A multivariate ARCH(∞) model with exogenous variables and dynamic conditional betas

Christian Francq*, Julien Royer†, Jean-Michel Zakoïan‡

Abstract

Factor models are highly common in the financial literature. Recent advances allow to relax the constancy of slope coefficients (the so-called betas) by considering conditional regressions. The theory on the estimation of these dynamic conditional betas however usually relies on short memory volatility models, which can be restrictive in empirical applications. Moreover, exogenous variables have proven useful in recent studies on volatility modeling. In this paper, we introduce a multivariate framework allowing for time-varying betas in which covolatilities can exhibit higher persistence than the standard exponential decay. Covariates are included in the dynamics of both conditional variances and betas. We establish stationarity conditions for the proposed model and prove the consistency and asymptotic normality of the QML estimator. Monte Carlo experiments are conducted to assess the performance of the estimation procedure in finite sample. Finally, we discuss the choice of potential relevant exogenous variables and illustrate the pertinence of the model on real data applications.

Keywords: Dynamic factors model, Long memory, Conditional betas

PRELIMINARY WORK, DO NOT CITE

*CREST, ENSAE and University Lille 3, BP60149, 59653 Villeneuve d'Ascq cedex, France; E-mail: christian.francq@univ-lille3.fr

†CREST, ENSAE and Institut Polytechnique de Paris, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France; E-mail: julien.royer@ensae.fr

‡CREST, ENSAE and University Lille 3, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France; E-mail: zakoian@ensae.fr

1 Introduction

An important area of the financial literature rests upon factor models. In static factor models, the regression coefficients are assumed to be constant over time. However, the justification for such parameters constancy are scarce and this assumption is often challenged in empirical studies. As wrongfully assuming betas constancy may lead to false financial conclusions, it is crucial to develop models accounting for their dynamic behavior. Time-varying parameter regressions were first considered through regime-switching models by Goldfeld and Quandt (1973) and Hamilton (1989) and more recently extended to the case of endogenous switching by Kim et al. (2008). However, in such models, the slope coefficients is considered constant in each state which can be restrictive when considering regimes with high duration. As an alternative, Shanken (1990) and Ferson and Harvey (1991) considered using financial or economics instrumental variables to recover time-varying betas. The obtained factor loadings are however highly sensitive to the choice of instrumental variables as noted by Ghysels (1998). In addition, Ang and Kristensen (2012) consider a family of deterministic paths for the regression coefficients while Hansen et al. (2014) considered a multivariate GARCH specification where asset returns are modeled through their conditional betas on the market that is assumed to follow a GARCH equation. Although the latter takes advantage of intraday information by using realized measures of volatilities and correlations, the beta is not directly modeled.

More recently, Engle (2016) proposed to recover the time-varying conditional betas indirectly from the conditional covariance matrix modeled using a dynamic conditional correlation specification. Building upon the seminal idea of Bollerslev et al. (1988) that conditional beta can be computed as the ratio of the conditional covariance to the conditional variance, introduce a general multivariate GARCH framework with dynamic betas. Although easy to implement, the model presents some drawbacks. First, the multivariate GARCH structure imposes an exponential decay of volatility shocks that might be problematic for empirical applications. Indeed, numerous studies have documented the high persistence of volatility in stock returns, see e.g. Ding and Granger (1996) or Conrad et al. (2011). More recently, Royer (2022) shows that peripheral assets, such as equity indices from emerging financial markets, exhibit higher persistence than allowed in standard GARCH models. In addition, the model proposed by Engle (2016) lacks tractability as the dynamics of slope coefficients are not directly specified, rendering statistical testing and financial interpretability difficult.

In this paper, we propose a new multivariate ARCH(∞) model that allows for a direct specification of the dynamic conditional betas. Theoretical analysis of multivariate long-memory volatility models have been scarce and to our knowledge,

none considers the case of time-varying regression models. Our model builds on the Cholesky-GARCH (CHAR) model of Darolles et al. (2018), relaxing its inherent short memory feature. Moreover, we allow the inclusion of exogenous variables in the dynamics of both conditional variances and betas. The inclusion of exogenous variables, and more precisely the inclusion of volatility measures computed intraday data, have proven useful in modeling individual GARCH (see e.g. Francq and Thieu (2019)) and multivariate GARCH models (Bollerslev et al. (2020b)). Our work is organized as follows. In Section 2, we present the Cholesky-ARCH-X(∞) model. We establish conditions for the existence of a stationary solution. In Section 3, we focus on statistical inference. We prove the strong consistency and asymptotic normality of the quasi maximum likelihood estimator (QMLE). Finally, Section 4 presents a preliminary financial application to illustrate our model.

2 ARCH-X(∞) extension of the CHAR model

Directly modeling betas dynamics provides numerous advantages as it allows to gain economic and financial interpretability on the time-varying dependency between assets or factors. To that extend, Darolles et al. (2018) proposed a method based on a Cholesky-GARCH (CHAR) model to obtain a direct specification of the conditional betas. More precisely, let $\boldsymbol{\varepsilon_t} = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})$ a vector of $m \geq 2$ return series verifying a general multivariate volatility model

$$\boldsymbol{\varepsilon_t} = \boldsymbol{\Sigma}_t^{1/2}\, \boldsymbol{\eta}_t \tag{1}$$

where $(\boldsymbol{\eta}_t)$ is iid $(\mathbf{0}, \boldsymbol{I}_m)$. Pourahmadi (1999) Cholesky decomposition of the covariance matrix yields

$$\boldsymbol{\Sigma}_t = \boldsymbol{L}_t \boldsymbol{G}_t \boldsymbol{L}_t' \tag{2}$$

where $\boldsymbol{G}_t = \mathrm{diag}(\boldsymbol{g}_t)$ is a diagonal matrix, and $\boldsymbol{L}_t$ is a lower unitriangular matrix. Define $\boldsymbol{v}_t = (v_{1,t}, \ldots, v_{m,t})$ with $v_{k,t} = \sqrt{g_{k,t}}\eta_{k,t}$ for $k = 1, \ldots, m$, be the recursively obtained orthogonal factors from $\boldsymbol{\varepsilon_t}$. By taking $\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{L}_t \boldsymbol{G}_t^{1/2}$, we obtain

$$\boldsymbol{\varepsilon_t} = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta_t} = \boldsymbol{L}_t \boldsymbol{v}_t$$

and thus $\boldsymbol{B}_t \boldsymbol{\varepsilon_t} = \boldsymbol{v}_t$ where $\boldsymbol{B}_t = \boldsymbol{L}_t^{-1}$. Let $\ell_{ij,t}$ (resp. $-\beta_{ij,t}$) denote the row i and column j element of $\boldsymbol{L}_t$ (resp. $\boldsymbol{B}_t$). Thus, the CHAR model allows a direct joint modeling of the dynamic conditional betas and the variance of the conditional orthogonal factors $v_{k,t}$.

Theoretical results for the CHAR model have been established under the assumption that the conditional volatilities $\boldsymbol{g}_t$ are given by a multivariate GARCH equation, such as the Extended Constant Conditional Correlation GARCH model of

Jeantheau (1998) where

$$\boldsymbol{g}_t = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{A}_i \boldsymbol{v}_{t-i}^2 + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{g}_{t-j} \tag{3}$$

with $\boldsymbol{v}_t^2 = (v_{1,t}^2, \ldots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ have nonnegative entries[1].

The CHAR model, as it is based on a multivariate GARCH equation, cannot account for the strong persistence of some financial returns series. In addition, although the betas dynamics are explicit, they do not exploit additional information than the one captured in the assets time series. Our proposed model alleviates the short-memory feature of model (3) and allows for exogenous variables. We consider a multivariate ARCH(∞) representation of the form

$$\boldsymbol{g}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \boldsymbol{A}_i \boldsymbol{v}_{t-i}^2 + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \boldsymbol{X}_{t-i} \tag{4}$$

where $\boldsymbol{v}_t^2 = (v_{1,t}^2, \ldots, v_{m,t}^2)$, and the vector $\boldsymbol{\omega} \in \mathbb{R}^m$ and the $m \times m$ matrices $\boldsymbol{A}_i$ and $m \times P$ matrices $\boldsymbol{\Pi}_i$, $i \in \{1, 2, \ldots\}$ have nonnegative entries. The vector $\boldsymbol{X}_t$ contains P positive exogenous variables. We assume that $(\boldsymbol{\eta}_t', \boldsymbol{X}_t')'$ is stationary and ergodic, with $\boldsymbol{\eta}_t$ independent of $\{\boldsymbol{X}_{t-i}, i > 0\}$. Note that Model (3) actually admits a representation of the form (4) as shown by Conrad and Karanasos (2010), however each term in matrices $\boldsymbol{A}_i$ is exponentially decaying as the ARCH(∞) kernel of a GARCH model.

