Incorporating Financial Big Data in Small Portfolio Risk Analysis: Market Risk Management Approach

Donggyu Kim* and Seunghyeon Yu

College of Business, Korea Advanced Institute of Science and Technology, Seoul, Korea

July 4, 2019

Abstract

When applying Value at Risk (VaR) procedures to specific positions or portfolios, we often focus on developing procedures only for the specific assets in the portfolio. However, since this small portfolio risk analysis ignores information from assets outside the target portfolio, there may be significant information loss. In this paper, we develop a dynamic process to incorporate the ignored information. We also study how to overcome the curse of dimensionality and discuss where and when benefits occur from a large number of assets, which is called the blessing of dimensionality. We find empirical support for the proposed method.

JEL classification: C13, C14, C32, C55, C58, G32.

Key words and phrases: Value at Risk, blessing of dimensionality, curse of dimensionality, high-dimensionality, principal component analysis, multivariate GARCH, factor model.

*corresponding author. Tel: +82 2 958 3448.
E-mail addresses: donggyukim@kaist.ac.kr (D. Kim), ysh93@kaist.ac.kr (S. Yu)

1 Introduction

Risk management has become more important than ever since the 2008 financial crisis. Several risk measurement methods, such as Value at Risk (VaR), expected shortfall (ES), entropic Value at Risk (EVaR), and superhedging price, have been developed (Acerbi and Tasche, 2002; Ahmadi-Javid, 2012; Bensaid et al., 1992; Jorion, 2000). VaR, in particular, is widely used to measure and control the level of risk by measuring the amount of loss for investments during a given period such as a day under normal market conditions (Jorion, 2000). When measuring VaR, we need to consider stylized market features such as the asymmetrical heavy-tailed log return distribution and the time-series dynamics (Christoffersen and Langlois, 2013; Cont, 2001; Longin and Solnik, 2001). To explain asymmetric and heavy-tailed log returns, researchers imposed more realistic distribution assumptions on the log returns. For example, Huisman et al. (1998) and Wu and Shieh (2007) considered student-t distribution in the VaR model; Zoia et al. (2018) employed the Gram-Charlier expansion, which is the polynomially modified Gaussian distribution; and Natarajan et al. (2008) developed asymmetry-robust VaR (ARVaR), which takes into account the asymmetric distribution of returns. On the other hand, in the stock market, we often observe dynamic structures. For example, large volatility tends to be followed by large volatility, and small volatility tends to be followed by small volatility; this is called *volatility clustering* (Mandelbrot, 1963). This stylized feature inspires the building of dynamic risk models. Examples include the dynamic copula model (Fantazzini, 2008), conditional autoregressive Value at Risk (CAViaR) (Engle and Manganelli, 2004), regime-switching VaR model (Billio and Pelizzon, 2000), and dynamic extreme value estimator (McNeil and Frey, 2000). Also, many researchers have attempted to apply dynamic volatility models to measure VaR based on the σ-based method (Brooks and Persand, 2003; Giot and Laurent, 2004; Hull and White, 1998; Jorion, 1996, 2000; Kuester et al., 2006; Sadorsky, 2005). Specifically, to account for the market dynamics, they employed dynamic volatility models such as portfolio univariate GARCH (Bollerslev, 1986), BEKK (Engle and Kroner, 1995), and constant conditional correlation (CCC) (Bollerslev, 1990).

Most of these dynamic risk models usually consider only the stock data in their target portfolio (Fan and Gu, 2003; Hendricks, 1996; Patton et al., 2019), and the portfolio size is usually relatively small. They often ignore the stock data that do not belong to the portfolio. We call this small portfolio risk analysis. However, it is doubtful whether small portfolio risk analysis is sufficient to capture portfolio risk dynamics. In fact, the factor model indicates that the individual asset risk can be decomposed into systematic risk (common risk) and idiosyncratic risk (firm-specific risk) (Bali et al., 2005; Malkiel and Xu, 1997; Shiller, 1995). Through portfolio diversification, we can nearly eliminate idiosyncratic risk, but systematic risk cannot be reduced. In light of this, portfolio risk dynamics may be primarily governed by systematic risk dynamics. Thus, it is important to model systematic risk. Systematic risk is known as common risk that affects the whole stock market, so every stock in the market may contain systematic risk information. Thus, employing financial big data can help to capture systematic risk dynamics. However, the multivariate VaR methods designed for small portfolio risk analysis are not feasible for analyzing financial big data due to the so-called *curse of dimensionality*. For instance, when the BEKK model (Engle and Kroner, 1995) is conducted on p-dimensional log return data, the number of parameters increases with the p^2 order. As the dimension p increases, the parameter estimation becomes computationally demanding due to the exploding of computation time, which is the NP hard problem. Furthermore, even if it is possible to estimate the parameters, they are not consistent estimators (Bickel and Levina, 2008; Engle et al., 2017; Pakel et al., 2017). Due to these practical and theoretical problems, even though the asset return data are easily accessible, risk managers cannot employ financial big data. This fact urges us to develop financial data analysis procedures that incorporate financial big data.

In this paper, we develop a VaR estimation procedure that incorporates financial big data into portfolio risk analysis and study how to overcome the curse of dimensionality. Specifically, we employ the approximate latent factor model (Ait-Sahalia and Xiu, 2017; Fan et al., 2013, 2019; Li et al., 2018) as the baseline model. Then the log return can be decomposed into the factor and idiosyncratic parts. For the factor part, we develop a dynamic model to account for systematic risk. For example, the conditional expected factor volatility given the current available information is a famous GARCH form and therefore is an autoregressive function of historical squared factor log returns. To evaluate the proposed dynamic factor model, we estimate the latent factor and idiosyncratic component using the principal orthogonal complement thresholding (POET) procedure (Fan et al., 2013). To do this, we further assume a sparse structure on the idiosyncratic volatility matrix, which is the immediate result of the approximate factor model. With the latent factor estimator, we propose a maximum quasi-likelihood estimation method for the GARCH parameter. The estimated GARCH parameters are employed to predict one-step ahead portfolio volatility. Finally, the one-step ahead VaR is predicted by the parametric and non-parametric σ-based approaches (Brooks and Persand, 2003; Giot and Laurent, 2004; Hull and White, 1998). We study their asymptotic properties and discuss when and where the gain is coming from incorporating financial big data, which is called the *blessing of dimensionality* (Donoho, 2000; Fan et al., 2013; Li et al., 2018). These theoretical results are examined by the simulation and empirical studies, and we compare the proposed method with various competing small portfolio risk analysis models.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the concept of Value at Risk and develop a dynamic risk model. In Section 3, we propose a maximum quasi-likelihood estimation method and study its asymptotic behavior. We also discuss how the proposed method can enjoy the blessing of dimensionality by employing financial big data. Section 4 shows how to predict the large volatility matrix and applies the proposed method to measure VaR. We also discuss how to solve the curse of dimensionality. In Section 5, we conduct Monte Carlo simulation studies to check the finite sample performance of the proposed methods, and in Section 6 we apply the proposed VaR measure to the empirical data. Finally, in Section 7 we provide concluding remarks. All technical proofs are given in the appendix.

2 A model setup

2.1 σ-based VaR method

One of the most popular risk measures is VaR, which estimates how much a portfolio of assets might lose under normal market conditions. Specifically, VaR at level α is defined as the α-percentile of a portfolio log return distribution as follows:

$$\mathrm{VaR}_{\alpha,t} = -\inf\big\{x : \mathbb{P}\{r_t \le x\} > \alpha\big\}, \tag{2.1}$$

where $\mathbb{P}$ denotes the distribution of the portfolio log returns r_t. When the portfolio log return distribution follows a location-scale family such as normal distribution and t-distribution, we can obtain VaR from the portfolio mean μ_t and volatility σ_t as follows:

$$\mathrm{VaR}_{\alpha,t} = -\mu_t - c_\alpha \sigma_t, \tag{2.2}$$

where c_α is an α-quantile value of the distributions of the standardized r_t. This method is called σ-based approach (Jorion, 2000). The performance of the σ-based approach depends on three components: the mean μ_t, volatility σ_t, and quantile c_α. The quantile c_α can be determined by the parametric and non-parametric methods. The parametric method involves the assumption of the parametric distribution for the standardized log return $(r_t - \mu_t)/\sigma_t$ such as normal distribution or student-t distribution, and c_α is determined by the α-quantile of each distributions. The non-parametric method uses a sample quantile of the historical standardized portfolio log returns $(r_t - \mu_t)/\sigma_t$. Detailed implementations are described in Section 4.2. On the other hand, several empirical studies indicate that the portfolio mean μ_t does not have strong time series patterns (Cajueiro and Tabak, 2004; Fama, 1998; Malkiel, 2003; Narayan and Smyth, 2004). Moreover, the mean μ_t gives a relatively small effect on VaR with extreme α, compared to $c_\alpha \sigma_t$. Meanwhile, in the stock market, we observe volatility time series structures such as the volatility clustering (Mandelbrot, 1963). From this point of view, it is crucial to determine the dynamic structure of the volatility σ_t, rather than μ_t. Thus, in this paper, we focus on how to incorporate the dynamic volatility structure in VaR.

