Article

High-Dimensional Conditional Covariance Matrices Estimation Using a Factor-GARCH Model

Xiaoling Li ^{1,2}, Xingfa Zhang ^{1,*} and Yuan Li ¹¹ School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China; mel0dy@scau.edu.cn (X.L.); mathly@gzhu.edu.cn (Y.L.)² College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

* Correspondence: xingfazhang@gzhu.edu.cn

Abstract: Estimation of a conditional covariance matrix is an interesting and important research topic in statistics and econometrics. However, modelling ultra-high dimensional dynamic (conditional) covariance structures is known to suffer from the curse of dimensionality or the problem of singularity. To partially solve this problem, this paper establishes a model by combining the ideas of a factor model and a symmetric GARCH model to describe the dynamics of a high-dimensional conditional covariance matrix. Quasi maximum likelihood estimation (QMLE) and least square estimation (LSE) methods are used to estimate the parameters in the model, and the plug-in method is introduced to obtain the estimation of conditional covariance matrix. Asymptotic properties are established for the proposed method, and simulation studies are given to demonstrate its performance. A financial application is presented to support the methodology.

Keywords: factor model; high-dimensional conditional covariance matrix; GARCH model

JEL Classification: G32; C53

Citation: Li, X.; Zhang, X.; Li, Y. High-Dimensional Conditional Covariance Matrices Estimation Using a Factor-GARCH Model. *Symmetry* **2022**, *14*, 158. <https://doi.org/10.3390/sym14010158>

Received: 11 November 2021

Accepted: 9 January 2022

Published: 13 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Estimation of a conditional covariance matrix is an important and popular topic in statistics and econometrics with wide applications in many disciplines, such as engineering, environmentology, psychology, economics and finance. A traditional approach to estimating conditional covariance matrices is based on multivariate GARCH models, e.g., the vectorized multivariate GARCH (VEC-GARCH) model of [1], the Baba, Engle, Kraft and Krone GARCH (BEKK-GARCH) model of [2], the constant conditional correlation GARCH (CCC-GARCH) model of [3] and the dynamic conditional correlation GARCH (DCC-GARCH) model of [4].

These GARCH models can be called symmetric GARCH models, which assume that the response of conditional covariance (volatility) to a shock is a function of the shock strength only without any correlation with the sign of the shock. To capture the leverage effects, some multivariate GARCH models take into account negative shocks with a larger impact on their conditional covariance compared with positive shocks of the same absolute value.

That is to say, the new impact curve, which traces the relation between conditional covariance and the previous shock, is asymmetric. These asymmetric multivariate GARCH models include the asymmetric dynamic covariance GARCH (ADC-GARCH) model of [5], the generalized diagonal VEC-GARCH model of [6] and related models of [7,8].

However, common multivariate GARCH models cannot be directly applied to ultra-high dimensional data, which is often the case when constructing a portfolio allocation in finance. This is because there are too many parameters to be estimated when the dimension is large, see, e.g., [9]. To overcome the problems caused by high dimensions, various methods have been developed to reduce the dimension of data and the number of parameters.

One of effective methods for dimension reduction is the factor analysis for time series, see [10–14] focusing on the conditional mean and [15–17] focusing on the conditional covariance. In addition, there are some variants of the factor model that introduce the GARCH structure to describe the conditional covariance, e.g., the full-factor multivariate GARCH (FF-GARCH) model introduced by [18], the orthogonal GARCH (O-GARCH) model proposed by [19] and the generalized O-GARCH (GO-GARCH) model of [20].

Ref. [21] investigated the asymptotic theory for the general factor GARCH model that nested the above three models. But the model proposed by [21] can't reduce the dimension if loading matrix is a square matrix. O-GARCH model as a special case of [21] is a successful multivariate GARCH model with a parsimonious form. This not only captures the dynamics of conditional covariance but can also reduce the dimensions effectively if the model is identified fairly.

However, the O-GARCH model treats all principal components as conditionally uncorrelated random variables, which may lead to a miss-specification and non-negligible errors in the fitting (see [22]). Moreover, when the number of factors is smaller than the dimension, the conditional covariance matrix obtained by the factor GARCH model is only non-negative definite but not positive-definite.

Recently, Ref. [23] proposed a dynamic structure and developed an estimation procedure for high-dimensional conditional covariance matrices. Their work is insightful but not applicable for the cases when the dimension p is equal to or larger than the sample size T . However, in practical applications, there are many cases facing $p > T$ when modelling the conditional covariance matrices. Hence, it makes sense to study the issue under ultra-high dimensional cases.

Motivated by the above literature, we propose an alternative model for estimating high-dimension conditional covariance matrix. The basic idea is to use a factor model to reduce the dimensions first, then fit the low-dimensional CCC-GARCH model to the common factor and use the univariate GARCH model to fit the idiosyncratic components.

The rest of the article is organised as follows. We begin in Section 2 with the proposed model, estimation methodology and the modelling procedure. In Section 3, asymptotic properties of the proposed estimators, including the necessary assumptions, are given. Section 4 constructs a portfolio allocation based on the formula for Markowitz's optimal portfolio using the proposed model as a conditional covariance matrix. To assess the performance of the proposed method, we illustrate simulation and a real example in Section 5. Section 6 provides our conclusions. All proofs of theoretical results are put in the Appendix A.

Throughout the paper, we use A^τ , $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the transpose and the minimum and maximum eigenvalues of a matrix A . Let $\text{vec}(A)$ be the column-stacking vector of a matrix A . The notations $\|A\|_F$, $\|A\|$ and $\|A\|_{\max}$ represent the Frobenius norm, spectral norm (operator norm) and element-wise norm of a matrix A , defined respectively by $\|A\|_F = \text{tr}^{1/2}(A^\tau A)$, $\|A\| = \lambda_{\max}^{1/2}(A^\tau A)$ and $\|A\|_{\max} = \max_{i,j} |a_{ij}|$.

When A is a vector, both $\|A\|_F$ and $\|A\|$ are equal to the Euclidean norm. For a random sequence a_n , $a_n = O_p(b_n)$ for some sequence b_n means that $P\{\|a_n\| > Cb_n\} = O(n^{-(1+\epsilon)})$, where ϵ is defined in Assumption 1(iii).

2. The Model and Methodology

2.1. Setting

In the proposed model, the symmetric GARCH model is used. Let y_t be an observed p -dimensional time series and assume $Ey_t = E\{y_t|\mathcal{F}_{t-1}\} = 0$. Then, y_t admits the following structure:

$$y_t = Bx_t + u_t, \quad (1)$$

where $B = (b_1, b_2, \dots, b_p)^\tau$, for $\{b_i\}_{K \times 1}$ is the unobservable factor loading vectors; $x_t = (x_{1t}, \dots, x_{Kt})^\tau$ is an observable K -dimensional vector of common factors, and K is fixed.

$u_t = (u_{1t}, \dots, u_{pt})^\tau$ is an unknown p -dimensional vector of idiosyncratic errors, which is independent of $\{x_t\}_{t \geq 1}$. We assume x_t satisfies that

$$\text{Var}(x_t | \mathcal{F}_{t-1}) \equiv \Sigma_x(t) = D_t \Gamma D_t, \quad (2)$$

where $D_t = \text{diag}(h_{1t}^{1/2}, \dots, h_{Kt}^{1/2})$, and $\underline{h}_t = (h_{1t}, \dots, h_{Kt})^\tau$ with

$$\underline{h}_t = \omega + \sum_{i=1}^{q_1} \Phi_i x_{t-i}^2 + \sum_{j=1}^{q_2} \Psi_j \underline{h}_{t-j},$$

$x_t^2 = (x_{1t}^2, \dots, x_{Kt}^2)^\tau$, and Γ is a symmetric matrix with $\gamma_{ij} = \gamma_{ji}$, for $1 \leq i, j \leq K$, that is

$$\Gamma = \begin{pmatrix} 1 & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & 1 & \cdots & \gamma_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{K1} & \gamma_{K2} & \cdots & 1 \end{pmatrix}.$$

We further assume that each component of u_t follows a univariate GARCH model,

$$E(u_t | \mathcal{F}_{t-1}) = 0, \quad \text{cov}(u_t | \mathcal{F}_{t-1}) \equiv \Sigma_u(t) = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{pt}^2),$$

where

$$\sigma_{\ell t}^2 = \alpha_{\ell 0} + \sum_{i=1}^m \alpha_{\ell i} u_{\ell t-i}^2 + \sum_{j=1}^s \beta_{\ell j} \sigma_{\ell t-j}^2, \quad \ell = 1, \dots, p. \quad (3)$$

The true parameters in the CCC-GARCH model of the common factor and univariate GARCH of the idiosyncratic components are denoted by $\phi = (\delta^\tau, \gamma^\tau)^\tau$ and $\theta = (\theta_1^\tau, \dots, \theta_p^\tau)^\tau$ respectively, where $\delta = \text{vec}(\omega, \Phi_1, \dots, \Phi_{q_1}, \Psi_1, \dots, \Psi_{q_2})$, $\gamma = (\gamma_{21}, \dots, \gamma_{K1}, \gamma_{32}, \dots, \gamma_{K2}, \dots, \gamma_{KK-1})^\tau$ and $\theta_\ell = (\alpha_{\ell 0}, \dots, \alpha_{\ell m}, \beta_{\ell 1}, \dots, \beta_{\ell s})^\tau$, for $\ell = 1, \dots, p$. The number of parameters is $[K + (q_1 + q_2) * K^2] + K * (K - 1)/2 + (1 + m + s) * p$.