In addition, the conditional betas dynamic is of the form

$$\boldsymbol{\beta}_t = \boldsymbol{c}(\boldsymbol{v}_{t-1}, \ldots, \boldsymbol{v}_{t-q}, \boldsymbol{g}_{t-1}^{1/2}, \ldots, \boldsymbol{g}_{t-q}^{1/2}, \boldsymbol{X}_{t-1}, \ldots, \boldsymbol{X}_{t-q}) + \sum_{k=1}^{p} \boldsymbol{C}_k \boldsymbol{\beta}_{t-k}, \tag{5}$$

for two integers p and q, where $\boldsymbol{c}$ is any measurable function from $\mathbb{R}^{(2m+P)q}$ to $\mathbb{R}^{m(m-1)/2}$. Alternatively, one can assume a similar dynamic model for the coefficients of the matrix $\boldsymbol{L}_t$, as

$$\boldsymbol{\ell}_t = \boldsymbol{c}(\boldsymbol{v}_{t-1}, \ldots, \boldsymbol{v}_{t-q}, \boldsymbol{g}_{t-1}^{1/2}, \ldots, \boldsymbol{g}_{t-q}^{1/2}, \boldsymbol{X}_{t-1}, \ldots, \boldsymbol{X}_{t-q}) + \sum_{k=1}^{p} \boldsymbol{C}_k \boldsymbol{\ell}_{t-k}. \tag{6}$$

[1] Conrad and Karanasos (2010) relax the hypothesis of nonnegativity for the components of the matrices $\boldsymbol{B}_j$ under the condition that the terms in the obtained ARCH(∞) kernel remain positive.

2.1 Stationarity conditions

We first note that (4) can be equivalently written as

$$\boldsymbol{g}_t = \boldsymbol{\omega}_t + \sum_{i=1}^{\infty} \boldsymbol{A}_i \boldsymbol{\Upsilon}_{t-i} \boldsymbol{g}_{t-i}, \tag{7}$$

where

$$\boldsymbol{\omega}_t = \boldsymbol{\omega} + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \boldsymbol{X}_{t-i}, \ \boldsymbol{v}_t^2 = \boldsymbol{\Upsilon}_t \boldsymbol{g}_t, \quad \boldsymbol{\Upsilon}_t = \begin{pmatrix} \eta_{1t}^2 & 0 & \dots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & & \eta_{mt}^2 \end{pmatrix}.$$

Let $\odot$ denote the Hadamard product and for any matrix $\boldsymbol{A} = (a_{ij})$ with nonnegative entries, let $\boldsymbol{A}^{\odot r} = (a_{ij}^r)$ for $r \in (0, 1]$. For any square matrix $\boldsymbol{B}$, let $\rho(\boldsymbol{B})$ the spectral radius of $\boldsymbol{B}$. Let the matrices

$$\mathcal{A}^{\odot r} = \sum_{i=1}^{\infty} \boldsymbol{A}_i^{\odot r}, \quad \mathcal{P}^{\odot r} = \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i^{\odot r},$$

with entries in $[0, \infty]$.

Theorem 1. *Suppose that for some $r \in (0, 1]$,*

(i) $\rho(\mathcal{A}^{\odot r} E \boldsymbol{\Upsilon}_t^{\odot r}) < 1$, $\mathcal{P}^{\odot r} < \infty$ *and* $E \boldsymbol{X}_t^{\odot r} < \infty$, *componentwise, and*

(ii) $\det \boldsymbol{C}(z) \neq 0$ *for all* $|z| < 1$, *where* $\boldsymbol{C}(z) = \left[\boldsymbol{I}_m - \sum_{k=1}^{s} \boldsymbol{C}_k z^k \right]$.

Then there exists a unique strictly stationary, ergodic and nonanticipative solution to the Cholesky-ARCH(∞) model (1)-(5) (or (1)-(4),(6)) satisfying $E(\boldsymbol{g}_t^{\odot r}) < \infty$.

Remark 2.1. *Previous results on the existence of a stationary solution to multivariate ARCH(∞) model can be found in Doukhan et al. (2006). Their condition however depends on the choice of a multiplicative matrix norm, which is not the case here.*

Finiteness of moments of the returns process $(\boldsymbol{\varepsilon_t})$ requires an additional assumption on the function $\boldsymbol{c}$ in (5). In the next result, $\| \cdot \|$ denotes the Euclidean norm.

Proposition 1. *Suppose that there exists $K > 0$ such that $\|\boldsymbol{c}(\boldsymbol{x})\| \leq K(\|\boldsymbol{x}\| + 1)$ for any $\boldsymbol{x} \in \mathbb{R}^{(2m+P)q}$. Then, under the assumptions of Theorem 1, the unique strictly stationary, ergodic and nonanticipative solution to the Cholesky-ARCH(∞) model (1)-(4) and (6) satisfies $\|\boldsymbol{\varepsilon}_t\|_r < \infty$ with $r \leq 1/2$.*

2.2 Particular specifications

In a recent article, Royer (2022) proposes an extension to the GARCH model to account for a higher persistence. A particular specification is given by the ARCH(∞) model

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{\infty} \left(\alpha\beta^{i-1} + \gamma i^{-d-1}\right) \varepsilon_{t-i}^2 \tag{8}$$

which nests the classical GARCH(1,1). Building upon this volatility specification, and taking into account possible exogenous covariates, a particular form of model (4) could be for any asset $k = 1, \ldots, m$

$$g_{k,t} = \underbrace{\omega + \sum_{i=1}^{\infty} \left(\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}\right) v_{k,t-i}^2}_{\text{specific ARCH}(\infty)\text{ effect}} + \underbrace{\sum_{j \neq k} \delta_{k,j} v_{j,t-1}^2}_{\text{spillover}} + \underbrace{\boldsymbol{\pi}_k' \boldsymbol{X}_{t-1}}_{\text{exogeneity}}$$

with $j = 1, \ldots, m$. Remark that in the case $\delta_{k,j} > 0$ for all k, j, each asset will have an effect on all the conditional volatilities of the system. Moreover, if one of the asset has long (*moderate*) memory, all of the conditional volatility processes will have a persistent component. This could be problematic when modeling liquid assets as they seem to present rather short memory, as noted by Royer (2022). To avoid this problem, it could be interesting to define a triangular system similar to the one proposed in Darolles et al. (2018). In the latter, the dynamics of the asset on the k-th row depends only on the previous rows, leading to

$$g_{k,t}^2 = \underbrace{\omega_k + \sum_{i=1}^{\infty} [\alpha_k \beta_k^{i-1} + \gamma_k i^{-d-1}] v_{k,t-i}^2}_{\text{ARCH}(\infty)\text{ effect}} + \underbrace{\sum_{j=1}^{k-1} \delta_{k,j} v_{j,t-1}^2}_{\text{spillover}} + \underbrace{\boldsymbol{\pi}_k' \boldsymbol{X}_{t-1}}_{\text{exogeneity}}. \tag{9}$$

Ordering the assets by their order of persistence would thus allow to have short memory for the assets on the top rows while having higher persistence for the assets on the bottom rows. This approach relates to the ordering based on liquidity proposed by Boudt et al. (2017) for the CholCov model.

Royer (2022) proposed an asymmetric extension of (8) to take into account the presence of both conditional asymmetry and high persistence of return volatility. In Model (9), we prefer to introduce asymmetry effects via the exogenous variables. For instance we could take

$$\boldsymbol{\pi}_k' \boldsymbol{X}_{t-1} = \pi_{k,+} X_{t-1} \mathbb{1}_{X_{t-1} \leq X_{t-2}} + \pi_{k,-} X_{t-1} \mathbb{1}_{X_{t-1} > X_{t-2}}$$

where X_t is an implied volatility index of the corresponding market (for example the VIX), or a realized measure of volatility as proposed by Barndorff-Nielsen and

Shephard (2002). Another popular approach is to use two exogenous variables, respectively capturing a downside or an upside risk. Possible candidates could be realized semivariances as advocated by Barndorff-Nielsen et al. (2010). The reason for introducing asymmetry via exogenous variables is both empirical and statistical. Indeed, when studying multivariate GARCH models, Bollerslev et al. (2020b) argue that modeling asymmetry through realized semicovariance matrices improve model fit compared to classical threshold models. Moreover, a volatility equation with terms of the form $\mathbb{1}_{v_{k,t-j}>0}$ would lead to major technical difficulties due to the non-differentiability with respect to the parameters when the factor $v_{k,t}$ is not observed (which is the case for $k > 1$).