2.2 Small portfolio risk analysis

When analyzing portfolio volatility, we usually investigate only the stock log return data that belong to the portfolio. For example, let $\mathbf{y}_{s,t}$ be the s-dimensional log return vector whose stocks belong to the portfolio. We also denote the co-volatility matrix of $\mathbf{y}_{s,t}$ by $\mathbf{\Sigma}_{s,t}$ and the s-dimensional portfolio weight vector by $\mathbf{w}_s$. Then the portfolio volatility can be obtained from $\mathbf{\Sigma}_{s,t}$ as follows:

$$\sigma_t = \sqrt{\mathbf{w}_s^\top \mathbf{\Sigma}_{s,t} \mathbf{w}_s}.$$

To account for the market dynamics of the portfolio return, several dynamic models have been introduced, such as the BEKK (Engle and Kroner, 1995), CCC (Bollerslev, 1990), and portfolio univariate GARCH (Bollerslev, 1986). See also Engle (2002), Kawakatsu (2006), and Van der Weide (2002). To be specific, the BEKK model with the mean vector $\boldsymbol{\mu}_{s,t}$ has the following structure:

$$\begin{aligned}
&\mathbf{y}_{s,t} = \boldsymbol{\mu}_{s,t} + \mathbf{\Sigma}_{s,t}^{\frac{1}{2}} \boldsymbol{\varepsilon}_{s,t}, \\
&\mathbf{\Sigma}_{s,t} = \mathbf{C}\mathbf{C}^\top + \mathbf{A}^\top (\mathbf{y}_{s,t-1} - \boldsymbol{\mu}_{s,t-1})(\mathbf{y}_{s,t-1} - \boldsymbol{\mu}_{s,t-1})^\top \mathbf{A} + \mathbf{B}^\top \mathbf{\Sigma}_{s,t-1} \mathbf{B}, \\
&\varepsilon_{s,it} \sim i.i.d.\ F(0,1),
\end{aligned}$$

where $\mathbf{C}$ is an $s \times s$ lower triangular matrix, $\mathbf{A}$ and $\mathbf{B}$ are $s \times s$ matrices, and $F(0,1)$ is some arbitrary distribution with mean zero and variance one. As the above BEKK model, the dynamic volatility models have the autoregressive form of the historical squared returns. Empirical study shows that the volatility is heterogeneous, and these models can explain the market dynamics (Bollerslev, 1986, 1990; Engle, 2002; Engle and Kroner, 1995; Kawakatsu, 2006; Van der Weide, 2002). Moreover, the dynamic volatility model helps to improve measuring VaR (Engle and Manganelli, 2004; Giot and Laurent, 2004; Kuester et al., 2006). That is, the performance of measuring VaR can be improved by identifying the market dynamics. In financial markets, we often observe that much of the market dynamic structure is coming from the systematic risk, which is related to the whole stock market, so the systematic risk information is contained in the whole stock. From this point of view, even if we consider only the small number of assets in the portfolio, to account for the market dynamics, we need to investigate the whole stock market and extract the systematic risk information from the stock returns. One of the naive extensions is to use the multivariate volatility model for the whole stock market such as the BEKK and CCC models. However, when applying these volatility models to incorporating a large number of assets, we face the curse of dimensionality (Bickel and Levina, 2008; Engle et al., 2008, 2017). Specifically, let p be the total number of assets, and then the number of parameters quadratically increases with p. Hence, with a large p, the estimated model is statistically inconsistent. Furthermore, in practice, the parameter optimization takes exponential computation time, which is known as the NP hard problem. Thus, it is not feasible to apply the usual multivariate volatility model to a large number of assets. In this paper, we discuss how to handle the curse of dimensionality issue and simultaneously how to account for the market dynamics.

2.3 High-dimensional dynamic volatility model with factor model

Consider a market consisting of p stocks. Let $\mathbf{y}_t$ denote the $\mathbb{R}^p$-valued random vector of the assets' log returns at time t. Then the portfolio is characterized by a vector of asset weights $\mathbf{w} \in \mathbb{R}^p$, for example, $r_t = \mathbf{w}^\top \mathbf{y}_t$. In financial analysis, we often employ the factor model (Ait-Sahalia and Xiu, 2017; Bai, 2003; Carhart, 1997; Fan et al., 2013, 2019; Fama and French, 1992, 1993, 2015) as follows:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{V}\mathbf{f}_t + \mathbf{u}_t, \tag{2.3}$$

where $\boldsymbol{\mu}$ is a constant mean vector, $\mathbf{f}_t$ is an r-dimensional factor, $\mathbf{V}$ is a $p \times r$ factor loading matrix, and $\mathbf{u}_t$ is an idiosyncratic component. We assume that the factor $\mathbf{f}_t$ and idiosyncratic component $\mathbf{u}_t$ are independent, and without loss of generality, we assume that the mean of $\mathbf{f}_t$ and $\mathbf{u}_t$ are zero. Usually the number of market factors is much smaller than the number of assets, so we assume that r is finite. Moreover, in this paper, we consider the latent factor model (Ait-Sahalia and Xiu, 2017; Fan et al., 2013, 2019; Li et al., 2018), so $\mathbf{f}_t$ and $\mathbf{u}_t$ are not observable. In this section, we propose a dynamic factor model under the latent factor structure in (2.3).

The factor model implies that the risk of assets stems from the factor and idiosyncratic components, which are called systematic and idiosyncratic risks, respectively. Idiosyncratic risk is related to the firm-specific risk, so it does not affect the whole market. Moreover, it can be mitigated by the portfolio diversification. In contrast, systematic risk arises from common market factor such as interest rate, inflation, oil price, and so on. Since systematic risk affects the whole market, we cannot mitigate this risk. From this point of view, we impose a dynamic structure on the factor to account for the market risk. For the idiosyncratic risk, we employ some martingale assumption, for example, the idiosyncratic volatility is constant over time. Then the conditional expected co-volatility matrix of $\mathbf{y}_t$ given the current available information $\mathcal{F}_t$ is expressed as follows:

$$\mathbb{E}\left[(\mathbf{y}_{t+1} - \boldsymbol{\mu})(\mathbf{y}_{t+1} - \boldsymbol{\mu})^\top \middle| \mathcal{F}_t\right] = \boldsymbol{\Sigma}_{t+1} = \mathbf{V}\boldsymbol{\Sigma}_{f,t+1}\mathbf{V}^\top + \boldsymbol{\Sigma}_u, \tag{2.4}$$

where $\boldsymbol{\Sigma}_{f,t+1} = \mathbb{E}\left[\mathbf{f}_{t+1}\mathbf{f}_{t+1}^\top \middle| \mathcal{F}_t\right]$, and $\boldsymbol{\Sigma}_u$ represents the idiosyncratic volatility matrix. Under the framework (2.4), the market dynamics can be explained by the factor volatility dynamics.

The latent factor model has an identification problem in $(\mathbf{V}, \mathbf{f}_t)$. For example, the pair $(\mathbf{V}, \mathbf{f}_t)$ is not distinguishable from $(\mathbf{V}\mathbf{H}^\top, \mathbf{H}\mathbf{f}_t)$ for any orthonormal matrix $\mathbf{H}$. To uniquely define the latent factor model, we often impose the following identifiability condition (Bai and Li, 2012; Bai and Ng, 2013; Fan et al., 2013; Kim and Fan, 2019; Li et al., 2018):

$$\mathbf{V}^\top\mathbf{V} = p\mathbf{I}_r \quad \text{and} \quad \boldsymbol{\Sigma}_{f,t} \text{ is diagonal.} \tag{2.5}$$

The identification condition (2.5) implies that the scaled factor loading matrix $p^{-1/2}\mathbf{V}$ and elements of $p\mathbf{\Sigma}_{f,t}$ are the eigenmatrix and eigenvalues of the conditional variance of the factor part. Thus, the market dynamics are explained by the dynamic structure of the eigenvalues, whereas its factor loading matrix is constant over time.

