Asymptotics of Cholesky GARCH models and time-varying conditional betas

Serge Darolles, Christian Francq, Sébastien Laurent

Asymptotics of Cholesky GARCH Models and Time-Varying Conditional Betas

Serge Darolles*, Christian Francq† and Sébastien Laurent‡

Abstract

This paper proposes a new observation-driven model with time-varying slope coefficients. Our model, called CHAR, is a Cholesky-GARCH model, based on the Cholesky decomposition of the conditional variance matrix introduced by Pourahmadi (1999) in the context of longitudinal data. We derive stationarity and invertibility conditions and proof consistency and asymptotic normality of the Full and equation-by-equation QML estimators of this model. We then show that this class of models is useful to estimate conditional betas and compare it to the approach proposed by Engle (2016). Finally, we use real data in a portfolio and risk management exercise. We find that the CHAR model outperforms a model with constant betas as well as the dynamic conditional beta model of Engle (2016).

JEL Classification: C13, C32, C53, C58.
Keywords: Multivariate-GARCH, conditional betas, covariance.

1 Introduction and motivations

The estimation of linear models on time series data is common in finance, especially in the context of factor models. The standard practice is to assume that the slope coefficients are constant over time. However, wrongly assuming the constancy of the parameters leads to misspecified models and potentially wrong financial decisions. For example, testing an asset pricing model such as the Capital Asset Pricing Model (CAPM) (Sharpe, 1964) with constant regression parameters may lead

*Université Paris-Dauphine, PSL Research University, CNRS, UMR 7088, DRM, Finance, France. E-mail: serge.darolles@dauphine.fr

†CREST, ENSAE, Université Paris-Saclay and University Lille 3, France. E-mail: christian.francq@univ-lille3.fr

‡Corresponding author: Sébastien Laurent, Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, Aix-Marseille Graduate School of Management – IAE, France. E-mail: sebastien.laurent@univ-amu.fr

Acknowledgments: The authors thank the participants of the CFE 2016 conference in Seville, of the 8th French Econometric Conference at ESSEC Business School, of the 18th OxMetrics User Conference at CASS Business School and in particular Lynda Khalaf, Siem Jan Koopman and Andrew Patton for useful comments. The first author gratefully acknowledge the financial support of the chair QuantValley/Risk Foundation "Quantitative Management Initiative". The first two authors are grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR-16-CE26-0015-02). The second author also thanks the labex ECODEC.

to a wrong rejection of the model if the parameters are changing over time. Similarly, attempting to evaluate the performance of an active management fund using the Sharpe Style Analysis Model (Sharpe, 1992) with constant parameters may lead to an over-valuation of the manager's skills. It is therefore important to develop simple statistical approaches capturing all the dynamic aspects of the financial series.

Using the terminology of Cox (1981), time series models with time-varying parameters can be categorized into two classes: parameter-driven and observation-driven models. In parameter-driven models, the slope coefficients are modeled as stochastic processes that are subject to their own source of error. These models, known as state space models, are however difficult to estimate, reason why they are less popular than the second class (see Durbin and Koopman, 2012, for a comprehensive review of state space models). The advantage of the observation-driven approach is that time variation of the parameters is introduced by allowing them to dependent on exogenous variables and past observations. Although parameters are time-varying, they are predictable given the information set at time $t-1$, which simplifies the derivation of the likelihood function. The simplest solution to address this question within this class of models is to rely on interaction variables. This requires of course assuming that the source of randomness of the slope coefficients depends only on exogenous variables, which rules out the interesting case of autoregressive dynamics.

Recently, Engle (2016) proposed a new model called Dynamic Conditional Beta (DCB in short) to obtain time-varying slope coefficients using an observation-driven model by extending Bollerslev, Engle and Wolldbridge (1988) to the case of more than one explanatory variable (or factor).[1] Assuming joint conditional normality of the variables appearing in the model (i.e., dependent and independent variables), Engle (2016) shows how to recover indirectly the time-varying slope coefficients of the independent variables using an estimate of the full conditional covariance matrix. In his application, he relies on a DCC model to obtain the conditional betas of the Fama-French three factor model. While this approach is easy to implement, testing and imposing the constancy of the conditional betas is cumbersome. Furthermore, identifying the variables determining the evolution of the betas is impossible because conditional betas are not modeled directly but recovered afterwards by a non-linear transformation of the elements of the estimated conditional covariance matrix.

[1] Maheu and Shamsi (2016) propose a nonparametric DCB model in a Bayesian framework.

In this paper, we propose a different approach essentially based on a natural orthogonalization of the observed time series, initially proposed by Pourahmadi (1999) in the context of longitudinal data. We first study an observation-driven model, called CHAR (for Cholesky-GARCH), that specifies directly the dynamics of time varying slope coefficients. We extend this work by considering time varying slope coefficients that depend on their lagged values and past shocks. We derive stationarity and invertibility conditions and proof consistency and asymptotic normality of the Full and equation-by-equation (EbE) QML estimators of this model. Our approach is very flexible because it allows to impose easily the constancy of some of the conditional betas. Furthermore, unlike conditional correlations, conditional betas can take any real values, which allows the use of general specifications.

We compare the numerical properties of the two estimators using a Monte-Carlo simulation study. Our results suggest that both the Full and EbE QML deliver satisfactory results in the sense that biases are very small for all parameters. In an empirical application, we show that this class of models is useful to estimate conditional betas in the context of the Fama-French three factor model. We consider four competing models to obtain one-step-ahead forecasts of the three conditional betas, i.e., two DCB models based on a CCC-GARCH(1,1) or a DCC-GARCH(1,1), our proposed model with time-varying betas and a special case of this model where the constancy of the betas is imposed. We use these forecasts to build tracking portfolios for the 12 US industry portfolios (Data are from Ken French's web site and cover the period spanning from February 1994 to August 2016). We rely on two loss functions (i.e., mean square error or mean absolute deviation) to quantify the magnitude of the tracking errors (i.e., difference between the observed returns and the tracking portfolios) of the four models and apply the Model Confidence Set (MCS) test of Hansen, Lunde and Nason (2011) to discriminate between them. We find that on the whole forecasting period (i.e., 2010-2016), our proposed model with time-varying betas belongs to the set of superior models in all cases while the DCB based on the DCC model belongs to this set in only 3 and 2 cases, when relying respectively on the MSE and MAD loss functions. Interestingly, the DCB-CCC model and the CHAR model with constant betas never show up in the MCS. We also show that our proposed model has forecasted betas that are much smoother than those obtained with the DCB-DCC model, which translates into smaller transaction costs.

The paper is organized as follows. In Section 2, we review the DCB model of Engle (2016) and present the CHAR model. Stationarity conditions for the CHAR model are studied in Section 3.

The Full and EbE QML estimators of this model are presented in Section 4 as well as invertibility conditions. Monte Carlo simulation results are reported in Section 5 and an application in Section 6. Finally, Section 7 concludes.

2 Direct versus indirect specification of conditional betas

Before studying in more detail the properties of our proposed model, let us present some notation and a competing approach for obtaining observation-driven time-series models with time-varying slope coefficients. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ be a vector of $m \geq 2$ returns with mean zero and satisfying a general volatility model of the form

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\vartheta}_0)\boldsymbol{\eta}_t, \ t = 1, \ldots, n, \tag{2.1}$$

where $(\boldsymbol{\eta}_t)$ is a sequence of independent and identically distributed (i.i.d.) random vectors with zero mean and identity covariance matrix, and

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\vartheta}_0)$$

is almost surely a positive definite $m \times m$ matrix, parametrized by a d-dimensional parameter $\boldsymbol{\vartheta}_0$ and depending on the information $\mathcal{F}_{t-1}$ generated by the past values of $\boldsymbol{\epsilon}_t$. Assuming that (2) admits a non anticipative stationary solution (the stationarity conditions are discussed in the sequel), $\boldsymbol{\Sigma}_t$ is the conditional variance-covariance matrix of $\boldsymbol{\epsilon}_t$.

Multivariate GARCH (MGARCH) models are very well suited for empirical applications needing an estimate or a forecast of $\boldsymbol{\Sigma}_t$ (see Bauwens, Laurent, and Rombouts, 2006 and Silvennoinen, and Teräsvirta, 2008 for a survey of MGARCH models). A natural example is the computation of the value-at-risk of a portfolio, when the portfolio's composition is observable and time-varying (see e.g., Francq and Zakoian, 2017). However, MGARCH models are also used in financial applications involving a non-linear transformation of $\boldsymbol{\Sigma}_t$. For instance, in the mean-variance framework (see e.g., Markowitz, 1952 in a static case), the optimal weights of a portfolio depend on an estimate of $\boldsymbol{\Sigma}_t^{-1}$. Another example is the estimation of conditional betas using the DCB model of Engle (2016) presented in the next subsection. In this paper, we advocate the use of a model that specifies directly the dynamics of the conditional betas instead $\boldsymbol{\Sigma}_t$.

2.1 Indirect specification of dynamic conditional betas

In some empirical finance applications, it is necessary to give a special role to one of the asset embedded in the vector $\boldsymbol{\epsilon}_t$, for example when a particular return series, say $y_t = \epsilon_{mt}$, is regressed on the other components of the vector $\boldsymbol{\epsilon}_t$, i.e., $\mathbf{x}_t = (\epsilon_{1t}, \ldots, \epsilon_{m-1t})'$. The coefficients of the multivariate regression of y_t on $\mathbf{x}_t$ are usually called betas. To obtain time-varying betas, Engle (2016) assumes that $\boldsymbol{\epsilon}_t = (\mathbf{x}_t, y_t)'$ has a a multivariate normal distribution (conditional on $\mathcal{F}_{t-1}$) with conditional covariance $\boldsymbol{\Sigma}_t$, i.e.,

$$\begin{pmatrix} \mathbf{x}_t \\ y_t \end{pmatrix} | \mathcal{F}_{t-1} \sim N \left(\begin{pmatrix} \mathbf{0}_{m-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx,t} & \boldsymbol{\Sigma}_{xy,t} \\ \boldsymbol{\Sigma}_{yx,t} & \Sigma_{yy,t} \end{pmatrix} \right),$$

where subscripts represent natural partitions. Engle (2016) uses the fact that the conditional distribution of y_t on $\mathbf{x}_t$ is

$$y_t | \mathbf{x}_t \sim N \left(\boldsymbol{\Sigma}_{yx,t} \boldsymbol{\Sigma}_{xx,t}^{-1} \mathbf{x}_t, \Sigma_{yy,t} - \boldsymbol{\Sigma}_{yx,t} \boldsymbol{\Sigma}_{xx,t}^{-1} \boldsymbol{\Sigma}_{xy,t} \right)$$

to derive an estimate of the conditional betas. Indeed, estimates of the time-varying coefficients of the regression of y_t on $\mathbf{x}_t$ can be recovered from $\boldsymbol{\Sigma}_t$ by the formula $\boldsymbol{\Sigma}_{xx,t}^{-1} \boldsymbol{\Sigma}_{xy,t}$ or $\Sigma_{xy,t}/\Sigma_{xx,t}$ if there is only one explanatory variable (like in Bollerslev, Engle and Wooldridge, 1988). Engle (2016) advocates the use of a Dynamic Conditional Correlation (DCC) model to estimate $\boldsymbol{\Sigma}_t$, but any multivariate GARCH model can be used. The DCC approach relies on the following decomposition of $\boldsymbol{\Sigma}_t$:

$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t, \tag{2.2}$$

where $\boldsymbol{D}_t$ and $\boldsymbol{R}_t$ are respectively a diagonal matrix with the conditional volatilities of $\boldsymbol{\epsilon}_t$ and its conditional correlation matrix. Matrix $\boldsymbol{D}_t$ is typically modeled using m univariate GARCH models on $\boldsymbol{\epsilon}_t$ and $\boldsymbol{R}_t$ using a scalar BEKK specification on the devolatilized series $\boldsymbol{D}_t^{-1} \boldsymbol{\epsilon}_t$ (which is transformed afterwards to get a correlation matrix). The Constant Conditional Correlation (CCC) model of Bollerslev (1990) is obtained by setting $\boldsymbol{R}_t = \boldsymbol{R}$.

2.2 Direct specification of dynamic conditional betas

The specificity of the previous approach is to deduce the conditional betas from an estimate of $\boldsymbol{\Sigma}_t$. Our approach is radically different because it allows to directly model these conditional betas. It is

based on a natural orthogonalization of $\boldsymbol{\epsilon}_t$, initially proposed by Pourahmadi (1999) in the case of longitudinal data. We extend his work by allowing the slope coefficients to vary over time using an observation-driven model with an autoregressive structure.

We follow Tsay (2010, Chapter 7) and introduce recursively the orthogonal factors obtained from $\boldsymbol{\epsilon}_t$. Let $\ell_{21,t}$ be the time-varying coefficient in the regression of ϵ_{2t} on $v_{1t} := \epsilon_{1t}$, conditional on $\mathcal{F}_{t-1}$. One can write

$$\epsilon_{2t} = \ell_{21,t} v_{1t} + v_{2t} = \beta_{21,t}\epsilon_{1t} + v_{2t},$$

where $\beta_{21,t} = \ell_{21,t} \in \mathcal{F}_{t-1}$ is the conditional beta in the regression of ϵ_{2t} on ϵ_{1t}, and v_{2t} is orthogonal to ϵ_{1t} conditionally on $\mathcal{F}_{t-1}$. More generally, we have

$$\epsilon_{it} = \sum_{j=1}^{i-1} \ell_{ij,t} v_{jt} + v_{it} = \sum_{j=1}^{i-1} \beta_{ij,t}\epsilon_{jt} + v_{it}, \qquad \text{for } i = 2, \ldots, m, \tag{2.3}$$

where v_{it} is uncorrelated to $v_{1t}, \ldots, v_{i-1,t}$, and thus uncorrelated to $\epsilon_{1t}, \ldots, \epsilon_{i-1,t}$, conditionally on $\mathcal{F}_{t-1}$. In particular, we obtain

$$\beta_{31,t} = \ell_{31,t} - \ell_{32,t}\ell_{21,t}, \qquad \beta_{32,t} = \ell_{32,t}.$$

In matrix form, (2.2) can be written

$$\boldsymbol{\epsilon}_t = \boldsymbol{L}_t \boldsymbol{v}_t \qquad \text{and} \qquad \boldsymbol{B}_t \boldsymbol{\epsilon}_t = \boldsymbol{v}_t,$$

where $\boldsymbol{L}_t$ and $\boldsymbol{B}_t = \boldsymbol{L}_t^{-1}$ are lower unitriangular (i.e., lower triangular with 1 on the diagonal) matrices, with $\ell_{ij,t}$ (resp. $-\beta_{ij,t}$) at the row i and column j of $\boldsymbol{L}_t$ (resp. $\boldsymbol{B}_t$) for $i > j$. For instance, for $m = 3$, we have

$$\boldsymbol{L}_t = \begin{pmatrix} 1 & 0 & 0 \\ \ell_{21,t} & 1 & 0 \\ \ell_{31,t} & \ell_{32,t} & 1 \end{pmatrix} \text{ and } \boldsymbol{B}_t = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_{21,t} & 1 & 0 \\ -\beta_{31,t} & -\beta_{32,t} & 1 \end{pmatrix}.$$

The vector $\boldsymbol{v}_t = (v_{1t}, \ldots, v_{mt})'$ can be interpreted as a vector of orthogonal factors, whose covariance matrix $\boldsymbol{G}_t$ is therefore diagonal but not necessarily constant over time, i.e., $\boldsymbol{G}_t = \text{diag}\,(g_{1t}, \ldots, g_{mt})$ with $g_{it} > 0$ a.s., for $i = 1, \ldots, m$ (since $\boldsymbol{\Sigma}_t$ is positive definite).[2]

[2] We follow the R language convention that $\text{diag}(\boldsymbol{v})$ is a diagonal matrix with the vector $\boldsymbol{v}$ on the diagonal, and $\text{diag}(\boldsymbol{M})$ is the vector of the diagonal of $\boldsymbol{M}$, when $\boldsymbol{M}$ is a square matrix.

We end up with the Cholesky decomposition of Pourahmadi (1999), i.e.,

$$\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\Sigma}_t = \text{Var}(\boldsymbol{L}_t\boldsymbol{v}_t) = \boldsymbol{L}_t\boldsymbol{G}_t\boldsymbol{L}_t', \tag{2.4}$$

which is an alternative to (2.1). Using (2.2), (2) can be rewritten as

$$\boldsymbol{\epsilon}_t = \boldsymbol{L}_t\boldsymbol{G}_t^{1/2}\boldsymbol{\eta}_t,$$

where $(\boldsymbol{\eta}_t)$ is defined as above.

Interestingly, when $i = m$, (2.2) corresponds to the regression of y_t on $\mathbf{x}_t$ with time-varying coefficients (in the case of demeaned series), i.e., $\epsilon_{mt} = \sum_{j=1}^{m-1} \beta_{mj,t}\epsilon_{jt} + v_{mt}$, that are directly comparable with those obtained indirectly with the DCB model of Engle (2016). However, our approach allows to directly specify a model for the $\beta_{ij,t}$ coefficients without having to impose any constraint appart from the stationarity constraint discussed in Section 3.

2.3 Examples of parameterization

An attractive feature of the Cholesky-GARCH (CHAR) model is that the dynamics of $\boldsymbol{\epsilon}_t$ can be defined by specifying successively the dynamics of the vector $\boldsymbol{v}_t$ of the orthogonal factors and the dynamics of the vector $\boldsymbol{\ell}_t = \text{vech}^0\boldsymbol{L}_t$ of the subdiagonal elements of $\boldsymbol{L}_t$, or alternatively the dynamics of $\boldsymbol{\beta}_t = -\text{vech}^0\boldsymbol{B}_t$.[3]

We illustrate our general framework by first considering, for the dynamics of this specification, the Extended Constant Conditional correlation GARCH model studied by Jeantheau (1998), Ling and McAleer (2003), He and Teräsvirta (2004), Aue, Hörmann, Horváth, and Reimherr (2009), Francq and Zakoïan (2010) and Pedersen (2017), among others. This model assumes that

$$\boldsymbol{g}_t = \boldsymbol{\omega}_0 + \sum_{i=1}^{q} \boldsymbol{A}_{0i}\boldsymbol{v}_{t-i}^2 + \sum_{j=1}^{p} \boldsymbol{B}_{0j}\boldsymbol{g}_{t-j}, \tag{2.5}$$

where $\boldsymbol{g}_t = (g_{1t}, \ldots, g_{mt})'$. To guarantee strict positivity of the components of $\boldsymbol{g}_t$, it is assumed that $\boldsymbol{\omega}_0$ is a vector of strictly positive coefficients, and that $\boldsymbol{A}_{0i}$ and $\boldsymbol{B}_{0j}$ are matrices of positive coefficients. Model (2.3) is an extension of the initial model introduced by Bollerslev (1990) for which the matrices $\boldsymbol{A}_{0i}$ and $\boldsymbol{B}_{0j}$ are assumed to be diagonal. The variances defined by (2.3) are functions of squared returns, and thus the model is not able to take into account the leverage effect

[3] $\overline{\text{vech}^0}$ denotes the operator stacking the sub-diagonal elements of a square matrix.

commonly observed on financial series of daily returns, i.e., the fact that negative returns tend to have an higher impact on the future volatility that positive returns of the same magnitude. An extension of (2.3) allowing asymmetric responses of positive and negative past returns on the future volatility is the Asymmetric GARCH (AGARCH) model studied by McAleer, Hoti and Chan (2009) and Francq and Zakoian (2012), among others. This model can be seen as a multivariate extension of the GJR model (Glosten, Jaganathan and Runkle, 1993). More precisely, we assume that

$$\boldsymbol{g}_t = \boldsymbol{\omega}_0 + \sum_{i=1}^{q} \left\{ \boldsymbol{A}_{0i,+} \boldsymbol{v}_{t-i}^{2+} + \boldsymbol{A}_{0i,-} \boldsymbol{v}_{t-i}^{2-} \right\} + \sum_{j=1}^{p} \boldsymbol{B}_{0j} \boldsymbol{g}_{t-j}, \tag{2.6}$$

where

$$\boldsymbol{v}_t^{2+} = \left(\left\{ v_{1t}^{+} \right\}^2, \cdots, \left\{ v_{mt}^{+} \right\}^2 \right)', \quad \boldsymbol{v}_t^{2-} = \left(\left\{ v_{1t}^{-} \right\}^2, \cdots, \left\{ v_{mt}^{-} \right\}^2 \right)'$$

with $x^+ = \max(x, 0) = (-x)^-$. To guarantee strict positivity of the components of $\boldsymbol{g}_t$, the parameters $\boldsymbol{\omega}_0$ and $\boldsymbol{B}_{0j}$ are as in (2.3), and $\boldsymbol{A}_{0i,+}$, $\boldsymbol{A}_{0i,-}$ are matrices of positive coefficients.

Note that for the above-mentioned traditional multivariate GARCH processes, a model of the form (2.3) or (2.3) is applied to the vector of the individual volatilities of $\boldsymbol{\epsilon}_t$ in a first step. In a second step, the model is completed by specifying either a fixed or a time-varying conditional correlation, leading to the classification of the CCC- and DCC-GARCH models (see Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009), Francq and Zakoïan (2010, Chapter 11), Tsay (2010, Chapter 7), Bauwens, Hafner and Laurent (2012) for general references). Models for the conditional correlation are complicated, in particular because the modeling of conditional correlations require imposing strong restrictions on the type of dynamics and in some cases the use of a non-linear transformation to recover the conditional correlations from quasi conditional correlations. This is the reason why no complete asymptotic theory is currently available for estimating DCC-GARCH models.

The Cholesky-GARCH approach avoids specifying a correlation structure but requires specifying how conditionally orthogonal factors $\boldsymbol{v}_t$ are obtained from $\boldsymbol{\epsilon}_t$ as well as the dynamics of their conditional variances. Moreover, the Cholesky-GARCH approach presents the strong advantage that the vector $\boldsymbol{\ell}_t$, of size $m_0 = m(m-1)/2$, is not constrained. One can thus consider a model of the form

$$\boldsymbol{\ell}_t = \boldsymbol{c}_0 \left(\boldsymbol{v}_{t-1}, \ldots, \boldsymbol{v}_{t-r}, \boldsymbol{g}_{t-1}^{1/2}, \ldots, \boldsymbol{g}_{t-r}^{1/2} \right) + \sum_{j=1}^{s} \boldsymbol{C}_{0j} \boldsymbol{\ell}_{t-j}, \tag{2.7}$$

where $\boldsymbol{c}_0$ is a general measurable function from $\mathbb{R}^m \times \cdots \times \mathbb{R}^m$ to $\mathbb{R}^{m_0}$ and where the $\boldsymbol{C}_{0j}$'s are $m_0 \times m_0$ matrices of coefficients, in particular not constrained to be positive. Alternatively to (2.3), one could specify a dynamic model directly on the beta's:

$$\boldsymbol{\beta}_t = \boldsymbol{c}_0\left(\boldsymbol{v}_{t-1}, \ldots, \boldsymbol{v}_{t-r}, \boldsymbol{g}_{t-1}^{1/2}, \ldots, \boldsymbol{g}_{t-r}^{1/2}\right) + \sum_{j=1}^{s} \boldsymbol{C}_{0j}\boldsymbol{\beta}_{t-j}. \tag{2.8}$$

Stationarity conditions are derived in Section 3 for Models (2.3)-(2.3) and (2.3)-(2.3) while consistency and asymptotic normality of the Full QMLE of these models are proved under very general assumptions in Section 4.

To get more explicit conditions for stationarity and invertibility of the CHAR model as well as consistency and asymptotic normality of the Full and EbE QMLE of this model, we consider a more specific parametrization than (2.3)-(2.3).

For the conditional variance of the factors we assume that

$$g_{it} = \omega_{0i} + \gamma_{0i+}\left(\epsilon_{1,t-1}^{+}\right)^2 + \gamma_{0i-}\left(\epsilon_{1,t-1}^{-}\right)^2 + \sum_{k=2}^{i} \alpha_{0i}^{(k)} v_{k,t-1}^2 + b_{0i} g_{i,t-1} \tag{2.9}$$

for $i = 1, \ldots, m$, with positivity constraints on the coefficients. By convention, any term of the form $\sum_{k=2}^{i}$ vanishes when $i < 2$. Note that the volatility of the i-th factor can depend on the past values of the factors of indices $k \leq i$. This model allows volatility spillovers between the factors. Note also that, although the factors are, by definition, conditionally uncorrelated, and thus marginally uncorrelated when second-order moments exist, they are not independent.

For the conditional betas, we adopt the following specification:

$$\beta_{ij,t} = \varpi_{0ij} + \varsigma_{0ij+}\epsilon_{1,t-1}^{+} + \varsigma_{0ij-}\epsilon_{1,t-1}^{-} + \sum_{k=2}^{i} \tau_{0ij}^{(k)} v_{k,t-1} + c_{0ij}\beta_{ij,t-1} \tag{2.10}$$

for any index (i, j) belonging to the set

$$\mathfrak{T}_m = \{(i, j) : i = 2, \ldots, m \text{ and } j = 1, \ldots, i-1\}.$$

Importantly, no further positivity constraint is required in (2.3) to ensure the positivite-definiteness of $\boldsymbol{\Sigma}_t$. The same equation could be considered for the $\ell_{ij,t}$'s but we decided not to study in detail such a specification because the main objective of this paper is to model conditional betas. Note that Tsay (2010) uses a specification where $\ell_{ij,t}$ depends on a constant, $\ell_{ij,t-1}$ and $\epsilon_{i,t-1}$. However, the statistical properties of this model have not been studied yet in the literature.

Equations (2.3)-(2.3) defines a triangular system in which the dynamics of the i-th row depends only on the dynamics of the previous rows. We will see in Subsection 4.2 that this structure facilitates the estimation of the system. In (2.3)-(2.3), the asymmetric effects are introduced via only the first factor, which plays a particular role in our applications. This allows keeping a reasonable number of parameters, and avoids differentiability problems entailed by the fact that the latent factors $v_{k,t}$ depend on parameters when $k > 1$.