We focus our attention on the conditional covariance matrix of y_t , which is based on the assumption that $p \geq T$:

$$\Sigma_y(t) = B \Sigma_x(t) B^\tau + \Sigma_u(t). \quad (4)$$

Compared to the existent factor GARCH models, there are two improved points for our model. One is that we assume the conditional covariance matrix consists of a non-negative positive matrix and a diagonal matrix, which is helpful to flexibly describe the dynamics and guarantee the positive-definiteness of the matrix even though the number of factor K is quite smaller than dimension p . Another is that our factor loading matrix B is only assumed as a full rank matrix and not necessarily to be orthogonal.

2.2. Estimation Procedure

In this study, we assume that $\{y_t, x_t\}_{t \geq 1}$ is observable. Similar to [24], we first use the least squares method to estimate the factor loading matrix B and obtain

$$\hat{B} = Y X^\tau (X X^\tau)^{-1},$$

where $Y = (y_1, \dots, y_T)$, $X = (x_1, \dots, x_T)$, and $\hat{B} = (\hat{b}_1, \dots, \hat{b}_p)^\tau$. Then, let

$$e_{\ell t} \equiv \hat{u}_{\ell t} = y_{\ell t} - \hat{b}_\ell^\tau x_t, \quad \ell = 1, \dots, p.$$

Following [23], let $u_{\ell t}$ in (3) be replaced by $e_{\ell t}$, and a synthetic GARCH model can be acquired as follows,

$$\sigma_{\ell t}^2 = \alpha_{\ell 0} + \sum_{i=1}^m \alpha_{\ell i} e_{\ell t-i}^2 + \sum_{j=1}^s \beta_{\ell j} \sigma_{\ell t-j}^2, \quad \ell = 1, \dots, p, \quad (5)$$

for each $\ell = 1, \dots, p$, assume $\theta_\ell = (\alpha_{\ell 0}, \dots, \alpha_{\ell m}, \beta_{\ell 1}, \dots, \beta_{\ell s})^\tau$. Note that the parameters estimating of (5) are based on the residuals rather than the observed data; hence, we call (5) a synthetic GARCH model. The QML approach is used to estimate θ_ℓ , and the negative quasi log-likelihood function of θ_ℓ is defined as follows

$$Q_{\ell T}(\theta_\ell) = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{e_{\ell t}^2}{\sigma_{\ell t}^2(\theta_\ell)} + \log \sigma_{\ell t}^2(\theta_\ell) \right\}, \quad (6)$$

where $\sigma_{\ell t}^2(\theta_\ell)$ can be recursively computed by (5) with initial values

$$e_{\ell 0}^2 = \dots = e_{\ell 1-m}^2 = \sigma_{\ell 0}^2 = \dots = \sigma_{\ell 1-s}^2 = \alpha_{\ell 0}. \quad (7)$$

Note that [25] demonstrated that the choice of the initial values is unimportant for the asymptotic properties of the QMLE. Hence, we minimise (6) corresponding to θ_ℓ on a compact set Λ defined in Assumption 1 and obtain the QMLE of $\hat{\theta}_\ell$ as follows,

$$\hat{\theta}_\ell = \arg \min_{\theta_\ell \in \Lambda} Q_{\ell T}(\theta_\ell), \quad \ell = 1, \dots, p.$$

Once θ_ℓ s are estimated, substitute them into (5) and set the initial values for $\sigma_{\ell t}^2$ and $e_{\ell t}^2$, e.g., in (7), an estimator $\hat{\sigma}_{\ell t}^2$ of $\sigma_{\ell t}^2$ can be acquired, namely an estimator of $\Sigma_u(t)$ is obtained.

In order to estimate $\Sigma_x(t)$, for simplicity, it is assumed that K in (1) is known in this study. We first estimate ϕ by QMLE based on the observable factors x_t . Then, obtain the estimator of $\Sigma_x(t)$ by substituting it into (2). The process is as the following two steps:
Step 1. We assume

$$\begin{aligned} x_t &= \Sigma_x^{1/2}(t) \eta_t, \quad \Sigma_x(t) = D_t \Gamma D_t, \\ D_t &= \text{diag}^{1/2}(h_t), \\ h_t &= \omega + \sum_{i=1}^{q_1} \Phi_i x_{t-i}^2 + \sum_{j=1}^{q_2} \Psi_j h_{t-j}, \quad x_t^2 = (x_{1t}^2, \dots, x_{Kt}^2)^\tau, \end{aligned} \quad (8)$$

where $\eta_t \sim \text{IID}(0, I_K)$. We construct the likelihood function

$$L_T(\phi) = \frac{1}{T} \sum_{t=1}^T l_t(\phi), \quad l_t(\phi) = -\frac{1}{2} \ln |D_t \Gamma D_t| - \frac{1}{2} x_t^\tau (D_t \Gamma D_t)^{-1} x_t, \quad (9)$$

with initial values

$$x_0^2 = \dots = x_{1-q_1}^2 = h_0 = \dots = h_{1-q_2} = \omega. \quad (10)$$

Maximize (9), we obtain the QMLE of ϕ as follows

$$\hat{\phi}_T = \arg \max_{\phi \in \Theta} L_T(\phi).$$

Step 2. By substituting the initial values (10) and the estimator $\hat{\phi}_T$ into (8), an estimator $\hat{h}_{kt}$ of h_{kt} , and then the estimator $\hat{\Sigma}_x(t)$ of $\Sigma_x(t)$ are acquired.

Similar to [24], by using the plug-in method, we substitute all corresponding estimators into (4) and obtain the estimator of conditional covariance matrix of y_t as follows

$$\hat{\Sigma}_y(t) = \hat{B} \hat{\Sigma}_x(t) \hat{B}^\tau + \hat{\Sigma}_u(t).$$

3. Theoretical Properties

In this section, we present the asymptotic properties of the proposed estimators. First, the strong mixing conditions are imposed to conduct asymptotic analysis of the LSE. Denote $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_T^\infty$ as the σ -algebras generated by $\{(x_t, u_t) : t \leq 0\}$ and $\{(x_t, u_t) : t \geq T\}$, respectively. The mixing coefficient is defined as

$$\alpha(T) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_T^\infty} |P(A)P(B) - P(AB)|. \quad (11)$$

Assumption 1.

- (i). For each $\ell = 1, \dots, p$, let $\{u_{\ell t}, \sigma_{\ell t}^2\}_{t \geq 1}$ be a strictly stationary process with GARCH(m, s) structure, for $\sup_{1 \leq \ell \leq p} E\sigma_{\ell 1}^{2\kappa} < \infty$ and $\kappa > 4$. $\{u_t\}_{t \geq 1}$ and $\{x_t\}_{t \geq 1}$ are assumed independent of each other.
- (ii). Assume $\zeta_{\ell t} = \sigma_{\ell t}^{-1} u_{\ell t}$, for each t and ℓ . Then, for each $\ell = 1, \dots, p$, $\zeta_{\ell t} \sim \text{iid}(0, 1)$ and $\sup_{\ell \leq p} E(\zeta_{\ell t}^{2\kappa}) \leq \infty$, where $0 < 2\epsilon \leq \kappa/2 - 2$, with a small constant ϵ .
- (iii). For each $\ell = 1, \dots, p$, the true value $\theta_{\ell,0}$ belongs to the interior of Λ , where Λ is a compact set, and $\Lambda \subset (c, +\infty) \times (c, +\infty)^{m+s}$ for a constant $c > 0$.
- (iv). Assume $\mathcal{A}_{\ell\theta_\ell}(z) = \sum_{i=1}^m \alpha_{\ell i} z^i$ and $\mathcal{B}_{\ell\theta_\ell}(z) = 1 - \sum_{j=1}^s \beta_{\ell j} z^j$, for $\ell = 1, \dots, p$. If $s > 0$, $\mathcal{A}_{\ell\theta_{\ell,0}}(z)$ and $\mathcal{B}_{\ell\theta_{\ell,0}}(z)$ have no common roots, $\mathcal{A}_{\ell\theta_{\ell,0}}(1) \neq 0$, and $\alpha_{\ell,0m} + \beta_{\ell,0s} \neq 0$.

Assumption 1 is a standard regularity assumption in the literature. The second part of condition (i) is standard in factor model literature (see [24,26]), allowing $\{u_t\}_{t \geq 1}$ to be weakly dependent with respect to time t . Other conditions are common assumptions of univariate GARCH models (see, e.g., [23]).

Assumption 2.

- (1). There is a constant L , such that $P\left\{\sup_{t \geq 1} \|x_t\| \leq L\right\} = 1 - O(1/T^{1+\epsilon})$.
- (2). There is a constant $M > 0$, for each i and j , such that $|b_{ij}| < M$.
- (3). There exist constants $v_1 > 0$ and $b_1 > 0$, for all $y > 0$ and $\ell \leq p$, such that

$$P(|u_{\ell t}| > y) \leq \exp\{-(y/b_1)^{v_1}\}.$$

Conditions (1) and (2) require that x_t and b_i are bounded. Condition (3) allows us to apply the Bernstein-type inequality for the weakly dependent data. This assumption is standard in the factor model as in [24].