To account for the factor dynamics, we employ the famous GARCH model (Bauwens et al., 2006; Engle and Kroner, 1995; Engle, 2002, 2016; Van der Weide, 2002). The conditional expected volatility of the factor is modeled by the following GARCH structure:

$$\begin{aligned}\mathbf{f}_t &= \left(\sqrt{h_{1t}(\boldsymbol{\theta})}\epsilon_{1t}, \dots, \sqrt{h_{rt}(\boldsymbol{\theta})}\epsilon_{rt}\right), \\ \mathbf{h}_t(\boldsymbol{\theta}) &= \boldsymbol{\omega} + \mathbf{A}\mathbf{f}_{t-1}^2 + \mathbf{B}\mathbf{h}_{t-1}(\boldsymbol{\theta}), \end{aligned} \tag{2.6}$$

where ϵ_{it}'s are i.i.d. random variables with mean zero and unit variance; $\mathbf{h}_t(\boldsymbol{\theta})$ is an r-dimensional vector of the conditional expected factor volatility $\mathbf{\Sigma}_{f,t}$, that is, $\mathbf{h}_t = \big(h_{1t}(\boldsymbol{\theta}), \dots, h_{rt}(\boldsymbol{\theta})\big) = \mathrm{diag}(\mathbf{\Sigma}_{f,t})$ where $\mathrm{diag}(\mathbf{X})$ is a vector whose elements are diagonal entries of $\mathbf{X}$; $\mathbf{f}_t^2$ is an element-wise squared vector of the factor log return $\mathbf{f}_t$; and $\boldsymbol{\theta} = (\boldsymbol{\omega}, \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{B}))$ is the GARCH parameters. The GARCH model in (2.6) indicates that the conditional expected volatility of the factor is the autoregressive form of the historical squared factor log returns. Thus, we expect that this model can capture the stylized market features such as the volatility clustering and heavy-tail. The empirical study in Section 6 also supports this. For the high-frequency financial data, Kim and Fan (2019) recently introduced the factor GARCH-Itô model to account for the market dynamics and they showed that the dynamic factor model can capture the market dynamics well. Thus, the proposed model is the discrete-time version of the factor GARCH-Itô model.

The idiosyncratic component is coming from the firm-specific risk, so their risk is not strongly connected. Empirical studies show that there are some local factors that affect few other idiosyncratic components (Ait-Sahalia and Xiu, 2017; Boivin and Ng, 2006; Kalnina and Tewou, 2015). In light of these, we allow a weak relationship between idiosyncratic components so that the idiosyncratic volatility $\mathbf{\Sigma}_u = ((\Sigma_{u,ij})_{i,j=1,\dots,p})$ satisfies the following sparse condition:

$$\max_{i \le p} \sum_{j=1}^{p} |\Sigma_{u,ij}|^q (\Sigma_{u,ii}\Sigma_{u,jj})^{(1-q)/2} \le s_p, \tag{2.7}$$

where $q \in [0,1)$ and the sparsity measure s_p diverges slowly with the dimension p, for example, $\log p$. This sparsity condition is widely employed in the large convariance matrix inferences (Ait-Sahalia and Xiu, 2017; Bickel and Levina, 2008; Cai and Liu, 2011; Fan et al., 2013, 2018, 2019; Kim and Fan, 2019). When the idiosyncratic volatility satisfies the exact sparsity, that is, $q = 0$, the sparsity condition indicates that each asset has at most s_p non-zero idiosyncratic correlations with other assets.

Under the volatility structure (2.4) and (2.6), the VaR of portfolios in (2.2) can be calculated as follows:

$$\mathrm{VaR}_{\alpha,t} = -\mathbf{w}^\top \boldsymbol{\mu} - c_\alpha \sqrt{\mathbf{w}^\top (\mathbf{V}\boldsymbol{\Sigma}_{f,t}\mathbf{V}^\top + \boldsymbol{\Sigma}_u)\mathbf{w}}.$$

To evaluate the above VaR value, we need to estimate the unobserved factor components and the idiosyncratic volatility. However, to incorporate the whole market assets information, we consider a large number of assets and consequently run into the curse of dimensionality problem. In the following section, we discuss how to overcome this issue and introduce an estimation procedure to incorporate the financial big data for risk analysis.

3 Factor dynamics estimation

First we define the notations. We denote $\|\cdot\|, \|\cdot\|_F$, and $\|\cdot\|_{\max}$ by the matrix spectral norm, Frobenius norm, and max norm, respectively. We use O_p as a big-O in probability, $\lambda_k(\mathbf{A})$ as the k^{th} largest eigenvalue of the square matrix $\mathbf{A}$, and $\mathrm{vec}(\mathbf{A})$ as the vectorization of $\mathbf{A}$. We also denote the true parameters by $\boldsymbol{\theta}_0 = (\boldsymbol{\omega}_0, \mathrm{vec}(\mathbf{A}_0), \mathrm{vec}(\mathbf{B}_0))$.

3.1 Latent factor estimation

To estimate the model parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\omega}_0, \mathrm{vec}(\mathbf{A}_0), \mathrm{vec}(\mathbf{B}_0))$, we first need to estimate the latent factor components $\mathbf{f}_t$ and $\mathbf{V}$. We recall that the conditional volatility matrix is

$$\boldsymbol{\Sigma}_{t+1} = \mathbf{V}\boldsymbol{\Sigma}_{f,t+1}\mathbf{V}^\top + \boldsymbol{\Sigma}_u,$$

and the mean conditional volatility matrix is

$$\bar{\boldsymbol{\Sigma}} = \mathbf{V}\bar{\boldsymbol{\Sigma}}_f\mathbf{V}^T + \boldsymbol{\Sigma}_u,$$

where $\bar{\boldsymbol{\Sigma}}_f = T^{-1}\sum_{t=1}^{T}\boldsymbol{\Sigma}_{f,t}$. Under the identification and sparsity conditions (2.4) and (2.7), the mean conditional volatility matrix has the low-rank plus sparse structure that is widely used in the high-dimensional factor analysis (Ait-Sahalia and Xiu, 2017; Fan et al., 2013, 2018, 2019; Kim and Fan, 2019). With the low-rank plus sparse structure, Fan et al. (2013) introduced the POET estimation procedure to estimate the latent factor volatility and idiosyncratic volatility matrices. To harness the POET procedure, we need a good proxy of the mean conditional volatility matrix. We use the sample covariance matrix $\widehat{\bar{\boldsymbol{\Sigma}}} = T^{-1}\sum_{t=1}^{T}(\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})^\top$ with sample mean $\bar{\mathbf{y}} = T^{-1}\sum_{t=1}^{T}\mathbf{y}_t$, and under some mild condition, the martingale convergence theorem implies that the sample covariance matrix converges to $\bar{\boldsymbol{\Sigma}}$. Thus, we apply the POET procedure with the sample covariance matrix to estimate the latent factor components. Specifically, the eigenvalue decomposition admits

$$\widehat{\bar{\boldsymbol{\Sigma}}} = \sum_{i=1}^{p}\widehat{\lambda}_i\widehat{\mathbf{q}}_i\widehat{\mathbf{q}}_i^\top, \tag{3.1}$$

where $\widehat{\lambda}_k$ and $\widehat{\mathbf{q}}_k$ are the k^{th} largest eigenvalues and eigenvectors of the sample covariance matrix $\widehat{\bar{\boldsymbol{\Sigma}}}$, respectively. Then the factor loading matrix estimator $\widehat{\mathbf{V}}$ is $\sqrt{p}(\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_r)$ and the mean factor volatility matrix estimator $\widehat{\bar{\boldsymbol{\Sigma}}}_f$ is $p^{-1}\mathrm{Diag}\big((\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)\big)$, where $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix whose diagonal entries are $\mathbf{x}$. Note that $\widehat{\mathbf{V}}$ has a multiplier $\sqrt{p}$ so that $\widehat{\mathbf{V}}$

satisfies the condition $\mathbf{V}^\top \mathbf{V} = p\mathbf{I}_r$. Then the latent factors can be obtained by

$$\widehat{\mathbf{f}}_t = \frac{1}{p}\widehat{\mathbf{V}}^\top (\mathbf{y}_t - \bar{\mathbf{y}}).$$

To derive the convergence rate of this latent factor component estimator, we need the following technical assumptions.

Assumption 1.