An advantage of introducing a dynamic model on the Cholesky decomposition of the conditional variance is that the positive definiteness of $\boldsymbol{\Sigma}_t$ is very easily guaranteed. Only the positivity of the elements of $\boldsymbol{g}_t$ is needed. There is no sign constraint on the elements of $\boldsymbol{\beta}_t$. We can thus consider conditional betas of the form

$$\beta_{ij,t} = \varpi_{ij} + \boldsymbol{\pi}'_{ij}\boldsymbol{X}_{ij,t-1} + \alpha_{ij}v^2_{i,t-1} + \gamma_{ij}v^2_{j,t-1} + c_{ij}\beta_{ij,t-1} \tag{10}$$

where $\boldsymbol{X}_{ij,t-1}$ is a vector of potential beta predictors.

An interesting specification of (10) could be to include realized betas lower frequencies in $\boldsymbol{X}_{ij,t-1}$. Building upon the realized volatility literature, realized betas were first introduced by Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2006) and are computed using higher frequency returns over fixed time intervals. Let $\boldsymbol{\varepsilon}_{w,t} = (\varepsilon_{w,1,t}, \ldots, \varepsilon_{w,m,t})'$ the vector of returns over the time interval w. The realized covariance matrix $\boldsymbol{\mathcal{C}}_t^{w,W}$ is then defined as the sum of the W outer products of the high-frequency return vectors. For example, w could be a day and W a month, or w could be a 5-minute interval and W a day.

$$\boldsymbol{\mathcal{C}}_t = \boldsymbol{\mathcal{C}}_t^{w,W} = \sum_{w=1}^{W} \boldsymbol{\varepsilon}_{w,t}\boldsymbol{\varepsilon}'_{w,t}$$

In addition, for $1 < k \leq m$, for any $m \times m$ square matrix $\boldsymbol{M}$, let $\boldsymbol{M}_{[kk]}$ denote the square sub-matrix formed by the first $k-1$ rows and columns and $\boldsymbol{M}_{[k]}$ the sub-vector formed by the first $k-1$ rows and the k-th column of $\boldsymbol{M}$. For $1 < i \leq m$, the realized betas between asset i and assets $j = 1, \ldots, i-1$ are then defined by

$$\overline{\boldsymbol{\beta}}_{i,t} = (\overline{\beta}_{ij,t})_{j=1,\ldots,i-1} = \left(\boldsymbol{\mathcal{C}}_{t[ii]}\right)^{-1}\boldsymbol{\mathcal{C}}_{t[i]}.$$

We denote $\overline{\beta}^{\mathcal{W}}_{ij,t}$ the weekly realized beta at time t computed over 5 trading days, and $\overline{\beta}^{\mathcal{M}}_{ij,t}$ the monthly realized beta at time t computed over 21 trading days.

Ignoring the squared terms in (10) and letting $\boldsymbol{X}_{ij,t-1} = (\overline{\beta}^{\mathcal{W}}_{ij,t-1},\ \overline{\beta}^{\mathcal{M}}_{ij,t-1})$ yields

$$\beta_{ij,t} = \varpi_{ij} + \pi^{\mathcal{W}}_{ij}\ \overline{\beta}^{\mathcal{W}}_{ij,t-1} + \pi^{\mathcal{M}}_{ij}\ \overline{\beta}^{\mathcal{M}}_{ij,t-1} + c_{ij}\beta_{ij,t-1} \tag{11}$$

allowing to model the conditional beta as an exponential smoothing of a combination of realized betas at lower frequencies. In that sense, Equation (11) is related to the fast growing literature on component volatility models and in particular to the HAR model of Corsi (2009)[2].

Alternatively, potential useful exogenous variables in the conditional betas dynamics could be the recently introduced realized semibetas of Bollerslev et al. (2021). Consider the realized covariance matrix decomposition

$$\boldsymbol{\mathcal{C}}_t = \boldsymbol{\mathcal{P}}_t + \boldsymbol{\mathcal{N}}_t + \boldsymbol{\mathcal{S}}^+_t + \boldsymbol{\mathcal{S}}^-_t$$

into the semicovariance components introduced by Bollerslev et al. (2020a)

$$\begin{aligned} \boldsymbol{\mathcal{P}}_t &= \textstyle\sum_{w=1}^{W} \boldsymbol{\varepsilon}^+_{w,t}\boldsymbol{\varepsilon}^{+\prime}_{w,t}, & \boldsymbol{\mathcal{S}}^+_t &= \textstyle\sum_{w=1}^{W} \boldsymbol{\varepsilon}^-_{w,t}\boldsymbol{\varepsilon}^{+\prime}_{w,t}, \\ \boldsymbol{\mathcal{N}}_t &= \textstyle\sum_{w=1}^{W} \boldsymbol{\varepsilon}^-_{w,t}\boldsymbol{\varepsilon}^{-\prime}_{w,t}, & \boldsymbol{\mathcal{S}}^-_t &= \textstyle\sum_{w=1}^{W} \boldsymbol{\varepsilon}^+_{w,t}\boldsymbol{\varepsilon}^{-\prime}_{w,t} \end{aligned}$$

where $\boldsymbol{\varepsilon}^{+(-)}_{w,t}$ denotes the componentwise positive (respectively negative) part of $\boldsymbol{\varepsilon}_{w,t}$. Each realized semicovariance component then yields realized semibetas

$$\begin{aligned} \overline{\boldsymbol{\beta}}^{\mathcal{P}}_{i,t} &= \left(\boldsymbol{\mathcal{C}}_{t[ii]}\right)^{-1} \boldsymbol{\mathcal{P}}_{t[i]}, & \overline{\boldsymbol{\beta}}^{\mathcal{S}^+}_{i,t} &= -\left(\boldsymbol{\mathcal{C}}_{t[ii]}\right)^{-1} \boldsymbol{\mathcal{S}}^+_{t[i]}, \\ \overline{\boldsymbol{\beta}}^{\mathcal{N}}_{i,t} &= \left(\boldsymbol{\mathcal{C}}_{t[ii]}\right)^{-1} \boldsymbol{\mathcal{N}}_{t[i]}, & \overline{\boldsymbol{\beta}}^{\mathcal{S}^-}_{i,t} &= -\left(\boldsymbol{\mathcal{C}}_{t[ii]}\right)^{-1} \boldsymbol{\mathcal{S}}^-_{t[i]} \end{aligned}$$

providing a four-way decomposition of the traditional realized betas

$$\overline{\boldsymbol{\beta}}_{i,t} = \overline{\boldsymbol{\beta}}^{\mathcal{P}}_{i,t} + \overline{\boldsymbol{\beta}}^{\mathcal{N}}_{i,t} - \overline{\boldsymbol{\beta}}^{\mathcal{S}^+}_{i,t} - \overline{\boldsymbol{\beta}}^{\mathcal{S}^-}_{i,t}.$$

Letting $\boldsymbol{X}_{ij,t-1} = (\overline{\beta}^{\mathcal{P}}_{ij,t-1},\ \overline{\beta}^{\mathcal{N}}_{ij,t-1},\ \overline{\beta}^{\mathcal{S}^+}_{ij,t-1},\ \overline{\beta}^{\mathcal{S}^-}_{ij,t-1})$, the conditional beta dynamics could then be given by

$$\beta_{ij,t} = \varpi_{ij} + \pi^{\mathcal{P}}_{ij}\ \overline{\beta}^{\mathcal{P}}_{ij,t-1} + \pi^{\mathcal{N}}_{ij}\ \overline{\beta}^{\mathcal{N}}_{ij,t-1} + \pi^{\mathcal{S}^+}_{ij}\ \overline{\beta}^{\mathcal{S}^+}_{ij,t-1} + \pi^{\mathcal{S}^-}_{ij}\ \overline{\beta}^{\mathcal{S}^-}_{ij,t-1} + c_{ij}\beta_{ij,t-1} \tag{12}$$

allowing for a more accurate contribution of downside and upside risk.

[2]Equation (11) can also be related to the Midas model of Ghysels et al. (2006) although the step function of the HAR model cannot be reproduced using the Beta function lag polynomial of Midas regressions.