Assumption 3. $\|p^{-1}B^\tau B - \Omega\| = o(1)$ for some $K \times K$ symmetric positive definite matrix Ω such that $\lambda_{\min}(\Omega)$ is bounded away from zero.

Assumption 3 requires that the factors should be pervasive, that is, impact every individual time series (e.g., [24,26]).

Assumption 4.

Let $\mathcal{A}_\phi(z) = \sum_{i=1}^{q_1} \Phi_i z^i$ and $\mathcal{B}_\phi(z) = I_K - \sum_{j=1}^{q_2} \Psi_j z^j$. Following convention, $\mathcal{A}_\phi(z) = 0$, if $q_1 = 0$, and $\mathcal{B}_\phi(z) = I_K$ if $q_2 = 0$.

- (a). The parameter space Θ is a compact subspace of Euclidean space, such that ϕ_0 is an interior point in Θ .
- (b). $\{x_t, \eta_t\}_{t \geq 1}$ is a strictly stationary and ergodic process.

- (c). The components of η_t are independent and their squares have non-degenerate distributions. In addition, there exist constants $v_3 > 0$ with $3v_3^{-1} + v_2^{-1} > 1$ and $b_3 > 0$, such that, for all $y > 0$ and $i \leq K$

$$P(|\eta_{it}| > y) \leq \exp(-(y/b_3)^{v_3}).$$

- (d). If $q_2 > 0$, then $\mathcal{A}_{\phi_0}(z)$ and $\mathcal{B}_{\phi_0}(z)$ are left coprime and $M_1(\mathcal{A}_{\phi_0}, \mathcal{B}_{\phi_0})$ has full rank K , where $M_1(\mathcal{A}_{\phi_0}, \mathcal{B}_{\phi_0}) = [\Phi_{q_1} \quad \Psi_{q_2}]$.
- (e). Γ is a positive correlation matrix for all $\phi \in \Theta$.
- (f). $E\|\eta_t \eta_t^\tau\|^2 < \infty$.

The above assumption is standard in CCC-GARCH models as in [25]. Conditions (a) and (b) are common conditions for GARCH model. Condition (c) allows us to apply the Bernstein-type inequality for the weakly dependent data. Condition (d) makes sure that $\mathcal{A}_{\phi_0}$ and $\mathcal{B}_{\phi_0}$ have no common roots. $\Sigma_x(t)$ is positive definite under Condition (e). Under Condition (f), $\{x_t\}_{t \geq 1}$ has finite d -order moments.

Remark 1. Under Assumptions 1 and 4, $\{u_t\}$ and $\{x_t\}$ are strictly stationary and ergodic, and α -mixing with geometric rate (see [27]).

Lemma 1. Suppose $0 < 2\epsilon \leq \kappa/2 - 2$ and $\kappa > 4$. Under Assumption 4, there exists a constant $C_1 > 0$, such that

$$P\left\{\|\hat{\Sigma}_x(t) - \Sigma_x(t)\|_F^2 > \frac{C_1 \log T}{T}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

Remark 2. The result of Lemma 1 can be proved in a similar fashion to Theorem 1 in [28].

Theorem 1. Under Assumptions 1–3, there exist a constant $C_2 > 0$ and a small constant ϵ defined in Lemma 1, such that

$$P\left\{\sup_{1 \leq \ell \leq p} \|\hat{\theta}_\ell - \theta_\ell\| > C_2 \sqrt{\frac{\log p}{T}}\right\} = O\left(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\right).$$

When p diverges, the Frobenius norm and operator norm can not fully employ the factor structure (see [29]). Hence, other norms are sought to help us understand the factor structure, and the entropy loss norm introduced by [30] is a good choice. Its formula is

$$\|\hat{A} - A\|_\Sigma = p^{-1/2} \|A^{-1/2}(\hat{A} - A)A^{-1/2}\|_F.$$

Next, we are going to give the asymptotic properties of the estimator for $\Sigma_y(t)$ in the following theorem, where $\nu^{-1} = 1.5\nu_1^{-1} + 1.5\nu_3^{-1} + \nu_2^{-1}$, ν_1 , ν_2 and ν_3 are defined in Assumptions 2 and 4, and Lemma A2.

Theorem 2. Suppose $(\log p)^{6/\nu-1} = o(T)$. Then, under Assumptions 1–4, there exist constants $C_3, C_4 > 0$ and a small constant ϵ defined in Lemma 1, such that

$$(i) \quad P\left\{\left\|\hat{\Sigma}_y(t) - \Sigma_y(t)\right\|_\Sigma^2 \leq \frac{C_3 p (\log p)^2}{T^2} + \frac{C_3 \log T}{T}\right\} = 1 - \left(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\right);$$

(ii)

$$P\left\{\left\|\hat{\Sigma}_y(t) - \Sigma_y(t)\right\|_{\max}^2 \leq \frac{C_4 \log p + C_4 \log T}{T}\right\} = 1 - O\left(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\right).$$

4. Portfolio Allocation

In this section, a brief introduction on the procedures of constructing the estimated portfolio allocation using our model is given. First, denote $E(y_t|\mathcal{F}_{t-1})$ as the conditional expectation of y_t , and take conditional expectation of (1), then we can acquire

$$E(y_t|\mathcal{F}_{t-1}) = BE(x_t|\mathcal{F}_{t-1}).$$

Hence,

$$\hat{E}(y_t|\mathcal{F}_{t-1}) = \hat{B}\hat{E}(x_t|\mathcal{F}_{t-1}) \quad (12)$$

is utilized as the estimate of $E(y_t|\mathcal{F}_{t-1})$, where $E(x_t|\mathcal{F}_{t-1})$ is estimated by applying a VAR(1) model (see [31]) as follows

$$\hat{E}(x_t|\mathcal{F}_{t-1}) = \hat{\mu} + \hat{\Xi}x_{t-1}.$$

Based on the mean-variance optimal portfolio proposed by [32,33], we construct the estimated optimal portfolio allocation similar to [23]: let w denote the allocation vector of p risky assets, to be held between times $t-1$ and t , solve the following minimization problem, and assume the solution as the estimate of w ,

$$\begin{aligned} \min_w w^\tau \Sigma_y(t) w, \\ \text{subject to } w^\tau 1_p = 1 \text{ and } w^\tau E(y_t|\mathcal{F}_{t-1}) = \pi, \end{aligned}$$

with π as the target return imposed on the portfolio, and 1_p is a p -dimensional column vector of 1. Existent result gives the explicit estimator of w as follows

$$\hat{w} = \frac{d_3 - d_2\delta}{d_1d_3 - d_2^2} \hat{\Sigma}_y^{-1}(t) 1_p + \frac{d_1\delta - d_2}{d_1d_3 - d_2^2} \hat{\Sigma}_y^{-1}(t) \hat{E}(y_t|\mathcal{F}_{t-1}), \quad (13)$$

where

$$\begin{aligned} d_1 &= 1_p^\tau \hat{\Sigma}_y^{-1}(t) 1_p, \quad d_2 = 1_p^\tau \hat{\Sigma}_y^{-1}(t) \hat{E}(y_t|\mathcal{F}_{t-1}), \\ d_3 &= \hat{E}^\tau(y_t|\mathcal{F}_{t-1}) \hat{\Sigma}_y^{-1}(t) \hat{E}(y_t|\mathcal{F}_{t-1}). \end{aligned}$$

5. Numerical Properties

The purpose of this section is to assess the finite-sample performance of the proposed model by using simulation and empirical studies.

5.1. Simulation Studies

In this section, we use a simulated example to illustrate the asymptotic properties, and show how well the proposed model works. In order to display the rates of convergence when p gradually increases, we use the following settings for $\{T, p\}$ with replication time $N = 200$: $\{T, p\} = \{500, 50\}, \{500, 250\}, \{500, 375\}, \{500, 500\}, \{500, 600\}$, and $\{1000, 50\}, \{1000, 250\}, \{1000, 500\}, \{1000, 750\}, \{1000, 1000\}, \{1000, 1200\}$. Let

$$K = 3, \quad q_1 = q_2 = 1, \quad m = s = 1.$$

In model $\sigma_{\ell t}^2 = \alpha_{\ell 0} + \alpha_{\ell 1} u_{\ell t-1}^2 + \beta_{\ell 1} \sigma_{\ell t-1}^2$, for $\ell = 1, \dots, p$, we assume

$$\alpha_{\ell 0} = 0.25, \quad \alpha_{\ell 1} = 0.15, \quad \beta_{\ell 1} = 0.1.$$

In the conditional covariance matrix $\Sigma_x(t) = D_t \Gamma D_t$ of x_t , we set

$$\Gamma = \begin{pmatrix} 1 & 0.1 & -0.2 \\ 0.1 & 1 & 0.15 \\ -0.2 & 0.15 & 1 \end{pmatrix},$$

and $D_t = \text{diag}^{1/2}(h_{1t}, h_{2t}, h_{3t})$, with

$$\begin{cases} h_{1t} = 0.01 + 0.1x_{1t-1}^2 + 0.2h_{1t-1}, \\ h_{2t} = 0.015 + 0.15x_{2t-1}^2 + 0.1h_{2t-1}, \\ h_{3t} = 0.02 + 0.1x_{3t-1}^2 + 0.1h_{3t-1}. \end{cases}$$

In order to define B , we generate its entries independently from a uniform distribution on $[-2, 2]$, and use these same values throughout all simulations.