(a) $\mathbb{E}\left[f_{it}^8\right], \mathbb{E}\left[h_{it}^4(\boldsymbol{\theta}_0)\right], \mathbb{E}\left[(u_{it}u_{jt} - \Sigma_{u,ij})^4\right]$*, and* $\mathbb{E}\left[\left(\mathbf{q}^\top \mathbf{u}_t / \sqrt{\mathbf{q}^\top \boldsymbol{\Sigma}_u \mathbf{q}}\right)^4\right]$ *are bounded by some constant C for all i, j and $\mathbf{q}$ s.t. $\|\mathbf{q}\| = 1$.*

(b) The minimum eigen-gap $\delta_r = \min_{k \le r} \left|\lambda_k(\overline{\boldsymbol{\Sigma}}_f) - \lambda_{k+1}(\overline{\boldsymbol{\Sigma}}_f)\right|$ *satisfies* $\delta_r \ge C$ *a.s.*

Remark 1. Assumption 1(b) is the pervasive condition that is widely used in the low-rank matrix inferences (Ait-Sahalia and Xiu, 2017; Fan et al., 2013, 2018, 2019; Kim and Fan, 2019; Stock and Watson, 2002). The pervasive condition with the sparsity condition (2.7) helps to distinguish the latent factor from the idiosyncratic volatility. Additionally, since the common market factor affects the whole asset, its proportion to the total variation is significant. Mathematically, this implies that the corresponding eigenvalues have the p order, so the pervasive condition is not restrictive.

The following theorem provides the convergence rates of the latent factor component estimators $\widehat{\mathbf{V}}$ and $\widehat{\mathbf{f}}_t$.

Theorem 3.1. *Under the model in Section 2.3, suppose Assumption 1. Then we have*

$$\min_{\mathbf{O}} \left\|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{O}\right\|^2 = O_p\left(\frac{p}{T} + \frac{s_p^2}{p}\right), \tag{3.2}$$

$$\left\|\widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2\right\| = O_p\left(\frac{1}{\sqrt{T}} + \sqrt{\frac{s_p}{p}}\right), \tag{3.3}$$

where $\mathbf{O}$ is a diagonal sign matrix which has value ± 1.

Remark 2. Eigenvectors have the sign problem; for example, $-\mathbf{v}_i$ and $\mathbf{v}_i$ are not distinguishable, where $\mathbf{v}_i$ is the i^{th} column vector of $\mathbf{V}$. So we put the sign matrix $\mathbf{O}$ in (3.2) to identify the eigenmatrix.

Remark 3. Theorem 3.1 shows that the convergence rate of $\widehat{\mathbf{f}}_t^2$ is $1/\sqrt{T} + \sqrt{s_p/p}$. The first term $1/\sqrt{T}$ is the usual optimal convergence rate for estimating the mean conditional volatility matrix $\bar{\mathbf{\Sigma}}$. The second term $\sqrt{s_p/p}$ is the cost to identify the latent factor.

When the factor is not observable and the number of assets is finite, it is impossible to estimate the latent factors. In contrast, Theorem 3.1 shows the blessing of financial big data in the latent factor estimation. Specifically, every asset contains the common market information. Thus, as the number of assets increases, the information for the latent factors also increases. Therefore, more stock sample exhibits a clearer latent signal, and we can estimate the latent factors consistently with the convergence rate $\sqrt{s_p/p}$.

3.2 Maximum quasi-likelihood estimation

In this section, we propose a model parameter estimation procedure. Specifically, we adopt the following quasi-maximum likelihood estimation (QMLE) method:

$$\min_{\boldsymbol{\theta}\in\mathbf{\Theta}} \sum_{t=1}^{T} \sum_{i=1}^{r} \left(\log h_{it}(\boldsymbol{\theta}) + h_{it}^{-1}(\boldsymbol{\theta}) f_{it}^2\right), \tag{3.4}$$

where $\mathbf{\Theta}$ is a compact parameter space. Then the well-developed asymptotic theorems for the QMLE provide the consistency (Bollerslev and Wooldridge, 1992; Comte and Lieberman, 2003; Lee and Hansen, 1994). However, the factors are not observable, so to evaluate the quasi-likelihood function, we use the non-parametric factor estimator $\widehat{\mathbf{f}}_t^2$. For example, the conditional co-volatility $\mathbf{h}_t(\boldsymbol{\theta})$ is estimated by

$$\widehat{\mathbf{h}}_t(\boldsymbol{\theta}) = \boldsymbol{\omega} + \mathbf{A}\widehat{\mathbf{f}}_{t-1}^2 + \mathbf{B}\widehat{\mathbf{h}}_{t-1}(\boldsymbol{\theta}), \tag{3.5}$$

and we use $\widehat{\mathbf{h}}_1(\boldsymbol{\theta}) = (\mathbf{I}_r - \mathbf{A} - \mathbf{B})^{-1}\boldsymbol{\omega}$ as the initial value. Note that $\widehat{\mathbf{h}}_1(\boldsymbol{\theta})$ is the unconditional volatility of the factor, and the effect of the initial value is negligible (see Lemma A.2). Then we calculate the quasi-likelihood function with the conditional co-volatility estimator $\widehat{\mathbf{h}}_t(\boldsymbol{\theta})$ and obtain the maximum quasi-likelihood estimator $\widehat{\boldsymbol{\theta}}$ as follows:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{t=1}^{T}\sum_{i=1}^{r}\Big(\log \widehat{h}_{it}(\boldsymbol{\theta}) + \widehat{h}_{it}^{-1}(\boldsymbol{\theta})\widehat{f}_{it}^{2}\Big). \tag{3.6}$$

To investigate the asymptotic properties for the QMLE $\widehat{\boldsymbol{\theta}}$, we need the following technical conditions.

Assumption 2.

(a) *The parameter space* $\boldsymbol{\Theta}$ *is a compact set such that every element* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *is positive;* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{E}[h_{it}^4(\boldsymbol{\theta})]$ *is bounded for all* $i = 1, \ldots, r$ *and* $t = 1, \ldots, T$*; the eigenvalues of* $\mathbf{B}$ *and* $\mathbf{A}$ *are positive and* $\|\mathbf{B}\| < 1$*; and* $\boldsymbol{\theta}_0$ *is the interior point.*

(b) $\mathbf{f}_t^2$*'s are non-degenerate random variables.*

The below theorem illustrates the convergence rate of the QMLE estimator $\widehat{\boldsymbol{\theta}}$.

Theorem 3.2. *Under the model in Section 2.3, suppose Assumptions 1 and 2. Then we have*

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\max} = O_p\left(\frac{1}{\sqrt{T}} + \sqrt{\frac{s_p}{p}}\right). \tag{3.7}$$

Remark 4. Theorem 3.2 shows that the QMLE estimator $\widehat{\boldsymbol{\theta}}$ has the convergence rate $1/\sqrt{T} + \sqrt{s_p/p}$. The term $\sqrt{s_p/p}$ originates from the latent factors estimation in (3.3), which is the cost to identify the latent factor. Thus, when the factors are observable, the convergence rate will be $1/\sqrt{T}$.

One of our main objectives is to predict the future volatility. With the QMLE estimator

$\widehat{\boldsymbol{\theta}}$, we can estimate the conditional volatility as follows:

$$\widehat{\boldsymbol{\Sigma}}_{f,t+1} = \mathrm{Diag}\left(\widehat{\mathbf{h}}_{t+1}(\widehat{\boldsymbol{\theta}})\right).$$

Then Theorems 3.1–3.2 immediately show the consistency of $\widehat{\boldsymbol{\Sigma}}_{f,t+1}$.