We also consider another specification of the CHAR model for which the EbE QMLE can be parallelized, i.e.,

$$g_{it} = \omega_{0i} + \gamma_{0i+}\left(\epsilon^{+}_{1,t-1}\right)^2 + \gamma_{0i-}\left(\epsilon^{-}_{1,t-1}\right)^2 + \alpha_{0i}v^2_{i,t-1} + b_{0i}g_{i,t-1}, \tag{2.11}$$

$$\beta_{ij,t} = \varpi_{0ij} + \varsigma_{0ij+}\epsilon^{+}_{1,t-1} + \varsigma_{0ij-}\epsilon^{-}_{1,t-1} + \tau_{0ij}v_{i,t-1} + \xi_{0ij}v_{i,t-1}v_{1,t-1} + c_{0ij}\beta_{ij,t-1}, \tag{2.12}$$

where the terms $\alpha_{0i}v^2_{i,t-1}$ and $\tau_{0ij}v_{i,t-1} + \xi_{0ij}v_{i,t-1}v_{1,t-1}$ appear only when $i > 1$. Because of the presence of the product term $v_{i,t-1}v_{1,t-1}$, (2.3) is not a particular case of (2.3).

3 Stationarity of the CHAR model

The objective of this section is to study the stationarity conditions for the CHAR model. Proofs are gathered in Section 8. Let $\boldsymbol{\eta}^{2+}_t = \left(\left\{\eta^{+}_{1t}\right\}^2, \cdots, \left\{\eta^{+}_{mt}\right\}^2\right)'$ and $\boldsymbol{\eta}^{2-}_t = \left(\left\{\eta^{-}_{1t}\right\}^2, \cdots, \left\{\eta^{-}_{mt}\right\}^2\right)'$, where $\boldsymbol{\eta}_t = (\eta_{1t}, \cdots, \eta_{mt})'$. For any integer k and any sequence of vectors $(\boldsymbol{x}_t)$, let $\boldsymbol{x}_{t:t-k} = (\boldsymbol{x}'_t, \ldots, \boldsymbol{x}'_{t-k})'$. With these notations, (2.3) can be rewritten as

$$\boldsymbol{z}_t = \boldsymbol{h}_t + \boldsymbol{H}_t\boldsymbol{z}_{t-1},$$

where $\boldsymbol{z}_t = \left(\boldsymbol{v}^{2+\prime}_{t:(t-q+1)}, \boldsymbol{v}^{2-\prime}_{t:(t-q+1)}, \boldsymbol{g}'_{t:(t-p+1)}\right)'$ and

$$\boldsymbol{h}_t = \left(\boldsymbol{\omega}'_0\boldsymbol{\Upsilon}^{+\prime}_t, 0'_{m(q-1)}, \boldsymbol{\omega}'_0\boldsymbol{\Upsilon}^{-\prime}_t, 0'_{(q-1)m}, \boldsymbol{\omega}'_0, 0'_{(p-1)m}\right)',$$

where $\boldsymbol{\Upsilon}^{+}_t = \text{diag}\left(\boldsymbol{\eta}^{2+}_t\right)$ and $\boldsymbol{\Upsilon}^{-}_t = \text{diag}\left(\boldsymbol{\eta}^{2-}_t\right)$. The matrix $\boldsymbol{H}_t$ depends on $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_t$. For example, when $p = q = 1$ we have

$$\boldsymbol{H}_t = \begin{pmatrix} \boldsymbol{\Upsilon}^{+}_t\boldsymbol{A}_{01,+} & \boldsymbol{\Upsilon}^{+}_t\boldsymbol{A}_{01,-} & \boldsymbol{\Upsilon}^{+}_t\boldsymbol{B}_{01} \\ \boldsymbol{\Upsilon}^{-}_t\boldsymbol{A}_{01,+} & \boldsymbol{\Upsilon}^{-}_t\boldsymbol{A}_{01,-} & \boldsymbol{\Upsilon}^{-}_t\boldsymbol{B}_{01} \\ \boldsymbol{A}_{01,+} & \boldsymbol{A}_{01,-} & \boldsymbol{B}_{01} \end{pmatrix}.$$

The top Lyapunov exponent of the sequence $(\boldsymbol{H}_t)$ is defined by

$$\gamma_0 = \inf_{t\geq 1} \frac{1}{t} E(\log \|\boldsymbol{H}_t \boldsymbol{H}_{t-1} \dots \boldsymbol{H}_1\|)$$

for any norm $\|\cdot\|$. Denote by $\varrho(\boldsymbol{M})$ the spectral radius of a square matrix $\boldsymbol{M}$, and let $\otimes$ be the Kronecker product of matrices.

Theorem 3.1 *Consider the CHAR model* (2)*. A sufficient condition for the existence of a strictly stationary, non anticipative*[4] *and ergodic CHAR process* $(\boldsymbol{\epsilon}_t)_t$ *satisfying Equations* (2.3)-(2.3)*, or Equations* (2.3)-(2.3)*, is i)* $\gamma_0 < 0$ *and ii)* $\det\left\{\boldsymbol{I}_{m_0} - \sum_{i=1}^{s} \boldsymbol{C}_{0i} z^i\right\} \neq 0$ *for all* $|z| \leq 1$*. Suppose in addition that* $\boldsymbol{c}_0$ *satisfies the Hölder condition iii)* $\|\boldsymbol{c}_0(\boldsymbol{x}) - \boldsymbol{c}_0(\boldsymbol{y})\| \leq K \|\boldsymbol{x} - \boldsymbol{y}\|^a$*, for some constant* $K > 0$*,* $a \in (0, 1]$ *and all* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* $\mathbb{R}^{2rm}$*. Then there exists* $s_0 > 0$ *such that* $E\|\boldsymbol{\epsilon}_1\|^{2s_0} < \infty$ *and* $E\|\boldsymbol{\Sigma}_1\|^{s_0} < \infty$*. If* $\gamma_0 \geq 0$ *then there exists no stationary solution to* (2.3)*.*

Let m_1 *be a positive integer. If*

$$E\|\boldsymbol{\eta}_1\|^{2m_1} < \infty \quad \text{and} \quad \varrho(E\boldsymbol{H}_1^{\otimes m_1}) < 1, \tag{3.1}$$

then $E\|\boldsymbol{v}_t\|^{2m_1} < \infty$*. If, in addition, conditions ii) and iii) hold, then* $E\|\boldsymbol{\epsilon}_1\|^{2m_1} < \infty$*.*

For models (2.3)-(2.3) and (2.3)-(2.3), the stationarity conditions are more explicit.

Corollary 3.1 *There exists a strictly stationary, non anticipative and ergodic process satisfying* (2) *with* (2.3)-(2.3) *or* (2.3)–(2.3) *when*

1) $E \log \left\{ \gamma_{01+} \left(\eta_{1,t-1}^{+}\right)^2 + \gamma_{01-} \left(\eta_{1,t-1}^{-}\right)^2 + b_{01} \right\} < 0,$

2) $E \log \left\{ \alpha_{0i}^{(i)} \eta_{it}^2 + b_{0i} \right\} < 0$ *for* $i = 2, \dots, m,$

3) $|c_{0ij}| < 1$ *for all* $(i, j) \in \mathfrak{T}_m$*.*

Moreover, the stationary solution satisfies $E\|\boldsymbol{\epsilon}_1\|^{2s_0} < \infty$*,* $E\|\boldsymbol{g}_1\|^{s_0} < \infty$*,* $E\|\boldsymbol{v}_1\|^{s_0} < \infty$*,* $E\|\boldsymbol{\beta}_1\|^{s_0} < \infty$ *and* $E\|\boldsymbol{\Sigma}_1\|^{s_0} < \infty$ *for some* $s_0 > 0$*.*

4 One-step and multi-step QMLE

The parameter $\boldsymbol{\vartheta}_0$ can be estimated by the standard Gaussian QMLE. For models with a "triangular structure", such as (2.3)-(2.3), a numerically more convenient equation-by-equation estimator (EbEE) can also be used. Proofs are gathered in Section 8.

[4] i.e $\boldsymbol{\epsilon}_t$ measurable with respect to the sigma-field generated by $\{\boldsymbol{\eta}_u, u \leq t\}$.

4.1 Full QMLE

Let Θ be a compact parameter space which contains $\boldsymbol{\vartheta}_0$. For all $\boldsymbol{\vartheta} \in \Theta$, assume that the variance $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})$ is well-defined, and consider its Cholesky decomposition $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}) = \boldsymbol{L}_t(\boldsymbol{\vartheta})\boldsymbol{G}_t(\boldsymbol{\vartheta})\boldsymbol{L}_t'(\boldsymbol{\vartheta})$. We also introduce the vector $\boldsymbol{g}_t(\boldsymbol{\vartheta})$ of generic element $g_{it}(\boldsymbol{\vartheta})$ such that $\boldsymbol{g}_t = \boldsymbol{g}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{g}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\vartheta}_0)$, and the vectors $\boldsymbol{\beta}_t(\boldsymbol{\vartheta})$ and $\boldsymbol{\ell}_t(\boldsymbol{\vartheta})$ such that $\boldsymbol{\beta}_t = \boldsymbol{\beta}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\beta}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\vartheta}_0)$ and $\boldsymbol{\ell}_t = \boldsymbol{\ell}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\ell}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\vartheta}_0)$. Given observations $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$, and arbitrary fixed initial values $\widetilde{\boldsymbol{\epsilon}}_i$ for $i \leq 0$, let the statistics

$$\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\vartheta}) = \boldsymbol{\Sigma}\left(\boldsymbol{\epsilon}_{t-1}, \ldots, \boldsymbol{\epsilon}_1, \widetilde{\boldsymbol{\epsilon}}_0, \widetilde{\boldsymbol{\epsilon}}_{-1}, \ldots; \boldsymbol{\vartheta}\right)$$

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

$$\widehat{\boldsymbol{\vartheta}}_n = \arg\min_{\boldsymbol{\vartheta}\in\Theta} \widetilde{O}_n(\boldsymbol{\vartheta}), \qquad \widetilde{O}_n(\boldsymbol{\vartheta}) = n^{-1}\sum_{t=1}^{n} \widetilde{q}_t(\boldsymbol{\vartheta}), \tag{4.1}$$

where

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

It is worth noting that for computing the QMLE, it is not necessary to inverse matrices ($\widetilde{\boldsymbol{G}}_t(\boldsymbol{\vartheta})$ being diagonal). This constitutes an attractive feature of the CHAR models.

4.1.1 General CHAR model

We now give high-level assumptions which entail the strong consistency and asymptotic normality (CAN) of the QML estimator for the general CHAR model. These assumptions will be made explicit for the particular model (2.3)-(2.3). In the sequel ρ denotes a generic constant belonging to $[0, 1)$, and K denotes a positive constant or a positive random variable measurable with respect to $\{\boldsymbol{\epsilon}_u, u < 0\}$ (and thus which does not depend on n).

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

A2: $\sup_{\boldsymbol{\vartheta}\in\Theta}\left\{\|\boldsymbol{g}_t(\boldsymbol{\vartheta}) - \widetilde{\boldsymbol{g}}_t(\boldsymbol{\vartheta})\| + \|\boldsymbol{\beta}_t(\boldsymbol{\vartheta}) - \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\vartheta})\|\right\} \leq K\rho_t$ where the random variable ρ_t satisfies $\sum_{t=1}^{\infty}\{E\rho_t^{s_1}\}^{1/3} < \infty$ for all $s_1 \in (0, s_0]$ and some $s_0 > 0$.

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

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

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

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

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

A8: For some neighbourhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$, there exists $s_0 > 0$ such that

$$\sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left\{ \left\| \frac{\partial \boldsymbol{g}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} - \frac{\partial \widetilde{\boldsymbol{g}}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\| + \left\| \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} - \frac{\partial \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right\| \right\} \leq K \rho_t$$

with ρ_t as in **A2**.

A9: For some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$, 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

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

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

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

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

Theorem 4.1 (CAN of the QMLE in the general CHAR case) *Let $(\boldsymbol{\epsilon}_t)$ be a non anticipative solution to the CHAR model* (2)*. Let $(\widehat{\boldsymbol{\vartheta}}_n)$ be a sequence of QML estimators satisfying (4.1). Under* **A1**-**A5** *we have*

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

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

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

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

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

with $\boldsymbol{K} = Evec(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')vec'(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')$ and

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

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

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

We also have the Bahadur representation

$$\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0 = \boldsymbol{J}^{-1}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\nabla}_t vec(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - \boldsymbol{I}_m) + o_P(n^{-1/2}), \tag{4.2}$$

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

4.1.2 QMLE for a specific parameterization

Now consider the case where $(\boldsymbol{\epsilon}_t)$ is a stationary and ergodic solution to model (2.3)-(2.3), whose unknown parameter $\boldsymbol{\vartheta}_0$ belongs to a compact set Θ of $\mathbb{R}^d$, with $d = m(m+1)(m+5)/3$. Let $\boldsymbol{\theta}_0^{(1)} = (\omega_{01}, \gamma_{01+}, \gamma_{01-}, b_{01})'$ and, for $i = 2, \ldots, m$, let

$$\boldsymbol{\theta}_0^{(i)} = (\omega_{0i}, \gamma_{0i+}, \gamma_{0i-}, \alpha_{0i}^{(2)}, \ldots, \alpha_{0i}^{(i)}, b_{0i})'$$

be the vector of the $i+3$ parameters involved in g_{it}. Assume that $\boldsymbol{\theta}_0^{(i)}$ belongs to a compact set $\Theta_{\boldsymbol{\theta}}^{(i)} \subset (0,\infty) \times [0,\infty)^{i+2}$. For $(i,j) \in \mathfrak{T}_m$, set

$$\boldsymbol{\varphi}_0^{(ij)} = (\varpi_{0ij}, \varsigma_{0ij+}, \varsigma_{0ij-}, \tau_{0ij}^{(2)}, \ldots, \tau_{0ij}^{(i)}, c_{0ij})',$$

and $\boldsymbol{\varphi}_0^{(i)} = (\boldsymbol{\varphi}_0^{(i1)'}, \ldots, \boldsymbol{\varphi}_0^{(i,i-1)'})'$. For $i \geq 2$, let $\Theta_{\boldsymbol{\varphi}}^{(i)}$ be the parameter space of $\boldsymbol{\varphi}_0^{(i)}$, a compact subspace of $\mathbb{R}^{(i-1)(i+3)}$. The parameter $\boldsymbol{\varphi}_0 := (\boldsymbol{\varphi}_0^{(2)'}, \ldots, \boldsymbol{\varphi}_0^{(m)'})'$ belongs to the compact subset $\Theta_{\boldsymbol{\varphi}} = \Theta_{\boldsymbol{\varphi}}^{(2)} \times \cdots \times \Theta_{\boldsymbol{\varphi}}^{(m)}$ of the Euclidean space of dimension $m(m-1)(2m+11)/6$. Let $\boldsymbol{\vartheta}_0^{(1)} = \boldsymbol{\theta}_0^{(1)}$ and $\boldsymbol{\vartheta}_0^{(i)} = (\boldsymbol{\theta}_0^{(i)'}, \boldsymbol{\varphi}_0^{(i)'})'$ for $i = 2, \ldots, m$. The parameter space $\Theta^{(i)} = \Theta_{\boldsymbol{\theta}}^{(i)} \times \Theta_{\boldsymbol{\varphi}}^{(i)}$ of $\boldsymbol{\vartheta}_0^{(i)}$ (with the convention $\Theta^{(1)} = \Theta_{\boldsymbol{\theta}}^{(1)}$) is a compact subset of $\mathbb{R}^{d_i}$, with $d_i = i(i+3)$. We have $\boldsymbol{\vartheta}_0 = (\boldsymbol{\vartheta}_0^{(1)'}, \ldots, \boldsymbol{\vartheta}_0^{(m)'})'$. Let $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}^{(1)'}, \ldots, \boldsymbol{\vartheta}^{(m)'})'$ be a generic element of the parameter space Θ and

$\boldsymbol{\varphi} = (\boldsymbol{\varphi}^{(2)'}, \ldots, \boldsymbol{\varphi}^{(m)'})'$ a generic element of Θ_φ. Using (2.2) to compute $v_{k,t-1}$ in (2.3), one can see that $\beta_{ij,t}$ actually depends on

$$\boldsymbol{\varphi}_0^{(+i)} = \left(\boldsymbol{\varphi}_0^{(i)'}, \boldsymbol{\varphi}_0^{(-i)'}\right)', \qquad \boldsymbol{\varphi}_0^{(-i)} = \left(\boldsymbol{\varphi}_0^{(i-1)'}, \ldots, \boldsymbol{\varphi}_0^{(2)'}\right)',$$

with the convention $\boldsymbol{\varphi}_0^{(+2)} = \boldsymbol{\varphi}_0^{(2)}$. We thus consider the function $\beta_{ij,t}(\boldsymbol{\varphi}^{(+i)})$ such that $\beta_{ij,t} = \beta_{ij,t}(\boldsymbol{\varphi}_0^{(+i)})$. Introduce also the notation

$$\boldsymbol{\vartheta}_0^{(+i)} = \left(\boldsymbol{\vartheta}_0^{(i)'}, \boldsymbol{\varphi}_0^{(-i)'}\right)' = \left(\boldsymbol{\theta}_0^{(i)'}, \boldsymbol{\varphi}_0^{(+i)'}\right)', \tag{4.3}$$

dropping the subscript "0" when considering a generic element of a parameter space denoted by $\Theta^{(+i)}$, and using the convention $\boldsymbol{\vartheta}^{(+i)} = \boldsymbol{\vartheta}^{(i)}$ for $i = 1, 2$. The QMLE of $\boldsymbol{\vartheta}_0$ is the solution to (4.1) where, with some abuse of notation,

$$\widetilde{q}_t(\boldsymbol{\vartheta}) = \sum_{i=1}^{m} \widetilde{q}_{it}(\boldsymbol{\vartheta}^{(+i)}), \tag{4.4}$$

with, for $t = 1, \ldots, n$,

$$\widetilde{q}_{it}(\boldsymbol{\vartheta}^{(+i)}) = \frac{\widetilde{v}_{it}^2(\boldsymbol{\varphi}^{(+i)})}{\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)})} + \log \widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)}), \tag{4.5}$$

$$\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)}) = \omega_{i,t-1} + \sum_{k=2}^{i} \alpha_i^{(k)} \widetilde{v}_{k,t-1}^2(\boldsymbol{\varphi}^{(+k)}) + b_i \widetilde{g}_{i,t-1}(\boldsymbol{\vartheta}^{(+i)}), \tag{4.6}$$

$$\widetilde{v}_{kt}(\boldsymbol{\varphi}^{(+k)}) = \epsilon_{kt} - \sum_{j=1}^{k-1} \widetilde{\beta}_{kj,t}(\boldsymbol{\varphi}^{(+k)}) \epsilon_{jt}, \tag{4.7}$$

$$\widetilde{\beta}_{ij,t}(\boldsymbol{\varphi}^{(+i)}) = \omega_{ij,t-1} + \sum_{k=2}^{i} \tau_{ij}^{(k)} \widetilde{v}_{k,t-1}(\boldsymbol{\varphi}^{(+k)}) + c_{ij} \widetilde{\beta}_{ij,t-1}(\boldsymbol{\varphi}^{(+i)}), \tag{4.8}$$

using the convention $\widetilde{v}_{1t}^2(\boldsymbol{\varphi}^{(+i)}) = \epsilon_{1t}^2$, the notations $\omega_{it} = \omega_i + \gamma_{i+} \left(\epsilon_{1t}^+\right)^2 + \gamma_{i-} \left(\epsilon_{1t}^-\right)^2$ and $\omega_{ijt} = \varpi_{ij} + \varsigma_{ij+} \epsilon_{1t}^+ + \varsigma_{ij-} \epsilon_{1t}^-$, and fixed initial values for $\widetilde{\beta}_{ij,0}(\cdot)$, $\boldsymbol{\epsilon}_0$ and $\widetilde{g}_{i,0}(\cdot)$.

4.1.3 Invertibility condition

Starting with the initial values and using (4.1.2) for $t = 0$ and successively (4.1.2) and (4.1.2) for $t = 1, 2, \ldots$, one can define $\widetilde{\beta}_{ij,t}(\boldsymbol{\varphi}^{(+i)})$ for any value of $\boldsymbol{\varphi}^{(+i)}$. One can expect that the initial values be asymptotically negligible, in the sense that, as $t \to \infty$, $\widetilde{\beta}_{ij,t}(\boldsymbol{\varphi}^{(+i)})$ be arbitrarily close to some function $\beta_{ij,t}(\boldsymbol{\varphi}^{(+i)})$ measurable with respect to the past observations $\mathcal{F}_{t-1}$. From (4.1.2), we then consistently approximate $\widetilde{v}_{it}(\boldsymbol{\varphi}^{(+i)})$ by a measurable function $v_{it}(\boldsymbol{\varphi}^{(+i)})$ of $\mathcal{F}_{t-1}$. By imposing

$|b_i| < 1$ in (4.1.2), one can also consistently approximate $\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)})$ by some function $g_{it}(\boldsymbol{\vartheta}^{(+i)})$ of $\mathcal{F}_{t-1}$. When these consistent $\mathcal{F}_{t-1}$-approximations hold for all values of i and j, the model (2.3)-(2.3) is said to be invertible. For prediction or estimation purposes, it is important that the model possesses this property. Indeed, if the invertibility does not hold, the approximate volatility $\widetilde{\boldsymbol{\Sigma}}_t(\boldsymbol{\vartheta})$ is likely to depend strongly on the initial values, even for large t, and thus the model would provide unstable predictions that depend much on the initial values. The noninvertibility would also entail estimation inconsistency, because the objective functions $\widetilde{O}_n^{(i)}(\boldsymbol{\vartheta}^{(+i)})$ could asymptotically depend on the initial values. Problems of invertibility have been documented in Wintenberger (2013) for the EGARCH model of Nelson (1991) and more recently Blasques et al. (2016) for the Beta-t GARCH of Harvey (2013). In both models, the conditional variance (or its log) depends in a very nonlinear way on past innovations.

We now give conditions for the uniform invertibility of Model (2.3)-(2.3).

For all $\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}$, let

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) = \left(\widetilde{\beta}_{21,t}(\boldsymbol{\varphi}^{(2)}), \widetilde{\beta}_{31,t}(\boldsymbol{\varphi}^{(+3)}) \ldots, \widetilde{\beta}_{m,m-1,t}(\boldsymbol{\varphi}^{(+m)})\right)' = -\text{vech}^0 \widetilde{\boldsymbol{B}}_t(\boldsymbol{\varphi}),$$

where $\widetilde{\boldsymbol{B}}_t(\boldsymbol{\varphi})$ is a lower unitriangular matrix. In vectorial form, the equations (4.1.2)-(4.1.2) write

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) = \boldsymbol{w}_{t-1} + \boldsymbol{T}\widetilde{\boldsymbol{B}}_{t-1}(\boldsymbol{\varphi})\boldsymbol{\epsilon}_{t-1} + \boldsymbol{C}\widetilde{\boldsymbol{\beta}}_{t-1}(\boldsymbol{\varphi}), \qquad t \geq 1, \tag{4.9}$$

with a fixed initial values for $\widetilde{\boldsymbol{\beta}}_0(\boldsymbol{\varphi})$. If the ℓ-th element of $\boldsymbol{\beta}_t$ contains $\beta_{ij,t}$, then the ℓ-th element of the random vector $\boldsymbol{w}_t$ is ω_{ijt}, row ℓ of $\boldsymbol{T}$ is $(0, \tau_{ij}^{(2)}, \ldots, \tau_{ij}^{(i)}, \boldsymbol{0}'_{m-i})$, and $\boldsymbol{C}$ is a diagonal matrix with c_{ij} as ℓ-th diagonal element. Let $\mathbf{D}_m^0$ be the matrix with elements 0 or 1 such that $\text{vec}\boldsymbol{A} = \mathbf{D}_m^0 \text{vech}^0 \boldsymbol{A}$ for any strictly lower triangular $m \times m$ matrix $\boldsymbol{A}$. Noting that $\text{vec}\widetilde{\boldsymbol{B}}_t(\boldsymbol{\varphi}) = -\mathbf{D}_m^0 \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) + \text{vec}\boldsymbol{I}_m$, we obtain

$$\boldsymbol{T}\widetilde{\boldsymbol{B}}_t(\boldsymbol{\varphi})\boldsymbol{\epsilon}_t = \boldsymbol{T}\boldsymbol{\epsilon}_t - (\boldsymbol{\epsilon}_t' \otimes \boldsymbol{T})\mathbf{D}_m^0 \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi})$$

and

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) = \boldsymbol{w}_{t-1}^* + \sum_{i=1}^{t-1} \left(\prod_{j=1}^{i} \boldsymbol{S}_{t-j}\right) \boldsymbol{w}_{t-i-1}^* + \left(\prod_{i=1}^{t-1} \boldsymbol{S}_{t-i}\right) \left(\boldsymbol{w}_0^* + \widetilde{\boldsymbol{\beta}}_0(\boldsymbol{\varphi})\right),$$

with $\boldsymbol{w}_t^* = \boldsymbol{w}_t + \boldsymbol{T}\boldsymbol{\epsilon}_t$ and $\boldsymbol{S}_t = \boldsymbol{S}_t(\boldsymbol{\varphi}) = \boldsymbol{C} - (\boldsymbol{\epsilon}_t' \otimes \boldsymbol{T})\mathbf{D}_m^0$. By the Cauchy rule, under the conditions

$$E \log^+ \sup_{\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}} \|\boldsymbol{w}_1 + \boldsymbol{T}\boldsymbol{\epsilon}_1\| < \infty, \tag{4.10}$$

$$\gamma_S := \limsup_{n \to \infty} \frac{1}{n} \log \sup_{\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}} \left\| \prod_{i=1}^{n} \boldsymbol{S}_{t-i} \right\| < 0, \tag{4.11}$$

where $\|\cdot\|$ denotes any multiplicative norm, the series

$$\boldsymbol{\beta}_t(\boldsymbol{\varphi}) = \boldsymbol{w}^*_{t-1} + \sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \boldsymbol{S}_{t-j} \right) \boldsymbol{w}^*_{t-i-1} \tag{4.12}$$

converges absolutely (and uniformly) with probability one. Moreover,

$$\sup_{\boldsymbol{\varphi}\in\Theta_{\boldsymbol{\varphi}}} \left\| \boldsymbol{\beta}_t(\boldsymbol{\varphi}) - \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) \right\| = \sup_{\boldsymbol{\varphi}\in\Theta_{\boldsymbol{\varphi}}} \left\| \left(\prod_{i=1}^{t-2} \boldsymbol{S}_{t-i} \right) \left\{ \boldsymbol{\beta}_1(\boldsymbol{\varphi}) - \widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\varphi}) \right\} \right\| \leq K\rho^t, \tag{4.13}$$

where K is a positive random variable measurable with respect to $\{\boldsymbol{\epsilon}_u, u \leq 0\}$ and $0 \leq \exp\{\gamma_S(\boldsymbol{\varphi})\} < \rho < 1$ (see Bougerol and Picard, 1992a and 1992b for conditions under which (4.1.3) is not only sufficient but also necessary for the existence of (4.1.3)).