Now, for each fixed p , the following steps give detailed simulation procedures:

- (1) Generate each component of $\{b_i\}_{i=1}^p$ independently from $U[-2, 2]$ and let $B = (b_1, \dots, b_p)^\tau$;
- (2) Generate $\zeta_{\ell t}$ from $N(0, 1)$, and let $u_{\ell t} = \sigma_{\ell t} \zeta_{\ell t}$, for each $\ell = 1, \dots, p$ and $t = 1, \dots, T + 1$;
- (3) Generate η_t from $N_K(0, I_K)$, where I_K is the K -dimensional identity matrix, and let $x_t = \Sigma_x^{1/2}(t) \eta_t$, for $t = 1, \dots, T + 1$;
- (4) Calculate $y_t = Bx_t + u_t$, for $t = 1, \dots, T + 1$.

Remark 3. We generate $T + 1$ samples but use only the first T data points to estimate $\Sigma_y(t)$. The additional point (x_{T+1}, y_{T+1}) helps us to get the 1-period simple return

$$R(\hat{w}) = \hat{w}^\tau y_{T+1} \quad (14)$$

of a portfolio allocation $\hat{w}$, where $\hat{w}$ is estimated at time T based on (x_t, y_t) , $t = 1, \dots, T$.

The main purpose of the proposed simulation is to describe the convergence rate of $\hat{\Sigma}_y(t)$ to $\Sigma_y(t)$ based on our model and O-GARCH model. We first calculate the means and standard deviations of $\hat{\Sigma}_y(t) - \Sigma_y(t)$ under the entropy norm $\|\cdot\|_\Sigma$ and the element-wise norm $\|\cdot\|_{\max}$. We also obtain the means and standard deviations of $\|\hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\|$ based on the proposed model. The conditional covariance matrix of O-GARCH model is singular when $K < p$, and hence we omit the corresponding estimate of $\Sigma_y^{-1}(t)$ based on O-GARCH here. The results are displayed in Tables 1–4.

Table 1. The mean and SD of the estimate errors, $N = 200$, $T = 500$.

Method	p	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _\Sigma$	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\max}$	$\ \hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\ $
FG	$p = 50$	0.1977(0.1526)	0.2084(0.1615)	0.8723(0.2370)
O-GARCH		0.9980(0.0007)	0.7145(0.1467)	-
FG	$p = 250$	0.1964(0.0967)	0.3073(0.1378)	1.1586(0.2554)
O-GARCH		0.9995(0.0001)	0.8907(0.1597)	-
FG	$p = 375$	0.1937(0.0782)	0.3419(0.1576)	1.2474(0.2593)
O-GARCH		0.9996(0.0001)	0.9461(0.1876)	-
FG	$p = 500$	0.2087(0.1025)	0.3633(0.1535)	1.2845(0.2567)
O-GARCH		0.9996(0.0001)	0.9721(0.1752)	-
FG	$p = 600$	0.2111(0.0987)	0.3887(0.1619)	1.3403(0.2446)
O-GARCH		0.9997(0.0001)	0.9913(0.1522)	-

Table 2. The mean and SD of the estimate errors, $N = 200, T = 1000$.

Method	p	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\Sigma}$	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\max}$	$\ \hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\ $
FG O-GARCH	$p = 50$	0.1511(0.1390) 0.9981(0.0008)	0.1733(0.1464) 0.7257(0.1479)	0.6287(0.1470) -
FG O-GARCH	$p = 250$	0.1323(0.0626) 0.9996(0.0001)	0.2259(0.1134) 0.8998(0.1609)	0.8200(0.1678) -
FG O-GARCH	$p = 500$	0.1365(0.0634) 0.9997(0.0001)	0.2694(0.1328) 0.9636(0.1608)	0.9267(0.1874) -
FG O-GARCH	$p = 750$	0.1355(0.0519) 0.9998(0.0001)	0.2836(0.1200) 1.0653(0.1974)	0.9729(0.2039) -
FG O-GARCH	$p = 1000$	0.1404(0.0581) 0.9998(0.0001)	0.3168(0.1495) 1.0609(0.1760)	1.0215(0.2101) -
FG O-GARCH	$p = 1200$	0.1446(0.0485) 0.9998(0.0001)	0.3203(0.1546) 1.0948(0.1924)	1.0313(0.1871) -

Table 3. The mean and SD of the estimate errors, $N = 500, T = 500$.

Method	p	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\Sigma}$	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\max}$	$\ \hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\ $
FG O-GARCH	$p = 50$	0.1979(0.1372) 0.9981(0.0007)	0.2104(0.1444) 0.7216(0.1462)	0.8543(0.2039) -
FG O-GARCH	$p = 250$	0.1863(0.0793) 0.9995(0.0002)	0.3061(0.1697) 0.9015(0.1850)	1.1623(0.2310) -
FG O-GARCH	$p = 375$	0.1973(0.0864) 0.9996(0.0001)	0.3407(0.1683) 0.9495(0.1824)	1.2277(0.2378) -
FG O-GARCH	$p = 500$	0.2010(0.0802) 0.9997(0.0001)	0.3604(0.1628) 0.9764(0.1681)	1.2807(0.2379) -
FG O-GARCH	$p = 600$	0.2122(0.0984) 0.9997(0.0001)	0.3972(0.1784) 1.0084(0.1908)	1.3286(0.2471) -

Table 4. The mean and SD of the estimate errors, $N = 500, T = 1000$.

Method	p	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\Sigma}$	$\ \hat{\Sigma}_y(t) - \Sigma_y(t)\ _{\max}$	$\ \hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\ $
FG O-GARCH	$p = 50$	0.1454(0.1279) 0.9982(0.0007)	0.1485(0.1097) 0.7151(0.1466)	0.6166(0.1652) -
FG O-GARCH	$p = 250$	0.1341(0.0659) 0.9996(0.0001)	0.2247(0.1149) 0.8835(0.1582)	0.8219(0.1789) -
FG O-GARCH	$p = 500$	0.1380(0.0592) 0.9997(0.0001)	0.2659(0.1227) 0.9779(0.1788)	0.9202(0.1940) -
FG O-GARCH	$p = 750$	0.1360(0.0494) 0.9998(0.0001)	0.2812(0.1141) 1.0466(0.1999)	0.9669(0.1824) -
FG O-GARCH	$p = 1000$	0.1403(0.0498) 0.9998(0.0001)	0.3109(0.1409) 1.0665(0.1962)	1.0116(0.1980) -
FG O-GARCH	$p = 1200$	0.1440(0.0561) 0.9998(0.0001)	0.3186(0.1342) 1.1031(0.2207)	1.0348(0.1827) -

We also use the Sharpe ratio

$$SR(\hat{w}) = \frac{E\{R(\hat{w})\}}{sd\{R(\hat{w})\}},$$

to assess how well $\hat{w}$ performs, where the standard deviation of $R(\hat{w})$ is denoted as $\text{sd}\{R(\hat{w})\}$. For convenience, let the risk-free rate to be zero. To depict the performance of the proposed portfolio allocation when p is large or even larger than T , we calculate the return presented in (13) with a target return $\pi = 0.01$. Furthermore, we compare our model with [29], in terms of portfolio allocations based on Markowitz's formula. The results are presented in Table 5, with replication $N = 200$. In this table, we denote our portfolio allocations by $\hat{w}$, and the portfolio allocations of [29] by $\hat{w}_F$.

Table 5. The means, standard deviations, Sharpe ratios, and risks.

	$T = 500$ $p = 250$	$T = 500$ $p = 450$	$T = 500$ $p = 600$	$T = 1000$ $p = 250$	$T = 1000$ $p = 450$	$T = 1000$ $p = 1200$
$E\{R(\hat{w})\}$	1.61%	1.21%	1.41%	1.81%	3.21%	1.13%
$E\{R(\hat{w}_F)\}$	−0.21%	1.01%	0.82%	1.75%	0.56%	1.33%
$\text{sd}\{R(\hat{w})\}$	0.1462	0.1134	0.1174	0.1632	0.1749	0.1751
$\text{sd}\{R(\hat{w}_F)\}$	0.2221	0.1734	0.1987	0.2511	0.3862	0.2778
$\text{SR}(\hat{w})$	0.1100	0.1071	0.1201	0.1110	0.1835	0.0648
$\text{SR}(\hat{w}_F)$	−0.0096	0.0582	0.0413	0.0698	0.0144	0.0480
$r(\hat{w})$	0.0210	0.0138	0.0189	0.0324	0.0312	0.0339
$r(\hat{w}_F)$	0.0611	0.0403	0.0550	0.0755	0.1525	0.0774

Based on Tables 1–5, we can make the following observations and conclusions:

- (1). Based on the proposed model, both the means and standard deviations are small and their values decrease as the sample size increases. The means of $\|\hat{\Sigma}_y(t) - \Sigma_y(t)\|_\Sigma$ and $\|\hat{\Sigma}_y(t) - \Sigma_y(t)\|_{\max}$ slightly increase as p increases, and the means of $\|\hat{\Sigma}_y^{-1}(t) - \Sigma_y^{-1}(t)\|$ increase a little faster, which is determined by the definitions of norms. Such similar results are also obtained in [29] and [24]. All the standard deviations are stable and significantly small when compared to their corresponding means, especially based on the entropy norm.
- (2). Based on the O-GARCH model, the means and standard deviations of $\|\hat{\Sigma}_y(t) - \Sigma_y(t)\|_\Sigma$ are not sensitive to p and T , and the standard deviations are significantly smaller than the means.
- (3). The results in Table 5 show that the Sharpe ratio of the proposed portfolio allocation is larger than those of the portfolio allocation based on [29]. The risks of the proposed allocation are smaller than the corresponding risks based on [29].