Corollary 3.1. *Under the assumptions in Theorem 3.2, we have*

$$\left\|\widehat{\boldsymbol{\Sigma}}_{f,t+1} - \boldsymbol{\Sigma}_{f,t+1}\right\|_F = O_p\left(\frac{1}{\sqrt{T}} + \sqrt{\frac{s_p}{p}}\right).$$

4 Large volatility matrix estimation and VaR forecast

4.1 One-step ahead large volatility matrix estimation

In the previous section, we find the blessing of dimensionality in the latent factor estimation. However, when it comes to estimating large volatility matrices, we still suffer from the curse of dimensionality. For example, sample volatility matrix estimators are inconsistent when both the number of assets and sample size go to infinity (Bickel and Levina, 2008; Cai and Liu, 2011; Marčenko and Pastur, 1967). To overcome the curse of dimensionality, we impose the sparse structure (2.7) on the idiosyncratic volatility matrix. As discussed in Section 2.3, in the stock market, the co-movement of stocks can be explained by the common factor, and the remaining idiosyncratic co-volatilities are weakly correlated. Thus, the sparsity condition is realistic. To estimate the sparse idiosyncratic volatility matrix, we employ the POET procedure (Fan et al., 2013) as follows. First we estimate the input idiosyncratic volatility matrix estimator by using the non-pervasive eigen-components as follows:

$$\widehat{\boldsymbol{\Sigma}}_u = \sum_{i=r+1}^{p} \widehat{\lambda}_i \widehat{\mathbf{q}}_i \widehat{\mathbf{q}}_i^\top,$$

where the eigenvalues $\widehat{\lambda}_i$'s and eigenvectors $\widehat{\mathbf{q}}_i$'s are defined in (3.1). Then we apply the thresholding method to the input idiosyncratic volatility matrix estimator as follows:

$$[\mathcal{T}(\widehat{\mathbf{\Sigma}}_u)]_{ij} = \begin{cases} \widehat{\Sigma}_{u,ii}, & \text{if } i = j \\ s_{ij}(\widehat{\Sigma}_{u,ij}) \mathbb{1}_{\left(|\widehat{\Sigma}_{u,ij}| \geq \tau_T \sqrt{\widehat{\Sigma}_{u,ii}\widehat{\Sigma}_{u,jj}}\right)}, & \text{if } i \neq j \end{cases}, \tag{4.1}$$

where $\mathbb{1}_{(\cdot)}$ is an indicator function and $s_{ij}(\cdot)$ is a shrinkage function satisfying $|s_{ij}(x) - x| \leq \tau_T \sqrt{\widehat{\Sigma}_{u,ii}\widehat{\Sigma}_{u,jj}}$. The examples of the shrink function $s_{ij}(x)$ are the soft thresholding function $s_{ij}(x) = x - sign(x)\tau_T \sqrt{\widehat{\Sigma}_{u,ii}\widehat{\Sigma}_{u,jj}}$ and the hard thresholding function $s_{ij}(x) = x$. The thresholding level τ_T will be given in Theorem 4.1. The working principle of the thresholding method is that the co-volatility is zero if the estimated correlation is weak. This makes the estimated idiosyncratic volatility matrix estimator sparse, so the estimated idiosyncratic volatility satisfies the sparse condition.

With the estimated factor volatility $\widehat{\mathbf{\Sigma}}_{f,t+1}$ and idiosyncratic volatility matrix $\mathcal{T}(\widehat{\mathbf{\Sigma}}_u)$, we estimate the conditional volatility matrix as follows:

$$\widehat{\mathbf{\Sigma}}_{t+1} = \widehat{\mathbf{V}}\widehat{\mathbf{\Sigma}}_{f,t+1}\widehat{\mathbf{V}}^\top + \mathcal{T}(\widehat{\mathbf{\Sigma}}_u). \tag{4.2}$$

We call the conditional volatility matrix estimator $\widehat{\mathbf{\Sigma}}_{t+1}$ the P-GARCH estimator. To investigate the asymptotic property of (4.2), we require the following technical conditions.

Assumption 3.

(a) The sample covariance estimator satisfies the concentration inequality, for any given $a > 0$,

$$\mathbb{P}\left\{ \max_{ij} \left| \widehat{\overline{\Sigma}}_{ij} - \overline{\Sigma}_{ij} \right| \geq C_a \sqrt{\frac{\log p}{T}} \right\} \leq p^{-a},$$

where C_a *is a constant only depending on* a.

(b) There is a constant C such that

$$\frac{1}{r}\max_{i\le p}\sum_{j=1}^{r}V_{ij}^2\le C.$$

(c) The smallest eigenvalue of $\mathbf{\Sigma}_u$ stays away from 0, and $\left|\Sigma_{u,ij}\right|\le C$ for all i, j.

(d) $p=o(T^2)$.

Remark 5. Assumption 3(a) is the sub-Gaussian condition. When investigating the high-dimensional inferences, the sub-Gaussian condition is essential. Additionally, when the log return $\mathbf{y}_t$ satisfies some sub-Gaussian property, we can obtain Assumption 3(a).

Remark 6. Assumption 3(b) is called the incoherence condition, which is widely assumed in analyzing the low-rank matrix (Candès and Recht, 2009; Fan et al., 2017). The basic intuition is that the factor loading matrix $\mathbf{V}$ is not to be sparse. That is, the factor affects almost all the stock returns. Thus, under the factor model, the incoherence condition is acceptable.

The following theorem shows the asymptotic behaviors for the P-GARCH estimator.

Theorem 4.1. *Under the model in Section 2.3, suppose that Assumptions 1–3 hold. Take the thresholding level as $\tau_T=C_\tau\left(\sqrt{\log p/T}+\sqrt{s_p/p}\right)$ for some positive constant C_τ. Then we have*

$$\begin{aligned}
\left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u)-\mathbf{\Sigma}_u\right\|_{\max}&=O_p\left(\tau_T\right),\\
\left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u)-\mathbf{\Sigma}_u\right\|&=O_p\left(s_p\tau_T^{1-q}\right),\\
\left\|\widehat{\mathbf{\Sigma}}_{t+1}-\mathbf{\Sigma}_{t+1}\right\|_{\max}&=O_p\left(\tau_T\right),\\
\left\|\widehat{\mathbf{\Sigma}}_{t+1}-\mathbf{\Sigma}_{t+1}\right\|_{\mathbf{\Sigma}_{t+1}}&=O_p\left(\frac{\sqrt{p}}{T}+s_p\tau_T^{1-q}\right),
\end{aligned}$$

where the relative Frobenius norm $\|\mathbf{G}_1\|_{\mathbf{G}_2}^2=p^{-1}\|\mathbf{G}_2^{-\frac{1}{2}}\mathbf{G}_1\mathbf{G}_2^{-\frac{1}{2}}\|_F^2$ for any given $p\times p$ matrices $\mathbf{G}_1$ and $\mathbf{G}_2$.

Remark 7. Theorem 4.1 shows that the P-GARCH is the consistent estimator as long as $p = o(T^2)$.

With the realistic low-rank plus sparse structure, we can enjoy the blessing of the dimensionality for estimating the factor volatility matrix. Moreover, using the regularization method, as shown in Theorem 4.1, we can overcome the curse of dimensionality.

4.2 One-step ahead VaR forecast from financial big data

In this section, we discuss how to measure the VaR value with the P-GARCH estimator $\widehat{\mathbf{\Sigma}}_{t+1}$ in (4.2). Using the plug-in method, we estimate the one-step ahead VaR as follows:

$$\widehat{\mathrm{VaR}}_{\alpha,t+1} = -\mathbf{w}^{\top}\bar{\mathbf{y}} - c_{\alpha}\sqrt{\mathbf{w}^{\top}\widehat{\mathbf{\Sigma}}_{t+1}\mathbf{w}}, \tag{4.3}$$

where c_{α} is an α-quantile and $\bar{\mathbf{y}}$ is a sample mean vector. To evaluate the VaR value, we need to determine the α-quantile value. To do this, we assume that the standardized portfolio log returns are i.i.d. Then when the standardized portfolio log returns follow the standard normal distribution, c_{α} is the α-quantile of the standard normal z_{α}. When the standardized portfolio log returns follow the multivariate t-distribution with the degrees of freedom ν, then $c_{\alpha} = t_{\nu,\alpha}\sqrt{(\nu-2)/\nu}$, where $t_{\nu,\alpha}$ is an α-quantile of the t-distribution with the degrees of freedom ν (Glasserman et al., 2002). We call them the parametric σ-based VaR estimator. The performance of the parametric σ-based VaR estimator depends on the distribution assumption. On the other hand, to obtain distribution robust VaR estimators, we use the non-parametric sample quantile method, which we call the non-parametric σ-based VaR estimator. For example, c_{α} is set to be $\lceil \alpha T \rceil$-th smallest value of $\{(\mathbf{w}^{\top}(\mathbf{y}_t - \bar{\mathbf{y}})/(\mathbf{w}^{\top}\widehat{\mathbf{\Sigma}}_t\mathbf{w})^{1/2}\}_{t=1}^{T}$, where $\lceil\cdot\rceil$ is a ceiling function. To derive convergence rate of the non-parametric σ-based VaR estimator, we require the following technical conditions.

Assumption 4.

(a) The estimated standardized return $\widehat{x}_t = \mathbf{w}^{\top}(\mathbf{y}_t - \bar{\mathbf{y}})/(\mathbf{w}^{\top}\widehat{\mathbf{\Sigma}}_t\mathbf{w})^{1/2}$ *satisfies the concen-*

tration inequality, for any given $a > 0$,

$$\max_{t \le T} \mathbb{P}\left\{ |\widehat{x}_t - x_t| \ge C_a \sqrt{\log T} \left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}} \right) \right\} \le T^{-a},$$

where $x_t = \mathbf{w}^\top (\mathbf{y}_t - \boldsymbol{\mu}) / (\mathbf{w}^\top \boldsymbol{\Sigma}_t \mathbf{w})^{1/2}$ and C_a is a constant only depending on a.