Note that for any $s_0 > 0$, there exists $K > 0$ such that $\log^+ \|\boldsymbol{w}_1\| \leq K + \|\boldsymbol{w}_1 + \boldsymbol{T}\boldsymbol{\epsilon}_1\|^{s_0}$. Therefore $E \log^+ \|\boldsymbol{w}_1 + \boldsymbol{T}\boldsymbol{\epsilon}_1\| < \infty$ if $E\|\boldsymbol{\epsilon}_1\|^{s_0} < \infty$ for some $s_0 > 0$, regardless of the value of $\boldsymbol{\varphi}$. Consequently, in view of Corollary 3.1, Condition (4.1.3) does not constrain the compact parameter space $\Theta_{\boldsymbol{\varphi}}$. By contrast, (4.1.3) entails non explicit constraints on $\boldsymbol{\varphi}$. By Jensen's inequality, the condition $\gamma_S < 0$ is satisfied if $E \sup_{\boldsymbol{\varphi}\in\Theta_{\boldsymbol{\varphi}}} \|\boldsymbol{S}_1\| < 1$ for some multiplicative norm $\|\cdot\|$. Take the spectral norm, defined by $\|\boldsymbol{A}\|_2 = \sqrt{\varrho(\boldsymbol{A}^*\boldsymbol{A})}$, with ϱ the spectral radius and other standard notations. Noting that $\left\|\mathbf{D}^0_m\right\|_2 = \sqrt{2}$ and $\|\boldsymbol{\epsilon}'_t \otimes \boldsymbol{T}\|_2 = \sqrt{\varrho(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t \otimes \boldsymbol{T}'\boldsymbol{T})} = \sqrt{\boldsymbol{\epsilon}'_t\boldsymbol{\epsilon}_t}\,\|\boldsymbol{T}\|_2$, and using the compactness of $\Theta_{\boldsymbol{\varphi}}$, we obtain that the uniform invertibility holds if

$$\max_{i,j} |c_{ij}| + \sqrt{2}\,\|\boldsymbol{T}\|_2\, E\sqrt{\boldsymbol{\epsilon}'_1\boldsymbol{\epsilon}_1} < 1, \qquad \forall \boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}. \tag{4.14}$$

Note that, although slightly more explicit, Condition (4.1.3) is more restrictive than Condition (4.1.3). In particular, (4.1.3) requires a finite second order moment for $\|\boldsymbol{\epsilon}_t\|$ when $\boldsymbol{T} \neq \mathbf{0}$. Note also that if the term $\sum_{k=2}^{i} \tau^{(k)}_{0ij} v_{k,t-1}$ does not appear in (2.3), then the model is invertible, i.e., (4.1.3) holds, if $\max_{i,j} |c_{ij}| < 1$ for all $\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}$.

4.1.4 Asymptotic property of the QMLE

Let us define the following empirical information matrices:

$$\boldsymbol{J}_n = \frac{\partial^2 \widetilde{O}_n(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} \qquad \text{and} \qquad \boldsymbol{I}_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \widetilde{q}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}} \frac{\partial \widetilde{q}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\vartheta}'}.$$

We need the following assumptions.

B1: Conditions 1), 2) and 3) of Corollary 3.1 are satisfied.

B2: For $i = 2, \ldots, m$, the distribution of η_{it}^2 conditionally on $\{\eta_{jt}, j \neq i\}$ is non degenerated. The support of η_{1t} contains at least two positive points and two negative points.

B3: For $(i, j) \in \mathfrak{T}_m$ and all $\boldsymbol{\vartheta}^{(i)} \in \Theta^{(i)}$ we have $\boldsymbol{\theta}^{(i)} \geq 0$ componentwise, $\omega_i > 0$, $|b_i| < 1$ and $|c_{ij}| < 1$.

B4: For $i = 1, \ldots, m$, we have $(\gamma_{0i+}, \gamma_{0i-}, \alpha_{0i}^{(2)}, \ldots, \alpha_{0i}^{(i)}) \neq 0$ and $(\varsigma_{0ij+}, \varsigma_{0ij-}, \tau_{0ij}^{(2)}, \ldots, \tau_{0ij}^{(i)}) \neq 0$.

B5: The uniform invertibility condition (4.1.3) holds true.

B6: There exist three pairs of Hölder conjugate numbers, (p_i, q_i) with $p_i^{-1} + q_i^{-1} = 1$ for $i \in \{1, 2, 3\}$, such that $E \left\| \boldsymbol{\epsilon}_t \right\|^{4p_1 q_2} < \infty$, $E \left\| \boldsymbol{\epsilon}_t \right\|^{2q_1 q_3} < \infty$, and there exists a neighbourhood $V(\boldsymbol{\varphi}_0)$ of $\boldsymbol{\varphi}_0$ such that

$$E \sup_{\boldsymbol{\varphi} \in V(\boldsymbol{\varphi}_0)} \left\| \boldsymbol{\beta}_t(\boldsymbol{\varphi}) \right\|^{4p_1 p_2} < \infty, \quad E \sup_{\boldsymbol{\varphi} \in V(\boldsymbol{\varphi}_0)} \left\| \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}'} \right\|^{2q_1 q_3} < \infty, \tag{4.15}$$

where $\boldsymbol{\beta}_t(\boldsymbol{\varphi})$ is defined by (4.1.3).

In the case where the conditional betas are constant, (4.1.4) holds true for any value of the Hölder conjugates. Therefore one can choose $q_1 = 3$ and q_2, q_3 arbitrarily close to 1, which shows that, in this case, **B6** is satisfied when $\|\boldsymbol{\epsilon}_t\|$ admits a moment larger than 6.

As shown by Corollary 3.1, **B1** ensures the existence of a strictly stationary solution to (2.3)-(2.3). Assumption **B2** is an identifiability assumption. Assumption **B3** guarantees the existence and positivity of $\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(i)})$, as well as the existence of $\widetilde{\beta}_{ij,t}(\boldsymbol{\vartheta}^{(i)})$. The first part of **B4** avoids the well-known identifiability issue $g_{it} = \omega_{0i}/(1 - b_{0i}) = \omega_{0i} - b_{0i} g_{i,t-1}$ that arises when the inequality does not hold. The second part is an analogous identifiability condition for $\beta_{ij,t}$. Assumption **B5** ensures that the initial values are asymptotically negligible.

The moment conditions in **B6** are required for the asymptotic normality (see the example given after the next theorem). It must be acknowledged that they are restrictive and not explicit. Indeed, Theorem 3.1 provides more explicit conditions for even-order moments only. Even for the simplest univariate ARCH processes ϵ_t, the precise conditions for the existence of $E|\epsilon_t|^{p_1}$ are unknown when p_1 is not an even integer. To our knowledge, the only asymptotic theory available for multivariate GARCH models with dynamic conditional correlation has been provided by Comte and Lieberman (2003) and concerns the BEKK model. These authors assume, in particular, a moment of order two

for the consistency and of order eight for the asymptotic normality. Actually, Avarucci, Beutner and Zaffaroni (2013) give evidence that moment conditions are necessary for the CAN of the QMLE in the case of the simplest VEC-GARCH model, which contrasts with the univariate GARCH models for which the CAN can be established without moment conditions on the observed process.

Theorem 4.2 (CAN of the QMLE) *Consider the CHAR model* (2) *satisfying* (2.3)-(2.3). *Under* **B1**-**B5**, *the QMLE* $\widehat{\boldsymbol{\vartheta}}_n$ *defined by* (4.1) *and* (4.1.2)-(4.1.2) *satisfies*

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

Under the additional assumptions **A6**, **A10** *and* **B6**, *the matrices* $\boldsymbol{I}_n$ *and* $\boldsymbol{J}_n$ *converge almost surely to positive-definite matrices* $\boldsymbol{I}$ *and* $\boldsymbol{J}$ *as* $n \to \infty$. *Moreover,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \boldsymbol{\Omega}\right), \quad \boldsymbol{\Omega} = \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}. \tag{4.16}$$

In order to emphasize the need of the moment conditions involved in **B6**, let us consider an elementary bivariate example. Assume $g_{1t} = \omega_{01} + \gamma_{01+}\left(\epsilon^+_{1,t-1}\right)^2 + \gamma_{01+}\left(\epsilon^+_{1,t-1}\right)^2 + b_{01}g_{1,t-1}$, $g_{2t} = \omega_{02}$ and $\beta_{21,t} = \varpi_0 + \varsigma_{0+}\epsilon^+_{1,t-1} + \varsigma_{0-}\epsilon^-_{1,t-1}$. There are 8 unknown parameters and, in view of (8) in the proof of Theorem 4.2, we have

$$I(8,8) = \mathrm{Var}\left(2\eta_{2t}\frac{1}{\sqrt{g_{2t}}}\frac{\partial v_{2t}}{\partial \varsigma_-}\right) = \frac{4}{\omega_{02}}E\epsilon^2_{1,t}\left(\epsilon^-_{1,t-1}\right)^2.$$

Therefore, even in this simplistic example, at least moments of order 4 are required.

4.2 Multi-step QMLE

Recall that (2.3)-(2.3) defines a triangular system in which the dynamics of the i-th row depends only on the dynamics of the previous rows. Therefore, instead of estimating simultaneously the d components of $\boldsymbol{\vartheta}_0$, one can estimate the parameters equation-by-equation (EbE) (see e.g., Francq and Zakoian, 2016 for an EbE estimator in another class of multivariate GARCH models).

4.2.1 Definition of the estimator

For simplicity, the EbE estimator (EbEE) of $\boldsymbol{\vartheta}_0$ is described in the case of Model (2.3)-(2.3), but the estimator can be readily extended to other models.

In a first step we estimate $\boldsymbol{\vartheta}_0^{(1)}$ by

$$\widehat{\boldsymbol{\vartheta}}_n^{(1)} = \underset{\boldsymbol{\vartheta}^{(1)}\in\Theta^{(1)}}{\arg\min}\, \widetilde{O}_n^{(1)}(\boldsymbol{\vartheta}^{(1)}), \qquad \widetilde{O}_n^{(1)}(\boldsymbol{\vartheta}^{(1)}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{q}_{1t}(\boldsymbol{\vartheta}^{(1)}), \tag{4.17}$$

where, according to (4.1.2), $\widetilde{q}_{1t}(\boldsymbol{\vartheta}^{(1)})$ is defined by

$$\widetilde{q}_{1t}(\boldsymbol{\vartheta}^{(1)}) = \frac{\epsilon_{1t}^2}{\widetilde{g}_{1t}(\boldsymbol{\vartheta}^{(1)})} + \log \widetilde{g}_{1t}(\boldsymbol{\vartheta}^{(1)}), \quad \widetilde{g}_{1t}(\boldsymbol{\vartheta}^{(1)}) = \omega_{1,t-1} + b_1 \widetilde{g}_{1,t-1}(\boldsymbol{\vartheta}^{(1)}).$$

The estimation of $\boldsymbol{\vartheta}_0^{(2)}$ can be done independently or in parallel to $\boldsymbol{\vartheta}_0^{(1)}$ by using

$$\widehat{\boldsymbol{\vartheta}}_n^{(2)} = \underset{\boldsymbol{\vartheta}^{(2)}\in\Theta^{(2)}}{\arg\min}\, \widetilde{O}_n^{(2)}(\boldsymbol{\vartheta}^{(2)}), \qquad \widetilde{O}_n^{(2)}(\boldsymbol{\vartheta}^{(2)}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{q}_{2t}(\boldsymbol{\vartheta}^{(2)}), \tag{4.18}$$

where, following again (4.1.2),

$$\begin{aligned}
\widetilde{q}_{2t}(\boldsymbol{\vartheta}^{(2)}) &= \frac{\widetilde{v}_{2t}^2(\boldsymbol{\varphi}^{(2)})}{\widetilde{g}_{2t}(\boldsymbol{\vartheta}^{(2)})} + \log \widetilde{g}_{2t}(\boldsymbol{\vartheta}^{(2)}),\\
\widetilde{g}_{2t}(\boldsymbol{\vartheta}^{(2)}) &= \omega_{2,t-1} + \alpha_2^{(2)} \widetilde{v}_{2,t-1}^2(\boldsymbol{\varphi}^{(2)}) + b_2 \widetilde{g}_{2,t-1}(\boldsymbol{\varphi}^{(2)}),\\
\widetilde{v}_{2t}(\boldsymbol{\varphi}^{(2)}) &= \epsilon_{2t} - \widetilde{\beta}_{21,t}(\boldsymbol{\varphi}^{(2)})\epsilon_{1t},\\
\widetilde{\beta}_{21,t}(\boldsymbol{\varphi}^{(2)}) &= \omega_{21,t-1} + \tau_{21}^{(2)} \widetilde{v}_{2,t-1}(\boldsymbol{\varphi}^{(2)}) + c_{21}\widetilde{\beta}_{21,t-1}(\boldsymbol{\varphi}^{(2)}).
\end{aligned}$$

For $i = 3, \ldots, m$, the estimation of $\boldsymbol{\vartheta}_0^{(i)}$ depends on the estimates $\widehat{\boldsymbol{\varphi}}_n^{(-i)} = \left(\widehat{\boldsymbol{\varphi}}_n^{(i-1)'}, \ldots, \widehat{\boldsymbol{\varphi}}_n^{(2)'}\right)'$ obtained in previous steps (i.e., $j < i$). We then estimate $\boldsymbol{\vartheta}_0^{(i)}$, for $i = 3, \ldots, m$, by

$$\widehat{\boldsymbol{\vartheta}}_n^{(i)} = \underset{\boldsymbol{\vartheta}^{(i)}\in\Theta^{(i)}}{\arg\min}\, \widetilde{O}_n^{(i)} \begin{pmatrix} \boldsymbol{\vartheta}^{(i)} \\ \widehat{\boldsymbol{\varphi}}_n^{(-i)} \end{pmatrix}, \qquad \widetilde{O}_n^{(i)}(\boldsymbol{\vartheta}^{(+i)}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{q}_{it}(\boldsymbol{\vartheta}^{(+i)}), \tag{4.19}$$

where $\widetilde{q}_{it}$ is defined in (4.1.2).

4.2.2 Asymptotic behavior of the EbEE

Let us define the following empirical information matrices:

$$\boldsymbol{J}_n^{(i)} = \frac{\partial^2 \widetilde{O}_n^{(i)}(\widehat{\boldsymbol{\vartheta}}_n^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)} \partial \boldsymbol{\vartheta}^{(i)'}}, \qquad \boldsymbol{I}_n^{(i)} = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \widetilde{q}_{it}(\widehat{\boldsymbol{\vartheta}}_n^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)}} \frac{\partial \widetilde{q}_{it}(\widehat{\boldsymbol{\vartheta}}_n^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)'}}, \qquad \boldsymbol{K}_n^{(i)} = \frac{\partial^2 \widetilde{O}_n^{(i)}(\widehat{\boldsymbol{\vartheta}}_n^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)} \partial \boldsymbol{\varphi}^{(-i)'}}$$

with the notation $\widehat{\boldsymbol{\vartheta}}_n^{(+i)} = \widehat{\boldsymbol{\vartheta}}_n^{(i)}$ for $i = 1, 2$ and $\widehat{\boldsymbol{\vartheta}}_n^{(+i)} = (\widehat{\boldsymbol{\vartheta}}_n^{(i)'}, \widehat{\boldsymbol{\varphi}}_n^{(-i)'})'$ for $i = 3, \ldots, m$. By definition, the components of $\boldsymbol{\eta}_t$ are uncorrelated. The form of the asymptotic variance of the EbEE can be simplified under the assumption that the components of $\boldsymbol{\eta}_t$ are independent, or more generally under the following assumption.

B7: The components of $\boldsymbol{\eta}_t$ are such that $E\eta_{it}\eta_{jt}^2 = 0$ and $E\eta_{it}^2\eta_{jt}^2 = 1$ when $i \neq j$.

Theorem 4.3 (CAN of the EbEE) *Consider Model* (2)*, satisfying* (2.3)-(2.3)*. Under* **B1**-**B5**, *the EbEE* $\widehat{\boldsymbol{\vartheta}}_n^{EbE} = \left(\widehat{\boldsymbol{\vartheta}}_n^{(1)'}, \ldots, \widehat{\boldsymbol{\vartheta}}_n^{(m)'}\right)'$ *defined by* (4.2.1), (4.2.1) *and* (4.2.1) *satisfies*

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

Under the additional assumptions **A6**, **A10** *and* **B6**, *for* $i = 1, \ldots, m$, *the matrices* $\boldsymbol{J}_n^{(i)}$, $\boldsymbol{I}_n^{(i)}$ *and* $\boldsymbol{K}_n^{(i)}$ *converge almost surely, respectively, to a positive-definite matrix* $\boldsymbol{J}^{(i)}$, *to a semi positive-definite matrix* $\boldsymbol{I}^{(i)}$, *and to a matrix* $\boldsymbol{K}^{(i)}$, *as* $n \to \infty$. *Moreover,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \boldsymbol{\Sigma}^{(i)} := \left(\boldsymbol{J}^{(i)}\right)^{-1} \boldsymbol{I}^{(i)} \left(\boldsymbol{J}^{(i)}\right)^{-1}\right\} \tag{4.20}$$

for $i = 1$ *and* $i = 2$. *According to* (4.1.2), *denote by* $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}^-}^{(+i)}$ *(or by* $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}^+}^{(+(i-1))}$*) the bottom-right sub-matrix of* $\boldsymbol{\Sigma}^{(+i)}$ *(or of* $\boldsymbol{\Sigma}^{(+(i-1))}$*) corresponding to the asymptotic variance of* $\widehat{\boldsymbol{\varphi}}_n^{(-i)}$ *(which is equal to* $\widehat{\boldsymbol{\varphi}}_n^{(+(i-1))}$*). Using this notation and the convention* $\boldsymbol{\Sigma}^{(+2)} = \boldsymbol{\Sigma}^{(2)}$, *for* $i = 3, \ldots, m$, *we have*

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(+i)} - \boldsymbol{\vartheta}_0^{(+i)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \boldsymbol{\Sigma}^{(+i)}\right) \tag{4.21}$$

with, under **B7**,

$$\boldsymbol{\Sigma}^{(+i)} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{(i)} & -\left(\boldsymbol{J}^{(i)}\right)^{-1} \boldsymbol{K}^{(i)} \boldsymbol{\Sigma}_{\boldsymbol{\varphi}^+}^{(+(i-1))} \\ -\boldsymbol{\Sigma}_{\boldsymbol{\varphi}^+}^{(+(i-1))} \boldsymbol{K}^{(i)'} \left(\boldsymbol{J}^{(i)}\right)^{-1} & \boldsymbol{\Sigma}_{\boldsymbol{\varphi}^+}^{(+(i-1))} \end{pmatrix}$$

and $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{(i)} = \left(\boldsymbol{J}^{(i)}\right)^{-1} \left\{\boldsymbol{I}^{(i)} + \boldsymbol{K}^{(i)} \boldsymbol{\Sigma}_{\boldsymbol{\varphi}^+}^{(+(i-1))} \boldsymbol{K}^{(i)'}\right\} \left(\boldsymbol{J}^{(i)}\right)^{-1}$.

4.3 Comparison between the Full QMLE and the EbEE

Consider the CHAR model (2) satisfying (2.3)-(2.3). If $m = 2$, the one-step full QMLE and the two-step EbEE are exactly the same because

$$\widetilde{Q}_n(\boldsymbol{\vartheta}) = \widetilde{Q}_n^{(1)}(\boldsymbol{\vartheta}^{(1)}) + \widetilde{Q}_n^{(2)}(\boldsymbol{\vartheta}^{(2)})$$

in this case. For $m \geq 3$, the two estimators are generally different. In order to compare their asymptotic accuracies, we consider a simplistic static model with $g_{it} = 1$ for $i \in \{1, 2, 3\}$, $\ell_{21,t} = \ell_{21}$,

$\ell_{32,t} = \ell_{32}$ and $\ell_{31,t} = 0$ (or equivalently $\beta_{21,t} = \beta_{21}$, $\beta_{32,t} = \beta_{32}$ and $\beta_{31,t} = -\beta_{21}\beta_{32}$). The unknown parameter is thus $\boldsymbol{\vartheta} = (\beta_{21}, \beta_{32})$. The full QMLE is

$$\widehat{\boldsymbol{\vartheta}}_n = \arg\min_{\boldsymbol{\vartheta}\in\Theta} \sum_{t=1}^{n} q_t(\boldsymbol{\vartheta}),$$

where $q_t(\boldsymbol{\vartheta}) = (\epsilon_{2t} - \beta_{21}\epsilon_{1t})^2 + (\epsilon_{3t} - \beta_{32}\epsilon_{2t} + \beta_{21}\beta_{32}\epsilon_{1t})^2$ and Θ is a compact parameter space such that $\boldsymbol{\vartheta}_0 \in \mathring{\Theta}$. We have

$$\frac{\partial q_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \begin{pmatrix} -2\epsilon_{1t}v_{2t}(\boldsymbol{\vartheta}) + 2\beta_{32}\epsilon_{1t}v_{3t}(\boldsymbol{\vartheta}) \\ -2v_{2t}(\boldsymbol{\vartheta})v_{3t}(\boldsymbol{\vartheta}) \end{pmatrix}$$

and

$$\frac{\partial^2 q_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} = \begin{pmatrix} 2\epsilon_{1t}^2 + 2\beta_{32}^2\epsilon_{1t}^2 & 2\epsilon_{1t}v_{3t}(\boldsymbol{\vartheta}) - 2\beta_{32}\epsilon_{1t}v_{2t}(\boldsymbol{\vartheta}) \\ 2\epsilon_{1t}v_{3t}(\boldsymbol{\vartheta}) - 2\beta_{32}\epsilon_{1t}v_{2t}(\boldsymbol{\vartheta}) & 2v_{2t}^2(\boldsymbol{\vartheta}) \end{pmatrix}$$

with

$$v_{2t}(\boldsymbol{\vartheta}) = \epsilon_{2t} - \beta_{21}\epsilon_{1t}, \qquad v_{3t}(\boldsymbol{\vartheta}) = \epsilon_{3t} - \beta_{32}\epsilon_{2t} + \beta_{21}\beta_{32}\epsilon_{1t}.$$

Assume for instance that the variable η_{2t} is independent of the vector $(\eta_{1t}, \eta_{3t})'$ and that this vector is distributed as the product $\eta\boldsymbol{u}$, where the random variable η and the vector $\boldsymbol{u}$ are independent, e.g., $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_2)$ and $E\eta^2 = 1$. In view of Theorem 4.1, and noting that $v_{it}(\boldsymbol{\vartheta}_0) = \eta_{it}$, we then have

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \overset{o_P(1)}{=} \left\{E\frac{\partial^2 q_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right\}^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial q_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \boldsymbol{\Omega} := \begin{pmatrix} \frac{1+\beta_{32}^2 E\eta^4}{(1+\beta_{32}^2)^2} & 0 \\ 0 & E\eta^4 \end{pmatrix}\right\}.$$

The EbEE of the parameter β_{21} is defined by

$$\widehat{\beta}_{21} = \arg\min_{\beta_{21}\in\Theta^{(1)}} \sum_{t=1}^{n}(\epsilon_{2t} - \beta_{21}\epsilon_{1t})^2$$

and satisfies

$$\sqrt{n}\left(\widehat{\beta}_{21} - \beta_{0,21}\right) = \frac{n^{-1/2}\sum_{t=1}^{n}\eta_{2t}\epsilon_{1t}}{n^{-1}\sum_{t=1}^{n}\epsilon_{1t}^2} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

It follows that when $E\eta^4$ is large enough, the EbEE of β_{21} is asymptotically more accurate than its Full QMLE. By contrast, it can be shown that the EbEE and Full QMLE of β_{32} have the same asymptotic distribution.