In conclusion. First, comparing the results between the proposed model and O-GARCH model in terms of the means and deviations of $\|\hat{\Sigma}_y(t) - \Sigma_y(t)\|_\Sigma$ and $\|\hat{\Sigma}_y(t) - \Sigma_y(t)\|_{\max}$, we find that the proposed model performs fairly better than the O-GARCH model. Although O-GARCH model also describes the dynamic feature of the conditional covariance matrix, this discards the idiosyncratic components, leading to the poor simulation performance.

Secondly, Table 5 suggests that the proposed allocation outperforms those evaluated by the method of [29], where only unconditional covariance is adopted without taking into account the dynamics of the covariance matrix. Lastly, all results are in good agreement with the theoretical properties in Section 3. In addition, the proposed estimated conditional covariance matrix is invertible no matter how large p is, which is quite important in finance and econometrics. Hence, it is seen that the proposed model works quite well in this simulation.

5.2. Real Data Analysis

In this section, we demonstrate the application of the proposed model to a real data set. We employ the data of daily returns of 49 Industrial Portfolios minus the risk-free rate as the response variable y_t , and take $R_m - R_f$, SMB and HML as the observable common

factors. The data was downloaded from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 3 August 2021). The data series span from 2 January 2001 to 31 December 2010 with a total of 2515 observations. Figures 1 and 2 depict the time plots of the three factors and the first three components of y_t (Agriculture, Food Products and Candy&Soda), respectively. It is clearly seen in the plots that the peaks is displayed evidently during the periods of the 2008–2009 financial crisis.

Figure 1. Time series plots of the three factors ($R_m - R_f$, SMB and HML).

Figure 2. Time series plots of y_{1t} , y_{2t} and y_{3t} .

In this example, we denote our portfolio allocation by FG , and compare it with the allocation proposed by [29] (denoted by $Fan2008$) in the terms of annualized Sharpe ratios. The calculation of annualized Sharpe ratio follows the trading strategy similar to [23]. Every year's trading days are approximately $T = 252$. All possible portfolio allocations are assumed attainable, short selling is admitted, and there is no transaction cost.

In the trading strategy, we form a portfolio allocation $\hat{w}$ at the end of each trading day, and hold it until the end of the next trading day. Between day $t - 1$ and t , we calculate the realized return as follows

$$R_t(\hat{w}) = \hat{w}^\tau y_t,$$

with $\hat{w}$ calculated based on (x_{t-i}, y_{t-i}) , for $i = 1, \dots, n$, and $t = 1, \dots, T$. Hence, the annualized Sharpe ratio is evaluated by

$$SR(\hat{w}) = \frac{\bar{R}(\hat{w})}{sd(\hat{w})} \sqrt{T},$$

where

$$\bar{R}(\hat{w}) = \frac{1}{T} \sum_{t=1}^T [R_t(\hat{w}) - R_{ft}], \quad sd(R) = \left\{ \frac{1}{T} \sum_{t=1}^T [R_t(\hat{w}) - R_{ft} - \bar{R}(\hat{w})]^2 \right\},$$

with R_{ft} as the risk-free rate of day t . We calculate the annualized Sharpe ratios at the end of the final trading day of each year, and repeat this by using sample size $n = 100$ and $n = 500$.

Figure 3 shows the plots of the proposed annualized Sharpe ratios based on $n = 100$ and 500 respectively. It is seen that the performance of the proposed annualized Sharpe ratio becomes better as sample size n increases.

Figure 3. Annualized Sharpe ratios.

Furthermore, the performances of the two allocations in terms of Sharpe ratio are depicted in Figure 4. It is found that *FG* outperforms *Fan2008* in the majority of the observations. This results can be attributed to the fact that there exists GARCH effect for the considered data. Consequently, the dynamic conditional covariance (*FG*) is expected to show certain superiority compared to the unconditional covariance (*Fan2008*).

Figure 4. Annualized Sharpe ratios for $n = 500$.

6. Conclusions

In this article, we proposed a factor-GARCH model for ultra-high dimensional conditional covariance. One advantage of the proposed model is that, when $p \geq T$ and both diverge, it can capture the dynamics of the conditional covariance matrix well and also guarantees that the estimated matrix is invertible. This plays an important role in practical applications. First, we use the popular factor model constructed by [34] to reduce the dimensions. Then, CCC-GARCH is imposed to fit the common factor vector.

Last, we use the univariate GARCH model to estimate the idiosyncratic errors of y_t . The loading factor matrix is estimated by the least square method, and QMLE is proposed to estimate the parameters in GARCH models. Then, we give the estimate of $\Sigma_y(t)$ by using the plug-in method. Asymptotic properties are derived for the estimation, and simulation studies are given to demonstrate its performance. A given financial application shows that the proposed model is of certain superiority compared to the existent models.

On the basis of this paper, several questions are worthy of further study. In the article, we investigated the case where the common factors are observable and K is known, and we assumed that idiosyncratic errors have univariate GARCH structure. However, in some occasions, the common factors may not be observable, the number of factors could be unknown, and the idiosyncratic errors might follow a more complicated structure. Hence, it makes sense to extend the proposed model to more general cases, and these are left for future studies.

Author Contributions: Methology, X.Z.; Simulation, X.L.; Empirical Analysis, X.L.; Supervision, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The work was partially supported by National Natural Science Foundation of China (11731015, 11701116), Innovative Team Project of Ordinary Universities in Guangdong Province (2020WCXTD018) (Y.L. and X.Z.) and Guangzhou University Research Fund (YG2020029, YH202108) (X.Z.).

Data Availability Statement: The data was downloaded at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 3 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Lemma A1. Suppose $(\log p)^2 = o(T)$. Under Assumptions 1–3, there exist a constant $C_5 > 0$ and a small constant ϵ defined in Lemma 1, such that

$$P\left\{\sup_{t \leq T, i \leq p} |\hat{u}_{it} - u_{it}| > C_5 \sqrt{\frac{\log p}{T}}\right\} \leq O\left(\frac{1}{T^{\epsilon+1}}\right).$$

Proof. By using Lemma 3.1(i) in [29], we have

$$P\left\{\sup_{i \leq p} \|\hat{b}_i - b_i\| > C' \sqrt{\frac{\log p}{T}}\right\} = O\left(\frac{1}{p^2} + \frac{1}{T^2}\right).$$

It is easy to derive that

$$\sup_{t \leq T, i \leq p} |\hat{u}_{it} - u_{it}| = \sup_{t \leq T, i \leq p} |(\hat{b}_i - b_i)^\tau x_t| \leq \sup_{i \leq p} \|\hat{b}_i - b_i\| \sup_{t \leq T} \|x_t\|.$$

By Assumption 2(1), this lemma follows. $\square$

Lemma A2. Under Assumptions 1 and 4, there exists constants v_2 and c , such that for all $t \in \mathcal{Z}^+$

$$\alpha(t) \leq \exp(-ct^{v_2}).$$

Proof. Under Assumption 4, $\{x_{kt}\}_{t=1}^T$ are strictly stationary and α -mixing with geometric rate (see [27]). By Assumption 1, x_t and u_t are independent, and u_t are α -mixing with geometric rate; therefore, there exist positive constant ν_2 and c , such that for all $t \in \mathcal{Z}^+$

$$\alpha(t) \leq \exp(-ct^{\nu_2}).$$

□

Proof of Theorem 1. Let the negative quasi-likelihood function

$$\tilde{Q}_{\ell T}(\theta_\ell) = \frac{1}{T} \sum_{t=1}^T \tilde{v}_{\ell t}(\theta_\ell), \quad \tilde{v}_{\ell t}(\theta_\ell) = \frac{u_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2(\theta_\ell)} + \log \tilde{\sigma}_{\ell t}^2(\theta_\ell),$$

where $\tilde{\sigma}_{\ell t}^2(\theta_\ell)$ is recursively defined, for $t \geq 1$, by

$$\tilde{\sigma}_{\ell t}^2(\theta_\ell) = \alpha_{\ell 0} + \sum_{i=1}^m \alpha_{\ell i} u_{\ell t-i}^2 + \sum_{j=1}^s \beta_{\ell j} \tilde{\sigma}_{\ell t-j}^2(\theta_\ell), \quad \ell = 1, \dots, p.$$