3 Statistical inference

We now assume that the conditional variance and its Cholesky decomposition

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0) = \boldsymbol{L}_t\boldsymbol{G}_t\boldsymbol{L}_t' = \boldsymbol{L}_t(\boldsymbol{\theta}_0)\boldsymbol{G}_t(\boldsymbol{\theta}_0)\boldsymbol{L}_t'(\boldsymbol{\theta}_0) \tag{13}$$

are parameterized by a d-dimensional parameter $\boldsymbol{\theta}_0$ belonging to a compact parameter space Θ, and that, for all $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{L}_t(\boldsymbol{\theta})$, $\boldsymbol{G}_t(\boldsymbol{\theta})$, and thus $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{L}_t(\boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{\theta})\boldsymbol{L}_t'(\boldsymbol{\theta})$, are $\mathcal{F}_{t-1}$ measurable, where $\mathcal{F}_t$ is the sigma-field generated by $\{\boldsymbol{\eta}_u, \boldsymbol{X}_u, u \leq t\}$.

We also introduce the vector $\boldsymbol{g}_t(\boldsymbol{\theta})$ of generic element $g_{it}(\boldsymbol{\theta})$ such that

$$\boldsymbol{g}_t = \boldsymbol{g}_t(\boldsymbol{\theta}_0) = \boldsymbol{g}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots; \boldsymbol{\theta}_0),$$

and the vectors $\boldsymbol{\beta}_t(\boldsymbol{\theta})$ and $\boldsymbol{\ell}_t(\boldsymbol{\theta})$ of $\mathbb{R}^{m_0}$, with $m_0 = (m-1)m/2$, such that

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_t(\boldsymbol{\theta}_0) = \boldsymbol{\beta}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots; \boldsymbol{\theta}_0)$$

and $\boldsymbol{\ell}_t = \boldsymbol{\ell}_t(\boldsymbol{\theta}_0) = \boldsymbol{\ell}(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots; \boldsymbol{\theta}_0)$. Given the observations $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, and arbitrary fixed initial values $\widetilde{\boldsymbol{\varepsilon}}_i$ and $\widetilde{\boldsymbol{X}}_i$ for $i \leq 0$, let the statistics

$$\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\left(\boldsymbol{\varepsilon}_{t-1}, \ldots, \boldsymbol{\varepsilon}_1, \widetilde{\boldsymbol{\varepsilon}}_0, \widetilde{\boldsymbol{\varepsilon}}_{-1}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots; \boldsymbol{\theta}\right)$$

and similarly define $\widetilde{\boldsymbol{L}}_t(\boldsymbol{\theta})$, $\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta}) = \operatorname{diag}\{\widetilde{\boldsymbol{g}}_t(\boldsymbol{\theta})\}$, $\widetilde{\boldsymbol{B}}_t(\boldsymbol{\theta})$, $\widetilde{\boldsymbol{\ell}}_t(\boldsymbol{\theta}) = \operatorname{vec}^0\widetilde{\boldsymbol{L}}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta}) = -\operatorname{vec}^0\widetilde{\boldsymbol{B}}_t(\boldsymbol{\theta})$. A QMLE of $\boldsymbol{\theta}_0$ is defined as any measurable solution $\widehat{\boldsymbol{\theta}}_n$ of

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} \widetilde{O}_n(\boldsymbol{\theta}), \qquad \widetilde{O}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^{n} \widetilde{q}_t(\boldsymbol{\theta}), \tag{14}$$

where

$$\widetilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{\Sigma}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \log\left|\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})\right| = \boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{B}}_t'(\boldsymbol{\theta})\widetilde{\boldsymbol{G}}_t^{-1}(\boldsymbol{\theta})\widetilde{\boldsymbol{B}}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \sum_{i=1}^{m}\log\widetilde{g}_{it}(\boldsymbol{\theta}).$$

Note that the Cholesky decomposition facilitates the computation of the QMLE, the diagonal matrix $\widetilde{\boldsymbol{G}}_t(\boldsymbol{\theta})$ being much easier to invert than $\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})$.

Darolles, Francq, and Laurent (2018) (DFL hereafter) gave a set of regularity conditions which entails the strong consistency and asymptotic normality (CAN) of the QML estimator $\widehat{\boldsymbol{\theta}}_n$. However, their framework does not allow for exogenous variables. Moreover their assumptions (in particular their assumption **A2**) preclude the ARCH(∞) models that we consider in the present paper.

We will thus give a set of alternative conditions for CAN of the QMLE. In the sequel ρ denotes a generic constant belonging to $[0,1)$, and K denotes a positive constant or a positive random variable which is $\mathcal{F}_0$-measurable. We take the spectral norm as matrix norm and the Euclidean norm as vector norm. Assume

A1: $\sup_{\boldsymbol{\theta}\in\Theta}\left\|\widetilde{\boldsymbol{G}}_t^{-1}(\boldsymbol{\theta})\right\| \leq K, \quad \sup_{\boldsymbol{\theta}\in\Theta}\left\|\boldsymbol{G}_t^{-1}(\boldsymbol{\theta})\right\| \leq K, \quad$ a.s.

A2: $\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{\beta}_t(\boldsymbol{\theta}) - \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})\| \leq K\rho_t$ where the random variable ρ_t satisfies $\sum_{t=1}^{\infty}\{E\rho_t^{s_1}\}^{1/p} < \infty$ for all $s_1 \in (0, s_0]$ and some $s_0 > 0$ and some $p > 1$.

A3: $E\left\{\|\boldsymbol{\ell}_t(\boldsymbol{\theta}_0)\|^{s_0} + \|\boldsymbol{g}_t(\boldsymbol{\theta}_0)\|^{s_0} + \sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{\beta}_t(\boldsymbol{\theta})\|^{s_0}\right\} < \infty$ for some $s_0 > 0$.

A4: For $\boldsymbol{\theta} \in \Theta$, $\{\boldsymbol{g}_t(\boldsymbol{\theta}), \boldsymbol{\beta}_t(\boldsymbol{\theta})\} = \{\boldsymbol{g}_t(\boldsymbol{\theta}_0), \boldsymbol{\beta}_t(\boldsymbol{\theta}_0)\}$ a.s. implies $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

A5: For any sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ of vectors of $\mathbb{R}^m$ and any sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ of vectors of $\mathbb{R}^r$, the functions $\boldsymbol{\theta} \mapsto \boldsymbol{g}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots; \boldsymbol{\theta})$ from Θ to $(0, +\infty)^m$ and $\boldsymbol{\theta} \mapsto \boldsymbol{\beta}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots; \boldsymbol{\theta})$ from Θ to $\mathbb{R}^{m_0}$ are continuous on Θ.

A6: $\boldsymbol{\theta}_0$ belongs to the interior $\mathring{\Theta}$ of Θ.

A7: For any sequence $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ of vectors of $\mathbb{R}^m$ and any sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ of vectors of $\mathbb{R}^r$, the functions $\boldsymbol{\theta} \mapsto \boldsymbol{g}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots; \boldsymbol{\theta})$ and $\boldsymbol{\theta} \mapsto \boldsymbol{\beta}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots; \boldsymbol{\theta})$ admit continuous second-order derivatives.

A8: For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, ρ_t as in **A2** and $s_0 > 0$

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial\boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \leq K\rho_t, \quad E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial\boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^{s_0} + \left\|\frac{\partial\boldsymbol{g}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^{s_0} < \infty.$$

A9: For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$, for all $i, j \in \{1, \ldots, m\}$ and $p > 0$, $q > 0$ and $r > 0$ such that $2q^{-1} + 2r^{-1} = 1$ and $p^{-1} + 2r^{-1} = 1$, we have

$$E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta})\frac{\partial^2\boldsymbol{\Sigma}_t(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\right\|^p < \infty,$$
$$E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\Sigma}_t(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\right\|^q < \infty,$$
$$E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\boldsymbol{\Sigma}_t^{1/2'}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\right\|^r < \infty,$$

where θ_i denotes the i-th element of the vector $\boldsymbol{\theta}$.

A10: $E\|\boldsymbol{\eta}_t\|^4 < \infty$.

A11: The matrices $\{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)/\partial\theta_i,\, i = 1,\dots,d\}$ are linearly independent with nonzero probability.

The previous assumptions are essentially the same as those of Darolles et al. (2018). Some of them have been extended to incorporate exogenous variables. To deal with these covariates we assume the following.

A1*: $(\boldsymbol{X}_t)$ is a sequence of positive random vectors such that $(\boldsymbol{\eta}_t', \boldsymbol{X}_t')'$ is stationary and ergodic, with $\boldsymbol{\eta}_t$ independent of $\{\boldsymbol{X}_{t-i}, i > 0\}$.