To illustrate the gain of efficiency of the EbE over the QML in this particular setting, Figure 1 reports the ratio between the asymptotic variance of the QML estimator of β_{21} and its EbE

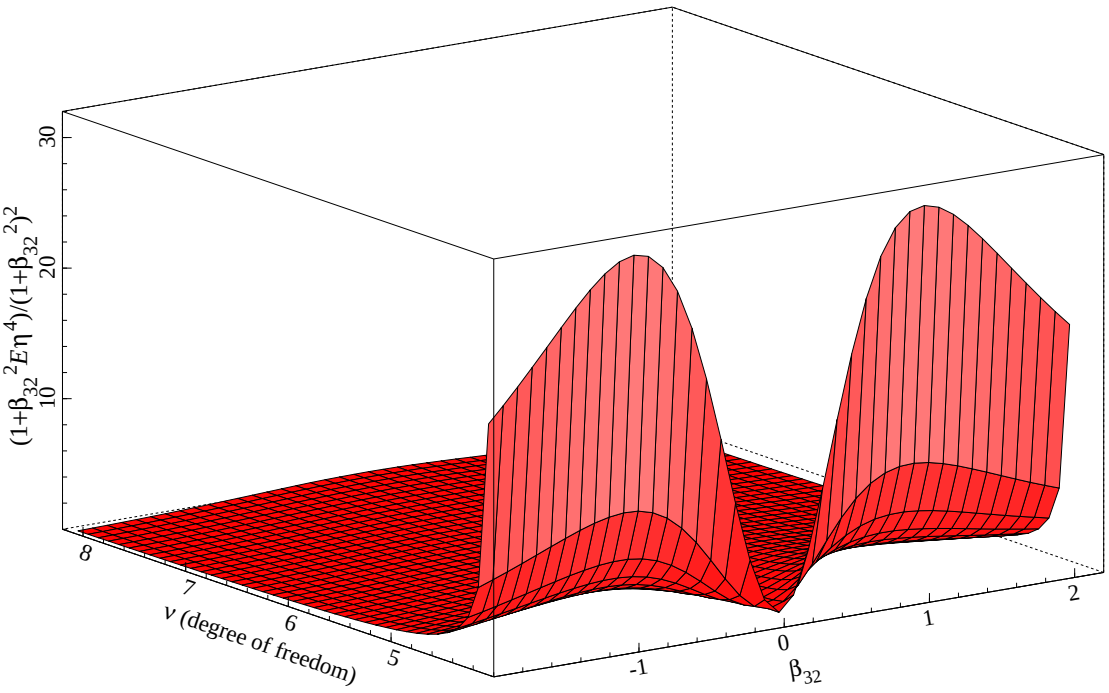

Figure 1: Ratio between the asymptotic variance of the QML estimator of β_{21} and its EbE counterpart as a function of the degree of freedom ν and β_{32}.

counterpart, i.e. $\frac{1+\beta_{32}^2 E\eta^4}{(1+\beta_{32}^2)^2}$, as a function of β_{32} and $E\eta^4$. We consider values of β_{32} ranging between -2 and 2. Instead of taking different values for $E\eta^4$, we assume that η follows a standardized Student-t distribution with degree of freedom ν such that $E\eta^4 = 3 + 6/(\nu - 4)$ for $\nu > 4$.

Figure 1 confirms that the gain of efficiency of the EbE estimator of β_{21} is inversely proportional to ν (and therefore increases with $E\eta^4$). The loss of efficiency of the Full QMLE can be very large for values of $\nu < 5$, i.e. $E\eta^4 > 9$.

Of course, we considered in the section a particular example, and one can easily find other examples for which the Full QMLE is more efficient than the EbEE.

4.4 Case where the estimation can be parallelized

Now consider the CHAR model (2.3)-(2.3). We adapt the notations used in Section 4.1.2 for model (2.3)-(2.3), by setting

$$\boldsymbol{\theta}^{(i)} = (\omega_i, \gamma_{i+}, \gamma_{i-}, \alpha_i, b_i)', \qquad \boldsymbol{\varphi}^{(ij)} = (\varpi_{ij}, \varsigma_{ij+}, \varsigma_{ij-}, \tau_{ij}, \xi_{ij}, c_{ij})',$$

for $(i,j) \in \mathfrak{T}_m$, keeping the convention of adding the subscript "0" to denote the true parameter value. Note that, with this new model, $\widetilde{v}_{it}(\boldsymbol{\varphi}^{(+i)})$ and $\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)})$ depend only on $\boldsymbol{\varphi}^{(i)}$ and $\boldsymbol{\vartheta}^{(i)}$, respectively. Indeed the QMLE is the solution to (4.1), with

$$\widetilde{q}_t(\boldsymbol{\vartheta}) = \sum_{i=1}^{m} \widetilde{q}_{it}(\boldsymbol{\vartheta}^{(i)}),$$

and, for $t = 1, \ldots, n$,

$$\begin{aligned}
\widetilde{q}_{it}(\boldsymbol{\vartheta}^{(i)}) &= \frac{\widetilde{v}_{it}^2(\boldsymbol{\varphi}^{(+i)})}{\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(i)})} + \log \widetilde{g}_{it}(\boldsymbol{\vartheta}^{(i)}), \\
\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(i)}) &= \omega_{i,t-1} + \alpha_i \widetilde{v}_{i,t-1}^2(\boldsymbol{\varphi}^{(i)}) + b_i \widetilde{g}_{i,t-1}(\boldsymbol{\vartheta}^{(i)}), \\
\widetilde{v}_{it}(\boldsymbol{\varphi}^{(i)}) &= \epsilon_{it} - \sum_{j=1}^{i-1} \widetilde{\beta}_{ij,t}(\boldsymbol{\varphi}^{(i)}) \epsilon_{jt}, \\
\widetilde{\beta}_{ij,t}(\boldsymbol{\varphi}^{(i)}) &= \omega_{ij,t-1} + \tau_{ij} \widetilde{v}_{i,t-1}(\boldsymbol{\varphi}^{(i)}) + \xi_{ij} \widetilde{v}_{i,t-1}(\boldsymbol{\varphi}^{(i)}) \epsilon_{1,t-1} + c_{ij} \widetilde{\beta}_{ij,t-1}(\boldsymbol{\varphi}^{(i)}),
\end{aligned}$$

and fixed initial values for $\widetilde{\beta}_{ij,0}(\boldsymbol{\varphi}^{(i)})$, $\boldsymbol{\epsilon}_0$ and $\widetilde{g}_{i,0}(\boldsymbol{\vartheta}^{(i)})$. The QMLE and EbEE of $\boldsymbol{\vartheta}_0^{(i)}$ thus coincide, and can be obtained by solving

$$\widehat{\boldsymbol{\vartheta}}_n^{(i)} = \underset{\boldsymbol{\vartheta}^{(i)} \in \Theta^{(i)}}{\arg\min} \sum_{t=1}^{n} \widetilde{q}_{it}(\boldsymbol{\vartheta}^{(i)}),$$

in parallel (or successively but without any particular order) for all $i \in \{1, \ldots, m\}$. The analog of Theorem 4.2 can be obtained, with a block-diagonal matrix $\boldsymbol{\Omega}$, if the conditions in **B4** are modified in an obvious way, and if the matrix $\boldsymbol{T}$ involved in $\boldsymbol{S}_t$ (see (4.1.3)-(4.1.3)), with ℓ-th row $(0, \tau_{ij}^{(2)}, \ldots, \tau_{ij}^{(i)}, \mathbf{0}'_{m-i})$, is replaced by the matrix $\boldsymbol{T}_t$ with ℓ-th row $(\mathbf{0}'_{i-1}, \tau_{ij} + \xi_{ij}\epsilon_{1t}, \mathbf{0}'_{m-i})$.

5 Monte-Carlo simulation

In this section, we first illustrate the invertibility condition and also confirm the good finite sample properties of the QML and EbE estimators with a Monte Carlo study.

5.1 Illustration of the invertibility condition

To illustrate the importance of the invertibility condition, consider a trivariate stochastic process $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})' = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t$, compatible with (2)-(2.3)-(2.3), with the additional restrictions that $\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{B}_t^{-1} \boldsymbol{G}_t^{1/2}$, with $\boldsymbol{G}_t = \boldsymbol{G} = \omega_0 I_3$, $\boldsymbol{\eta}_t \overset{i.i.d.}{\sim} N(\mathbf{0}, \boldsymbol{I}_3)$ and $\beta_{ij,t} = \varpi + \tau_{ij}^{(i)} v_{i,t-1}$ for $(i,j) \in \mathfrak{T}_3$.

This model corresponds to a CHAR model with *i.i.d.* Gaussian innovations, homoscedastic orthogonal factors $(v_{1t}, v_{2t}, v_{3t})'$ and time varying betas depending only on past shocks $v_{i,t-1}$. For simplicity, the constants in the variance of the factors and the conditional betas are restricted to be the same in all equations. The three equations describing the conditional betas can be written more compactly as follows:

$$\begin{aligned}
\boldsymbol{\beta}_t := \begin{pmatrix} \beta_{21,t} \\ \beta_{31,t} \\ \beta_{32,t} \end{pmatrix} &= \begin{pmatrix} \varpi \\ \varpi \\ \varpi \end{pmatrix} + \begin{pmatrix} 0 & \tau_{21}^{(2)} & 0 \\ 0 & 0 & \tau_{31}^{(3)} \\ 0 & 0 & \tau_{32}^{(3)} \end{pmatrix} \boldsymbol{v}_{t-1} \\
&= \begin{pmatrix} \varpi \\ \varpi \\ \varpi \end{pmatrix} + \begin{pmatrix} 0 & \tau_{21}^{(2)} & 0 \\ 0 & 0 & \tau_{31}^{(3)} \\ 0 & 0 & \tau_{32}^{(3)} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\beta_{21,t-1} & 1 & 0 \\ -\beta_{31,t-1} & -\beta_{32,t-1} & 1 \end{pmatrix} \boldsymbol{\epsilon}_{t-1} \\
&= \boldsymbol{w}_{t-1}^* + \boldsymbol{S}_{t-1}\boldsymbol{\beta}_{t-1},
\end{aligned}$$

with $\boldsymbol{w}_t^* = \begin{pmatrix} \varpi + \tau_{21}^{(2)}\epsilon_{2,t} \\ \varpi + \tau_{31}^{(3)}\epsilon_{3,t} \\ \varpi + \tau_{32}^{(3)}\epsilon_{3,t} \end{pmatrix}$ and $\boldsymbol{S}_t = \begin{pmatrix} -\tau_{21}^{(2)}\epsilon_{1,t} & 0 & 0 \\ 0 & -\tau_{31}^{(3)}\epsilon_{1,t} & -\tau_{31}^{(3)}\epsilon_{2,t} \\ 0 & -\tau_{32}^{(3)}\epsilon_{1,t} & -\tau_{32}^{(3)}\epsilon_{2,t} \end{pmatrix}$.

Figure 2 plots $n = 1,000$ observations of $\boldsymbol{\epsilon}_t$ generated using the above model with parameters $\omega_0 = 2, \varpi = 0.5$ and $\tau_{21}^{(1)} = \tau_{31}^{(2)} = \tau_{32}^{(2)} = 0.5$.

Figure 3 plots the simulated betas and the filtered betas obtained using the following recursive formulas:

$$\begin{aligned}
v_{kt} &= \epsilon_{kt} - \sum_{j=1}^{k-1} \beta_{kj,t}\epsilon_{jt} \\
\beta_{ij,t} &= \varpi + \tau_{ij}^{(i)} v_{i,t-1}
\end{aligned}$$

for $k = 1, 2$ and 3 and $(i, j) \in \mathfrak{T}_3$. Importantly, the parameters ϖ and $\tau_{ij}^{(i)}$ are not estimated but set to their true values and the same initial value $\beta_{ij,0} = \varpi$ is used in the simulation of the data and in the computation of the betas using the above formulas. Note that this model implies that $\beta_{31,t} = \beta_{32,t} \forall t$ because they both depend only on $v_{3,t-1}$ and furthermore $\tau_{31}^{(2)} = \tau_{32}^{(2)}$.

The invertibility condition stated in (4.1.3) implies that when $n \to \infty$, $\tilde{\gamma}_S = \frac{1}{n}\log\|\prod_{i=1}^{n} \boldsymbol{S}_{n-i+1}\| < 0$ for all $\boldsymbol{\varphi} \in \Theta_{\boldsymbol{\varphi}}$ and in particular for $\boldsymbol{\varphi}_0$. Using the Frobenius norm for $\|\cdot\|$, we obtain a value of $\tilde{\gamma}_S = -0.41 < 0$ satisfying the invertibility condition. Note that in this case, filtered and true conditional betas are indistinguishable.

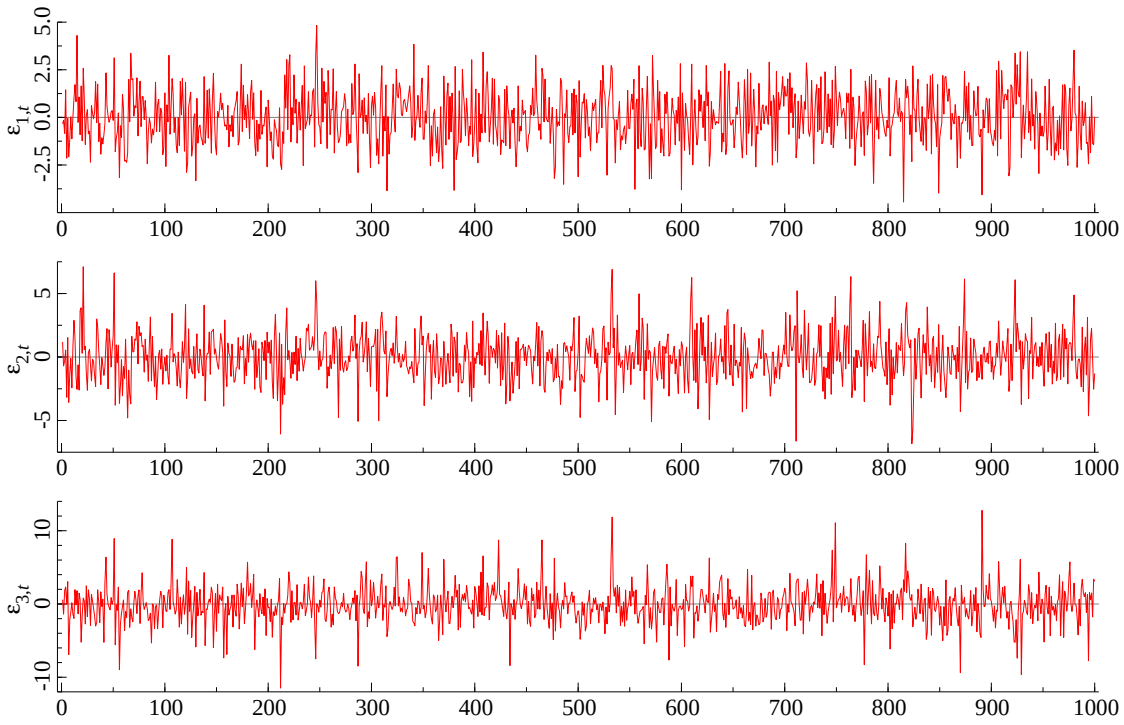

Figure 2: Simulated returns for the case satisfying the invertibility condition

Figure 4 plots 1,000 observations generated using the same model as above except that $\tau_{21}^{(1)} = \tau_{31}^{(2)} = \tau_{32}^{(2)} = 0.85$. For this simulation, we have an estimate of $\tilde{\gamma}_S = 0.09 > 0$ implying $\gamma_S > 0$ (provided that n is sufficiently large), which implies that the invertibility condition is not satisfied in this case. Interestingly, the filtered values of $\beta_{31,t}$ and $\beta_{32,t}$ plotted in Figure 5 diverge completely at the end of the sample, which illustrates the impact of the rejection of the invertibility condition.

5.2 Finite sample properties of the QML and EbE estimators

To study the finite sample properties of the QML and EbE estimators of the CHAR model, we perform a Monte Carlo simulation.

Simulation setup: We generate $T = 1000, 2000$ or 4000 observations of a $m(= 5)$-dimensional return process $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\eta}_t$ $(t = 1, \ldots, T)$, where $\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{B}_t^{-1}\boldsymbol{G}_t^{1/2}$. In the first simulation (i.e., Gaussian case) $\boldsymbol{\eta}_t \overset{i.i.d.}{\sim} N(\mathbf{0}, \boldsymbol{I}_5)$ while in the second simulation (i.e., Student-t case) innovations follow independent standardized Student-t distributions with 7 degrees of freedom, i.e., $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{mt})$, where $\eta_{it} \overset{i.i.d.}{\sim} T(\mathbf{0}, 1, 7)$ for $i = 1, \ldots, m$ with $\eta_{it} \perp \eta_{jt}$ for $i \neq j$.

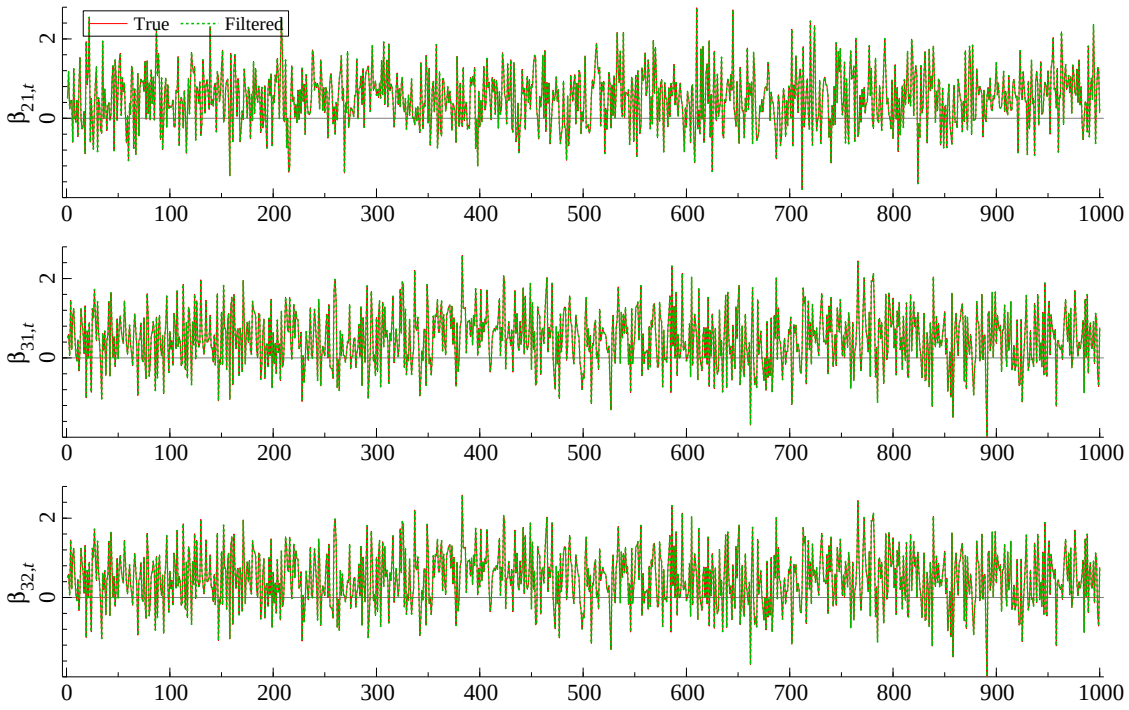

Figure 3: True and filtered conditional beta for the case satisfying the invertibility condition

The respectively m and $m(m-1)/2$ elements g_{it} and $\beta_{ij,t}$ defining $\boldsymbol{G}_t$ and $\boldsymbol{B}_t$ are specified as restricted versions of (2.3)-(2.3), i.e.

$$g_{it} = \omega_{0i} + \alpha_{0i} v_{i,t-1}^2 + b_{0i} g_{i,t-1}$$

$$\beta_{ij,t} = \varpi_{0ij} + \tau_{0ij} v_{i,t-1} + c_{0ij} \beta_{ij,t-1}.$$

We set the parameters governing the dynamics in the conditional variances to $\omega_{0i} = 0.1$, $\alpha_{0i} = 0.1$ and $b_{0i} = 0.8$ (for $i = 1, \ldots, M$) so that all elements of $\boldsymbol{v}_t$ have a unit unconditional variance and GARCH effects. Parameters of the conditional betas are set to $\varpi_{0ij} = 0.1$ $\tau_{0ij} = 0.2$ and $c_{0ij} = 0.8$. Results reported below are based on programs written by the authors using Ox version 7.1 (Doornik, 2012).

Results: Summary statistics on the Monte Carlo simulation are reported in Tables 1 and 2. The left part of both tables corresponds to the Full QMLE while the right part is for the EbEE. Due to the high number of parameters, results are not reported for each parameter separately. Instead, we report averages across the number of series for the GARCH parameters entering in $\boldsymbol{\theta}$ (i.e., ω, α and β) or across the number of conditional betas for the parameters entering in $\boldsymbol{\varphi}$ (i.e., ϖ, τ and

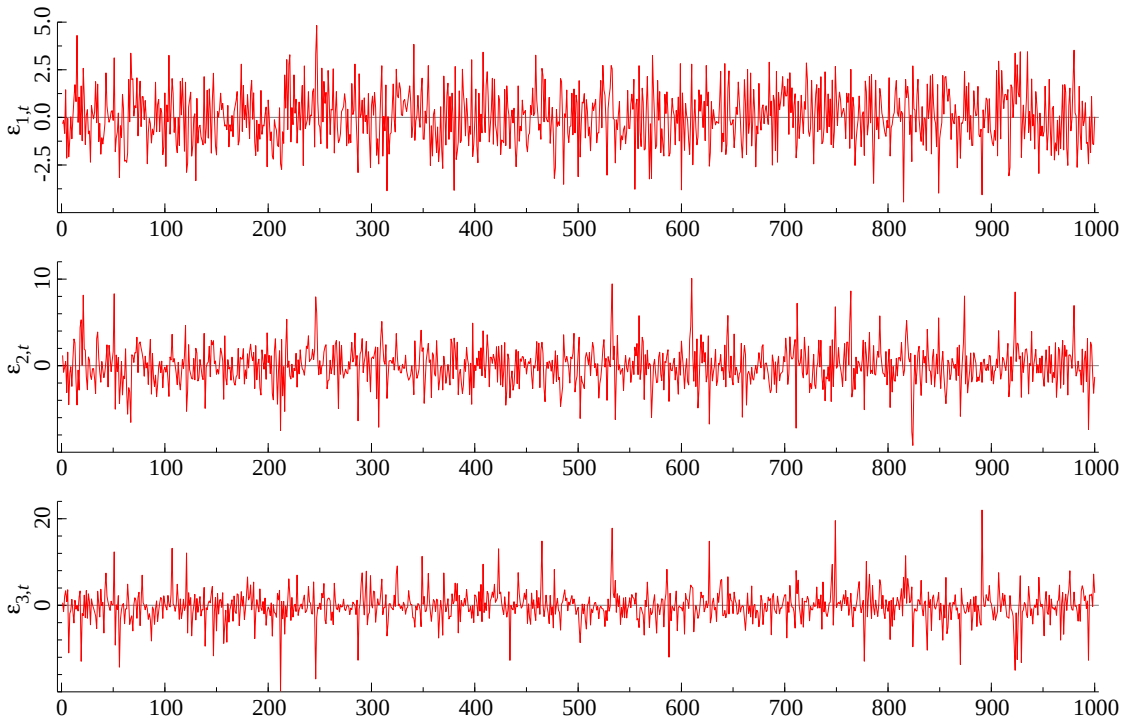

Figure 4: Simulated returns for the case not satisfying the invertibility condition

c). Note that for ω, α and β, averages are taken over $m = 5$ values while for ϖ, τ and c, they are taken over $m(m-1)/2 = 10$ values. Averages across all parameters are also reported in rows labelled "ALL" .

Columns "BIAS" correspond to the empirical bias of the estimates over 1,000 replications. More specifically, the biases reported in rows "ω" are computed as follows:

$$\text{BIAS of } \hat{\omega} = \frac{1}{1,000} \sum_{r=1}^{1,000} \frac{1}{5} \sum_{i=1}^{5} (\hat{\omega}_i^{(r)} - \omega_{0i}),$$

where $\hat{\omega}_i^{(r)}$ is either the Full QMLE or EbEE of ω_i obtained at the r-th replication. Similarly, the RMSE of $\hat{\omega}$ is computed as follows:

$$\text{RMSE of } \hat{\omega} = \sqrt{\frac{1}{1,000} \sum_{r=1}^{1,000} \frac{1}{5} \sum_{i=1}^{5} (\hat{\omega}_i^{(r)} - \omega_{0i})^2}.$$

The relative efficiency of the EbEE with respect to the Full QMLE is reported in column "RE". For each row of the tables, RE is defined as the ratio between the RMSE of the EbE divided by the RMSE of the Full QML. A value of RE greater than 1 therefore means that the Full QML is more

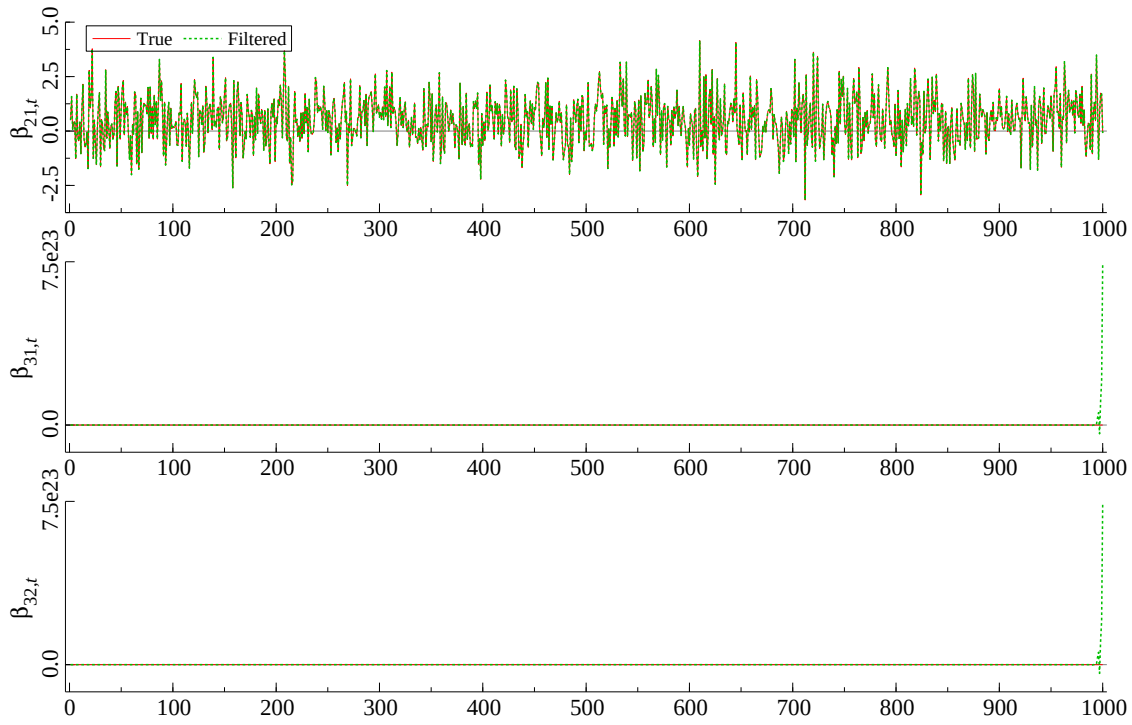

Figure 5: True and filtered conditional beta for the case not satisfying the invertibility condition

efficient than the EbE. Finally, columns "5% CB" and "95% CB" correspond to the frequencies of true parameters falling below the bands of a 90% confidence interval computed using Theorems 4.2 and 4.3, respectively for the Full QMLE and EbEE. Again these frequencies are averaged over the number of GARCH equations or conditional betas.