Let the true value of θ_ℓ denoted by $\theta_{\ell 0}$ for simplicity. First, we demonstrate that $\hat{\theta}_\ell$ is consistent. The negative observed quasi log-likelihood function can be defined as its counterpart

$$Q_{\ell T}(\theta_\ell) = \frac{1}{T} \sum_{t=1}^T v_{\ell t}(\theta_\ell), \quad v_{\ell t}(\theta_\ell) = \frac{e_{\ell t}^2}{\sigma_{\ell t}^2(\theta_\ell)} + \log \sigma_{\ell t}^2(\theta_\ell),$$

where $\sigma_{\ell t}^2(\theta_\ell)$ is defined in Section 2. Following the proof of Theorem 7.1 of [25], we only need to prove

$$\sup_{1 \leq \ell \leq p} \sup_{\theta_\ell \in \Lambda} |Q_{\ell T}(\theta_\ell) - \tilde{Q}_{\ell T}(\theta_\ell)| \rightarrow_p 0, \quad \text{as } n \rightarrow \infty.$$

Denote

$$\underline{\tilde{\sigma}}_{\ell t}^2(\theta_\ell) = \begin{pmatrix} \tilde{\sigma}_{\ell t}^2(\theta_\ell) \\ \tilde{\sigma}_{\ell t-1}^2(\theta_\ell) \\ \vdots \\ \tilde{\sigma}_{\ell t-s+1}^2(\theta_\ell) \end{pmatrix}, \quad \tilde{\mathcal{C}}_{\ell t}(\theta_\ell) = \begin{pmatrix} \alpha_{\ell 0} + \sum_{j=1}^m \alpha_{\ell j} u_{\ell t-j}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad B_\ell = \begin{pmatrix} \beta_{\ell 1} & \beta_{\ell 2} & \cdots & \beta_{\ell s} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then, we have

$$\underline{\tilde{\sigma}}_{\ell t}^2 = \tilde{\mathcal{C}}_{\ell t} + B_\ell \underline{\tilde{\sigma}}_{\ell t-1}^2. \quad (\text{A1})$$

The condition of Assumption 1(ii) and the compactness of Λ imply that $\rho = \sup_{\theta_\ell \in \Lambda} \rho(B_\ell) < 1$, where $\rho(B)$ is the spectral radius of B . Iterating (A1), we thus obtain

$$\underline{\tilde{\sigma}}_{\ell t}^2 = \sum_{i=0}^{t-1} B_\ell^i \tilde{\mathcal{C}}_{\ell t-i} + B_\ell^t \underline{\tilde{\sigma}}_{\ell 0}^2.$$

Acquire its counterpart $\underline{\sigma}_{\ell t}^2(\theta_\ell)$ by replacing $\tilde{\sigma}_{\ell t-i}^2(\theta_\ell)$ by $\sigma_{\ell t-i}^2(\theta_\ell)$ in $\underline{\tilde{\sigma}}_{\ell t}^2(\theta_\ell)$, and $u_{\ell t-i}^2$ by $e_{\ell t-i}^2$ in $\tilde{\mathcal{C}}_{\ell t}$ respectively, with $e_{\ell 0}^2, \dots, e_{\ell 1-m}^2$ as the initial values, and obtain

$$\underline{\sigma}_{\ell t}^2 = \sum_{i=0}^{t-1} B_\ell^i \mathcal{C}_{\ell t-i} + B_\ell^t \underline{\sigma}_{\ell 0}^2.$$

Let $r_\ell = \sup_{t \leq T} |e_{\ell t} - u_{\ell t}|$. Then, if $t \geq m + 1$, there exists a constant $c' > 0$, such that

$$\|\tilde{c}_{\ell t} - c_{\ell t}\| \leq \left| \sum_{i=1}^m \alpha_{\ell i} (e_{\ell t-i}^2 - u_{\ell t-i}^2) \right| \leq c' \cdot r_\ell^2 + 2r_\ell \sum_{i=1}^m \alpha_{\ell i} |u_{\ell t-i}|.$$

Hence, for some constants c_1 and c_2 , we have

$$\begin{aligned} \|\tilde{\sigma}_{\ell t}^2 - \sigma_{\ell t}^2\| &\leq \left\| \sum_{i=0}^{t-(m+1)} B_\ell^i (\tilde{c}_{\ell t-i} - c_{\ell t-i}) \right\| + \left\| \sum_{i=t-m}^{t-1} B_\ell^i (\tilde{c}_{\ell t-i} - c_{\ell t-i}) \right\| + \|B_\ell^t (\tilde{\sigma}_{\ell 0}^2 - \sigma_{\ell 0}^2)\| \\ &\leq c_1 \sum_{k=0}^{t-m-1} \rho^k \left[r_\ell^2 + 2r_\ell \sum_{j=1}^m \alpha_{\ell j} |u_{\ell t-k-j}| \right] + c_1 \rho^{t-m} \frac{1 - \rho^m}{1 - \rho} + \rho^t \|\tilde{\sigma}_{\ell 0}^2 - \sigma_{\ell 0}^2\| \\ &\leq c_2 (r_\ell^2 + r_\ell + \rho^t), \end{aligned}$$

for $x > 0$, there exists $\log x \leq x - 1$. It follows that, for $x, y > 0$, $\left| \log \frac{x}{y} \right| \leq \frac{|x-y|}{\min(x,y)}$. Then, for some constants $c'_1, c'_2, c'_3, C'_1, C'_2$ and C'_3 , we obtain

$$\begin{aligned} \sup_{\theta_\ell \in \Lambda} |Q_{\ell T}(\theta_\ell) - \tilde{Q}_{\ell T}(\theta_\ell)| &\leq \frac{1}{T} \sum_{t=1}^T \sup_{\theta_\ell \in \Lambda} \left\{ \left| \frac{r_{\ell t}^2}{\sigma_{\ell t}^2} - \frac{u_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right| + \left| \log \frac{\sigma_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right| \right\} \\ &\leq \frac{1}{T} \sum_{t=1}^T \sup_{\theta_\ell \in \Lambda} \left\{ \left| \frac{\tilde{\sigma}_{\ell t}^2 (r_{\ell t}^2 - u_{\ell t}^2) + (\tilde{\sigma}_{\ell t}^2 - \sigma_{\ell t}^2) u_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2 \sigma_{\ell t}^2} \right| + \left| \log \frac{\sigma_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right| \right\} \\ &\leq \frac{1}{T} \sum_{t=1}^T \left\{ c'_1 \cdot \frac{1}{\alpha_L} r_\ell + c'_2 \cdot \frac{1}{\alpha_L^2} (r_\ell^2 + r_\ell + \rho^t) u_{\ell t}^2 + c'_3 \cdot \frac{1}{\alpha_L} \{ r_\ell^2 + r_\ell + \rho^t \} \right\} \\ &\leq C'_1 \cdot (r_\ell^2 + r_\ell) + C'_2 \cdot \frac{1}{T} \sum_{t=1}^T \rho^t u_{\ell t}^2 + C'_3 \cdot \frac{1}{T} \sum_{t=1}^T \rho^t, \end{aligned}$$

where $\alpha_L = \inf_{\theta_\ell \in \Lambda} |\alpha_{\ell 0}|$. By Lemma A.1(i), $r_\ell = O_p(\sqrt{\frac{\log p}{T}})$, and $\sup_{\ell \leq p} E u_{\ell t}^{2\kappa} < \infty$ imply that $\rho^t u_{\ell t}^2 \rightarrow_p 0$. Then, $\sup_{1 \leq \ell \leq p} \sup_{\theta_\ell \in \Lambda} |Q_{\ell T}(\theta_\ell) - \tilde{Q}_{\ell T}(\theta_\ell)| \rightarrow_p 0$. Hence, by the proof of Theorem 7.1 of [25], we obtain

$$\hat{\theta}_\ell \rightarrow_p \theta_\ell, \quad T \rightarrow \infty.$$

Following the proof of Theorem 1(IV) in [23], we acquire the convergence rate of $\sup_{1 \leq \ell \leq p} \|\hat{\theta}_\ell - \theta_\ell\|$. Hence Theorem 1 is proved. $\square$

Lemma A3. Under Assumptions 1–4, There exist a constant $C_6 > 0$ and a small constant ϵ defined in Lemma 1, such that

$$P \left\{ \sup_{1 \leq \ell \leq p} \left| \frac{\hat{\sigma}_{\ell t}^2 - \tilde{\sigma}_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right| > C_6 \sqrt{\frac{\log p}{T}} \right\} \leq O \left(\frac{1}{T^{1+\epsilon}} \right).$$

Proof. Denote $B(i, j)$ as the (i, j) th element of the matrix B and $c(i)$ as the i th entry of a vector c . The expression of the conditional covariance $\hat{\sigma}_{\ell t}^2$ is given as

$$\hat{\sigma}_{\ell t}^2 = \sum_{j=0}^{t-1} \hat{B}_\ell^j(1, 1) \hat{c}_{\ell t-j}(1) + \sum_{i=1}^s \hat{B}_\ell^t(1, i) \hat{\sigma}_{\ell 0}^2(i).$$