(b) The cumulative density function of x_t and its inverse function are continuous.

(c) The variance of the portfolio return r_t is bounded and strictly positive.

Remark 8. The sub-Gaussian condition Assumption 4(a) is true under some sub-Gaussian condition for $\mathbf{y}_t$. Assumption 4(b) is usually imposed to study quantile estimations (Chen and Tang, 2005).

Then the following theorem shows the convergence rates of VaR.

Theorem 4.2. *Under the model in Section 2.3, suppose that Assumptions 1–4 hold. Then for an arbitrary portfolio weight $\mathbf{w}$ with the gross exposure constraint $\|\mathbf{w}\|_1 \le C$, the parametric σ-based VaR estimator has*

$$\left| \widehat{VaR}_{\alpha,t+1} - VaR_{\alpha,t+1} \right| = O_p \left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}} \right).$$

Moreover, the non-parametric σ-based VaR estimator is

$$\left| \widehat{VaR}_{\alpha,t+1} - VaR_{\alpha,t+1} \right| = O_p \left(\sqrt{\log T} \left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}} \right) \right).$$

In this paper, we focus on investigating the effect of the VaR estimator with the financial big data for a small portfolio. When we do not incorporate financial big data, that is, p is finite, the absolute VaR error is not consistent for all parametric and non-parametric estimator. This is because the term $\sqrt{s_p/p}$ does not converge and so the latent factor estimation error is dominant. Therefore, Theorem 4.2 supports that incorporating financial big data leads to the blessing in the VaR forecast by capturing common factor dynamics.

5 Simulation study

In this section, we conducted Monte Carlo simulations to check the finite sample performances of the proposed P-GARCH and corresponding VaR model. The data generating process is analogous to the model (2.3) and (2.4). The number of factors was chosen to be $r = 3$; the mean of $\mathbf{y}_t$ is set to be zero $\boldsymbol{\mu} = \mathbf{0}$; and the factor loading matrix $\mathbf{V}$ was randomly sampled from the first r right singular vector of random matrix which has the elements i.i.d. Unif$(0, 1)$. The latent factors $\mathbf{f}_t$ were generated by the multivariate normal distribution with the conditional expected volatility $\boldsymbol{\Sigma}_{f,t}$ as follows:

$$\boldsymbol{\Sigma}_{f,t} = \mathrm{Diag}(\mathbf{h}_t(\boldsymbol{\theta}_0)) \quad \text{and} \quad \mathbf{h}_t(\boldsymbol{\theta}_0) = \boldsymbol{\omega}_0 + \mathbf{A}_0\mathbf{f}_{t-1}^2 + \mathbf{B}_0\mathbf{h}_{t-1}(\boldsymbol{\theta}_0),$$

where $\mathbf{h}_1(\boldsymbol{\theta}_0) = (\mathbf{I}_r - \mathbf{A}_0 - \mathbf{B}_0)^{-1}\boldsymbol{\omega}_0$,

$$\boldsymbol{\omega}_0 = \begin{pmatrix} 0.003 \\ 0.002 \\ 0.001 \end{pmatrix}, \qquad \mathbf{A}_0 = \begin{pmatrix} 0.2 & 0.3 & 0.4 \\ 0.15 & 0.12 & 0.2 \\ 0.1 & 0.1 & 0.1 \end{pmatrix}, \qquad \mathbf{B}_0 = \begin{pmatrix} 0.2 & 0.1 & 0.1 \\ 0.2 & 0.05 & 0.07 \\ 0.1 & 0 & 0.05 \end{pmatrix}.$$

The idiosyncratic risk $\mathbf{u}_t$ was generated from the multivariate normal distribution with the following sparse co-volatility

$$\Sigma_{u,ij} = 0.01 \times 0.5^{|i-j|}. \tag{5.1}$$

We varied p from 20 to 500 and T from 500 to 10,000, and employed the QMLE procedure in Section 3.2 to estimate the GARCH parameters. We repeated the whole procedure 500 times. We calculated the mean absolute error based on the 500 simulations.

Table 1 reports the mean absolute error (MAE) of the QMLE estimate $\widehat{\boldsymbol{\theta}}$ for various p and T. To save space, we only documented 9 parameters estimation results out of 21 parameters, but the rest of the parameters show similar behavior. Table 1 shows that MAEs decreases as T or p increases. This result supports the theoretical findings in Theorem 3.2.

Table 1: Mean absolute error (MAE) of the QMLE estimate $\widehat{\boldsymbol{\theta}}$ with 500 replications. To save space, only the first 3 elements of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ are reported. The other results have a similar pattern.

p	T	MAE$\times 10^2$								
		ω_1	ω_2	ω_3	A_{11}	A_{12}	A_{13}	B_{11}	B_{12}	B_{13}
20	500	0.138	0.088	0.068	5.706	9.483	12.637	14.279	18.768	21.801
	2000	0.060	0.047	0.042	2.816	4.343	6.830	11.338	13.448	15.169
	4000	0.046	0.040	0.034	2.033	3.359	5.621	9.738	10.703	14.345
	6000	0.039	0.032	0.031	1.736	2.887	5.436	8.871	9.450	12.538
	10000	0.030	0.031	0.029	1.333	2.825	4.947	8.279	8.912	12.923
100	500	0.124	0.079	0.060	5.816	8.651	12.264	13.739	16.993	19.134
	2000	0.056	0.040	0.028	2.594	4.306	5.933	10.340	12.012	13.912
	4000	0.040	0.027	0.020	1.846	2.925	3.978	9.869	10.098	12.730
	6000	0.029	0.021	0.015	1.550	2.263	3.147	8.539	8.335	11.924
	10000	0.023	0.018	0.012	1.047	1.874	2.606	7.993	7.731	10.932
500	500	0.118	0.079	0.055	5.788	8.539	11.931	14.059	16.716	18.482
	2000	0.058	0.039	0.027	2.734	4.208	5.619	10.161	11.007	13.436
	4000	0.039	0.027	0.019	1.786	2.916	3.948	8.975	9.451	12.525
	6000	0.031	0.023	0.015	1.546	2.448	3.571	8.615	8.533	11.790
	10000	0.024	0.018	0.012	1.057	1.764	2.456	7.932	7.477	10.843

With the estimated parameter $\widehat{\boldsymbol{\theta}}$, we validated the one-step ahead volatility prediction $\widehat{\boldsymbol{\Sigma}}_{t+1}$ with several matrix norms. We condcuted the thresholding procedure for the idiosyncratic volatility in (4.1) with the tuning parameters $C_\tau = 1$ and $s_p = 1$. Figure 1 draws the prediction errors of $\widehat{\boldsymbol{\Sigma}}_{t+1} - \boldsymbol{\Sigma}_{t+1}$ with the Frobenius, spectral, max, and relative Frobenius norms. For the Frobenius and spectral norms, the errors are large at $p = 500$. This is because the Frobenius and spectral norms depend on the dimensionality p. However, for the relative Frobenius norm, the error decreases as the dimension p increases. From this result, we can find the blessing of the dimensionality. The max norm does not show any significant difference for different p. This may be because the large volatility matrix has a greater chance to have a large max norm error, even though the element-wise error decreases. In all cases, the error decreases as the sample size T increases. These results support the