Results suggest that both the Full QML and EbE deliver satisfactory results in the sense that biases are very small for all parameters (particularly for $T = 2,000$ and $4,000$). In the Gaussian case (i.e. Table 1), the Full QML is about 9%, 16% and 30% more efficient than the EbE, respectively for T=$1,000, 2,000$ and $4,000$. In the case of Student-t, innovations the RE of the Full QML is slightly greater.

About the validity of Theorems 4.2 and 4.3 in finite samples, the rejection frequencies reported in columns "5% CB" and "95% CB" are close to the theoretical values for all parameters both for Gaussian and Student-t innovations.

Unreported simulation results (available upon request) suggest very similar finite sample behaviors of the Full and EbE QML for a specification of the conditional betas of the type

Table 1: Monte-Carlo simulation – Gaussian case

	Full QML				EbE				
	BIAS	RMSE	5% CB	95% CB	BIAS	RMSE	5% CB	95% CB	RE
$T = 1,000$									
ω	0.0160	0.0614	3.5988	93.4440	0.0190	0.0672	4.070	92.251	1.10
α	0.0030	0.0295	3.2265	92.9470	0.0036	0.0322	3.659	91.696	1.09
β	-0.0196	0.0804	6.2461	95.1600	-0.0230	0.0876	7.379	94.347	1.09
ϖ	0.0006	0.0121	4.2192	93.4950	0.0007	0.0132	4.604	92.559	1.08
τ	0.0009	0.0149	5.6256	95.2640	0.0011	0.0165	6.208	94.779	1.10
c	-0.0013	0.0230	6.1531	95.3980	-0.0016	0.0250	7.009	94.913	1.09
ALL	0.0000	0.0301	5.0075	94.4300	0.0000	0.0329	5.639	93.644	1.09
$T = 2,000$									
ω	0.0068	0.0320	2.4869	95.4020	0.0087	0.0370	3.305	93.138	1.16
α	0.0013	0.0185	2.6750	95.2560	0.0017	0.0217	3.766	93.410	1.17
β	-0.0082	0.0439	4.2633	96.6140	-0.0104	0.0510	6.130	95.460	1.16
ϖ	0.0001	0.0067	3.8454	94.5140	0.0002	0.0079	4.550	93.410	1.17
τ	0.0002	0.0088	4.6290	95.6010	0.0004	0.0102	5.690	94.812	1.17
c	-0.0003	0.0126	5.0888	96.2800	-0.0004	0.0148	6.266	95.690	1.18
ALL	0.0000	0.0167	4.0613	95.5630	0.0000	0.0195	5.135	94.426	1.17
$T = 4,000$									
ω	0.0024	0.0172	2.1475	96.3990	0.0039	0.0225	3.442	93.831	1.31
α	0.0005	0.0113	2.2993	96.8980	0.0006	0.0148	3.831	94.372	1.30
β	-0.0030	0.0248	3.4924	97.2450	-0.0046	0.0321	5.649	95.909	1.29
ϖ	0.0000	0.0037	2.8850	96.1710	0.0000	0.0049	4.058	94.177	1.31
τ	0.0001	0.0049	3.4924	96.7140	0.0001	0.0065	5.336	95.249	1.33
c	-0.0001	0.0071	3.5792	96.9310	-0.0001	0.0092	5.487	95.747	1.30
ALL	0.0000	0.0094	3.0947	96.6860	0.0000	0.0123	4.743	94.940	1.30

Note: Monte-Carlo simulation results based on 1,000 replications. DGP is a $m(= 5)$-dimensional return process $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\eta}_t$, where $\boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{B}_t^{-1}\boldsymbol{G}_t^{1/2}$, $\boldsymbol{\eta}_t \overset{i.i.d.}{\sim} N(\mathbf{0}, \boldsymbol{I}_5)$, $g_{it} = \omega_{0i} + \alpha_{0i}v_{i,t-1}^2 + b_{0i}g_{i,t-1}$ (with $\omega_{0i} = 0.1$, $\alpha_{0i} = 0.1$, $b_{0i} = 0.8$) for $i = 1, \ldots, m$, $\beta_{ij,t} = \varpi_{0ij} + \tau_{0ij}v_{i,t-1} + c_{0ij}\beta_{ij,t-1}$ (with $\varpi_{0ij} = 0.1$ $\tau_{0ij} = 0.2$ and $c_{0ij} = 0.8$) for $i = 1, \ldots, m$, $j < i$. Columns "BIAS" and "RMSE" correspond to the empirical bias and root mean square error. Columns "5% CB" and "95% CB" correspond to the frequencies of true parameters falling below the bands of a 90% confidence interval. Column "RE" is the ratio between the RMSE of the EbE and that of Full QML. Entries in this table are averages across the m series for ω, α and β, across the $m(m-1)/2$ conditional betas for the parameters ϖ, τ and c and across all parameters for rows labelled "All".

Table 2: Monte-Carlo simulation – Student-t case

	Full QML				EbE				
	BIAS	RMSE	5% CB	95% CB	BIAS	RMSE	5% CB	95% CB	RE
T=1,000									
ω	0.0179	0.0689	3.3720	92.4130	0.0225	0.0771	4.227	90.284	1.12
α	0.0039	0.0355	2.2339	91.9070	0.0057	0.0414	2.860	90.074	1.17
β	-0.0233	0.0919	6.9968	94.5840	-0.0293	0.1031	8.980	93.270	1.12
ϖ	0.0008	0.0135	4.0569	93.9520	0.0010	0.0144	4.532	92.650	1.06
τ	0.0010	0.0155	4.9104	95.9750	0.0012	0.0178	5.910	95.152	1.15
c	-0.0016	0.0245	5.9536	95.7960	-0.0018	0.0271	7.308	95.226	1.10
ALL	-0.0001	0.0337	4.7161	94.4830	-0.0001	0.0378	5.730	93.298	1.12
T=2,000									
ω	0.0064	0.0337	2.1758	94.7250	0.0093	0.0416	3.141	91.734	1.23
α	0.0009	0.0215	1.7363	93.8460	0.0020	0.0266	2.748	91.189	1.24
β	-0.0080	0.0478	5.0989	96.7470	-0.0120	0.0588	7.546	94.984	1.23
ϖ	0.0002	0.0065	3.0659	95.7030	0.0003	0.0080	4.308	94.079	1.22
τ	0.0003	0.0083	3.6813	96.6810	0.0005	0.0101	4.984	95.507	1.22
c	-0.0005	0.0124	4.2088	96.9120	-0.0007	0.0150	5.965	95.725	1.21
ALL	-0.0001	0.0175	3.4359	95.9900	0.0000	0.0215	4.883	94.281	1.23
T=4,000									
ω	0.0023	0.0179	1.7156	97.0390	0.0049	0.0259	3.125	93.310	1.45
α	0.0002	0.0123	1.3631	96.6630	0.0008	0.0177	3.194	92.755	1.44
β	-0.0028	0.0259	3.0552	97.7910	-0.0060	0.0374	6.505	95.370	1.44
ϖ	0.0001	0.0034	2.8555	96.7100	0.0001	0.0050	4.630	94.051	1.46
τ	0.0000	0.0045	2.7850	97.4150	0.0002	0.0065	4.792	95.567	1.45
c	-0.0002	0.0064	3.4078	97.1210	-0.0002	0.0094	5.787	95.486	1.46
ALL	0.0000	0.0094	2.6923	97.1090	0.0000	0.0136	4.805	94.627	1.45

Note: See Table 1 expect that $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{mt})$, where $\eta_{it} \overset{i.i.d.}{\sim} T(\mathbf{0}, 1, 7)$ for $i = 1, \ldots, m$ with $\eta_{it} \perp \eta_{jt}$ for $i \neq j$.

$\beta_{ij,t} = \varpi_{0ij} + \tau_{0ij} v_{j,t-1} + c_{0ij}\beta_{ij,t-1}$, $\beta_{ij,t} = \varpi_{0ij} + \tau_{0ij} v_{i,t-1} v_{1,t-1} + c_{0ij}\beta_{ij,t-1}$ and $\beta_{ij,t} = \varpi_{0ij} + \tau_{0ij} v_{i,t-1} v_{j,t-1} + c_{0ij}\beta_{ij,t-1}$. The first two specifications are nested respectively in (2.3) and (2.3). The third one satisfies the stationarity conditions stated in Section 3. When $m > 2$ it is not embedded in the theoretical setting of Section 4 but this specification is found to be empirically relevant (see next section).

6 Empirical Applications

The aim of this section is to illustrate the usefulness of the CHAR model in a portfolio and risk management exercice. We start with the same dataset used in Engle (2016). We consider the 12 US industry portfolios (i.e., BusEq, Chems, Durbl, Enrgy, Hlth, Manuf, Money, NoDur, Other, Shops, Telcm and Utils) and the three factors introduced in an asset pricing context by Fama and French (1992, 2004). Daily data are obtained from Ken French's website (French, 2014) and cover the period 1994-August 2016. The three risk factors are the market factor MKT proxied by the excess log-returns on the SP500 index as well as the standard Fama French size and value factors SMB (*Small Minus Big*) and HML (*High Minus Low*). The 3 factors MKT_t, SMB_t, HML_t and the 12 industry portfolios log-returns in excess to the risk free rate (i.e., r_{kt} for $k = 1, \ldots, 12$) are plotted in Figures 6 and 7, respectively. All the familiar asset pricing results can be easily expressed in terms of conditional betas, and in particular

$$E_{t-1}(r_{kt}) = \beta_{k,MKT,t} E_{t-1}(MKT_t) + \beta_{k,SMB,t} E_{t-1}(SMB_t) + \beta_{k,HML,t} E_{t-1}(HML_t), \qquad (6.1)$$

where $\beta_{k,MKT,t}$, $\beta_{k,SMB,t}$ and $\beta_{k,HML,t}$ are the three time-varying parameters of interest to be estimated from the data. In this asset pricing context, expected returns on any asset is linear in the betas and only depend upon the risk premiums embedded in the factors. In other terms, there is no alpha or intercept in (6). This property can be directly used to build an in-sample test of the asset pricing model (see e.g., Engle, 2016).

We follow the approach initially proposed by Patton and Verardo (2012) that consists in building hedged portfolios to offset some unwanted exposures to predetermined factors. Let us consider an industry portfolio with a clear small caps biais. The over/underperformance of this portfolio compared to the market can be explained by a statistically significant $\beta_{k,SMB,t}$ coefficient. The only way to control for this effect is to build a portfolio that buys the industry portfolio and sells the

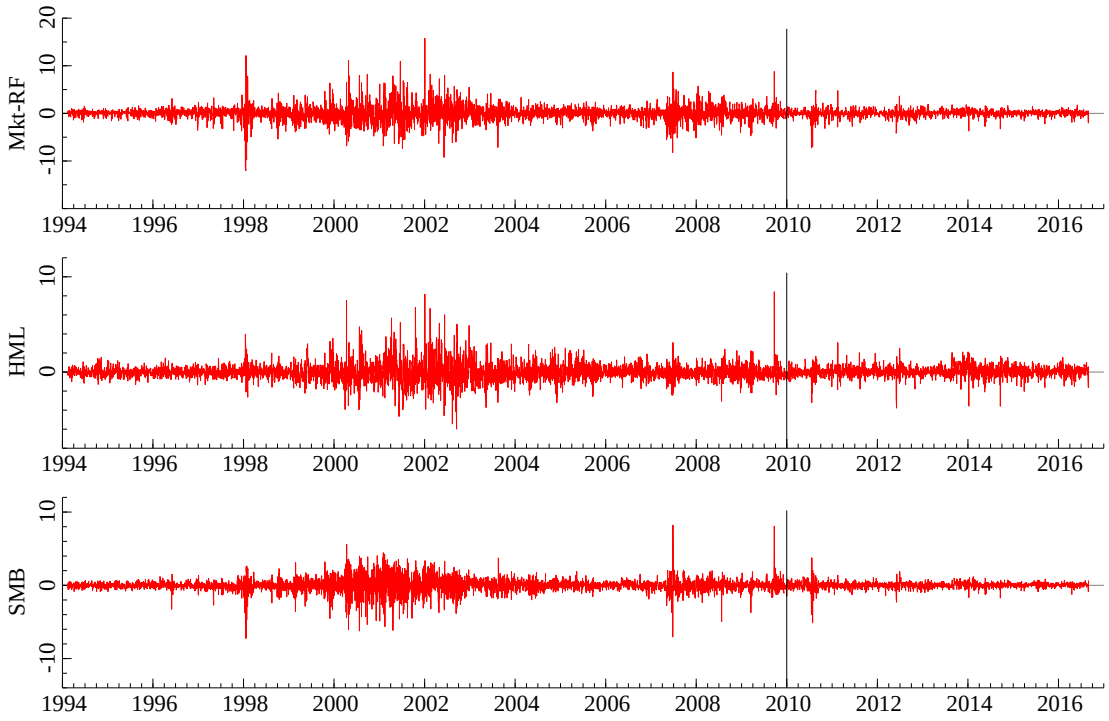

Figure 6: Time series graph of the three factors (MKT_t, SMB_t and HML_t) for the whole sample (i.e., in- and out-of-sample period)

implied SMB exposure. If $\beta_{k,SMB,t}$ is time-varying, out-of-sample forecasts are needed to adjust the position. Generalizing this idea to the three factors, Equation (6) leads to an hedged portfolio or equivalently a tracking error series with zero mean.

To implement this strategy, four competing models are used to estimate the three conditional betas and obtain one-step-ahead forecasts, i.e., two DCB models and two CHAR models. For the DCB models, we consider a CCC-GARCH(1,1) model and a DCC-GARCH(1,1) estimated on 4-dimensional systems $\boldsymbol{\epsilon}_t = (\mathbf{x}_t, y_t)'$ with $\mathbf{x}_t = (MKT_t, SMB_t, HML_t)'$ and $y_t = r_{kt}$ using the same notation as in Section 2. In-sample conditional betas estimates of the DCB models are obtained using the formula $\boldsymbol{\Sigma}_{yx,t}\boldsymbol{\Sigma}_{xx,t}^{-1}$ while one-step-ahead forecasts are obtained as $\boldsymbol{\Sigma}_{yx,t+1|t}\boldsymbol{\Sigma}_{xx,t+1|t}^{-1}$.

The third model is a CHAR model with constant betas (denoted C-CHAR) and a GARCH(1,1) specification for the conditional variance of the factors, i.e.,

$$g_{it} = \omega_i + \alpha_i v_{i,t-1}^2 + b_i g_{i,t-1}, \qquad \beta_{ij,t} = \varpi_{ij}. \tag{6.2}$$

The main pitfall of dynamic specifications of the elements of a Cholesky decomposition is that

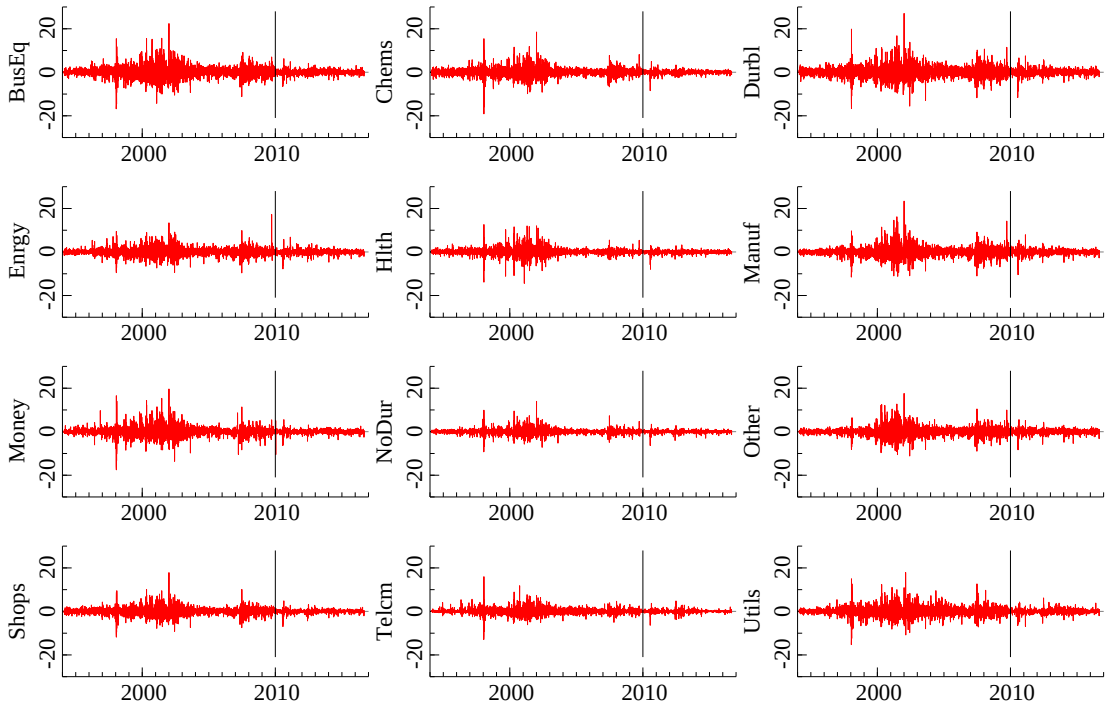

Figure 7: Time series graph of the log-returns of the 12 industry portfolios for the whole sample (i.e., in- and out-of-sample period)

the order of the series matters and that most of the times all possible permutations of the series are plausible.[5] In our case, the last series of the vector $\boldsymbol{\epsilon}_t$ is known and corresponds to the k-th industry portfolio so that $\beta_{41,t}$, $\beta_{42,t}$ and $\beta_{43,t}$ correspond to $\beta_{k,MKT,t}, \beta_{k,SMB,t}$ and $\beta_{k,HML,t}$ (in a certain order). The choice of $\epsilon_{1,t}$ is also natural because it makes more sense to dynamically orthogonalize the SMB and the HML factors with respect to the MKT factor rather than the opposite. Indeed, the size and value factors are generated using long/short dollar neutral strategies to get approximatively market neutral returns. The order between SMB and HML is however not known apriori and has been chosen using the Schwartz Information Criterion (SIC). We fixed the order as follows: $\boldsymbol{\epsilon}_t = (MKT_t, SMB_t, HML_t, r_{kt})'$.

The last model (simply denoted CHAR) relaxes the assumption of constancy of the conditional betas. We start by discussing the specification of the conditional betas of the first three components of $\boldsymbol{\epsilon}_t$, as they are identical whatever the industry portfolio we consider. We tested several specifications for the $\beta_{ij,t}$'s and found that the best one is

$$\beta_{ij,t} = \varpi_{ij} + \tau_{ij} v_{i,t-1} v_{j,t-1} + c_{ij} \beta_{ij,t-1} \tag{6.3}$$

for (i, j) belonging to the set $\mathfrak{T}_3$. With this choice, we can empirically check for instance whether the factors are already orthogonal on average (i.e, $\varpi_{ij} = 0$), or need to be conditionally orthogonalized (i.e, $\varpi_{ij}, \tau_{ij}, c_{ij} \neq 0$).

In a second step, and to model the remaining $\beta_{k,MKT,t}, \beta_{k,SMB,t}$ and $\beta_{k,HML,t}$ terms relative to a given industry portfolio, several specifications are considered. We allow each conditional beta to be either constant like in (6) or time-varying according to (6). The best specification (chosen using the SIC) is used to obtain one-step ahead forecasts of the three conditional betas of interest. Note that in almost all cases, the best model is found to be the one where all conditional betas follow Equation (6). One-step-ahead forecasts, denoted $\beta_{ij,t+1|t}$, are obtained using either (6) or by translating (6) one step into the future because (6) only depends on quantities that are observed at time t.

The first two models are estimated by Gaussian QML while the two CHAR models are estimated

[5] The CholCov of Boudt et al. (2017) is a nonparametric estimator of the quadratic variation of log-prices under asynchronicity and microstructure noise. Like the CHAR model it is relies on the Cholesky decomposition and allows a sequential estimation of the covariance matrix. The order of the series also matters in this case. To make optimal use of the data, Boudt et al. (2017) propose an ordering of the series in terms of decreasing liquidity.

by the EbE QMLE described in Section 4. All models are estimated on demeaned log-returns (the empirical means are computed on in-sample observations and not the whole sample) on rolling windows of 4,000 observations. Models are reestimated every 3 steps (parameters are therefore kept constant to produce three consecutive forecasts before being reevaluated).

Before presenting the results of the forecasting exercise, we report in Table 3 the estimation results of a CHAR model for series BusEq and the period spanning from 2000-02-08 to 2015-12-31, i.e., the values used to produce the forecasts of the CHAR model for the first working day in 2016. The left part of the table corresponds to the Full QMLE and the right part to the EbE QMLE (i.e., the method used to produce the forecasts). We choose to comment this example because it is one of the few cases where not all conditional betas are time-varying.

Table 3: Full and EbE QMLE of the CHAR model for BusEq

	Full QML			EbE QML		
	Coefficient	Std-Err	p-value	Coefficient	Std-Err	p-value
ω_1	0.01924	0.00495	0.00010	0.01924	0.00495	0.00010
α_1	0.09558	0.01134	0.00000	0.09558	0.01134	0.00000
b_1	0.89050	0.01184	0.00000	0.89050	0.01184	0.00000
ω_2	0.00575	0.00173	0.00090	0.00575	0.00173	0.00090
α_2	0.06644	0.01176	0.00000	0.06644	0.01176	0.00000
b_2	0.91329	0.01574	0.00000	0.91329	0.01573	0.00000
ϖ_{21}	0.00057	0.00043	0.18980	0.00058	0.00043	0.17750
τ_{21}	0.02128	0.00334	0.00000	0.02128	0.00305	0.00000
c_{21}	0.99620	0.00162	0.00000	0.99621	0.00160	0.00000
ω_3	0.00163	0.00071	0.02160	0.00163	0.00071	0.02070
α_3	0.08838	0.01965	0.00000	0.08838	0.01942	0.00000
b_3	0.90699	0.02020	0.00000	0.90699	0.01996	0.00000
ϖ_{31}	0.00004	0.00015	0.79040	0.00003	0.00015	0.85730
τ_{31}	0.02336	0.00779	0.00270	0.02336	0.00689	0.00070
c_{31}	0.99546	0.00150	0.00000	0.99547	0.00148	0.00000
ϖ_{32}	-0.24824	0.06639	0.00020	-0.24824	0.07138	0.00050
τ_{32}	0.09610	0.01117	0.00000	0.09610	0.01076	0.00000
c_{32}	-0.30902	0.31845	0.33190	-0.30902	0.35103	0.37870
ω_4	0.00106	0.00060	0.07700	0.00106	0.00061	0.07940
α_4	0.04480	0.01468	0.00230	0.04480	0.01481	0.00250
b_4	0.95173	0.01554	0.00000	0.95173	0.01566	0.00000
ϖ_{41}	0.00394	0.00140	0.00480	0.00394	0.00143	0.00600
τ_{41}	0.01036	0.00223	0.00000	0.01036	0.00223	0.00000
c_{41}	0.99630	0.00126	0.00000	0.99630	0.00130	0.00000
ϖ_{42}	0.08221	0.01696	0.00000	0.08221	0.01734	0.00000
ϖ_{43}	-0.00242	0.00176	0.16800	-0.00242	0.00192	0.20820
τ_{43}	0.04332	0.02636	0.10040	0.04332	0.02665	0.10420
c_{43}	0.99509	0.00365	0.00000	0.99509	0.00405	0.00000

Note: Full QMLE (left panel) and EbE QMLE (right panel) of the CHAR model for BusEq and the period spanning from 2000-02-08 to 2015-12-31. Standard errors are computed using the formulas described in Theorems 4.2 and 4.3.

Importantly, results of the two estimation methods are found to be almost identical, reason

why we rely on the EbE estimator in the forecasting exercise because convergence is achieved much faster with this multi-step method. First, Table 3 suggests that all factors have GARCH effects as the α_i and b_i, $i = 1, ..., 4$ coefficients are statistically significant. Second, the two SMB and HML factors are marginally orthogonal to the MKT factor as the parameters ϖ_{21} and ϖ_{31} are not statistically significant and the estimates of the unconditional betas are 0.15 and -0.19, respectively for $E(\beta_{21,t})$ and $E(\beta_{31,t})$. However, the two corresponding conditional market betas show a time-varying behavior with long term persistence (the two autoregressive coefficients c_{21} and c_{31} are close to 1), meaning that we can use current information to predict non zero conditional market betas for the value and size factors. Concerning the projection of HML on the SMB factor, we observe a negative marginal value but a simpler time-varying behavior with no long term persistance as the parameter $c_{3,2}$ parameter is not statistically different from zero. This last result on SMB and HML justifies our choice to orthogonalize the three factors. Third, we now study the dynamic properties of the industry portfolio returns. While $\beta_{BusEq,SMB,t}$ is constant and slightly positive ($\beta_{42,t} = \varpi_{42} = 0.08$), time-varying behaviors are detected in $\beta_{BusEq,MKT,t}$ and $\beta_{BusEq,HML,t}$, which are found to be very persistent (with estimates of the autoregressive coefficients c_{41} and c_{43} again close to 1).