DFL made the assumptions **A2** and **A8** by replacing $\boldsymbol{\beta}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\theta})$ by $\boldsymbol{g}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{g}}_t(\boldsymbol{\theta})$. Such assumptions are not satisfied for volatilities of the form (9). We thus replace these assumptions by the following.

A2*: As $t \to \infty$, $\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{g}_t(\boldsymbol{\theta}) - \widetilde{\boldsymbol{g}}_t(\boldsymbol{\theta})\| \to 0$ a.s.

To simplify the notation we omit "$(\boldsymbol{\theta})$", and write for instance $\boldsymbol{B}_t$ and $\boldsymbol{L}_{0t}$ instead of $\boldsymbol{B}_t(\boldsymbol{\theta})$ and $\boldsymbol{L}_t(\boldsymbol{\theta}_0)$.

A3*: There exist three conjugate numbers $p > 0$, $q > 0$ and $r > 0$ (such that $p^{-1} + q^{-1} + r^{-1} = 1$) and a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E\left\{\|\boldsymbol{L}_{0t}\|^{2p} + \|\boldsymbol{G}_{0t}\|^{2q} + \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \|\boldsymbol{B}_t\|^{2r}\right\} < \infty$$

A4*: For $i = 1,\dots,m$, there exist $\xi > 0$, $\rho \in (0,1]$, $a > 1 - \rho/2$, three conjugate numbers $p > 0$, $q > 0$, $r > 0$, and a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$E \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left\{\|\boldsymbol{B}_t\boldsymbol{L}_{0t}\|^{p\rho\left(\frac{1+\xi}{\xi}\right)} + \left\|\boldsymbol{G}_t^{-1}\frac{\partial \boldsymbol{G}_t}{\partial\theta_i}\right\|^{q\rho\left(\frac{1+\xi}{\xi}\right)} + \|\boldsymbol{G}_{0t}\|^{2r\rho\left(\frac{1+\xi}{\xi}\right)}\right\} < \infty \quad (15)$$

and

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left\|\|\boldsymbol{g}_t - \widetilde{\boldsymbol{g}}_t\|^{\rho}\right\|_{1+\xi} \le \frac{K}{t^a}.$$

A5*: **A4*** holds when (15) is replaced by

$$E \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} \left\{\|\boldsymbol{B}_t\boldsymbol{L}_{0t}\|^{p\rho\left(\frac{1+\xi}{\xi}\right)} + \left\|\frac{\partial \boldsymbol{B}_t}{\partial\theta_i}\boldsymbol{L}_{0t}\right\|^{q\rho\left(\frac{1+\xi}{\xi}\right)} + \|\boldsymbol{G}_{0t}\|^{2r\rho\left(\frac{1+\xi}{\xi}\right)}\right\} < \infty$$

Theorem 2 (CAN of the QMLE in the general case)**.** *Let $(\boldsymbol{\varepsilon}_t)$ be a volatility model (1) satisfying the $\mathcal{F}_{t-1}$-measurable Cholesky decomposition (13).*[3] *Let $(\widehat{\boldsymbol{\theta}}_n)$*

[3] In particular, under **A1***, $\boldsymbol{\varepsilon}_t$ is stationary and ergodic as fixed measurable function of $\{\boldsymbol{\eta}_u, \boldsymbol{X}_u, u < t\}$

be a sequence of QML estimators satisfying (14). Under **A1-A5** *and* **A1*****-A3*** *we have*

$$\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0, \quad \textit{almost surely as } n \to \infty.$$

Under the additional assumptions **A6-A10** *and* **A4*****-A5****, we have the existence of the* $d \times d$ *matrix*

$$\boldsymbol{J} = E\boldsymbol{D}_t' \left\{ \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0) \right\} \boldsymbol{D}_t, \qquad \boldsymbol{D}_t = \frac{\partial vec \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'},$$

and of the $d \times d$ *matrix* $\boldsymbol{I}$ *of generic term*

$$\boldsymbol{I}(i,j) = Tr\left\{ \boldsymbol{K} E \boldsymbol{C}_{j,t} \boldsymbol{C}_{i,t}' \right\},$$

with $\boldsymbol{K} = Evec(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')vec'(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')$*,* $\boldsymbol{I}_m$ *the identity matrix of size* m *and*

$$\boldsymbol{C}_{i,t} = \left\{ \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \right\} vec \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)}{\partial \theta_i}.$$

Moreover, under the additional assumption **A11**, $\boldsymbol{J}$ *is invertible and*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right\} \ \textit{as } n \to \infty. \tag{16}$$

We also have the Bahadur representation

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = \boldsymbol{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\nabla}_t vec(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - \boldsymbol{I}_m) + o_P(1), \tag{17}$$

where $\boldsymbol{\nabla}_t = \boldsymbol{D}_t' \left\{ \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\theta}_0) \right\}$.

4 Simulations

In order to assess the finite sample properties of the QMLE, we carry out some Monte Carlo experiments. In the following simulations, we use Gaussian innovations ($\boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$).

To conduct our experiments, we first focus on specification (9)-(10) where we do not include exogenous variables in the volatility and where we consider the monthly realized beta between the Agricultural industry and the Market portfolio as an exogenous variable in the conditional betas dynamics. We simulate a thousand samples of size 5000 of the Cholesky-ARCH(∞)-X and on each sample we compute the QMLE using (i) the realized beta as exogenous variables (ii) the realized semibetas as exogenous variables. The empirical mean and Bias of the obtained QMLE are reported in Table 1. We see that the QMLE performs well.

	$\boldsymbol{\theta}_0$	$\widehat{\boldsymbol{\theta}}_n$	Bias	$\widehat{\boldsymbol{\theta}}_n$	Bias
ω_1	0.10	0.1014	0.0014	0.1014	0.0014
α_1	0.10	0.0940	-0.0060	0.0940	-0.0060
β_1	0.85	0.8541	0.0041	0.8541	0.0041
γ_1	0.00	0.0108	0.0108	0.0108	0.0108
ω_2	0.20	0.2007	0.0007	0.2005	0.0005
α_2	0.10	0.1073	0.0073	0.1074	0.0074
β_2	0.70	0.6969	-0.0031	0.6967	-0.0033
γ_2	0.10	0.0917	-0.0083	0.0917	-0.0083
δ_{21}	0.10	0.0999	-0.0001	0.0998	-0.0002
$\overline{\omega}_{21}$	0.70	0.7342	0.0342	0.7398	0.0398
c_{21}	0.70	0.6852	-0.0148	0.6823	-0.0177
π_{21}	0.20	0.2098	0.0098	–	–
$\pi_{21}^{\mathcal{P}}$	–	–	–	0.2121	0.0121
$\pi_{21}^{\mathcal{N}}$	–	–	–	0.2123	0.0123
$\pi_{21}^{\mathcal{S}^+}$	–	–	–	-0.2074	-0.0074
$\pi_{21}^{\mathcal{S}^-}$	–	–	–	-0.2081	-0.0081

Table 1: *Estimation results for 1000 simulations of size 5000 of a Cholesky ARCH(∞) with realized betas*

In particular, since the realized can be decomposed into semibetas, we see that the four coefficients estimated on the exogenous variables are well estimated and equal in absolute value to the coefficient π_{21} used in the simulation.

We then focus on specification (9)-(12) where we use semibetas between the Agricultural industry and the Market portfolio as an exogenous variable in the conditional betas dynamics. Results are reported in Table 2 and show similar good properties of the estimator.

5 Application: Do semibetas matter when modeling conditional betas?

In this section we consider the modeling of the conditional betas between 35 industries portfolios and the Market portfolio. Our dataset contains daily returns from January 2000 to December 2020 and are obtained from Kenneth French's website.