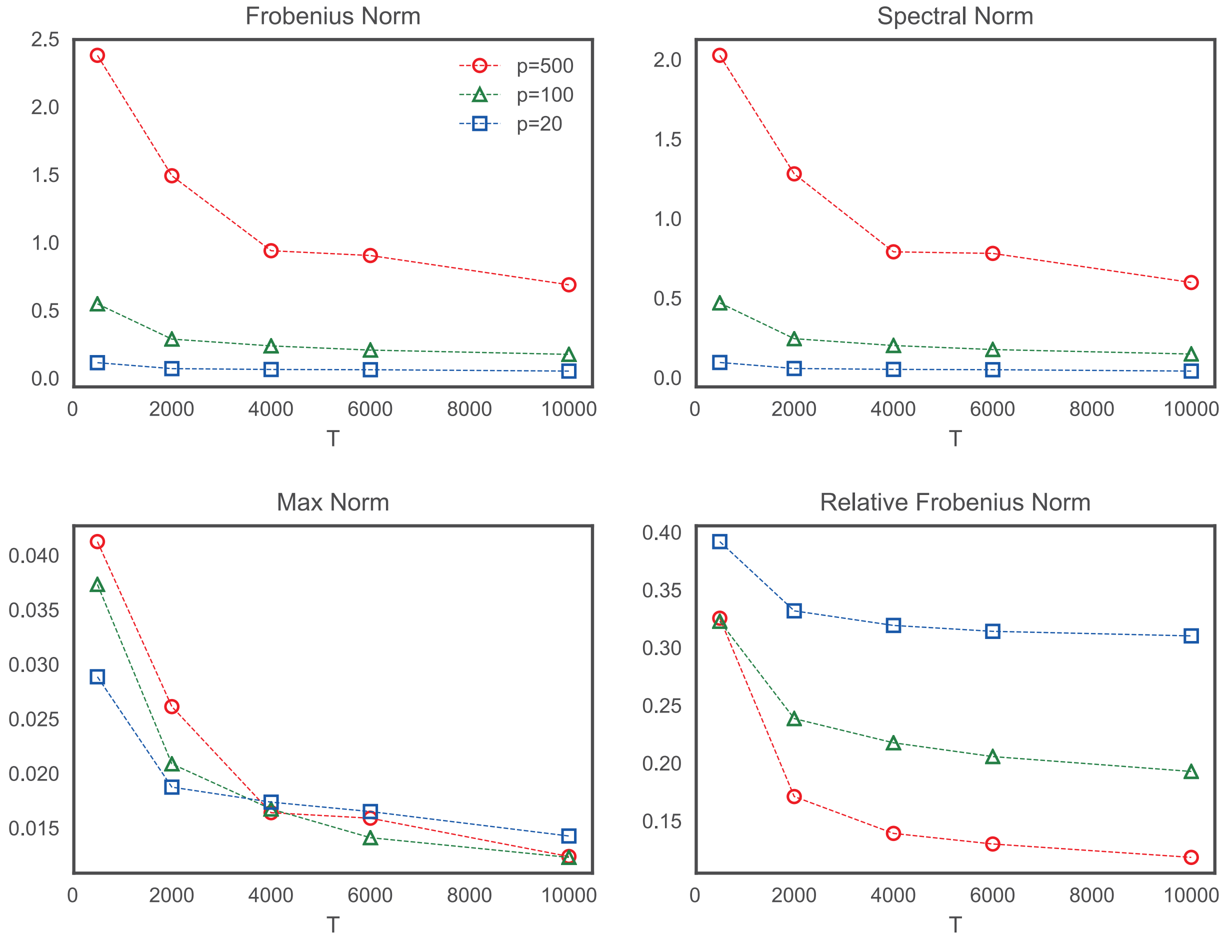

Figure 1: Average predicted volatility errors of $\widehat{\mathbf{\Sigma}}_{t+1} - \mathbf{\Sigma}_{t+1}$ against T under the Frobenius, spectral, max, and relative Frobenius norms with $p = 20, 100, 500$.

theoretical findings in Theorem 4.1.

One of the main purposes of this paper is to check the benefit of incorporating financial big data in small portfolio risk analysis. To do this, we compared the performance of predicting one-step ahead volatility matrices of the assets in the portfolio. We chose the portfolio size as $s = 5$ and the total number of assets as $p = 500$. To capture the market dynamics using only the assets in the portfolio, we considered the CCC model (Bollerslev, 1990), BEKK with diagonal-constraint and variance targeting (BEKK) (Engle and Kroner, 1995; Pedersen and Rahbek, 2014). Additionally, in the empirical study, we often impose the static or slow time varying covariance assumption, and under this condition, we adopted POET (Fan

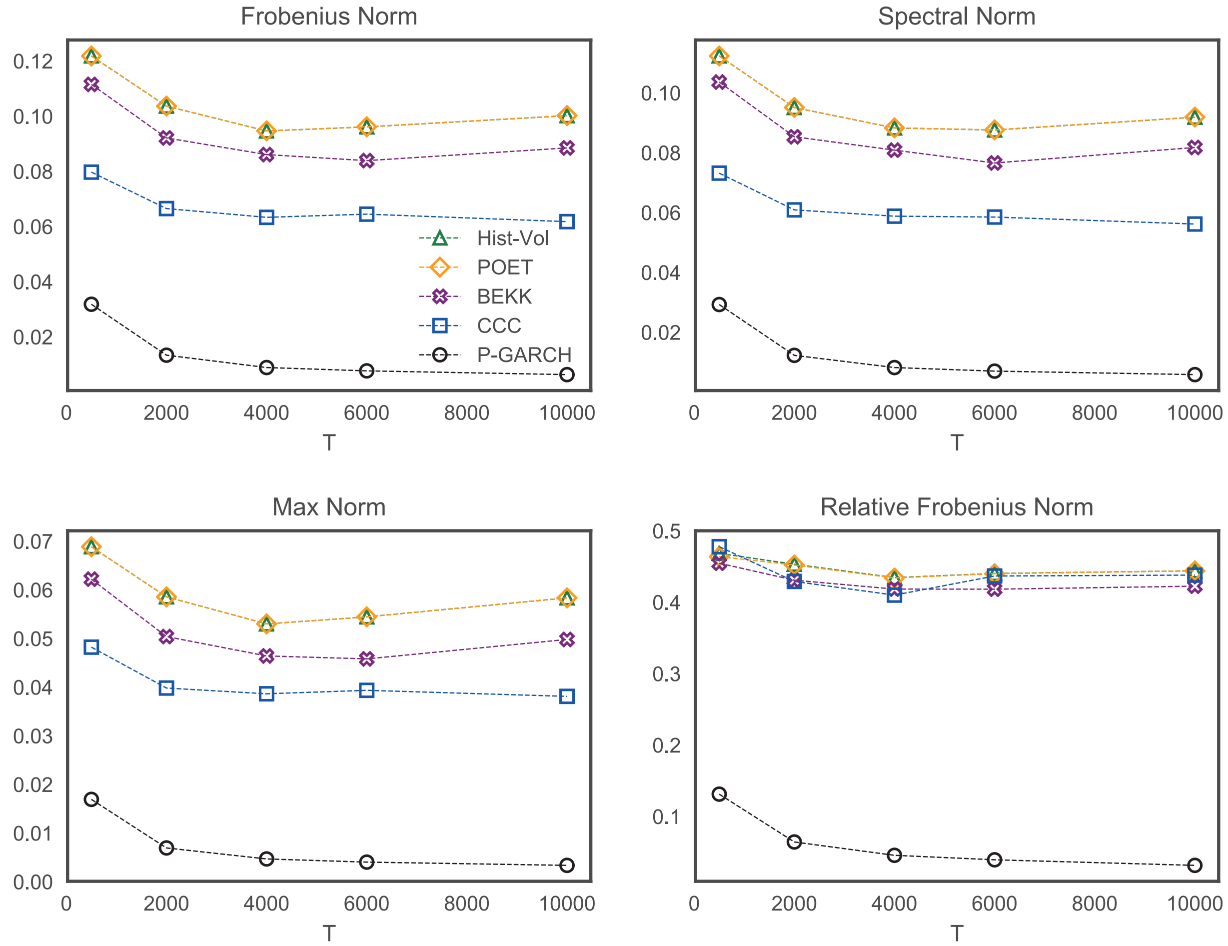

Figure 2: Average predicted volatility errors of the portfolio volatility matrix $\widehat{\Sigma}_{s,t+1} - \Sigma_{s,t+1}$ against T under the Frobenius, spectral, max, and relative Frobenius norms with $p = 500$ and portfolio size 5.

et al., 2013) and historical sample covariance (Hist-Vol). Figure 2 plots the average predicted volatility errors of the portfolio volatility matrix $\widehat{\Sigma}_{s,t+1} - \Sigma_{s,t+1}$ against the sample size T under the Frobenius, spectral, max, and relative Frobenius norms. Figure 2 illustrates that the P-GARCH model is superior to the other multivariate volatility matrix estimates. This may be because using only small portfolio data, they cannot explain the market dynamics fully.