To replace $\hat{\beta}_{\ell j}$ by $\beta_{\ell j}$ in B_ℓ , we acquire the matrix $\hat{B}_\ell$. $\hat{c}_{\ell t}$ and $\hat{\sigma}_{\ell 0}^2$ can be defined in the same way. The expression of true conditional variance can be given in the similar way as the counterpart

$$\tilde{\sigma}_{\ell t}^2 = \sum_{j=0}^{t-1} B_{\ell}^j(1,1) \tilde{\mathcal{C}}_{\ell t-j}(1) + \sum_{i=1}^s B_{\ell}^t(1,i) \tilde{\sigma}_{\ell 0}^2(i).$$

Hence,

$$\begin{aligned} \hat{\sigma}_{\ell t}^2 - \tilde{\sigma}_{\ell t}^2 &= \sum_{j=0}^{t-1} \hat{B}_{\ell}^j(1,1) [\hat{\mathcal{C}}_{\ell t-j}(1) - \tilde{\mathcal{C}}_{\ell t-j}(1)] + \sum_{j=0}^{t-1} [\hat{B}_{\ell}^j(1,1) - B_{\ell}^j(1,1)] \tilde{\mathcal{C}}_{\ell t-j}(1) \\ &\quad + \sum_{i=1}^s \hat{B}_{\ell}^t(1,i) [\hat{\sigma}_{\ell 0}^2(i) - \tilde{\sigma}_{\ell 0}^2(i)] + \sum_{i=1}^s [\hat{B}_{\ell}^t(1,i) - B_{\ell}^t(1,i)] \tilde{\sigma}_{\ell 0}^2(i) \\ &\equiv U_{\ell 1} + U_{\ell 2} + U_{\ell 3} + U_{\ell 4}. \end{aligned}$$

(a) First, we take into account the term $U_{\ell 1}$ and observe that

$$\begin{aligned} \|\hat{\mathcal{C}}_{\ell t} - \tilde{\mathcal{C}}_{\ell t}\| &\leq |\hat{\alpha}_{\ell 0} - \alpha_{\ell 0}| + \sum_{j=1}^m |\hat{\alpha}_{\ell j} e_{\ell t-j}^2 - \alpha_{\ell j} u_{\ell t-j}^2| \\ &= |\hat{\alpha}_{\ell 0} - \alpha_{\ell 0}| + \sum_{j=1}^m |\hat{\alpha}_{\ell j} e_{\ell t-j}^2 - \hat{\alpha}_{\ell j} u_{\ell t-j}^2 + \hat{\alpha}_{\ell j} u_{\ell t-j}^2 - \alpha_{\ell j} u_{\ell t-j}^2| \\ &\leq |\hat{\alpha}_{\ell 0} - \alpha_{\ell 0}| + \sum_{j=1}^m \hat{\alpha}_{\ell j} |e_{\ell t-j}^2 - u_{\ell t-j}^2| + \sum_{j=1}^m |\hat{\alpha}_{\ell j} - \alpha_{\ell j}| u_{\ell t-j}^2, \end{aligned}$$

under Assumption 1(i) and Theorem 1, there exists a constant $c_1 > 0$, such that

$$\|\hat{\mathcal{C}}_{\ell t} - \tilde{\mathcal{C}}_{\ell t}\| \leq c_1 \left\{ \max_{0 \leq i \leq m} |\hat{\alpha}_{\ell i} - \alpha_{\ell i}| + r_{\ell}^2 + r_{\ell} \sum_{i=1}^m |u_{\ell t-i}| \right\}.$$

$\sum_{j=1}^m \sum_{i=0}^{t-1} \rho^j |u_{\ell t-i-j}|$ is bounded and $\tilde{\sigma}_{\ell t}^2 \geq \alpha_{\ell 0}$, hence for a constant $c_2 > 0$, we acquire

$$\begin{aligned} \left| \frac{U_{\ell 1}}{\tilde{\sigma}_{\ell t}^2} \right| &\leq c_2 \left\{ \max_{0 \leq i \leq m} |\hat{\alpha}_{\ell i} - \alpha_{\ell i}| + r_{\ell}^2 + r_{\ell} \sum_{i=1}^m \sum_{j=0}^{t-1} \rho^j |u_{\ell t-i-j}| \right\} \\ &\leq c_2 \left\{ \max_{0 \leq i \leq m} |\hat{\alpha}_{\ell i} - \alpha_{\ell i}| + r_{\ell} \right\}. \end{aligned}$$

As a result, there is a large constant $C'_5 > 0$, such that

$$P \left\{ \sup_{1 \leq \ell \leq p} \left| \frac{U_{\ell 1}}{\tilde{\sigma}_{\ell t}^2} \right| > C'_5 \sqrt{\frac{\log p}{T}} \right\} = O \left(\frac{1}{T^{1+\epsilon}} \right).$$

(b) For $U_{\ell 2}$, let $\hat{\delta}_{\ell} = \sup_{1 \leq i \leq s} |\hat{\beta}_{\ell i} - \beta_{\ell i}| / \beta_{\ell i}$. Following the definition of $\hat{B}_{\ell}$ and B_{ℓ} , we have

$$\left| \frac{\hat{B}_{\ell}^j(1,1) - B_{\ell}^j(1,1)}{B_{\ell}^j(1,1)} \right| \leq \max\{|(1 - \hat{\delta}_{\ell})^j - 1|, |(1 + \hat{\delta}_{\ell})^j - 1|\} \leq 2\hat{\delta}_{\ell}^j (1 + \hat{\delta}_{\ell})^{j-1},$$

for small value $\hat{\delta}_{\ell}$. $\tilde{\sigma}_{\ell t}^2 \geq \alpha_{\ell 0} + B_{\ell}^j(1,1) \mathcal{C}_{\ell t-j}(1)$ with the inequation $x/(1+x) \leq x^{\vartheta}$ for $x \geq 0$ and $\vartheta \in (0,1)$. We acquire

$$\begin{aligned} \left| \frac{U_{\ell 2}}{\tilde{\sigma}_{\ell t}^2} \right| &\leq \sum_{j=0}^{t-1} \left| \frac{(\hat{B}_{\ell}^j - B_{\ell}^j)(1, 1)}{B_{\ell}^j(1, 1)} \right| \left| \frac{B_{\ell}^j(1, 1) \tilde{c}_{\ell t-j}(1)}{\alpha_{\ell 0} + B_{\ell}^j(1, 1) \tilde{c}_{\ell t-j}(1)} \right| \\ &\leq 2\hat{\delta}_{\ell} \sum_{j=0}^{t-1} j(1 + \hat{\delta}_{\ell})^{j-1} \rho^{j\vartheta} \tilde{c}_{\ell t-j}^{\vartheta}(1). \end{aligned}$$

As a result, by selecting a suitable but small ϑ , and following Theorem 1, there is a large positive constant c_3 , such that

$$P \left\{ \sup_{1 \leq \ell \leq p} \left| \frac{U_{\ell 2}}{\tilde{\sigma}_{\ell t}^2} \right| > c_3 \sqrt{\frac{\log p}{T}} \right\} \leq P \left\{ \sup_{1 \leq \ell \leq p} \hat{\delta}_{\ell} > c_3 \sqrt{\frac{\log p}{T}} \right\} = O \left(\frac{1}{T^{1+\epsilon}} \right).$$

(c) Similar to (a) and (b), it is not hard to prove that $|U_{\ell 3}|/\tilde{\sigma}_{\ell t}^2$ and $|U_{\ell 4}|/\tilde{\sigma}_{\ell t}^2$ are bounded. Hence, Lemma A3 follows. $\square$

Proof of Theorem 2. (i) First, by Lemma A3, we have

$$\begin{aligned} \left\| \hat{\Sigma}_u(t) - \Sigma_u(t) \right\|_{\Sigma}^2 &\leq p^{-1} \left\| \Sigma_y^{-1/2}(t) \left(\hat{\Sigma}_u(t) - \Sigma_u(t) \right) \Sigma_y^{-1/2}(t) \right\|_F^2 \\ &\leq p^{-1} \cdot p^{-1} \left\| \Sigma_y^{-1/2}(t) \left(\hat{\Sigma}_u(t) - \Sigma_u(t) \right) \Sigma_y^{-1/2}(t) \right\|^2 \\ &\leq \sup_{1 \leq \ell \leq p} \left| \frac{\hat{\sigma}_{\ell t}^2 - \tilde{\sigma}_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right|^2 \cdot O(1). \end{aligned}$$

So

$$P \left\{ \left\| \hat{\Sigma}_u(t) - \Sigma_u(t) \right\|_{\Sigma}^2 > \frac{c_4 \log p}{T} \right\} = O \left(\frac{1}{T^{1+\epsilon}} \right).$$

Define

$$\begin{aligned} C_T &= \hat{B} - B, \\ G_T &= \hat{\Sigma}_x(t) - \Sigma_x(t). \end{aligned}$$