We fit the Cholesky ARCH(∞) model defined for asset i by

$$g_{i,t}^2 = \omega_i + \sum_{k=1}^{\infty}[\alpha_i\beta_i^{k-1} + \gamma_i k^{-2}]v_{i,t-k}^2 + \sum_{j=1}^{i-1}\delta_{ij}v_{j,t-1}^2$$

	$\boldsymbol{\theta}_0$	$\widehat{\boldsymbol{\theta}}_n$	Bias
ω_1	0.10	0.1026	0.0026
α_1	0.10	0.0994	-0.0006
β_1	0.85	0.8491	-0.0009
ω_2	0.20	0.2015	0.0015
α_2	0.10	0.0991	-0.0009
β_2	0.70	0.6996	-0.0004
δ_{21}	0.10	0.1004	0.0004
$\overline{\omega}_{21}$	0.70	0.7059	0.0059
c_{21}	0.70	0.6970	-0.0030
$\pi_{21}^{\mathcal{P}}$	0.20	0.2022	0.0022
$\pi_{21}^{\mathcal{N}}$	0.50	0.5059	0.0059
$\pi_{21}^{\mathcal{S}^+}$	0.00	0.0018	0.0018
$\pi_{21}^{\mathcal{S}^-}$	0.00	0.0025	0.0025

Table 2: *Estimation results for 1000 simulations of size 5000 of a Cholesky GARCH with semibetas*

and consider the conditional beta dynamic

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}}\overline{\beta}_{ij,t}^{\mathcal{P}} + \pi_{ij}^{\mathcal{N}}\overline{\beta}_{ij,t}^{\mathcal{N}} + \pi_{ij}^{\mathcal{S}^+}\overline{\beta}_{ij,t}^{\mathcal{S}^+} + \pi_{ij}^{\mathcal{S}^-}\overline{\beta}_{ij,t}^{\mathcal{S}^-} + c_{ij}\beta_{ij,t}$$

wich reduces to

$$\beta_{ij,t} = \varpi_{ij} + \pi_{ij}^{\mathcal{P}}\overline{\beta}_{ij,t} + c_{ij}\beta_{ij,t}$$

if $\pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^+} = -\pi_{ij}^{\mathcal{S}^-}$. Testing the informational benefit of including realized semibetas instead of standard realized betas can thus be achieved by testing

$$\begin{cases} H_0: & \pi_{ij}^{\mathcal{P}} = \pi_{ij}^{\mathcal{N}} \quad \text{and} \quad \pi_{ij}^{\mathcal{P}} = -\pi_{ij}^{\mathcal{S}^+} \quad \text{and} \quad \pi_{ij}^{\mathcal{N}} = -\pi_{ij}^{\mathcal{S}^-} \\ H_0: & \pi_{ij}^{\mathcal{P}} \neq \pi_{ij}^{\mathcal{N}} \quad \text{or} \quad \pi_{ij}^{\mathcal{P}} \neq -\pi_{ij}^{\mathcal{S}^+} \quad \text{or} \quad \pi_{ij}^{\mathcal{N}} \neq -\pi_{ij}^{\mathcal{S}^-} \end{cases}$$

which can be rewritten as the linear constraint on the parameter vector $\boldsymbol{\theta}_0$

$$\begin{cases} H_0: & \boldsymbol{R}\boldsymbol{\theta}_0 = (0,0,0)' \\ H_0: & \boldsymbol{R}\boldsymbol{\theta}_0 \neq (0,0,0)'. \end{cases}$$

We thus can use the standard Wald statistic

$$W_n = n(\boldsymbol{R}\,\widehat{\boldsymbol{\theta}}_n' \left(\boldsymbol{R}\boldsymbol{J}_n^{-1}\boldsymbol{I}_n\boldsymbol{J}_n^{-1}\boldsymbol{R}'\right)^{-1} (\boldsymbol{R}\,\widehat{\boldsymbol{\theta}}_n)$$

which asymptotically follows a χ^2 distribution with 3 degrees of freedom from standard results (see for example Gouriéroux and Monfort (1995)).

Table 3 shows p-values for each industry portfolio and provides a strong argument in favor of using semibetas rather than realized betas as covariates when modeling conditional betas.

Industry	W_n	Industry	W_n	Industry	W_n
Agric	0.000	Print	0.005	Manuf	0.000
Mines	0.000	Chems	0.082	Trans	0.000
Oil	0.000	Ptrlm	0.000	Phone	0.000
Stone	0.009	Rubbr	0.000	TV	0.091
Cnstr	0.000	Lethr	0.000	Utils	0.000
Food	0.000	Glass	0.000	Garbg	0.000
Smoke	0.000	Metal	0.000	Whlsl	0.002
Txtls	0.000	MtlPr	0.030	Rtail	0.000
Apprl	0.000	Machn	0.000	Money	0.000
Wood	0.000	Elctr	0.000	Srvc	0.000
Chair	0.000	Cars	0.979	Other	0.054
Paper	0.958	Instr	0.977		

Table 3: *p-values of the Wald test on the importance of the semibetas*

In addition, Table 4 gives estimated values of γ_{Mkt} and γ_{indus} parameters for each industry portfolio. The γ parameter in Equation (9) is of particular interest as it allows to differentiate between a GARCH(1,1) and a more persistent volatility process. Interestingly, we see that the Market portfolio seems to exhibit short memory as γ_{Mkt} is equal to 0. This result is similar to the finding of Royer (2022) when considering capitalization-weighted equity indices from developed markets. However the picture is very different for industry portfolios once dynamically hedged from the Market effect. Indeed, most of the estimated γ_{indus} are far from 0, which seems to validate the use of an ARCH(∞) equation.

	γ_{Mkt}	γ_{indus}		γ_{Mkt}	γ_{indus}		γ_{Mkt}	γ_{indus}
Agric	0.000	0.122	Print	0.000	0.078	Manuf	0.000	0.001
Mines	0.000	0.050	Chems	0.000	0.143	Trans	0.000	0.091
Oil	0.000	0.065	Ptrlm	0.000	0.038	Phone	0.000	0.102
Stone	0.000	0.108	Rubbr	0.000	0.042	TV	0.000	0.075
Cnstr	0.000	0.053	Lethr	0.000	0.070	Utils	0.000	0.113
Food	0.000	0.105	Glass	0.000	0.086	Garbg	0.000	0.218
Smoke	0.000	0.166	Metal	0.000	0.034	Whlsl	0.000	0.115
Txtls	0.000	0.198	MtlPr	0.000	0.154	Rtail	0.000	0.029
Apprl	0.000	0.080	Machn	0.000	0.080	Money	0.000	0.164
Wood	0.000	0.101	Elctr	0.000	0.103	Srvc	0.000	0.121
Chair	0.000	0.119	Cars	0.000	0.121	Other	0.000	0.067
Paper	0.000	0.123	Instr	0.000	0.078			

Table 4: γ_{Mkt} *and* γ_{indus} *parameters for each industry*

6 Conclusion

Models accounting for the time variation of slope coefficients in linear regression have recently attracted substantial attention in the econometric literature. However, most of this literature relies on short memory models for the conditional volatilities in the system. Such hypothesis may prove inadequate on empirical applications. In this paper, we propose a novel multivariate ARCH(∞) model with dynamic betas where exogenous variables can be introduced in the dynamics of both the conditional covariances and the dynamic slope coefficients. In that sense, we provide two major extensions to the Cholesky GARCH model of Darolles et al. (2018). We prove the existence of a stationary solution which contributes to the scarce literature on multivariate ARCH(∞) models. Additionally, we establish statistical inference results and prove the consistency and asymptotic normality of Quasi Maximum Likelihood estimator. Finite sample properties of the QMLE are assessed through Monte Carlo experiments, while an application on various industry portfolios investigates the role of the recently introduced semibetas of Bollerslev et al. (2021) in the dynamic of conditional betas. This application provides arguments in favor of these new realized measures while underlying the need to consider more persistent volatility processes than standard multivariate GARCH models allow. An additional application focusing on forecasts comparison is ongoing. In this application we aim at capturing the effects of a global equity index and a country-specific equity index on individual stocks, which has been the subject of a vast financial literature on emerging markets (see for example, Bekaert and Harvey (1995) , Bekaert et al. (2005) or Chaieb et al. (2021)).

References

T. G. Andersen, T. Bollerslev, F. X. Diebold, and G. Wu. Realized beta: Persistence and predictability. *Advances in Econometrics*, 20:1–40, 2006.

A. Ang and D. Kristensen. Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156, 2012.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004.

O. E. Barndorff-Nielsen, S. Kinnebrouk, and N. Shephard. Measuring downside risk: realised semivariance. In J. R. T. Bollerslev and M. Watson, editors, *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pages 117–136. Oxford University Press, 2010.

G. Bekaert and C. R. Harvey. Time-varying world market integration. *The Journal of Finance*, 50(2):403–444, 1995.

G. Bekaert, C. Harvey, and A. Ng. Market integration and contagion. *The Journal of Business*, 78(1):39–69, 2005.

T. Bollerslev, R. F. Engle, and J. M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.

T. Bollerslev, J. Li, A. J. Patton, and R. Quaedvlieg. Realized semicovariances. *Econometrica*, 88(4):1515–1551, 2020a.

T. Bollerslev, A. J. Patton, and R. Quaedvlieg. Multivariate leverage effects and realized semicovariance GARCH models. *Journal of Econometrics*, 217(2):411–430, 2020b.

