j,X}^{(1)}(M_\theta,\Theta)\Big\}\\&=C\sum_{j\geq k}\sum_{i=0}^{1}\left\{\alpha_{j,Y}^{(i)}(f_\theta,\Theta)+\alpha_{j,X}^{(i)}(f_\theta,\Theta)+\alpha_{j,Y}^{(i)}(M_\theta,\Theta)+\alpha_{j,X}^{(i)}(M_\theta,\Theta)\right\}.\end{aligned}$$

Then, from the condition (3.4), we obtain

$$\sum_{k\geq 2}\frac{1}{\sqrt{k\log\log k}}\mathbb{E}\left[\left\|\frac{\partial\widehat{q}_k(\theta)}{\partial\theta}-\frac{\partial q_k(\theta)}{\partial\theta}\right\|_{\Theta}\right]$$
$$\leq\sum_{k\geq 2}\frac{1}{\sqrt{k\log\log k}}\sum_{j\geq k}\sum_{i=0}^{1}\left\{\alpha_{j,Y}^{(i)}(f_\theta,\Theta)+\alpha_{j,X}^{(i)}(f_\theta,\Theta)+\alpha_{j,Y}^{(i)}(M_\theta,\Theta)\right.$$
$$\left.+\alpha_{j,X}^{(i)}(M_\theta,\Theta)\right\}<\infty.$$

Hence, (8.22) is satisfied, and Lemma 5 holds. ■

Let us prove the part (i.) of the theorem.

(i.) We have

$$P(\widehat{m}_n=m^*)=1-P(\widehat{m}_n\supsetneq m^*)-P(\widehat{m}_n\not\supseteq m^*).$$

Therefore, it suffices to show that

$$\lim_{n\to\infty}P(\widehat{m}_n\supsetneq m^*)=\lim_{n\to\infty}P(\widehat{m}_n\not\supseteq m^*)=0. \tag{8.23}$$

1. Let $m\in\mathcal{M}$ such as $m\supsetneq m^*$. We have,

$$\frac{1}{\sqrt{\log\log n}}\big(\widehat{C}(m^*)-\widehat{C}(m)\big)$$
$$=\frac{2}{\sqrt{\log\log n}}\big(\widehat{L}_n\big(\widehat{\theta}(m)\big)-\widehat{L}_n\big(\widehat{\theta}(m^*)\big)-\frac{\kappa_n}{\sqrt{\log\log n}}(|m|-|m^*|). \tag{8.24}$$

Let us establish that

$$\frac{1}{\sqrt{\log\log n}}\big(\widehat{L}_n\big(\widehat{\theta}(m)\big)-\widehat{L}_n\big(\widehat{\theta}(m^*)\big)\big)=O_P(1). \tag{8.25}$$

From the Taylor expansion of $\widehat{L}_n$, we can find $\overline{\theta}(m)$ between $\widehat{\theta}(m)$ and θ^* such that

$$\widehat{L}_n\big(\widehat{\theta}(m)\big)-\widehat{L}_n\big(\theta^*\big)=\frac{\partial L_n(\theta^*)}{\partial\theta}\big(\widehat{\theta}(m)-\theta^*\big)$$
$$-\frac{1}{2}\sqrt{n}\big(\widehat{\theta}(m)-\theta^*\big)'F(\theta^*,m)\sqrt{n}\big(\widehat{\theta}(m)-\theta^*\big)+nR_n'(m), \tag{8.26}$$

where

$$R_n'(m)=\frac{1}{n}\Big\{\frac{\partial\widehat{L}_n(\theta^*)}{\partial\theta'}-\frac{\partial L_n(\theta^*)}{\partial\theta'}\Big\}\big(\widehat{\theta}(m)-\theta^*\big)$$
$$+\frac{1}{2}\big(\widehat{\theta}(m)-\theta^*\big)'\Big(\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\widehat{L}_n(\overline{\theta}(m))+F(\theta^*,m)\Big)\big(\widehat{\theta}(m)-\theta^*\big)$$

and

$$F(\theta^*, m) = \Big(\mathbb{E}\Big[\frac{\partial^2 q_0(\theta^*)}{\partial\theta_i\partial\theta_j}\Big]\Big)_{i,j\in m}.$$