The estimated conditional betas $\beta_{BusEq,MKT,t}$, $\beta_{BusEq,SMB,t}$ and $\beta_{BusEq,HML,t}$ are plotted in Figure 8. Each graph contains the estimated betas for the four competing models. It is interesting to notice that conditional betas filtered with the CHAR model are much smoother than those obtained with the DCC-DCB model. In the case of $\beta_{BusEq,MKT,t}$ for example, we expect to get a conditional beta stable and close to 1. We observe this result for all models, except at the burst of the internet bubble (i.e., 2000-2002). In particular, the CCC model gives very erratic conditional betas during this period, while conditional market betas of the DCC and CHAR models are much more realistic.

Next, we compute one-step-ahead forecasts $\beta_{k,MKT,t+1|t}, \beta_{k,SMB,t+1|t}$ and $\beta_{k,HML,t+1|t}$, for the four models and the 12 industry portfolios. The sample-size has been chosen so that the first forecast corresponds to the first working day of year 2010 (i.e., 4th of January 2010), which is highlighted by a vertical line in Figures 6 and 7. The total number of forecasts is 1,678. For the sake of illustration, the predicted betas of the first portfolio (BusEq) are plotted in Figure 9. Recall that the models are reevaluated every 3 steps, reason why the forecasts of the model with constant betas (i.e., C-CHAR) are slowly changing over-time. Several comments are in order.

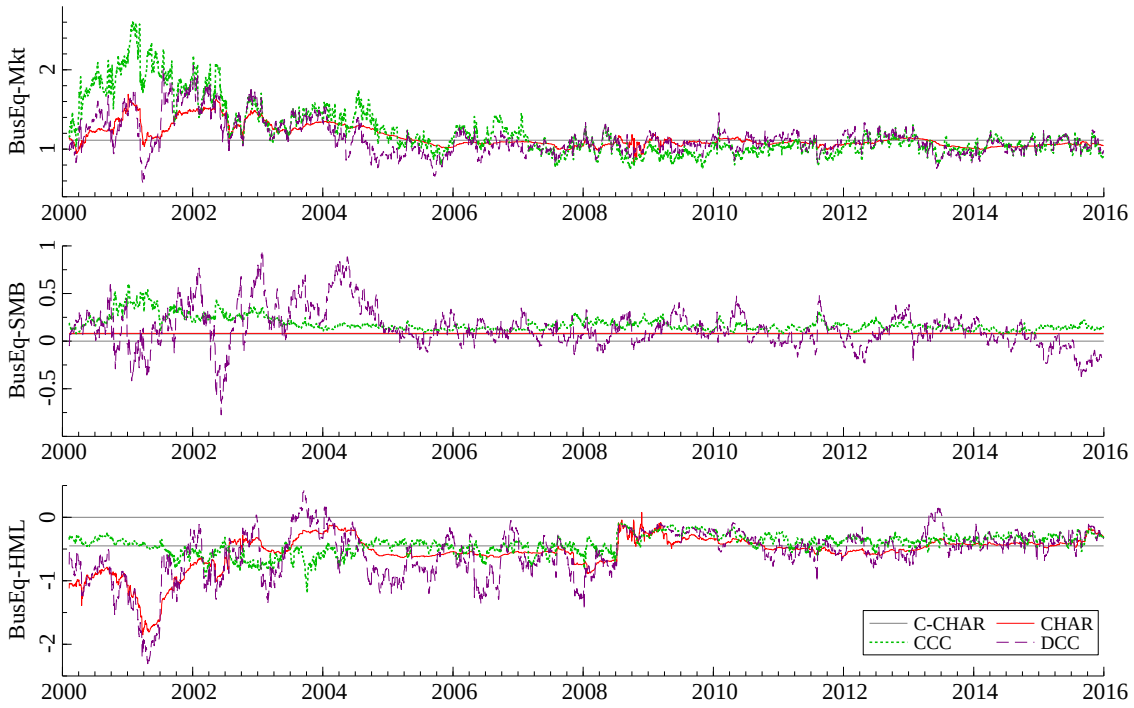

Figure 8: Conditional betas ($\beta_{BusEq,MKT,t}$, $\beta_{BusEq,SMB,t}$ and $\beta_{BusEq,HML,t}$) of the four competing models for the period spanning from 2000-02-08 to 2015-12-31

Figure 9: One-step ahead forecasts of the conditional betas of BusEq for the 4 competing models

First, we observe huge differences between the forecasts produced by the four models. Focusing on the conditional market beta (i.e., the top graph in Figure 9), the CHAR forecasts (red continuous line) are much smoother than those of the CCC-DCB and DCC-DCB models (green and purple dotted lines, respectively). The same comment applies to the other industry portfolios (graphs are not reported to save place but are available upon request).

Second, the CHAR model gives a tradeoff between the too smooth behavior of the C-CHAR model and the shaky behavior of the two DCB models. Interestingly, forecasts of the conditional HML beta of the DCC model deviate much from the forecasts of the other models between March-2013 and August-2013. Indeed, the DCC predicts an almost zero beta while the other 3 models keep predicting conditional betas close to their sample mean. In a risk management perspective, if the objective is to offset the impact of the HML factor onto the Buseq portfolio, the message given by the DCC during this period is to cut the hedging component of the portfolio, while the message of the other three models is to continue with the same risk management policy.

Finally, for each model, the predicted conditional betas are used to build a hedging portfolio

used to immunize industry portfolios against all factors. The returns of this portfolio are easily obtained using the forecasted conditional betas, i.e.,

$$Z_{k,t+1|t} = \beta_{k,MKT,t+1|t} MKT_{t+1} + \beta_{k,SMB,t+1|t} SMB_{t+1} + \beta_{k,HML,t+1|t} HML_{t+1},$$

where MKT_{t+1}, SMB_{t+1} and HML_{t+1} are the realized (non-demeaned) log-returns of the three factors. Economically, this hedging portfolio corresponds to the portfolio invested in the risk factors that optimally tracks the corresponding industry portfolio return. The term hedging means that we can short sell this portfolio to cover the main risks of a given portfolio.

For each of the 12 industry portfolios and the four different hedging portfolios, we compute the realized tracking error as follows:

$$TE_{k,t+1} = r_{k,t+1} - Z_{k,t+1|t}$$

and seek for the model with the smallest sample mean square error (MSE) or mean absolute deviation (MAD) over the 1,678 values of the tracking error by means of the Model Confidence Set of Hansen et al. (2011). A similar approach has been used by Hansen et al. (2014) in the context of Realized Beta GARCH models and more recently by Boudt et al. (2017) on the one-step-ahead forecasts of the CholCov estimator. The Fama French three factor model, predicts that these factors are sufficient to price industry portfolios and therefore the tracking error is expected to have a zero mean (or zero alpha). Before reporting the results of the MCS test, we test the validity of this assumption for the four competing model by regressing $TE_{k,t+1}$ on a constant (on the whole out-of-sample period). Robust (HAC) t-statistics for the null hypothesis that this constant is zero are reported in Table 4. Results are in favor of this assumption in all but one case for the C-CHAR, CHAR and DCC-DCB models and for 9 cases for the CCC-DCB. The only industry portfolios for which this assumption is violated at the 5% significance level is NoDur.

Table 5 contains the results of the MCS test with a MSE loss function, with a significance level of 20%, and 10,000 bootstrap samples (with a block length of 5 observations). Models highlighted with the symbol ✓ are contained in the model confidence set (or set of superior models). The CHAR tracking portfolios always belong to the set of superior models while the DCC-DCB appears in the MCS in only 3 cases.[6] Interestingly, the C-CHAR is always rejected from the MCS suggesting

[6] While summary statistics on the realized tracking errors are not reported in Table 5, it is worth to mention that the CHAR model has the smallest MSE (and MAD) in 11 cases.

Table 4: Robust (HAC) t-statistic of the regression of the tracking errors on a constant

	C-CHAR	CHAR	CCC	DCC
BusEq	-1.025	-0.442	0.133	-0.889
Chems	0.683	0.150	1.187	0.252
Durbl	-0.059	-0.402	-0.123	-0.720
Enrgy	-1.191	-1.865	-0.528	-1.665
Hlth	1.664	1.123	1.777	1.325
Manuf	-0.101	-0.269	0.064	-0.197
Money	-0.699	-0.648	-0.837	-0.541
NoDur	**2.731**	**2.180**	**2.961**	**2.343**
Other	-0.193	-0.436	-0.394	-0.415
Shops	1.746	1.947	**2.447**	1.768
Telcm	1.392	1.604	**2.247**	1.858
Utils	1.053	0.858	1.549	1.010

Note: Robust (HAC) t-statistics for the null hypothesis that the coefficient in the regression of $TE_{k,t+1}$ on a constant is zero. Values in bold are greater (in absolute value) than the critical value at the 5% significance level.

that the assumption of constant betas leads to inferior tracking portfolios. The CCC-DCB models is also rejected from the MCS in all cases despite the fact that although this model imposes the constancy of the correlations it allows the conditional betas to be time-varying. Very similar results are obtained (but not reported to save place) with a MAD loss function (the only difference is that the DCC model appears in the MCS in only 2 cases instead of 3). The main conclusion drawn from Table 5 is that the conditional beta forecasts produced by the CHAR model give the best tracking portfolios and that the DCC-DCB model is not statistically inferior in 3 cases when relying on a MSE loss function.

A natural question arises on whether, the conditional beta forecasts of the CHAR model lead to more or less transaction costs than the DCC-DCB model. To answer this question let $\Delta_{\beta_{k,MKT}} =$

$\sum_{t=2}^{1,678} |\beta_{k,MKT,t+1|t} - \beta_{k,MKT,t|t-1}|$ be the sum of the absolute value of the variations of the predicted conditional stock market betas of the k-th industry portfolio for a given model. A higher value of $\Delta_{\beta_{k,MKT}}$ translates naturally into more transaction costs. Column MKT in Table 6 corresponds to the ratio between the value of $\Delta_{\beta_{k,MKT}}$ obtained for the CHAR and the DCC-DCB models. Columns SMB and HML correspond to the same ratio but for the other two factors. A value close to 1 means that the two models deliver equally stable beta forecasts while a value smaller (resp. greater) than 1 means that the CHAR (resp. DCC-DCB) model delivers more stable beta forecasts and therefore less (resp. more) transaction costs. Interestingly, all ratios are well below 1 (i.e., between about 0.3 and 0.7) suggesting that the conditional beta forecasts of the CHAR model are much smoother than those of the DCC-DCB model and therefore the CHAR model induces much less transaction costs.

7 Conclusion

This paper introduces a new model to estimate time series regressions with time-varying coefficients, called conditional betas in some financial applications. The CHAR model (for Cholesky-GARCH model), is inspired by the model of Pourahmadi (1999), originally proposed in a longitudinal data framework. Unlike in the case of conditional correlations, conditional betas do not need to be constrained except for the stationarity condition, which makes this approach very appealing.

Our model is flexible enough to directly specify the dynamics of conditional betas and impose the constancy of some of these coefficients. We study several specifications of the conditional beta dynamics and derive some stationarity and invertibility conditions. We also prove consistency and asymptotic normality of the Full and Equation-by-Equation QML estimators. The finite sample behavior of these two estimators is also numerically investigated by means of a Monte Carlo simulation.

We illustrate the usefulness of the CHAR model in a portfolio and risk management exercise. We follow Patton and Verardo (2012) and build hedged portfolios to offset some unwanted exposures to the three factors introduced in an asset pricing context by Fama and French (1992, 2004). We use four different models to forecast conditional betas, build tracking portfolios and compute tracking errors between forecasted and realized portfolio returns. We find that the CHAR model with time-varying betas is the one tracking the best the 12 US industry portfolios. Indeed, this model

outperforms the three competing models and delivers much smoother conditional beta forecasts, which translates into smaller transaction costs.

We limit our investigations to dynamic specifications only involving past shocks and past conditional betas. Adding exogenous explanatory variables like in Patton and Verardo (2012) would allow the identification of variables influencing betas over time. As this would complicate the derivation of its statistical properties, we leave this for future work.

8 Proofs

Proof of Theorem 3.1. First note that the top Lyapunov γ_0 is well defined in $[-\infty, \infty)$ because $E \log^+ \|\boldsymbol{H}_1\| < \infty$. Using Bougerol and Picard (1992a), it is shown in Francq and Zakoian (2012) that (2.3) admits a strictly stationary (and non anticipative) solution if and only if $\gamma_0 < 0$. Assume $\gamma_0 < 0$. The stationary and non anticipative solution is unique and ergodic, and is given by the $(2q+1)$-th component of

$$\boldsymbol{z}_t = \boldsymbol{h}_t + \sum_{k=1}^{\infty} \left(\prod_{i=0}^{k-1} \boldsymbol{H}_{t-i} \right) \boldsymbol{h}_{t-k}. \tag{8.1}$$

The process $(\boldsymbol{v}_t', \boldsymbol{g}_t')'$ is then stationary and ergodic. By the ergodic theorem, the process $(\boldsymbol{c}_{0t})_t$, where $\boldsymbol{c}_{0t} = \boldsymbol{c}_0 \left(\boldsymbol{v}_t, \ldots, \boldsymbol{v}_{t-r+1}, \boldsymbol{g}_t, \ldots, \boldsymbol{g}_{t-r+1}\right)$, is also stationary and ergodic. Under the condition *ii*), a strictly stationary non anticipative and ergodic solution to (2.3) (or (2.3)) and is defined by $\boldsymbol{\ell}_t$ (or $\boldsymbol{\beta}_t$) equal to

$$\left\{ \boldsymbol{I}_{m_0} - \sum_{i=1}^{s} \boldsymbol{C}_{0i} B^i \right\}^{-1} \boldsymbol{c}_{0,t-1} = \boldsymbol{c}_{0,t-1} + \sum_{k=1}^{\infty} \boldsymbol{\Pi}_k \boldsymbol{c}_{0,t-k-1},$$

where B denotes the Backward operator. The solution of the CHAR model is then defined by $\boldsymbol{\epsilon}_t = \boldsymbol{L}_t \boldsymbol{v}_t$.

Lemma 2.3 in Berkes, Horváth and Kokoszka (2003) (see also Corollary 2.3 in Francq and Zakoian, 2010) shows that the strict stationarity condition *i*) implies the existence of $s_0 > 0$ such that $E\|\boldsymbol{z}_1\|^{s_0} < \infty$. The condition *iii*) then entails $E\left\|\boldsymbol{c}_{01}\right\|^{2s_0} < \infty$. Under the condition *ii*), we have $\|\boldsymbol{\Pi}_k\| \leq K\rho^k$ for some $K > 0$ and $\rho \in [0, 1)$. We thus have $E\left\|\boldsymbol{\ell}_1\right\|^{s_0} < \infty$ for some $s_0 > 0$. By the Hölder inequality, we conclude that $\|\boldsymbol{\epsilon}_t\|$ and $\|\boldsymbol{\Sigma}_t\|$ admit a small order moment.

Without loss of generality, we now take the matrix norm defined by $\|\boldsymbol{M}\| = \sum_{i,j} |m_{i,j}|$ for any matrix $\boldsymbol{M}$ whose generic element is $m_{i,j}$. Noting that $\|\boldsymbol{A}\|\|\boldsymbol{B}\| = \|\boldsymbol{A} \otimes \boldsymbol{B}\|$ and using elementary

properties of the Kroneker product, we obtain

$$E\left\|\prod_{i=0}^{k-1}\boldsymbol{H}_{t-i}\boldsymbol{h}_{t-k}\right\|^{m}=E\left\|\prod_{i=0}^{k-1}\boldsymbol{H}_{t-i}^{\otimes m_1}\boldsymbol{h}_{t-k}^{\otimes m_1}\right\|=\left\|(E\boldsymbol{H}_1^{\otimes m_1})^k E\boldsymbol{h}_1^{\otimes m_1}\right\|.$$

Therefore, under the conditions (3.1), the L^{m_1}-norm of the k-th term of the sum (8) is bounded by $K\rho^k$, and thus $E\left\|\boldsymbol{z}_t\right\|^{m_1}<\infty$, which entails $E\left\|\boldsymbol{v}_t\right\|^{2m_1}<\infty$ and $E\left\|\boldsymbol{g}_t\right\|^{m_1}<\infty$. Under the condition iii), we then have $E\left\|\boldsymbol{c}_{0t}\right\|^{2m_1}<\infty$, and thus $E\left\|\boldsymbol{\epsilon}_1\right\|^{2m_1}<\infty$ under ii). □

Proof of Corollary 3.1. We write the proof for Model (2.3)-(2.3), the proof for Model (2.3)–(2.3) being identical. Under the Condition *1)* it is well known that g_{1t} is well defined and that $v_{1t}=\sqrt{g_{1t}}\eta_{1t}$ is a stationary GJR process. Assume that there exist stationary, non anticipative and ergodic processes $(v_{1,t},\ldots,v_{i_0-1,t})$ and $(g_{1,t},\ldots,g_{i_0-1,t})$ satisfying (2.3) for $i<i_0$ then, by the ergodic theorem, the process

$$z_{i_0,t}=\omega_{0i_0}+\gamma_{0i_0+}\left(\epsilon_{1,t-1}^{+}\right)^2+\gamma_{0i_0-}\left(\epsilon_{1,t-1}^{-}\right)^2+\sum_{k=2}^{i_0-1}\alpha_{0i_0}^{(k)}v_{k,t-1}^2$$

and, under *3)*, the processes

$$\beta_{ij}=\sum_{\ell=0}^{\infty}c_{0ij}^{\ell}u_{ij,t-1-\ell},\quad u_{ij,t}=\varpi_{0ij}+\varsigma_{0ij+}\epsilon_{1,t-1}^{+}+\varsigma_{0ij-}\epsilon_{1,t-1}^{-}+\sum_{k=2}^{i}\tau_{0ij}^{(k)}v_{k,t-1},$$

for $i<i_0$ and $(i,j)\in\mathfrak{T}_m$ inherit the property of being stationary, non anticipative and ergodic. By Cauchy's rule, under Condition *2)*, the process

$$g_{i_0,t}=z_{i_0,t}+\sum_{\ell=1}^{\infty}z_{i_0,t-\ell}\prod_{k=1}^{\ell}\left\{\alpha_{0i_0}^{(i_0)}\eta_{i_0,t-k}^2+b_{0i_0}\right\}$$

is also stationary, non anticipative and ergodic. The existence of the strictly stationary solution is thus obtained by mathematical induction on i_0. The existence of the moments is shown as in Theorem 3.1. □

Proof of Theorem 4.1. By compactness of Θ, using the arguments of Wald (1949), to establish the strong consistency it suffices to show that for all $\boldsymbol{\vartheta}\neq\boldsymbol{\vartheta}_0$, there exists a neighborhood $V(\boldsymbol{\vartheta})$ of $\boldsymbol{\vartheta}$ such that

$$\liminf_{n\to\infty}\inf_{\boldsymbol{\vartheta}^*\in V(\boldsymbol{\vartheta})}\widetilde{O}_n(\boldsymbol{\vartheta}^*)>\lim_{n\to\infty}\widetilde{O}_n(\boldsymbol{\vartheta}_0),\quad\text{a.s.}\tag{8.2}$$

We first show that $\widetilde{O}_n$ can be replaced by O_n, where

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

Omitting the subscript "$(\boldsymbol{\vartheta})$" in the notations, we have

$$\sup_{\boldsymbol{\vartheta}\in\Theta}|O_n-\widetilde{O}_n|\leq\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\boldsymbol{\epsilon}_t'\left(\boldsymbol{B}_t\boldsymbol{G}_t^{-1}\boldsymbol{B}_t'-\widetilde{\boldsymbol{B}}_t\widetilde{\boldsymbol{G}}_t^{-1}\widetilde{\boldsymbol{B}}_t\right)\boldsymbol{\epsilon}_t\right| \tag{8.3}$$
$$+\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\sup_{\boldsymbol{\vartheta}\in\Theta}|\log g_{it}-\log\widetilde{g}_{it}|\,.$$

Note that we have the decomposition $\boldsymbol{B}_t\boldsymbol{G}_t^{-1}\boldsymbol{B}_t'-\widetilde{\boldsymbol{B}}_t\widetilde{\boldsymbol{G}}_t^{-1}\widetilde{\boldsymbol{B}}_t=A_t+B_t+C_t$, with

$$A_t=(\boldsymbol{B}_t-\widetilde{\boldsymbol{B}}_t)\boldsymbol{G}_t^{-1}\boldsymbol{B}_t',\qquad B_t=\widetilde{\boldsymbol{B}}_t\boldsymbol{G}_t^{-1}(\widetilde{\boldsymbol{G}}_t-\boldsymbol{G}_t)\widetilde{\boldsymbol{G}}_t^{-1}\boldsymbol{B}_t,\qquad C_t=\widetilde{\boldsymbol{B}}_t\widetilde{\boldsymbol{G}}_t^{-1}(\boldsymbol{B}_t-\widetilde{\boldsymbol{B}}_t).$$

In the sequel, we take the spectral norm as matrix norm and the Euclidean norm as vector norm. Using **A1**-**A2**, we obtain

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\boldsymbol{\epsilon}_t'A_t\boldsymbol{\epsilon}_t\right|=\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\ \mathrm{Tr}\ \left\{(\boldsymbol{B}_t-\widetilde{\boldsymbol{B}}_t)\boldsymbol{G}_t^{-1}\boldsymbol{B}_t'\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right\}\right|$$
$$\leq K\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\vartheta}\in\Theta}\|\boldsymbol{\beta}_t-\widetilde{\boldsymbol{\beta}}_t\|\|\boldsymbol{g}_t^{-1}\|\|\boldsymbol{\beta}_t\|\|\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\|\leq K\frac{1}{n}\sum_{t=1}^{n}\rho_t\|\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\|\sup_{\boldsymbol{\vartheta}\in\Theta}\|\boldsymbol{\beta}_t\|.$$

Moreover $\sum_{t=1}^{\infty}\rho_t\|\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\|\sup_{\boldsymbol{\vartheta}\in\Theta}\|\boldsymbol{\beta}_t\|$ is finite a.s. since

$$E\left(\sum_{t=1}^{\infty}\rho_t\left\|\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t\right\|\sup_{\boldsymbol{\vartheta}\in\Theta}\|\boldsymbol{\beta}_t\|\right)^{s_0}\leq\left\{E\left\|\boldsymbol{\epsilon}_1\right\|^{6s_0}\right\}^{1/3}\left\{E\left(\sup_{\boldsymbol{\vartheta}\in\Theta}\|\boldsymbol{\beta}_1\|\right)^{3s_0}\right\}^{1/3}\sum_{t=1}^{\infty}\left\{E\rho_t^{3s_0}\right\}^{1/3}<\infty$$

for s_0 small enough, by **A2** and **A3**. The same bounds are obtained when A_t is replaced by B_t or C_t. Therefore the first term of the right-hand side of the inequality (8) is of order $O(n^{-1})$ a.s. Now, consider the second term. The elementary inequality $\log(x)\leq x+1$ and **A1**-**A2** entail

$$\sup_{\boldsymbol{\vartheta}\in\Theta}|\log g_{it}-\log\widetilde{g}_{it}|=\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\log\left(1+\frac{\widetilde{g}_{it}-g_{it}}{g_{it}}\right)\right|\leq K\rho_t,$$

noting that $g_{it}^{-1}\leq\|\boldsymbol{g}_t^{-1}\|$. We thus have shown that

$$\sup_{\boldsymbol{\vartheta}\in\Theta}|O_n(\boldsymbol{\vartheta})-\widetilde{O}_n(\boldsymbol{\vartheta})|=O(n^{-1})\quad\text{a.s.} \tag{8.4}$$

We now show that $Eq_t(\boldsymbol{\vartheta})$ is well defined in $\mathbb{R}\cup\{+\infty\}$ for all $\boldsymbol{\vartheta}$, and in $\mathbb{R}$ for $\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0$. By **A1** we have

$$Eq_t^-(\boldsymbol{\vartheta})\leq\textstyle\sum_{i=1}^{m}E\log^-g_{it}(\boldsymbol{\vartheta})<\infty.$$

At $\boldsymbol{\vartheta}_0$, Jensen's inequality and **A3** entail

$$Eq_t(\boldsymbol{\vartheta}_0) = E\boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \sum_{i=1}^{m} \frac{1}{s} E \log g_{it}^s(\boldsymbol{\vartheta}_0) < \infty.$$

The previous results and the ergodic theorem then entail that

$$\lim_{n\to\infty} \widetilde{O}_n(\boldsymbol{\vartheta}_0) = \lim_{n\to\infty} O_n(\boldsymbol{\vartheta}_0) = Eq_t(\boldsymbol{\vartheta}_0).$$

Similarly, (8) and the ergodic theorem applied to the stationary process (X_t) with $X_t = \inf_{\boldsymbol{\vartheta}^* \in V_m(\boldsymbol{\vartheta})} q_t(\boldsymbol{\vartheta}^*)$ show that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\vartheta}^* \in V_m(\boldsymbol{\vartheta})} \widetilde{O}_n(\boldsymbol{\vartheta}^*) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \inf_{\boldsymbol{\vartheta}^* \in V_m(\boldsymbol{\vartheta})} q_t(\boldsymbol{\vartheta}^*) = E \inf_{\boldsymbol{\vartheta}^* \in V_m(\boldsymbol{\vartheta})} q_t(\boldsymbol{\vartheta}^*),$$

where $V_m(\boldsymbol{\vartheta})$ denotes the ball of center $\boldsymbol{\vartheta}$ and radius $1/m$. If $E|q_t(\boldsymbol{\vartheta})| < \infty$, by Fatou's lemma and **A5**, for any $\varepsilon > 0$ there exists m sufficiently large such that

$$E \inf_{\boldsymbol{\vartheta}^* \in V_m(\boldsymbol{\vartheta})} q_t(\boldsymbol{\vartheta}^*) > Eq_t(\boldsymbol{\vartheta}) - \varepsilon.$$

If $Eq_t^+(\boldsymbol{\vartheta}) = \infty$, then the left-hand side of the previous inequality can be made arbitrarily large.