T. Bollerslev, A. J. Patton, and R. Quaedvlieg. Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics*, 2021.

K. Boudt, S. Laurent, A. Lunde, R. Quaedvlieg, and O. Sauri. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 196(2):347–367, 2017.

I. Chaieb, H. Langlois, and O. Scaillet. Factors and risk premia in individual international stock returns. *Journal of Financial Economics*, 141(2):669–692, 2021.

C. Conrad and M. Karanasos. Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, pages 838–862, 2010.

C. Conrad, M. Karanasos, and N. Zeng. Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study. *Journal of Empirical Finance*, 18(1):147–159, 2011.

F. Corsi. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196, 02 2009.

S. Darolles, C. Francq, and S. Laurent. Asymptotics of Cholesky GARCH models and time-varying conditional betas. *Journal of Econometrics*, 204(2):223–247, 2018.

Z. Ding and C. W. Granger. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215, 1996.

P. Doukhan, G. Teyssiere, and P. Winant. A LARCH(∞) vector valued process. In P. Bertail, P. Doukhan, and P. Soulier, editors, *Dependence in probability and statistics*, pages 245–258. Springer, 2006.

R. F. Engle. Dynamic conditional beta. *Journal of Financial Econometrics*, 14 (4):643–667, 2016.

W. E. Ferson and C. R. Harvey. The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415, 1991.

C. Francq and L. Q. Thieu. QML inference for volatility models with covariates. *Econometric Theory*, 35(1):37–72, 2019.

E. Ghysels. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573, 1998.

E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95, 2006.

S. M. Goldfeld and R. E. Quandt. A Markov model for switching regressions. *Journal of Econometrics*, 1(1):3–15, 1973.

C. Gouriéroux and A. Monfort. *Statistics and Econometric Models*, volume 2 of *Themes in Modern Econometrics*. Cambridge University Press, 1995.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, pages 357–384, 1989.

P. R. Hansen, A. Lunde, and V. Voev. Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics*, 29(5):774–799, 2014.

T. Jeantheau. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, pages 70–86, 1998.

C.-J. Kim, J. Piger, and R. Startz. Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, 143(2): 263–273, 2008.

M. Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.

J. Royer. Conditional asymmetry in Power ARCH(∞) models. *Journal of Econometrics*, 2022.

J. Shanken. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2):99–120, 1990.

Appendix A Proofs and technical results

Proof of Theorem 1. In view of (7), consider the random vector with (possibly infinite) nonnegative components,

$$\boldsymbol{S}_t = \boldsymbol{\omega}_t + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \boldsymbol{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \ldots \boldsymbol{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\ldots-i_k} \boldsymbol{\omega}_{t-i_1-\ldots-i_k}. \tag{18}$$

For any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with positive entries and conformable dimensions, the c_r-inequality entails $(\boldsymbol{AB})^{\odot r} \leq \boldsymbol{A}^{\odot r}\boldsymbol{B}^{\odot r}$ and $(\boldsymbol{A}+\boldsymbol{B})^{\odot r} \leq \boldsymbol{A}^{\odot r}+\boldsymbol{B}^{\odot r}$ for $r \in (0,1]$. We thus have,

$$E\boldsymbol{\omega}_t^{\odot r} = \boldsymbol{\omega}^{\odot r} + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i^{\odot r} E\boldsymbol{X}_t^{\odot r} < \infty,$$

componentwise, and for $\boldsymbol{S}_t$ satisfying (18),

$$\begin{aligned} E\boldsymbol{S}_t^{\odot r} &\leq E\boldsymbol{\omega}_t^{\odot r} + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \boldsymbol{A}_{i_1}^{\odot r} E\boldsymbol{\Upsilon}_{t-i_1}^{\odot r} \ldots \boldsymbol{A}_{i_k}^{\odot r} E\boldsymbol{\Upsilon}_{t-i_1-\ldots-i_k}^{\odot r} E\boldsymbol{\omega}_t^{\odot r} \\ &= \left[\boldsymbol{I}_m + \sum_{k=1}^{\infty} (\mathcal{A}^{\odot r} E\boldsymbol{\Upsilon}_t^{\odot r})^k \right] E\boldsymbol{\omega}_t^{\odot r} < \infty, \end{aligned}$$

componentwise. It follows that $\boldsymbol{S}_t$ is finite a.s. The process $(\boldsymbol{S}_t)$ is thus strictly stationary and nonanticipative as a function of the past values of $\boldsymbol{\eta}_t$. In addition, it is clear that $(\boldsymbol{S}_t)$ satisfies the stochastic recurrence equation

$$\boldsymbol{S}_t = \boldsymbol{\omega}_t + \sum_{i=1}^{\infty} \boldsymbol{A}_i \boldsymbol{\Upsilon}_{t-i} \boldsymbol{S}_{t-i}. \tag{19}$$

Now we turn to uniqueness. Denote by $(\boldsymbol{\varepsilon}_t^*)$ any strictly stationary and nonanticipative solutions of the model. For all $N \geq 1$, we obtain

$$\begin{aligned} \boldsymbol{g}_t^* - \boldsymbol{S}_t &= \left\{ \boldsymbol{\omega}_t + \sum_{k=1}^{N} \sum_{i_1,\ldots,i_k \geq 1} \boldsymbol{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \ldots \boldsymbol{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\ldots-i_k} \boldsymbol{\omega}_{t-i_1-\ldots-i_k} - \boldsymbol{S}_t \right\} \\ &\quad + \sum_{i_1,\ldots,i_{N+1} \geq 1} \boldsymbol{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \ldots \boldsymbol{A}_{i_{N+1}} \boldsymbol{\Upsilon}_{t-i_1-\cdots-i_{N+1}} \boldsymbol{g}_{t-i_1-\cdots-i_{N+1}}^* \\ &:= \{\boldsymbol{g}_{t,N}^* - \boldsymbol{S}_t\} + \boldsymbol{R}_{t,N}. \end{aligned}$$

We have

$$\boldsymbol{S}_t - \boldsymbol{g}_{t,N}^* = \sum_{k=N+1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \boldsymbol{A}_{i_1} \boldsymbol{\Upsilon}_{t-i_1} \ldots \boldsymbol{A}_{i_k} \boldsymbol{\Upsilon}_{t-i_1-\ldots-i_k} \boldsymbol{\omega}_{t-i_1-\ldots-i_k},$$

which has nonnegative terms. Thus, by arguments already given,

$$E(\boldsymbol{S}_t - \boldsymbol{g}^*_{t,N})^{\odot r} \leq \sum_{k=N+1}^{\infty} (\mathcal{A}^{\odot r} E \boldsymbol{\Upsilon}_t^{\odot r})^k \to 0, \quad \text{as } N \to \infty.$$

Similarly, the entries of $\boldsymbol{R}_{t,N}$ are nonnegative and, using the fact that the process $\boldsymbol{g}^*_t$ is nonanticipative and satisfies $E\{(\boldsymbol{g}^*_t)^{\odot r}\} < \infty$, we have

$$E(\boldsymbol{R}_{t,N})^{\odot r} \leq (\mathcal{A}^{\odot r} E \boldsymbol{\Upsilon}_t^{\odot r})^{N+1} E\{(\boldsymbol{g}^*_t)^{\odot r}\} \to 0, \quad \text{as } N \to \infty.$$

We thus have $\boldsymbol{g}^*_t = \boldsymbol{S}_t$ a.s. The proof of Theorem 1 is now complete. □

We now turn to the proof of Proposition 1.

Proof of Proposition 1. In view of (6), we have

$$\boldsymbol{\ell}_t = \sum_{k=0}^{\infty} \boldsymbol{D}_k \boldsymbol{c}_{t-k},$$

where $\boldsymbol{c}_t = \boldsymbol{c}(\boldsymbol{v}_{t-1}, \ldots, \boldsymbol{v}_{t-q}, \boldsymbol{g}^{1/2}_{t-1}, \ldots, \boldsymbol{g}^{1/2}_{t-q}) := \boldsymbol{c}(\boldsymbol{x}_{t-1})$, and $(\boldsymbol{D}_k)$ is the sequence of matrices obtained in the inversion of the matrix polynomial $\boldsymbol{C}(z)$. It follows that, for $r \leq 1/2$

$$E\|\boldsymbol{\ell}_t\|^{2r} = \sum_{k=0}^{\infty} \|\boldsymbol{D}_k\|^{2r} E\|\boldsymbol{c}_t\|^{2r} \leq K(E\|\boldsymbol{x}_t\|^{2r} + 1) < \infty,$$

and thus $E\|\boldsymbol{L}_t\|^{2r} < \infty$. By Hölder's inequality, we deduce

$$E\|\boldsymbol{\varepsilon}_t\|^r \leq \{E\|\boldsymbol{L}_t\|^{2r}\}^{1/2} \{E\|\boldsymbol{G}_t^{1/2}\|^{2r}\}^{1/2} E\|\boldsymbol{\eta}_t\|^r < \infty$$

which concludes the proof. □

Before developing the proof of Theorem 2, we first state the following elementary lemma.