It is easy to obtain

$$\begin{aligned} \left\| \hat{\Sigma}_y(t) - \Sigma_y(t) \right\|_{\Sigma}^2 &\leq \left\| \left(\hat{B} \hat{\Sigma}_x(t) \hat{B}^{\tau} - B \Sigma_x(t) B^{\tau} \right) + \left(\hat{\Sigma}_u(t) - \Sigma_u(t) \right) \right\|_{\Sigma}^2 \\ &\leq 8 \left\| C_T \hat{\Sigma}_x(t) C_T^{\tau} \right\|_{\Sigma}^2 + 8 \left\| B G_T B^{\tau} \right\|_{\Sigma}^2 + 16 \left\| B \hat{\Sigma}_x(t) C_T^{\tau} \right\|_{\Sigma}^2 + 2 \left\| \hat{\Sigma}_u(t) - \Sigma_u(t) \right\|_{\Sigma}^2. \end{aligned}$$

Similar to the proof of Lemma B.3 in [24], we have

$$P \left\{ \left\| B G_T B^{\tau} \right\|_{\Sigma}^2 + \left\| B \hat{\Sigma}_x(t) C_T^{\tau} \right\|_{\Sigma}^2 \geq \frac{C_3 \log p}{T} + \frac{C_3 \log T}{T p} \right\} = O \left(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}} \right),$$

and

$$P \left\{ \left\| C_T \hat{\Sigma}_x(t) C_T^{\tau} \right\|_{\Sigma}^2 \geq \frac{C_3 p (\log p)^2}{T^2} \right\} = O \left(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}} \right).$$

By Lemma A3, Theorem 2 (i) follows.

(ii) By Lemma A3, we have

$$\left\| \hat{\Sigma}_u(t) - \Sigma_u(t) \right\|_{\max} \leq \sup_{1 \leq \ell \leq p} \left| \frac{\hat{\sigma}_{\ell t}^2 - \tilde{\sigma}_{\ell t}^2}{\tilde{\sigma}_{\ell t}^2} \right|,$$

then

$$P\left\{\left\|\widehat{\Sigma}_u(t) - \Sigma_u(t)\right\|_{\max}^2 \geq \frac{c_5 \log p}{T}\right\} = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

In addition, we have

$$\begin{aligned}\left\|\widehat{\Sigma}_y(t) - \Sigma_y(t)\right\|_{\max} &= \left\|\left(\widehat{B}\widehat{\Sigma}_x(t)\widehat{B}^\tau - B\Sigma_x(t)B^\tau\right) + \left(\widehat{\Sigma}_u(t) - \Sigma_u(t)\right)\right\|_{\max} \\ &\leq \left\|\widehat{B}\widehat{\Sigma}_x(t)\widehat{B}^\tau - B\Sigma_x(t)B^\tau\right\|_{\max} + \left\|\widehat{\Sigma}_u(t) - \Sigma_u(t)\right\|_{\max} \\ &\leq 2\left\|C_T\Sigma_x(t)B^\tau\right\|_{\max} + \left\|BG_TB^\tau\right\|_{\max} + \left\|C_T\Sigma_x(t)C_T^\tau\right\|_{\max} \\ &\quad + 2\left\|BG_TC_T^\tau\right\|_{\max} + \left\|C_TG_TC_T^\tau\right\|_{\max} + \left\|\widehat{\Sigma}_u(t) - \Sigma_u(t)\right\|_{\max}.\end{aligned}$$

Similar to the proof of part (i) of Theorem 2 in [28], we obtain Theorem 2 (ii). $\square$

References

1. Bollerslev, T.; Engle, R.F.; Wooldridge, J.M. A Capital Asset Pricing Model With Time-varying Covariances. *J. Politic. Econ.* **1988**, *96*, 116–131. [\[CrossRef\]](#)
2. Engle, R.; Kroner, K.F. Multivariate Simultaneous Generalized ARCH. *Econom. Theor.* **1995**, *11*, 122–150. [\[CrossRef\]](#)
3. Bollerslev, T. Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Rev. Econ. Statist.* **1990**, *72*, 498–505. [\[CrossRef\]](#)
4. Engle, R. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *J. Bus. Econ. Statist.* **2002**, *20*, 339–350. [\[CrossRef\]](#)
5. Kroner, F.K.; Ng, V.K. Modelling asymmetric comovements of asset returns. *Rev. Financ. Stud.* **1998**, *11*, 817–844. [\[CrossRef\]](#)
6. Hansson, B.; Hordahl, P. Testing the conditional CAPM using multivariate GARCH-M. *Appl. Financ. Econm.* **1998**, *8*, 377–388. [\[CrossRef\]](#)
7. Bauwens, I.; Laurent, S.; Rombouts, J.V.K. Multivariate GARCH model: A survey. *J. Appl. Econ.* **2006**, *21*, 79–109. [\[CrossRef\]](#)
8. Hafner, C.; Herwartz, H. Time-varying market price of risk in the CAPM. Approaches, empirical evidence and implications. *Finance* **1998**, *9*, 93–112.
9. Francq, C.; Zakoian, J.-M. Estimating multivariate GARCH models equation by equation. *J. R. Stat. Soc.-Ser. B* **2016**, *78*, 613–635. [\[CrossRef\]](#)
10. Gao, Z.; Tsay, R. A structural-factor approach to modeling high-dimensional time series and space-time data. *J. Time Ser. Anal.* **2019**, *40*, 343–362. [\[CrossRef\]](#)
11. Gao, Z.; Tsay, R. Modeling high-dimensional time series: A factor model with dynamically dependent factors and diverging eigenvalues. *J. Am. Statist. Assoc.* **2020**, 1–44. [\[CrossRef\]](#)
12. Lam, C.; Yao, Q.; Bathia, N. Estimation for latent factor models for high-dimensional time series. *Biometrika* **2011**, *98*, 901–918. [\[CrossRef\]](#)
13. Lam, C.; Yao, Q. Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **2012**, *40*, 694–726. [\[CrossRef\]](#)
14. Pan, J.; Yao, Q. Modelling multiple time series via common factors. *Biometrika* **2008**, *95*, 356–379. [\[CrossRef\]](#)
15. Engle, R.; Ng, V.; Rothschild, M. Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bill. *J. Econometr.* **1990**, *45*, 213–238. [\[CrossRef\]](#)
16. Bollerslev, T.; Engle, R. Common persistence in conditional variances. *Econometrica* **1993**, *61*, 167–186. [\[CrossRef\]](#)
17. Lin, W. Alternative estimators for factor GARCH model—A Monte Carlo comparison. *J. Appl. Econ.* **1992**, *7*, 259–279. [\[CrossRef\]](#)
18. Vrontos, I.; Dellaportas, P.; Politis, D. A full-factor multivariate GARCH model. *Econometr. J.* **2003**, *6*, 311–333. [\[CrossRef\]](#)
19. Alexander, C. Orthogonal GARCH. In *Mastering Risk*; Financial Times-Prentice-Hall: London, UK, 2001; Volume 2, pp. 21–38.
20. van der Weide, R. GO-GARCH: A multivariate generalized orthogonal GARCH model. *J. Appl. Econ.* **2002**, *17*, 549–564. [\[CrossRef\]](#)
21. Hafner, C.; Preminger, A. Asymptotic theory for a factor GARCH model. *Econom. Theory* **2009**, *25*, 336–363. [\[CrossRef\]](#)
22. Fan, J.; Wang, M.; Yao, Q. Modelling multivariate volatilities via conditionally uncorrelated components. *J. R. Statist. Soc. B* **2008**, *70*, 679–702. [\[CrossRef\]](#)
23. Guo, S.; Box, J.L.; Zhang, W. A Dynamic Structure for High-Dimensional Covariance Matrices and Its Application in Portfolio Allocation. *J. Am. Statist. Assoc.* **2017**, *112*, 235–253. [\[CrossRef\]](#)
24. Fan, J.; Mincheva, L.M. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **2011**, *39*, 3320–3356. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Francq, C.; Zakoian, J.-M. *GARCH Modes: Structure, Statistical Inference and Financial Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2019.

26. Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc.-Ser. B* **2013**, *75*, 603–680. [[CrossRef](#)]
27. Lindner, A.M. Stationarity, mixing, distributional properties and moments of GARCH(p, q) processes. In *Handbook of Financial Time Series*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 43–69.
28. Li, X.; Li, Y.; Pan, J.; Zhang, X. A Factor-GARCH Model for High Dimensional Volatilities. Guangzhou University, Guangzhou, China. Unpublished manuscript.
29. Fan, J.; Fan, Y.; Lv, J. High dimensional covariance matrix estimation using a factor model. *J. Econometr.* **2008**, *147*, 186–197. [[CrossRef](#)]
30. James, W.; Stein, C. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*; University of California Press: Berkeley, CA, USA, 1961; pp. 361–379.
31. Tsay, R.S. *Multivariate Time Series Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
32. Markowitz, H.M. Portfolio selection. *J. Financ.* **1952**, *7*, 77–91.
33. Markowitz, H.M. *Portfolio Selection: Efficient Diversification of Investments*; John Wiley & Sons: Hoboken, NJ, USA, 1959.
34. Fama, E.; French, K. The cross-section of expected stock returns. *J. Financ.* **1992**, *47*, 427–465. [[CrossRef](#)]