To show (8), it thus remains to show that $Eq_t(\boldsymbol{\vartheta})$ is minimized at $\boldsymbol{\vartheta}_0$. Without loss of generality, assume that $Eq_t^+(\boldsymbol{\vartheta}) < \infty$. Let $\lambda_{i,t}$ be the eigenvalues of $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta})$, which are positive. We have

$$\begin{aligned} Eq_t(\boldsymbol{\vartheta}) - Eq_t(\boldsymbol{\vartheta}_0) &= E \log\{|\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}_0)|\} + E\left(\mathrm{Tr}\left\{[\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) - \boldsymbol{I}_m]\right\}\right) \\ &= E\left\{\sum_{i=1}^{m} (\lambda_{it} - 1 - \log \lambda_{it})\right\} \geq 0, \end{aligned}$$

where the inequality is strict unless if $\lambda_{it} = 1$ a.s. for all i, that is iff $\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)$ a.s., which is equivalent to $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ under **A4**. The consistency follows.

Elementary matrix derivative computations yield

$$\begin{aligned} \frac{\partial}{\partial \vartheta_i} q_t(\boldsymbol{\vartheta}_0) &= \mathrm{Tr}\left\{(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t')\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\vartheta}_0)\frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)}{\vartheta_i}\boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\vartheta}_0)\right\} \\ &= \frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}_t(\boldsymbol{\vartheta}_0)}{\partial \vartheta_i}\left\{\boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\vartheta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\vartheta}_0)\right\}\mathrm{vec}(\boldsymbol{I}_m - \boldsymbol{\eta}_t\boldsymbol{\eta}_t'), \end{aligned}$$

where "Tr", "vec" and "$\otimes$" of denote respectively the trace, vec and Kronecker operators. We also have

$$\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} q_t(\boldsymbol{\vartheta}) = \sum_{i=1}^{5} c_i,$$

with

$$
\begin{aligned}
c_1 &= \boldsymbol{\epsilon}_t' \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_j} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \boldsymbol{\epsilon}_t, \\
c_2 &= \boldsymbol{\epsilon}_t' \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_j} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \boldsymbol{\epsilon}_t \\
c_3 &= -\boldsymbol{\epsilon}_t' \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial^2 \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \boldsymbol{\epsilon}_t \\
c_4 &= -\mathrm{Tr}\left(\frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_j} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \right) \\
c_5 &= \mathrm{Tr}\left(\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial^2 \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \right).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
E\left\{ \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} q_t(\boldsymbol{\vartheta}_0) \right\} &= \mathrm{Tr}\left(\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_j} \boldsymbol{\Sigma}_t^{-1/2'}(\boldsymbol{\vartheta}) \right) \\
&= \mathrm{vec}\left(\frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \right)' \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}) \mathrm{vec}\left(\frac{\partial \boldsymbol{\Sigma}_t(\boldsymbol{\vartheta})}{\partial \vartheta_i} \right),
\end{aligned}
$$

using elementary properties of the vec and Kronecker operators "vec" and "$\otimes$". By the consistency and **A6**, we have $\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0 \in \mathring{\Theta}$, and thus almost surely $\partial \sum_{t=1}^n \widetilde{q}_t(\widehat{\boldsymbol{\vartheta}}_n)/\partial \boldsymbol{\vartheta} = \mathbf{0}$ for n large enough. Taylor expansions and **A7**- **A8** thus show that almost surely

$$
\mathbf{0} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial q_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} + \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} q_t(\boldsymbol{\vartheta}_{ij}^*) \right) \sqrt{n} \left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0 \right) + o_P(1),
$$

where the $\boldsymbol{\vartheta}_{ij}^*$'s are between $\widehat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$ componentwise. To show that the previous matrix into brackets converges almost surely to $\boldsymbol{J}$ it suffices to use the ergodic theorem, the continuity of the derivatives, and to show that

$$
E \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0)} \left| \frac{\partial^2 q_t(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \right| < \infty
$$

for some neighborhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$, which follows from **A7** and **A9**.

If $\boldsymbol{J}$ was not invertible, there would exist some nonzero $\boldsymbol{\lambda} \in \mathbb{R}^d$ such that $\boldsymbol{\lambda}' \boldsymbol{J} \boldsymbol{\lambda} = 0$. Since $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}_0) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\vartheta}_0)$ is almost surely positive definite, this entails that $\boldsymbol{D}_t \boldsymbol{\lambda} = \mathbf{0}$ with probability one, which is excluded by **A11**. The Bahadur linearization (4.1) easily follows.

It is clear from (8), **A10** and already given arguments that $\left\{ \frac{\partial}{\partial \boldsymbol{\vartheta}} q_t(\boldsymbol{\vartheta}_0) \right\}_t$ is a square integrable martingale difference. The conclusion then follows from the central limit theorem in Billingsley (1961). □

Proof of Theorem 4.2. First note that Corollary 3.1 and the condition $|b_i| < 1$ uniformly in Θ

entail that **A3** is satisfied. Now recall that **B5** entails (4.1.3). With the notation $\boldsymbol{v}_t(\boldsymbol{\varphi}) = \boldsymbol{B}_t(\boldsymbol{\varphi})\boldsymbol{\epsilon}_t$ and $\widetilde{\boldsymbol{v}}_t(\boldsymbol{\varphi}) = \widetilde{\boldsymbol{B}}_t(\boldsymbol{\varphi})\boldsymbol{\epsilon}_t$, we thus have

$$\sup_{\boldsymbol{\varphi}\in\Theta_{\boldsymbol{\varphi}}} \|\boldsymbol{v}_t(\boldsymbol{\varphi}) - \widetilde{\boldsymbol{v}}_t(\boldsymbol{\varphi})\| \leq K\rho^t \|\boldsymbol{\epsilon}_t\|. \tag{8.5}$$

Thanks to the condition $|b_i| < 1$ in **B3**, one can then define $\boldsymbol{g}_t(\boldsymbol{\vartheta}) = \left(g_{1t}(\boldsymbol{\vartheta}^{(+1)}, \ldots, g_{mt}(\boldsymbol{\vartheta}^{(+m)})\right)'$ by

$$g_{it}(\boldsymbol{\vartheta}^{(+i)}) = \sum_{j=0}^{\infty} b_i^j \left\{\omega_{i,t-j-1} + \sum_{k=2}^{i} \alpha_i^{(k)} v_{k,t-j-1}^2(\boldsymbol{\varphi}^{(+k)})\right\}.$$

One can also define $q_t(\boldsymbol{\vartheta}) = \sum_{i=1}^m q_{it}(\boldsymbol{\vartheta}^{(+i)})$, where $q_{it}(\boldsymbol{\vartheta}^{(+i)})$ is obtained by replacing $\widetilde{v}_{kt}(\boldsymbol{\varphi}^{(+i)})$ and $\widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)})$ by $v_{kt}(\boldsymbol{\varphi}^{(+i)})$ and $g_{it}(\boldsymbol{\vartheta}^{(+i)})$ in $\widetilde{q}_{it}(\boldsymbol{\vartheta}^{(+i)})$.

Using (8) and the compactness of Θ, we have

$$\begin{aligned}
&\left|g_{it}(\boldsymbol{\vartheta}^{(+i)}) - \widetilde{g}_{it}(\boldsymbol{\vartheta}^{(+i)})\right| \\
=&\left|\sum_{k=2}^{i} \alpha_i^{(k)} \left\{v_{k,t-1}^2(\boldsymbol{\varphi}^{(+k)}) - \widetilde{v}_{k,t-1}^2(\boldsymbol{\varphi}^{(+k)})\right\} + b_i \left\{g_{i,t-1}(\boldsymbol{\vartheta}^{(+i)}) - \widetilde{g}_{i,t-1}(\boldsymbol{\vartheta}^{(+i)})\right\}\right| \\
\leq& K\rho^{t-1} \|\boldsymbol{\epsilon}_{t-1}\| \left(\|\boldsymbol{v}_{t-1}\| + \|\boldsymbol{\epsilon}_{t-1}\|\right) + \rho \left|g_{i,t-1}(\boldsymbol{\vartheta}^{(+i)}) - \widetilde{g}_{i,t-1}(\boldsymbol{\vartheta}^{(+i)})\right|.
\end{aligned}$$

It follows that

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \|\boldsymbol{g}_t(\boldsymbol{\vartheta}) - \widetilde{\boldsymbol{g}}_t(\boldsymbol{\vartheta})\| \leq K\rho_t, \quad \rho_t = \rho^t \sum_{i=1}^{t-1} \|\boldsymbol{\epsilon}_i\| \left(\|\boldsymbol{v}_i\| + \|\boldsymbol{\epsilon}_i\|\right).$$

For $s_1 < s_0/2$ and $s_0 < 1$ satisfying **A3**, we have

$$\sum_{t=1}^{\infty} \{E\rho_t^{s_1}\}^{1/3} \leq \sum_{t=1}^{\infty} \left\{t\rho^{s_1 t}\left(E\|\boldsymbol{\epsilon}_1\|^{2s_1} + \sqrt{E\|\boldsymbol{\epsilon}_1\|^{2s_1} E\|\boldsymbol{v}_1\|^{2s_1}}\right)\right\}^{1/3} < \infty.$$

It follows that **A2** is satisfied. The positivity constraints in **B3** entail that **A1** holds true. We then obtain the equation (8) that shows the asymptotic irrelevance of the initial values. Using the ergodic theorem, it follows that $\widetilde{O}_n(\boldsymbol{\vartheta})$ converges almost surely to $Eq_t(\boldsymbol{\vartheta}) = \sum_{i=1}^m Eq_{it}(\boldsymbol{\vartheta}^{(+i)})$, which is well defined in $\mathbb{R} \cup \{+\infty\}$ because

$$\inf_{\boldsymbol{\vartheta}^{(+i)}\in\Theta^{(+i)}} g_{it}(\boldsymbol{\vartheta}^{(+i)}) > 0 \tag{8.6}$$

from **B3**. Since $E\|\boldsymbol{g}_1\|^{s_0} < \infty$ for some $s_0 > 0$ by Corollary 3.1, we have $E \log g_{it}(\boldsymbol{\vartheta}_0^{(+i)}) < \infty$, and

thus $E\left|q_{it}(\boldsymbol{\vartheta}_0^{(+i)})\right| < \infty$. When $Eq_{it}(\boldsymbol{\vartheta}^{(+i)}) < \infty$, we have

$$\begin{aligned} Eq_{it}(\boldsymbol{\vartheta}^{(+i)}) - Eq_{it}(\boldsymbol{\vartheta}_0^{(+i)}) =& E\left\{\log\frac{g_{it}(\boldsymbol{\vartheta}^{(+i)})}{g_{it}(\boldsymbol{\vartheta}_0^{(+i)})} + \frac{g_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{g_{it}(\boldsymbol{\vartheta}^{(+i)})} - 1\right\} + E\frac{\left\{v_{it}(\boldsymbol{\varphi}^{(+i)}) - v_{it}(\boldsymbol{\varphi}_0^{(+i)})\right\}^2}{g_{it}(\boldsymbol{\vartheta}^{(+i)})} \\ &+ 2E\frac{v_{it}(\boldsymbol{\varphi}_0^{(+i)})\left\{v_{it}(\boldsymbol{\varphi}^{(+i)}) - v_{it}(\boldsymbol{\varphi}_0^{(+i)})\right\}}{g_{it}(\boldsymbol{\vartheta}^{(+i)})} \end{aligned} \tag{8.7}$$

with the convention $\boldsymbol{\varphi}^{(+2)} = \boldsymbol{\varphi}^{(2)}$ and $v_{1t}(\boldsymbol{\varphi}^{(+1)}) = \epsilon_{1t}$. The last expectation in (8) is equal to zero because, when $i > 1$,

$$v_{it}(\boldsymbol{\varphi}^{(+i)}) - v_{it}(\boldsymbol{\varphi}_0^{(+i)}) = -\sum_{j=1}^{i-1}\left\{\beta_{ij,t}(\boldsymbol{\varphi}^{(+i)}) - \beta_{ij,t}(\boldsymbol{\varphi}_0^{(+i)})\right\}\epsilon_{jt} \tag{8.8}$$

is orthogonal to $v_{it} = v_{it}(\boldsymbol{\varphi}_0^{(+i)})$. By the inequality $x - 1 \geq \log x$ for $x > 0$, it follows that $Eq_{it}(\boldsymbol{\vartheta}^{(+i)}) \geq Eq_{it}(\boldsymbol{\vartheta}_0^{(+i)})$ with equality if and only if

$$v_{it}(\boldsymbol{\varphi}^{(+i)}) = v_{it}(\boldsymbol{\varphi}_0^{(+i)}) \quad \text{and} \quad g_{it}(\boldsymbol{\vartheta}^{(+i)}) = g_{it}(\boldsymbol{\vartheta}_0^{(+i)}) \quad \text{a.s.} \tag{8.9}$$

In view of (2.3) and (4.1.2), these equalities entail

$$0 = \omega_{0i,t-1} - \omega_{i,t-1} + \sum_{k=2}^{i}\left(\alpha_{0i}^{(k)} - \alpha_i^{(k)}\right)v_{k,t-1}^2 + (b_{0i} - b_i)\,g_{i,t-1}.$$

We thus have

$$\left(\alpha_{0i}^{(i)} - \alpha_i^{(i)}\right)\eta_{i,t-1}^2 = R_{t-1}^{(-i)},$$

where $R_{t-1}^{(-i)}$ is a random variable measurable with respect to the σ-field $\mathcal{F}_{t-1}^{(-i)}$ generated by $\mathcal{F}_{t-2}$ and $\{\eta_{j,t-1}, j \neq i\}$. It follows that the distribution of $\left(\alpha_{0i}^{(i)} - \alpha_i^{(i)}\right)\eta_{i,t-1}^2$ conditionally to $\mathcal{F}_{t-1}^{(-i)}$ is degenerated. Since $\eta_{i,t-1}^2$ is independent of $\mathcal{F}_{t-2}$, the distribution of $\left(\alpha_{0i}^{(i)} - \alpha_i^{(i)}\right)\eta_{i,t-1}^2$ conditionally to $\{\eta_{j,t-1}, j \neq i\}$ is degenerated. By the first part of **B2**, it follows that $\alpha_i^{(i)} = \alpha_{0i}^{(i)}$. Repeating the argument, we show that $\alpha_i^{(k)} = \alpha_{0i}^{(k)}$ for $k = 2, \ldots, i$. Now, noting that

$$\begin{aligned} \omega_{0i,t-1} - \omega_{i,t-1} = \omega_{0i} - \omega_i &+ (\gamma_{0i+} - \gamma_{i+})g_{1,t-1}\eta_{1,t-1}^2 1_{\{\eta_{1,t-1}>0\}} \\ &+ (\gamma_{0i-} - \gamma_{i-})g_{1,t-1}\eta_{1,t-1}^2 1_{\{\eta_{1,t-1}<0\}} \end{aligned}$$

we obtain

$$(\gamma_{0i+} - \gamma_{i+})\eta_{1,t-1}^2 1_{\{\eta_{1,t-1}>0\}} + (\gamma_{0i-} - \gamma_{i-})\eta_{1,t-1}^2 1_{\{\eta_{1,t-1}<0\}} = R_{t-2},$$

where R_t is a random variable measurable with respect to $\mathcal{F}_t$. An equation of the form $ax^2 1_{\{x>0\}} + b = 0$ cannot admit two positive solutions when $a \neq 0$. Therefore the last part of **B2** entails that $\gamma_{0i+} = \gamma_{i+}$ and $\gamma_{0i-} = \gamma_{i-}$. We then obtain

$$0 = \omega_{0i} - \omega_i + (b_{0i} - b_i)\, g_{i,t-1},$$

which entails $\omega_{0i} = \omega_i$ and $b_{0i} = b_i$, when $g_{i,t-1}$ is not constant with probability one (which is guaranteed by **B4** and **B2**). We have shown that (8) implies $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}_0^{(i)}$. Without loss of generality, assume $i > 1$. In view of (8), the first equality of (8) entails

$$\sum_{j=1}^{i-1} \left\{ \beta_{ij,t}(\boldsymbol{\varphi}^{(+i)}) - \beta_{ij,t}(\boldsymbol{\varphi}_0^{(+i)}) \right\} \sqrt{g_{jt}} \eta_{jt} = 0 \quad \text{a.s.}$$

Since the variables $\eta_{1t}, \ldots, \eta_{i-1,t}$ are not linearly dependent, and the variables g_{jt} are strictly positive (by **B2** and **B3**), it follows that

$$\beta_{ij,t}(\boldsymbol{\varphi}^{(+i)}) = \beta_{ij,t}(\boldsymbol{\varphi}_0^{(+i)}) \quad \text{a.s.}, \quad j = 1, \ldots, i-1. \tag{8.10}$$

Using $v_{it}(\boldsymbol{\varphi}^{(+i)}) = v_{it}(\boldsymbol{\varphi}_0^{(+i)})$ in (8), (2.3), (4.1.2) and (8) entails

$$\begin{aligned} 0 =& \varpi_{ij} - \varpi_{0ij} + (\varsigma_{ij+} - \varsigma_{0ij+})\, \epsilon_{1,t-1}^{+} + (\varsigma_{ij-} - \varsigma_{0ij-})\, \epsilon_{1,t-1}^{-} \\ &+ \sum_{k=2}^{i} \left(\tau_{ij}^{(k)} - \tau_{0ij}^{(k)} \right) v_{k,t-1} + (c_{ij} - c_{0ij})\, \beta_{ij,t-1}. \end{aligned}$$

We thus have

$$\left(\tau_{ij}^{(i)} - \tau_{0ij}^{(i)} \right) \eta_{i,t-1} = R_{t-1}^{(-i)},$$

where $R_{t-1}^{(-i)}$ is $\mathcal{F}_{t-1}^{(-i)}$-measurable. **B2** then entails $\tau_{ij}^{(i)} = \tau_{0ij}^{(i)}$. Continuing in this way, by **B2** and **B4** it follows that $\boldsymbol{\varphi}^{(ij)} = \boldsymbol{\varphi}_0^{(ij)}$, and finally we have shown that (8) implies $\boldsymbol{\vartheta}_0^{(i)} = \boldsymbol{\vartheta}^{(i)}$. The remainder of the proof of the consistency of $\widehat{\boldsymbol{\vartheta}}_n$ then follows from a standard compactness argument, as in the proof of Theorem 4.1.

Now we show the convergence in distribution (4.2). We first show that the effect of the initial values becomes asymptotically negligible. In Section 4.1.3, we have seen that, under **B5**, for $t \geq 2$

$$\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}) = \boldsymbol{w}_{t-1}^{*} + \boldsymbol{S}_{t-1} \widetilde{\boldsymbol{\beta}}_{t-1}(\boldsymbol{\varphi}), \quad \boldsymbol{\beta}_t(\boldsymbol{\varphi}) = \boldsymbol{w}_{t-1}^{*} + \boldsymbol{S}_{t-1} \boldsymbol{\beta}_{t-1}(\boldsymbol{\varphi}),$$

with a fixed initial value $\widetilde{\boldsymbol{\beta}}_1(\boldsymbol{\varphi})$. Denoting by φ_k the k-th element of $\boldsymbol{\varphi}$, we thus have

$$\frac{\partial \widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi})}{\partial \varphi_k} = \widetilde{\boldsymbol{w}}_{t-1}^{(k)} + \boldsymbol{S}_{t-1} \frac{\partial \widetilde{\boldsymbol{\beta}}_{t-1}(\boldsymbol{\varphi})}{\partial \varphi_k}, \quad \frac{\partial \boldsymbol{\beta}_t(\boldsymbol{\varphi})}{\partial \varphi_k} = \boldsymbol{w}_{t-1}^{(k)} + \boldsymbol{S}_{t-1} \frac{\partial \boldsymbol{\beta}_{t-1}(\boldsymbol{\varphi})}{\partial \varphi_k},$$

with

$$\widetilde{\boldsymbol{w}}_t^{(k)} = \frac{\partial \boldsymbol{w}_t^*}{\partial \varphi_k} + \frac{\partial \boldsymbol{S}_t}{\partial \varphi_k}\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi}), \quad \boldsymbol{w}_t^{(k)} = \frac{\partial \boldsymbol{w}_t^*}{\partial \varphi_k} + \frac{\partial \boldsymbol{S}_t}{\partial \varphi_k}\boldsymbol{\beta}_t(\boldsymbol{\varphi}).$$

In view of (4.1.3), this entails that

$$\sup_{\boldsymbol{\varphi}\in\Theta_\varphi}\left\|\widetilde{\boldsymbol{w}}_t^{(k)} - \boldsymbol{w}_t^{(k)}\right\| \le K\rho_t, \quad \sup_{\boldsymbol{\varphi}\in\Theta_\varphi}\left\|\frac{\partial\widetilde{\boldsymbol{\beta}}_t(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}'} - \frac{\partial\boldsymbol{\beta}_t(\boldsymbol{\varphi})}{\partial\boldsymbol{\varphi}'}\right\| \le K\rho_t,$$

where ρ_t is as in **A2**. It follows that

$$\sup_{\boldsymbol{\vartheta}\in\Theta}\left\|\frac{\partial O_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} - \frac{\partial\widetilde{O}_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\right\| = O(n^{-1}) \quad \text{a.s.} \tag{8.11}$$

By the consistency and **A6**, we thus have

$$0 = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\vartheta}}\widetilde{q}_t(\widehat{\boldsymbol{\vartheta}}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\vartheta}}q_t(\widehat{\boldsymbol{\vartheta}}_n) + o_P(1).$$

A Taylor expansion then gives

$$o_P(1) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\vartheta}}q_t(\boldsymbol{\vartheta}_0) + \left[\frac{1}{n}\sum_{t=1}^n \frac{\partial^2}{\partial\vartheta_k\partial\vartheta_j}q_t(\boldsymbol{\vartheta}_{kl})\right]\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right),$$

where, for the generic element (k,j) of the matrix into brackets, $\boldsymbol{\vartheta}_{kl}$ is between $\boldsymbol{\vartheta}_0$ and $\widehat{\boldsymbol{\vartheta}}_n$. Omitting the argument "$(\boldsymbol{\vartheta}^{(+i)})$" in $g_{it}(\boldsymbol{\vartheta}^{(+i)})$ and "$(\boldsymbol{\varphi}^{(+i)})$" in $v_{it}(\boldsymbol{\varphi}^{(+i)})$, for $i \ge 2$

$$\frac{\partial q_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial\boldsymbol{\vartheta}^{(+i)}} = \left(1 - \frac{v_{it}^2}{g_{it}}\right)\frac{1}{g_{it}}\frac{\partial g_{it}}{\partial\boldsymbol{\vartheta}^{(+i)}} + 2\frac{v_{it}}{\sqrt{g_{it}}}\frac{1}{\sqrt{g_{it}}}\frac{\partial v_{it}}{\partial\boldsymbol{\vartheta}^{(+i)}}, \tag{8.12}$$

and thus

$$\frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{\partial\boldsymbol{\vartheta}^{(+i)}} = \left(1 - \eta_{it}^2\right)\frac{1}{g_{it}}\frac{\partial g_{it}}{\partial\boldsymbol{\vartheta}^{(+i)}}(\boldsymbol{\vartheta}_0^{(+i)}) + 2\eta_{it}\frac{1}{\sqrt{g_{it}}}\frac{\partial v_{it}}{\partial\boldsymbol{\vartheta}^{(+i)}}(\boldsymbol{\vartheta}_0^{(+i)}). \tag{8.13}$$

To establish that

$$\lim_{n\to\infty}\boldsymbol{I}_n = \boldsymbol{I} := E\frac{\partial q_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}}\frac{\partial q_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\vartheta}'} \qquad \text{a.s.},$$

it suffices to use the ergodic theorem, the consistency of $\widehat{\boldsymbol{\vartheta}}_n$, the arguments of Exercise 7.9 in Francq and Zakoian (2010), and to show that for some neighbourhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$

$$E\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)}\left\|\frac{\partial q_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\frac{\partial q_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}'}\right\| < \infty. \tag{8.14}$$

Note that, by the arguments given in the proof of Corollary 3.1, under **B5** we have

$$E\sup_{\boldsymbol{\varphi}\in\Theta_\varphi}\|\boldsymbol{\beta}_t(\boldsymbol{\varphi})\|^{s_0} < \infty, \quad E\sup_{\boldsymbol{\varphi}\in\Theta_\varphi}\|\boldsymbol{v}_t(\boldsymbol{\varphi})\|^{s_0} < \infty$$

for some $s_0 > 0$. Note that **A6** entails that the first $i+3$ components $\boldsymbol{\vartheta}^{(+i)} \in V(\boldsymbol{\vartheta}_0^{(+i)})$ are strictly positive, for any sufficiently small neighbourhood $V(\boldsymbol{\vartheta}_0^{(+i)})$ of $\boldsymbol{\vartheta}_0^{(+i)}$. By an extension of (5.20) in Hamadeh and Zakoïan (2011) and (42) in Francq and Thieu (2016), we then have

$$E \sup_{\boldsymbol{\vartheta}^{(+i)} \in V(\boldsymbol{\vartheta}_0^{(+i)})} \left\| \frac{1}{g_{it}} \frac{\partial g_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\theta}^{(i)}} \right\|^d < \infty, \qquad E \sup_{\boldsymbol{\vartheta}^{(+i)} \in V(\boldsymbol{\vartheta}_0^{(+i)})} \left\| \frac{1}{g_{it}} \frac{\partial^2 g_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\theta}^{(i)} \partial \boldsymbol{\theta}^{(i)\prime}} \right\|^d < \infty, \tag{8.15}$$

for any integer d and some neighbourhood $V(\boldsymbol{\vartheta}_0^{(+i)})$ of $\boldsymbol{\vartheta}_0^{(+i)}$. Denote by $\varphi_j^{(+i)}$ the components of $\boldsymbol{\varphi}^{(+i)}$, for $j = 1, \ldots, i(i-1)(2i+11)/6$. Now, note that for $i \geq 2$

$$\frac{\partial g_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \varphi_j^{(+i)}} = 2 \sum_{k=0}^{\infty} b_i^k \sum_{\ell=2}^{i} \alpha_i^{(\ell)} v_{\ell,t-k-1} \frac{\partial v_{\ell,t-k-1}(\boldsymbol{\varphi}^{(+\ell)})}{\partial \varphi_j^{(+i)}}.$$

For all $p_1 \geq 1$, using the inequality $x/(1+x^2) \leq 1/2$ for all x, we thus have

$$\begin{aligned}
\left\| \frac{1}{g_{it}} \frac{\partial g_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \varphi_j^{(i)}} \right\|_{p_1} \leq & 2 \sum_{k=0}^{\infty} \sum_{\ell=2}^{i} \left\| \frac{b_i^k \alpha_i^{(\ell)} v_{\ell,t-k-1} \frac{\partial v_{\ell,t-k-1}(\boldsymbol{\varphi}^{(+\ell)})}{\partial \varphi_j^{(+i)}}}{\omega_i + b_i^k \alpha_i^{(\ell)} v_{\ell,t-k-1}^2} \right\|_{p_1} \\
\leq & K \sum_{k=0}^{\infty} \rho^k \sum_{\ell=2}^{i} \left\| \frac{\partial v_{\ell,t-k-1}(\boldsymbol{\varphi}^{(+\ell)})}{\partial \varphi_j^{(+i)}} \right\|_{p_1}.
\end{aligned}$$

In view of (8), the inequalities $\|\boldsymbol{v}_t(\boldsymbol{\varphi})\| \leq \|\boldsymbol{B}_t(\boldsymbol{\varphi})\| \|\boldsymbol{\epsilon}_t\|$ and $\|\frac{\partial \boldsymbol{v}_t(\boldsymbol{\varphi})}{\partial \varphi_k}\| \leq \|\frac{\partial \boldsymbol{B}_t(\boldsymbol{\varphi})}{\partial \varphi_k}\| \|\boldsymbol{\epsilon}_t\|$, and **B6**, for some neighbourhood $V(\boldsymbol{\varphi}_0)$ of $\boldsymbol{\varphi}_0$ we have

$$E \sup_{\boldsymbol{\varphi} \in V(\boldsymbol{\varphi}_0)} \|\boldsymbol{v}_t(\boldsymbol{\varphi})\|^{4p_1} < \infty, \tag{8.16}$$

$$E \sup_{\boldsymbol{\varphi} \in V(\boldsymbol{\varphi}_0)} \left\| \frac{\partial \boldsymbol{v}_t(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} \right\|^{2q_1} < \infty,$$

$$E \sup_{\boldsymbol{\vartheta}^{(+i)} \in V(\boldsymbol{\vartheta}_0^{(+i)})} \left\| \frac{1}{g_{it}} \frac{\partial g_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\vartheta}^{(+i)}} \right\|^{2q_1} < \infty. \tag{8.17}$$

In view of (8), we thus obtain (8) from the Hölder inequality, (8) and (8)-(8).