Moreover, since $\widehat{\theta}(m)$, $\overline{\theta}(m) \xrightarrow[n\to\infty]{a.s.} \theta^*$, in this case of overfitting, the same arguments as in the proof of Lemma 3 (ii.) lead to

$$-\frac{1}{n}\frac{\partial^2 \widehat{L}_n\big(\overline{\theta}(m)\big)}{\partial\theta\partial\theta'} \xrightarrow[n\to\infty]{a.s.} F(\theta^*, m).$$

Then, one can show as in the proof of Theorem 2.2 that $nR'_n(m) = o_P(1)$. Also, we have $\sqrt{n}\big(\widehat{\theta}(m) - \theta^*\big) = O_P(1)$. In addition, $\big\{\frac{\partial q_t(\theta^*)}{\partial\theta}|\mathcal{F}_{t-1},\, t\in\mathbb{Z}\big\}$ is a stationary ergodic square integrable martingale difference sequence (see above). Hence, from the law of iterative logarithm for martingales (see for instance [27, 28]), we get,

$$\frac{1}{\sqrt{n\log\log n}}\frac{\partial L_n(\theta^*)}{\partial\theta} = O(1).$$

Thus, we have from (8.26),

$$\begin{aligned}
&\frac{1}{\sqrt{\log\log n}}\big(\widehat{L}_n\big(\widehat{\theta}(m)\big) - \widehat{L}_n\big(\theta^*\big)\big)\\
&= \frac{1}{\sqrt{n\log\log n}}\frac{\partial L_n(\theta^*)}{\partial\theta}\sqrt{n}\big(\widehat{\theta}(m) - \theta^*\big)\\
&+ \frac{1}{2\sqrt{\log\log n}}\sqrt{n}\big(\widehat{\theta}(m) - \theta^*\big)' F(\theta^*, m)\sqrt{n}\big(\widehat{\theta}(m) - \theta^*\big)\\
&\qquad\qquad + \frac{1}{\sqrt{\log\log n}} nR'_n(m)\\
&= O(1)O_P(1) + o(1)O_P(1)O_P(1) + o_P(1) = O_P(1). \qquad (8.27)
\end{aligned}$$

By using the same arguments with $m = m^*$, we get

$$\frac{1}{\sqrt{\log\log n}}\big(\widehat{L}_n\big(\widehat{\theta}(m^*)\big) - \widehat{L}_n\big(\theta^*\big)\big) = O_P(1). \qquad (8.28)$$

Hence, (8.25) holds from (8.27) and (8.28).

Therefore, since $\kappa_n/\sqrt{\log\log n} \xrightarrow[n\to\infty]{} \infty$ and $|m| > |m^*|$, then (8.24) and (8.25) lead to

$$\frac{1}{\sqrt{\log\log n}}\big(\widehat{C}(m^*) - \widehat{C}(m)\big) \xrightarrow[n\to\infty]{\mathcal{P}} -\infty \ .$$

This implies that, for large n,

$$\widehat{C}(m) - \widehat{C}(m^*) > 0$$

with probability one; that is, $P(\widehat{m}_n \supsetneq m^*) \underset{n\to\infty}{\longrightarrow} 0$.

2. Let $m \in \mathcal{M}$ such as $m \not\supseteq m^*$. We have,

$$\frac{1}{n}\big(\widehat{C}(m^*)-\widehat{C}(m)\big) = \frac{2}{n}\big(\widehat{L}_n\big(\widehat{\theta}(m)\big)-\widehat{L}_n\big(\widehat{\theta}(m^*)\big)\big)-\frac{\kappa_n}{n}(|m|-|m^*|). \quad (8.29)$$