Lemma 1. *Let (u_t), (x_t) and (w_t) be three sequences of random variables such that u_t is independent of $x_t w_t$, and for $\xi > 0$, $\rho \in (0, 1]$ and $a > 1 - \rho/2$, we have $\sup_t E|u_t|^\rho < \infty$, $\sup_t \||x_t|^\rho\|_{\frac{1+\xi}{\xi}} < \infty$ and $\||w_t|^\rho\|_{1+\xi} \leq K/t^a$. Then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t x_t w_t \to 0 \text{ in probability as } n \to \infty.$$

Proof of Lemma 1. The result comes from the Markov and Hölder inequalities:

$$E\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}u_tx_tw_t\right|^{\rho}\leq\frac{K}{n^{\rho/2}}\sum_{t=1}^{n}\||x_t|^{\rho}\|_{\frac{1+\xi}{\xi}}\,\||w_t|^{\rho}\|_{1+\xi}\leq Kn^{-\rho/2}(n^{-a+1}+1)\to 0.$$

□

Proof of Theorem 2. It suffices to modify a few steps of the proof of Theorem 4.1 in DFL. Since we are considering volatility models with a larger memory than standard GARCH with finite orders, the main difficulty is to show that the initial values are asymptotically unimportant. More precisely, we have to show that

$$\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\Theta}|O_n(\boldsymbol{\theta})-\widetilde{O}_n(\boldsymbol{\theta})|=0\quad\text{a.s.}\tag{20}$$

where

$$O_n(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^{n}q_t(\boldsymbol{\theta}),\quad q_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t+\sum_{i=1}^{m}\log g_{it}.$$

Using **A1**–**A3**, DFL showed that one can replace $\widetilde{\boldsymbol{B}}_t$ by $\boldsymbol{B}_t$ in $\widetilde{q}_t(\boldsymbol{\theta})$ to establish (20). It thus remains to prove that

$$a_n:=\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left|\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}(\widetilde{\boldsymbol{G}}_t-\boldsymbol{G}_t)\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\right|\to 0\quad\text{a.s.}\tag{21}$$

Let $\varepsilon>0$. By **A2***, almost surely there exists N such that $\sup_{\boldsymbol{\theta}\in\Theta}\|\widetilde{\boldsymbol{G}}_t-\boldsymbol{G}_t\|<\varepsilon$ for all $t>N$. We thus have almost surely

$$\begin{aligned}a_n&=\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left|\ \mathrm{Tr}\ \left\{(\widetilde{\boldsymbol{G}}_t-\boldsymbol{G}_t)\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\right\}\right|\\&\leq\varepsilon\frac{K}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\right\|+\frac{K}{n}\sum_{t=1}^{N}\sup_{\boldsymbol{\theta}\in\Theta}\|\widetilde{\boldsymbol{g}}_t-\boldsymbol{g}_t\|\left\|\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\right\|.\end{aligned}$$

By **A1** and **A3***, we have $c:=E\sup_{\boldsymbol{\theta}\in\Theta}\left\|\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\right\|<\infty$. By the ergodic theorem, it follows that $\lim_{n\to\infty}a_n\leq\varepsilon Kc$. Since ε can be chosen arbitrarily small, we obtain (21). The consistency follows by the arguments given in DFL.
To prove asymptotic normality, the main difficulty is again to show that the initial values are asymptotically unimportant. More precisely we have to show that there exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial q_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}-\frac{\partial\widetilde{q}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\|=0\quad\text{a.s.}\tag{22}$$

Using standard matrix derivative computations, we have

$$\frac{\partial}{\partial\theta_i}q_t(\boldsymbol{\theta}) = \mathrm{Tr}\left\{\left(\boldsymbol{I}_m - \boldsymbol{G}_t^{-1}\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\right)\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{G}_t}{\partial\theta_i}\right\} + 2\mathrm{Tr}\left\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{B}_t}{\partial\theta_i}\right\}.$$

We thus have

$$\frac{\partial}{\partial\theta_i}q_t(\boldsymbol{\theta}) - \frac{\partial}{\partial\theta_i}\widetilde{q}_t(\boldsymbol{\theta}) = \sum_{j=1}^{6} a_{jt}(\boldsymbol{\theta}), \qquad a_{jt}(\boldsymbol{\theta}) = \mathrm{Tr}\boldsymbol{A}_{jt},$$

with

$$\begin{aligned}
\boldsymbol{A}_{1t} =& \left(\widetilde{\boldsymbol{G}}_t^{-1} - \boldsymbol{G}_t^{-1}\right)\boldsymbol{B}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{G}_t}{\partial\theta_i}, & \boldsymbol{A}_{2t} =& \widetilde{\boldsymbol{G}}_t^{-1}\left(\widetilde{\boldsymbol{B}}_t - \boldsymbol{B}_t\right)\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{B}_t'\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{G}_t}{\partial\theta_i},\\
\boldsymbol{A}_{3t} =& \widetilde{\boldsymbol{G}}_t^{-1}\widetilde{\boldsymbol{B}}_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\left(\widetilde{\boldsymbol{B}}_t' - \boldsymbol{B}_t'\right)\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{G}_t}{\partial\theta_i}, & \boldsymbol{A}_{4t} =& 2\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\left(\boldsymbol{B}_t' - \widetilde{\boldsymbol{B}}_t'\right)\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{B}_t}{\partial\theta_i},\\
\boldsymbol{A}_{5t} =& 2\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{B}}_t'\left(\boldsymbol{G}_t^{-1} - \widetilde{\boldsymbol{G}}_t^{-1}\right)\frac{\partial\boldsymbol{B}_t}{\partial\theta_i}, & \boldsymbol{A}_{6t} =& \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{B}}_t'\widetilde{\boldsymbol{G}}_t^{-1}\left(\frac{\partial\boldsymbol{B}_t}{\partial\theta_i} - \frac{\partial\widetilde{\boldsymbol{B}}_t}{\partial\theta_i}\right).
\end{aligned}$$

Using **A1**–**A3** and **A8**, the elementary inequalities $(\sum_i |a_i|)^s \le \sum_i |a_i|^s$ for $s \in (0,1]$ and $|\mathrm{Tr}(\boldsymbol{AB})| \le K\|\boldsymbol{A}\|\|\boldsymbol{B}\|$ with obvious notations, together with the Hölder inequality, we note that

$$E\left(\sum_{t=1}^{\infty}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}|a_{4t}(\boldsymbol{\theta})|\right)^s \le KE\|\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\|^s\sum_{t=1}^{\infty}E\left\|\boldsymbol{\ell}_t\right\|^{2s}\left\|\boldsymbol{g}_t\right\|^{2s}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial\boldsymbol{\beta}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^s\rho_t^s$$

is finite when s is sufficiently small. Therefore $\sum_{t=1}^{\infty}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}|a_{4t}(\boldsymbol{\theta})|$ is a.s. finite and we have shown

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}|a_{it}(\boldsymbol{\theta})| = o(1) \quad \text{a.s.} \tag{23}$$

for $i = 4$. Similarly, it can be seen that (23) holds for $i = 2, 3, 6$. The case $i = 1$ in (23) follows from **A4***, applying Lemma 1 with $u_t = \|\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\|$, $w_t = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\widetilde{\boldsymbol{G}}_t^{-1} - \boldsymbol{G}_t^{-1}\right\|$, and $x_t = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\|\boldsymbol{B}_t\boldsymbol{L}_{0t}\|^2\left\|\boldsymbol{G}_t^{-1}\frac{\partial\boldsymbol{G}_t}{\partial\theta_i}\right\|\|\boldsymbol{G}_{0t}\|^2$. The case $i = 5$ follows from **A5***. We thus have shown (22). The rest of the proof follows that of Theorem 4.1 in DFL. □