We now show that $\boldsymbol{I}$ is invertible. If this is not the case, then there exists a non zero vector $\boldsymbol{\lambda} \in \mathbb{R}^d$ such that $\sum_{i=1}^m \boldsymbol{\lambda}' \frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{\partial \boldsymbol{\vartheta}} = 0$ a.s., which implies

$$\sum_{i=1}^{m} \left(1 - \eta_{it}^2\right) \frac{1}{g_{it}} \boldsymbol{\lambda}' \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0^{(+i)}) + 2\eta_{it} \frac{1}{\sqrt{g_{it}}} \boldsymbol{\lambda}' \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\varphi}_0^{(+i)}) = 0 \quad \text{a.s.}$$

By **B2**, this entails

$$\boldsymbol{\lambda}' \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0^{(+i)}) = 0 \quad \text{and} \quad \boldsymbol{\lambda}' \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\varphi}_0^{(+i)}) = 0 \quad \text{a.s. for } i = 1, \ldots, m. \tag{8.18}$$

Let the elimination matrices $\mathbf{E}^0_{\boldsymbol{\theta}^{(i)}}$ and $\mathbf{E}^0_{\boldsymbol{\varphi}^{(ij)}}$ (whose elements are zero or one) such that $\boldsymbol{\theta}^{(i)} = \mathbf{E}^0_{\boldsymbol{\theta}^{(i)}}\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi}^{(ij)} = \mathbf{E}^0_{\boldsymbol{\varphi}^{(ij)}}\boldsymbol{\vartheta}$. Let $\boldsymbol{\lambda}_{\boldsymbol{\theta}^{(i)}} = \mathbf{E}^0_{\boldsymbol{\theta}^{(i)}}\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{\boldsymbol{\varphi}^{(ij)}} = \mathbf{E}^0_{\boldsymbol{\varphi}^{(ij)}}\boldsymbol{\lambda}$. With this notation, from (2.3) we have

$$\begin{aligned}\boldsymbol{\lambda}'\frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0^{(+i)}) =&\boldsymbol{\lambda}'_{\boldsymbol{\theta}^{(i)}}\frac{\partial}{\partial \boldsymbol{\theta}^{(i)}}\left\{\omega_{0i} + \gamma_{0i+}\left(\epsilon^+_{1,t-1}\right)^2 + \gamma_{0i-}\left(\epsilon^-_{1,t-1}\right)^2 + \sum_{k=2}^{i}\alpha_{0i}^{(k)}v^2_{k,t-1}(\boldsymbol{\varphi}_0^{(+k)})\right\}\\ &+ 2\sum_{k=2}^{i}\alpha_{0i}^{(k)}v_{k,t-1}\boldsymbol{\lambda}'\frac{\partial v_{k,t-1}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\varphi}_0^{(+k)}) + \boldsymbol{\lambda}'\frac{\partial g_{i,t-1}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_0^{(+i)}).\end{aligned}$$

Therefore (8) entails

$$\left(1\ (\epsilon^+_{t-1})^2\ (\epsilon^-_{t-1})^2\ v^2_{2,t-1}\cdots\ v^2_{i,t-1}\ g_{i,t-1}\right)\boldsymbol{\lambda}_{\boldsymbol{\theta}^{(i)}} = 0.$$

Then, as in Page 163 of Francq and Zakoian (2010), it can be shown that **B2** and **B4** imply that $\boldsymbol{\lambda}_{\boldsymbol{\theta}^{(i)}} = \mathbf{0}$. Now note that the second equality of (8) implies

$$0 = \boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial v_{it}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}) = -\sum_{\ell=1}^{i-1}\epsilon_{\ell,t}\boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial \beta_{i\ell,t}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}).$$

When $\ell \neq j$, we have

$$\begin{aligned}\boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial \beta_{i\ell,t}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}) =&\tau^{(i)}_{0i\ell}\boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial v_{i,t-1}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}) + c_{0i\ell}\boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial \beta_{i\ell,t-1}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)})\\ =&c_{0i\ell}\boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial \beta_{i\ell,t-1}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}) = 0\end{aligned}$$

by stationarity. By **B2**, we thus have

$$0 = \boldsymbol{\lambda}'_{\boldsymbol{\varphi}^{(ij)}}\frac{\partial \beta_{ij,t}}{\partial \boldsymbol{\varphi}^{(ij)}}(\boldsymbol{\varphi}_0^{(+i)}) = \left(1\ \epsilon^+_{t-1}\ \epsilon^-_{t-1}\ v_{2,t-1}\cdots\ v_{i,t-1}\ \beta_{ij,t-1}\right)\boldsymbol{\lambda}_{\boldsymbol{\varphi}^{(ij)}}.$$

We then show that $\boldsymbol{\lambda}_{\boldsymbol{\varphi}^{(ij)}} = \mathbf{0}$ from **B2** and **B4**. Since the previous results are true for all $i \in \{1,\ldots,m\}$ and all $(i,j) \in \mathfrak{T}_m$, we have shown that $\boldsymbol{\lambda} = \mathbf{0}$, which proves the invertibility by contradiction.

Differentiating (8), we obtain

$$\begin{aligned}\frac{\partial^2 q_{it}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\vartheta}^{(+i)}\partial \boldsymbol{\vartheta}^{(+i)'}} =&\left(1 - \frac{v_{it}^2}{g_{it}}\right)\frac{1}{g_{it}}\frac{\partial^2 g_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}\partial \boldsymbol{\vartheta}^{(+i)'}} + \left(\frac{2v_{it}^2}{g_{it}} - 1\right)\frac{1}{g_{it}}\frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}}\frac{1}{g_{it}}\frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(+i)'}}\\ &+ \frac{2}{g_{it}}\frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}}\frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(+i)'}} + \frac{2v_{it}}{g_{it}}\frac{\partial^2 v_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}\partial \boldsymbol{\vartheta}^{(+i)'}}\\ &- \frac{2v_{it}}{g_{it}}\left\{\frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}}\frac{1}{g_{it}}\frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(+i)'}} + \frac{1}{g_{it}}\frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(+i)}}\frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(+i)'}}\right\}.\end{aligned}$$

We thus have

$$\boldsymbol{J} := \sum_{i=1}^{m} E \frac{\partial^2 q_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} = \sum_{i=1}^{m} E \frac{1}{g_{it}} \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}} \frac{1}{g_{it}} \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}'} + 2E \frac{1}{g_{it}} \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}} \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}'}.$$

The invertibility of $\boldsymbol{J}$ and the convergence of $\boldsymbol{J}_n$ to $\boldsymbol{J}$ is shown by already given arguments. The conclusion follows from the central limit theorem for squared integrable martingale differences. □

Proof of Theorem 4.3. The consistency of $\widehat{\boldsymbol{\vartheta}}_n^{(1)}$ and $\widehat{\boldsymbol{\vartheta}}_n^{(2)}$ is obtained exactly as the consistency in Theorem 4.2. We then continue the proof by induction on i, and assume that $\widehat{\boldsymbol{\vartheta}}_n^{(k)} \to \boldsymbol{\vartheta}_0^{(k)}$ a.s. for $k = 1, \ldots, i-1$. Let $V^{(-i)}(\boldsymbol{\varphi}_0^{(-i)})$ and $V^{(i)}(\boldsymbol{\vartheta}_0^{(i)})$ be arbitrary neighbourhoods of $\boldsymbol{\varphi}_0^{(-i)}$ and $\boldsymbol{\vartheta}_0^{(i)}$, and let

$$V^{(+i)}(\boldsymbol{\vartheta}_0^{(+i)}) = V^{(i)}(\boldsymbol{\vartheta}_0^{(i)}) \times V^{(-i)}(\boldsymbol{\varphi}_0^{(-i)})$$

be the corresponding neighbourhood of $\boldsymbol{\vartheta}_0^{(+i)}$. For any $\boldsymbol{\vartheta}_1^{(i)} \neq \boldsymbol{\vartheta}_0^{(i)}$, by (8) and arguments given in the proof of Theorem 4.1, there exists a neighbourhood $V^{(i)}(\boldsymbol{\vartheta}_1^{(i)})$ of $\boldsymbol{\vartheta}_1^{(i)}$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \inf_{\boldsymbol{\vartheta}^{(i)} \in V^{(i)}(\boldsymbol{\vartheta}_1^{(i)})} q_{it}(\boldsymbol{\vartheta}^{(i)}, \widehat{\boldsymbol{\varphi}}_n^{(-i)}) \geq E \inf_{\boldsymbol{\vartheta}^{(+i)} \in V^{(i)}(\boldsymbol{\vartheta}_1^{(i)}) \times V^{(-i)}(\boldsymbol{\varphi}_0^{(-i)})} q_{it}(\boldsymbol{\vartheta}^{(+i)}) > E q_{it}(\boldsymbol{\vartheta}_0^{(+i)}).$$

By compactness of the parameter space, this entails that, asymptotically, $\widehat{\boldsymbol{\vartheta}}_n^{(i)}$ belongs to $V^{(i)}(\boldsymbol{\vartheta}_0^{(i)})$, which completes the proof of the consistency.

Now we show the convergence in distribution (4.3)-(4.3). For $i = 1$, the result comes from already known results on the QMLE of univariate APARCH models (see Hamadeh and Zakoïan, 2011). Note that in the case $i = 2$ the EbEE still coincides with the one-step QMLE (see Subsection 4.2). Therefore (4.3) follows from Theorem 4.2 applied with $m = 2$. More precisely, in the proof of this theorem, we have shown that for $j = 2$ one has the Bahadur representation

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\vartheta}}_n^{(j)} - \boldsymbol{\vartheta}_0^{(j)} \\ \widehat{\boldsymbol{\varphi}}_n^{(-j)} - \boldsymbol{\varphi}_0^{(-j)} \end{pmatrix} \overset{o_P(1)}{=} \boldsymbol{M}^{(j)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \frac{\partial q_{jt}(\boldsymbol{\vartheta}_0^{(+j)})}{\partial \boldsymbol{\vartheta}^{(j)}} \\ \frac{\partial q_{j-1,t}(\boldsymbol{\vartheta}_0^{(+(j-1))})}{\partial \boldsymbol{\varphi}^{(j-1)}} \\ \vdots \\ \frac{\partial q_{2t}(\boldsymbol{\vartheta}_0^{(2)})}{\partial \boldsymbol{\varphi}^{(2)}} \end{pmatrix} \tag{8.19}$$

for some matrix of the form

$$\boldsymbol{M}^{(j)} = \begin{pmatrix} -\left(\boldsymbol{J}_{\boldsymbol{\vartheta}}^{(j)}\right)^{-1} & -\left(\boldsymbol{J}_{\boldsymbol{\vartheta}}^{(j)}\right)^{-1} \boldsymbol{K}^{(j)} \boldsymbol{M}_{\boldsymbol{\varphi}^-}^{(j)} \\ \boldsymbol{0} & \boldsymbol{M}_{\boldsymbol{\varphi}^-}^{(j)} \end{pmatrix},$$

and the convention $\boldsymbol{M}^{(2)} = -\left(\boldsymbol{J}_{\boldsymbol{\vartheta}}^{(2)}\right)^{-1}$.

We now show (8) by induction on j, for $3 \leq j \leq m$. Similarly to (8), one can show that

$$\sup_{\boldsymbol{\vartheta}^{(+i)} \in \Theta^{(+i)}} \left\| \frac{\partial O_n^{(i)}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\vartheta}^{(+i)}} - \frac{\partial \widetilde{O}_n^{(i)}(\boldsymbol{\vartheta}^{(+i)})}{\partial \boldsymbol{\vartheta}^{(+i)}} \right\| = O(n^{-1}) \quad \text{a.s.} \tag{8.20}$$

We have

$$0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\vartheta}^{(i)}} \widetilde{q}_{it}(\widehat{\boldsymbol{\vartheta}}_n^{(i)}, \widehat{\boldsymbol{\varphi}}_n^{(-i)}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\vartheta}^{(i)}} q_{it}(\widehat{\boldsymbol{\vartheta}}_n^{(i)}, \widehat{\boldsymbol{\varphi}}_n^{(-i)}) + o_P(1),$$

where the first equality comes from **A6** and the consistency, and the second one from (8). A first Taylor expansion thus gives

$$o_P(1) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(i)}, \widehat{\boldsymbol{\varphi}}_n^{(-i)})}{\partial \boldsymbol{\vartheta}^{(i)}} + \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 q_{it}(\boldsymbol{\vartheta}_{kl}^{(+i)})}{\partial \vartheta_k^{(i)} \partial \vartheta_j^{(i)}} \right] \sqrt{n} \left(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)} \right),$$

where, for the generic element (k, j) of the matrix into brackets, $\boldsymbol{\vartheta}_{kl}^{(+i)}$ is between $\boldsymbol{\vartheta}_0^{(+i)}$ and $\widehat{\boldsymbol{\vartheta}}_n^{(+i)}$. Another Taylor expansion yields

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(i)}, \widehat{\boldsymbol{\varphi}}_n^{(-i)})}{\partial \boldsymbol{\vartheta}^{(i)}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)}} + \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 q_{it}(\boldsymbol{\vartheta}_0^{(i)}, \boldsymbol{\varphi}_{kl}^{(-i)})}{\partial \vartheta_k^{(i)} \partial \varphi_j^{(-i)}} \right] \sqrt{n} \left(\widehat{\boldsymbol{\varphi}}_n^{(-i)} - \boldsymbol{\varphi}_0^{(-i)} \right),$$

where the $\boldsymbol{\varphi}_{kl}^{(-i)}$'s are between $\boldsymbol{\varphi}_0^{(-i)}$ and $\widehat{\boldsymbol{\varphi}}_n^{(-i)}$. By the arguments used to prove that $\boldsymbol{I}_n$ and $\boldsymbol{J}_n$ converge to $\boldsymbol{I}$ and $\boldsymbol{J}$ in Theorem 4.2, one can show that

$$\lim_{n \to \infty} \boldsymbol{J}_n^{(i)} = \lim_{n \to \infty} \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 q_{it}(\boldsymbol{\vartheta}_{kl}^{(+i)})}{\partial \vartheta_k^{(i)} \partial \vartheta_j^{(i)}} \right] = \boldsymbol{J}^{(i)}$$

and

$$\lim_{n \to \infty} \boldsymbol{K}_n^{(i)} = \lim_{n \to \infty} \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 q_{it}(\boldsymbol{\vartheta}_0^{(i)}, \boldsymbol{\varphi}_{kl}^{(-i)})}{\partial \vartheta_k^{(i)} \partial \varphi_j^{(-i)}} \right] = \boldsymbol{K}^{(i)},$$

where

$$\boldsymbol{J}^{(i)} = E \frac{1}{g_{it}} \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(i)}} \frac{1}{g_{it}} \frac{\partial g_{it}}{\partial \boldsymbol{\vartheta}^{(i)'}} (\boldsymbol{\vartheta}_0^{(+i)}) + 2E \frac{1}{g_{it}} \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(i)}} \frac{\partial v_{it}}{\partial \boldsymbol{\vartheta}^{(i)'}} (\boldsymbol{\vartheta}_0^{(+i)})$$

is a positive definite matrix, and

$$\boldsymbol{K}^{(i)} = E \frac{\partial^2 q_{it}(\vartheta_0^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)} \partial \boldsymbol{\varphi}^{(-i)}}.$$

We thus have

$$\sqrt{n} \left(\widehat{\boldsymbol{\vartheta}}_n^{(i)} - \boldsymbol{\vartheta}_0^{(i)} \right) \overset{o_P(1)}{=} -\left\{ \boldsymbol{J}^{(i)} \right\}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial q_{it}(\boldsymbol{\vartheta}_0^{(+i)})}{\partial \boldsymbol{\vartheta}^{(i)}} + \boldsymbol{K}^{(i)} \sqrt{n} \left(\widehat{\boldsymbol{\varphi}}_n^{(-i)} - \boldsymbol{\varphi}_0^{(-i)} \right) \right\}.$$

By induction, one can assume that

$$\sqrt{n}\left(\widehat{\boldsymbol{\varphi}}_n^{(-i)}-\boldsymbol{\varphi}_0^{(-i)}\right)=\sqrt{n}\left(\widehat{\boldsymbol{\varphi}}_n^{(+(i-1))}-\boldsymbol{\varphi}_0^{(+(i-1))}\right)=\boldsymbol{M}_{\boldsymbol{\varphi}^-}^{(i)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\frac{\partial q_{i-1,t}(\boldsymbol{\vartheta}_0^{(+(i-1))})}{\partial\boldsymbol{\varphi}^{(i-1)}}\\ \vdots\\ \frac{\partial q_{2t}(\boldsymbol{\vartheta}_0^{(2)})}{\partial\boldsymbol{\varphi}^{(2)}}\end{pmatrix},$$

which holds true for $i-1=2$. It follows that (4.1) holds for $j=i$. By (8), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\frac{\partial q_{jt}(\boldsymbol{\vartheta}_0^{(+j)})}{\partial\boldsymbol{\vartheta}^{(j)}}\\ \frac{\partial q_{i-1,t}(\boldsymbol{\vartheta}_0^{(+(i-1))})}{\partial\boldsymbol{\varphi}^{(i-1)}}\\ \vdots\\ \frac{\partial q_{2t}(\boldsymbol{\vartheta}_0^{(2)})}{\partial\boldsymbol{\varphi}^{(2)}}\end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\boldsymbol{I}^{(+i)}\right).$$

Under **B7**, the matrix $\boldsymbol{I}^{(+i)}$ is block-diagonal, and the conclusion follows. □

References

Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009) Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* 37, 4046–4087.

Avarucci, M., Beutner, E. and Zaffaroni, P. (2013) On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models. *Econometric Theory* 29, 545–566.

Bauwens, L., Hafner, C. and Laurent, S. (2012) Volatility Models. *Handbook of Volatility Models and Their Applications* eds L. Bauwens, C. Hafner and S. Laurent, John Wiley & Sons, Inc., Hoboken, NJ, USA.

Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2006) Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.

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

Billingsley, P. (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Blasques, F., Gorgi, P., Koopman, S.J. and Wintenberger O. (2016) Feasible invertibility conditions and maximum likelihood estimation for observation-driven Models, Tinbergen Institute Discussion Paper 16-082/III

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

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* 72, 498–505.

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

Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1729.

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

Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.

Comte, F. and Lieberman, O. (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

Doornik, J. (2012) Object-oriented matrix programming using Ox. Timberlake Consultants Press.

Durbin J. and Koopman, S.J. (2012) Time series analysis by state space methods, Volume 38, Oxford Statistical Science Series, Oxford University Press.

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

Fama, E. and French, K. (1992) The cross-section of expected returns. *Journal of Finance* 47, 427–465.

Fama, E. and French, K. (2004) The capital asset pricing model: theory and evidences. *Journal of Economic Perspectives* 18, 25–46.

Francq, C. and Thieu, L.Q. (2015) QML inference for volatility models with covariates. MPRA preprint No. 63198.

Francq, C. and Zakoïan, J-M. (2010) *GARCH models: structure, statistical inference and financial applications.* John Wiley.

Francq, C. and Zakoïan, J-M. (2012) QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* 28, 179–206.

Francq, C. and Zakoïan, J-M. (2016) Estimating multivariate volatility models equation by equation. *Journal of the Royal Statistical Society B* 78, 613–635.

Francq, C. and Zakoïan, J-M. (2017) Estimation risk for the VaR of portfolios driven by semi-parametric multivariate models. *Preprint*.

Glosten, L.R., Jaganathan, R. and Runkle, D. (1993) On the relation between the expected values and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

Hamadeh, T. and Zakoïan, J-M. (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141, 488–507.

Hansen, P.R., Lunde, A. and Nason, J.M. (2011) The model confidence set. *Econometrica* 79, 453–497.

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

Harvey, A. (2013). Dynamic models for volatility and heavy tails: With applications to financial and economic time series. New York: Cambridge University Press.

He, C. and Teräsvirta, T. (2004) An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory* 20, 904–926.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.

Ling, S. and McAleer, M. (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

Maheu, J.M and Shamsi, A.Z. (2016) Nonparametric dynamic conditional beta. MPRA Paper No. 73764.

McAleer, M., Hoti, S., and Chan, F. (2009) Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Reviews* 28, 422–440.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Patton, A. and Verardo, M. (2012) Does beta move with news? Firm-specific information flows and learning about profitability. *The Review of Financial Studies* 25, 2789–2839.

Pedersen, R.S. (2017) Inference and testing on the boundary in extended constant conditional correlation GARCH models. *Journal of Econometrics* 196, 23–36.

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

Silvennoinen, A. and Teräsvirta, T. (2009) Multivariate GARCH models. *Handbook of Financial Time Series* T.G. Andersen, R.A. Davis, J-P. Kreiss and T. Mikosch, eds. New York: Springer.

Tsay, R. S. (2010) *Analysis of financial time series*, 3rd Edition, John Wiley & Sons.

Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics* 20, 595–601.

Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics*, 40(4):846–867.

Table 5: Results of the MCS test

	C-CHAR	CHAR	CCC	DCC
BusEq		✓		
Chems		✓		
Durbl		✓		
Enrgy		✓		✓
Hlth		✓		✓
Manuf		✓		
Money		✓		✓
NoDur		✓		
Other		✓		
Shops		✓		
Telcm		✓		
Utils		✓		

The table shows the models included in the Model Confidence Set in the beta hedging exercise. Models highlighted with the symbol ✓ are contained in the model confidence set using a MSE loss function. The significance level for the MCS is set to 20%, and 10,000 bootstrap samples (with a block length of 5 observations). The tracking error is defined as $TE_{k,t+1} = r_{k,t+1} - r_{0,t+1} - Z_{k,t+1|t}$, where $Z_{k,t+1|t} = \beta_{k,MKT,t+1|t} MKT_{t+1} + \beta_{k,SMB,t+1|t} SMB_{t+1} + \beta_{k,HML,t+1|t} HML_{t+1}$.

Table 6: Transaction costs

	MKT	SMB	HML
BusEq	0.356	0.380	0.341
Chems	0.310	0.263	0.376
Durbl	0.419	0.464	0.693
Enrgy	0.373	0.337	0.456
Hlth	0.461	0.667	0.397
Manuf	0.442	0.402	0.430
Money	0.390	0.397	0.366
NoDur	0.414	0.383	0.296
Other	0.273	0.343	0.335
Shops	0.344	0.297	0.395
Telcm	0.334	0.414	0.640
Utils	0.465	0.408	0.431

Note: let $\Delta_{\beta_{k,j}} = \sum_{t=2}^{1,678} |\beta_{k,j,t+1|t} - \beta_{k,j,t|t-1}|$ be the sum of the absolute value of the variations of the predicted conditional betas between the k-th industry portfolio and the j-th (where j = MKT, SMB and HML) factor for a given model. For each column, the figures correspond to the ratio between the value of $\Delta_{\beta_{k,j}}$ obtained for the CHAR and the DCC-DCB models.