Using the same arguments in the proof of Theorem 3.1 of Bardet *et al.* (2020) [3], we get

$$\frac{1}{n}\big(\widehat{L}_n\big(\widehat{\theta}(m)\big) - \widehat{L}_n\big(\widehat{\theta}(m^*)\big)\big) = L(\theta^*(m)) - L(\theta^*) + o(1) \quad a.s.,$$

where $L(\theta) = -\mathbb{E}[q_0(\theta)]$, for all $\theta \in \Theta$. Note that, the function $L : \Theta \to \mathbb{R}$ has a unique maximum at θ^* (see the proof of Theorem 2.1). Since $m \not\supseteq m^*$, it holds that $\theta^* \notin \Theta(m)$; and consequently, $L(\theta^*(m))-L(\theta^*) < 0$ $a.s.$ Thus, according to (8.29) and since $\kappa_n/n \underset{n\to\infty}{\longrightarrow} 0$, we get

$$\lim_{n\to\infty}\frac{1}{n}\big(\widehat{C}(m^*)-\widehat{C}(m)\big)<0 \quad a.s. \text{ and } \quad \widehat{C}(m)-\widehat{C}(m^*)>0 \quad a.s. \text{ for large } n.$$

This implies that $P(\widehat{m}_n \not\supseteq m^*) \underset{n\to\infty}{\longrightarrow} 0$. Hence, the condition (8.23) holds; and the part (i.) of the theorem is established.

(ii.) Let $m \in \mathcal{M}$ such as $m \supsetneq m^*$. We have

$$\frac{1}{\log\log n}\big(\widehat{C}(m) - \widehat{C}(m^*)\big) = \frac{2}{\log\log n}\big(\widehat{L}_n\big(\widehat{\theta}(m^*)\big) - \widehat{L}_n\big(\widehat{\theta}(m)\big)\big) + \frac{\kappa_n}{\log\log n}(|m| - |m^*|).$$

Moreover, from the same arguments as in the proof of Theorem 3.1 in [18], one can show that

$$\frac{1}{\log\log n}\big(\widehat{L}_n\big(\widehat{\theta}(m^*)\big) - \widehat{L}_n\big(\widehat{\theta}(m)\big)\big) = O(1) \quad a.s.$$

Thus, we can find a constant c such that if $\liminf_{n\to\infty}\kappa_n/\log\log n > c$, then

$$\liminf_{n\to\infty}\frac{1}{\log\log n}\big(\widehat{C}(m) - \widehat{C}(m^*)\big) > 0 \quad a.s.$$

This implies that

$$\widehat{C}(m) - \widehat{C}(m^*) > 0 \quad a.s. \quad \text{for large } n. \quad (8.30)$$

Note that, the inequality (8.30) also holds when $m \not\supseteq m^*$ (see the part 2. of the proof of (i.)). Hence, we deduce that $\widehat{m}_n = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\, \widehat{C}(m) = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\big(\widehat{C}(m) - \widehat{C}(m^*)\big) \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} m^*$; which establishes the strong consistency of $\widehat{m}_n$.

(iii.) Using Lemma 5, this part can be proved by going along similar lines as in [18]. ■

Acknowledgments

The authors are very grateful to the Editor, the Associate Editor and the anonymous Referee for many relevant suggestions and comments which helped to improve the contents of this article.

References

[1] Bardet, J. M., Kengne, K. and Wintenberger, O. Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics 6*, (2012), 435-477. MR2988415

[2] Bardet, J.-M. and Wintenberger, O. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics 37, 5B*, (2009), 2730-2759. MR2541445

[3] Bardet, J.-M., Kamila, K. and Kengne, W. Consistent model selection criteria and goodness-of-fit test for common time series models. *Electron. J. Stat. 14*, (2020), 2009-2052. MR4091861

[4] Bierens, H. J. Topics in advanced econometrics: estimation, testing, and specification of cross-section and time series models. *Cambridge University Press*, (1996). MR1291390

[5] Deistler, M. The properties of the parameterization of ARMAX systems and their relevance for structural estimation and dynamic specification. *Econometrica: Journal of the Econometric Society*, (1983), 1187-1207. MR0710226

[6] Doukhan, P. and Wintenberger, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl. 118*, (2008), 1997-2013. MR2462284

[7] Francq, C. and Sucarrat, G. An equation-by-equation estimator of a multivariate log-GARCH-X model of financial returns. *Journal of Multivariate Analysis, 153*, (2017), 16-32. MR3578836

[8] Francq, C. and Thieu, L.Q. QML inference for volatility models with covariates. *Econometric Theory, 35*, (2019), 37-72. MR3904171

[9] Grønneberg, S. and Holcblat, B. On partial-sum processes of ARMAX residuals. *Annals of Statistics 47(6)*, (2019), 3216–3243. MR4025740

[10] Guo, S., Ling S. and Zhu, K. Factor double autoregressive models with application to simultaneous causality testing. *Journal of Statistical Planning and Inference 148*, (2014), 82-94. MR3174149

[11] Han, H. and Kristensen, D. Asymptotic Theory for the QMLE in GARCHX Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics 32*, (2014), 416-429. MR3238595

[12] Han, H. Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics, 13*, (2015), 188-221

[13] Hannan, E.J. The identification and parametrization of ARMAX and state space forms. *Econometrica: Journal of the Econometric Society*, (1976), 713–723. MR0479509

[14] Hannan, E. J. and Deistler, M. The statistical theory of linear systems. *SIAM*, (2012). MR3397291

[15] Ing, C.-K. Multistep prediction in autoregressive processes. *Econometric theory, 19(02)*, (2003), 254-279. MR1966030

[16] Ing, C.-K. and Wei, C.-Z. On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis, 85(01)*, (2003), 130-155. MR1978181

[17] Ing, C.-K. and Wei, C.-Z. Order selection for same-realization predictions in autoregressive processes. *The Annals of Statistics, 33(05)*, (2005), 2423-2474. MR2211091

[18] Kengne, W. Strongly consistent model selection for general causal time series. *Statistics and Probability Letters*, (2021). MR4207366

[19] Kounias, E.G. and Weng, T.-S. An inequality and almost sure convergence. *Annals of Mathematical Statistics 33*, (1969), 1091-1093. MR0245058

[20] Ling, S. and McAleer, M. Asymptotic theory for a vector ARMA-GARCH model. *Econometric theory 19(02)* (2003), 280-310. MR1966031

[21] Ling, S. Adouble AR(p) model: structure and estimation. *Statist. Sinica 17*, (2007), 161-175. MR2352507

[22] Nana, G.N., Korn, R. and Elwein-Sayer, C. GARCH-extended models: theoretical properties and applications. *arXiv:*1307.6685v1, (2013).

[23] Ng, K. Y. Statistical Modelling For Forecasting PM_{10} Concentrations In Peninsular Malaysia. *PhD Thesis, Universiti Sains Malaysia*, (2017).

[24] Ng, K. Y. and Awang, N. Multiple linear regression and regression with time series error models in forecasting PM 10 concentrations in Peninsular Malaysia. *Environmental monitoring and assessment 190(2)*, (2018), 1-11.

[25] Pedersen, R.S. and Rahbek, A. Testing Garch-X Type Models. *Econometric Theory 35(5)*, (2018), 1-36. MR4010504

[26] Souza, J. B., Reisen, V. A., Franco, G. C., Ispány, M., Bondon, P., and Santos, J. M. Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 67*, 2 (2018), 453–480. MR3758775

[27] Stout, W. F. The Hartman-Wintner law of the iterated logarithm for martingales. *The Annals of Mathematical Statistics 41* (1970), 2158-2160.

[28] Stout, W. F. Almost sure convergence. *Academic press* (1974). MR0455094

[29] Sucarrat, Genaro and Grønneberg, S. and Escribano, A. Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown. *Computational Statistics & Data Analysis 100* (2016), 582-594. MR3505819

[30] Tong, H. Non-linear time series: a dynamical system approach. *Oxford University Press* (1990). MR1079320

[31] Zambom, A. Z. and Gel, Y. R. Testing for local covariate trend effects in volatility models. *Electronic Journal of Statistics 14(2)* (2020), 2529-2550. MR4121798