Finally, we compared the VaR forecasting performance with the small portfolio risk analysis methods. We set α as 1% and fixed $p = 500$. The portfolio size varied from 1 to 20. We

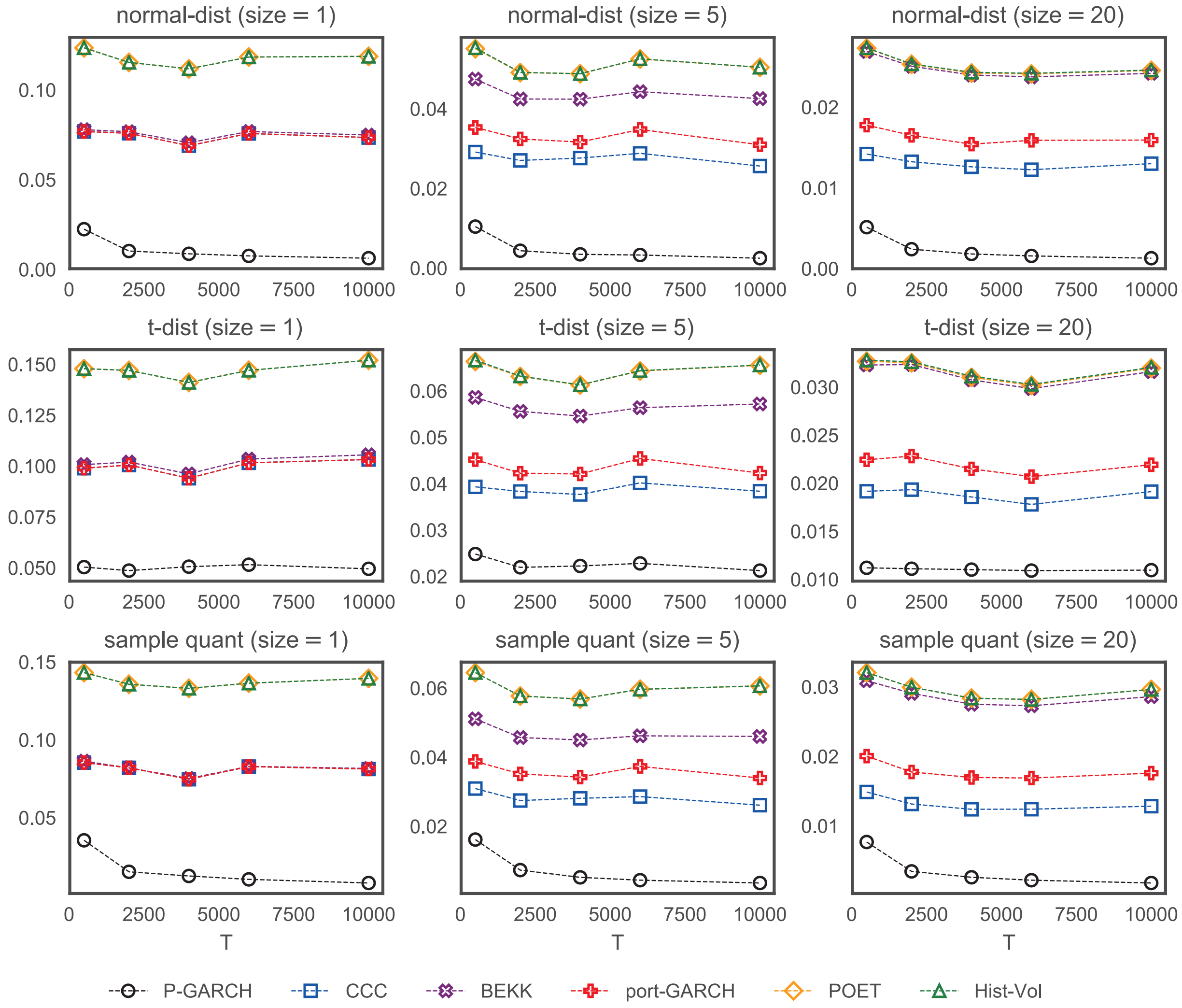

Figure 3: MAE of the 1%-level VaR forecasts for $p = 500$ with 500 replications. From the left, the plot shows the MAE of VaR with the parametric methods under normal and t-distribution ($\nu = 6$) and non-parametric method with the sample quantile with the portfolio sizes 1, 5, and 20.

also calculated the small portfolio matrix estimators $\widehat{\Sigma}_{s,t+1}$ based on the P-GARCH, CCC, BEKK, POET, and Hist-Vol for the volatility forecast. We additionally added the portfolio univariate GARCH model (Bollerslev, 1986) for the comparison. Specifically, we applied the usual GARCH(1,1) model to the portfolio return. For the α-quantile, we considered the normal and student-t distribution with 6 degrees of freedom ($\nu = 6$) for the parametric method, whereas we used the non-parametric method with $\lceil \alpha T \rceil$-th smallest values of the standardized portfolio log returns, as in Section 4.2. The portfolio was set to be equally weighted. The true VaR is $z_\alpha(\mathbf{w}^\top \boldsymbol{\Sigma}_{t+1}\mathbf{w})^{1/2}$, where z_α is the α-quantile for the standard normal distribution. Figure 3 draws the MAEs of the 1%-level VaR forecasts for $p = 500$ with 500 replications. From Figure 3, we find that the P-GARCH outperforms the other models for any σ-based methods and portfolio size.

6 Empirical study

In this section, we examined the one-step ahead VaR forecasting performance with the empirical data. The data comprise 18 years CRSP daily percentage log returns of S&P 500 constituents from January 1, 2000, to December 31, 2017 (4,523 days). The log returns are calculated from the percentage return of the last sale or closing bid/ask price including dividend. We selected every firm that was ever a constituent of the S&P 500 during this period and filtered out the stocks which have any missing data, leaving 492 firms ($p = 492$).

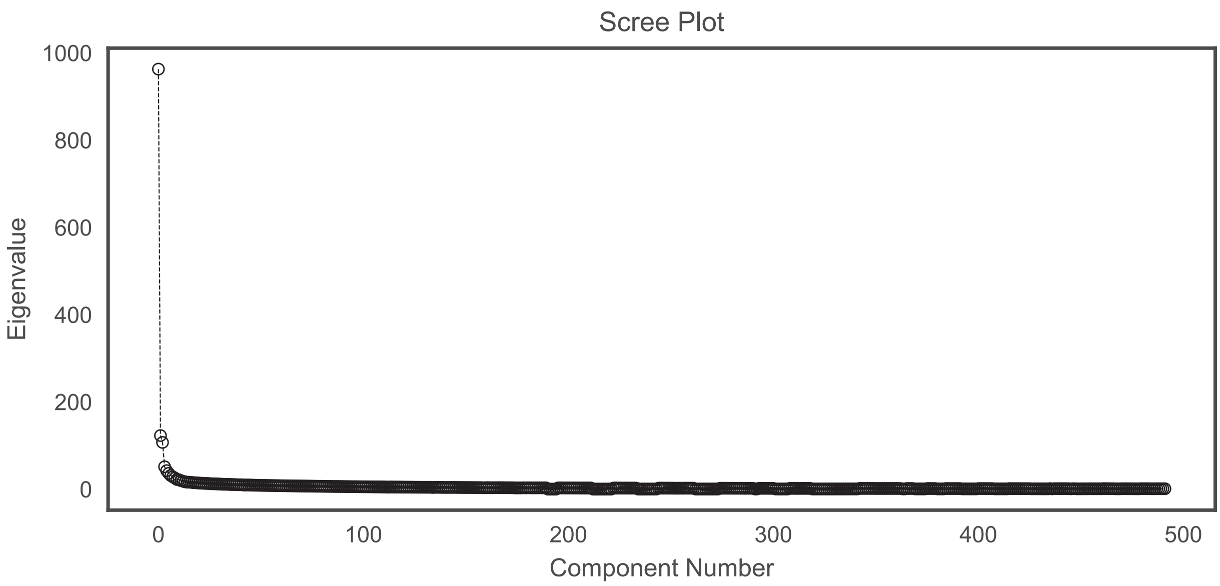

Figure 4: The scree plot of the sample covariance matrix of 492 stocks.

To employ the proposed P-GARCH model, the number of factors r is required. To do this, we drew the scree plot in Figure 4. Figure 4 shows that r is less than 3. So we determined the possible values of r as 1, 2, and 3. This is also matched with the previous empirical studies (Chan et al., 1998; Fama and French, 1992, 1993). For the thresholding step, we used the global industry classification standard (GICS) sector (Fan et al., 2016). For example, we maintained within-sector volatilities but set others to zero. The equally weighted portfolios of size 5 and 20 are randomly sampled from the 492 stocks with 500 repetitions, and the single-asset portfolio contains only one stock from 492 assets. With these portfolios, we predicted VaR with significance level $\alpha = 10\%$, 5%, 2%, and 1% using the parametric and non-parametric methods, as discussed in Section 4.2. For the model fitting, we employed a rolling window scheme with the window size $T = 252$ (1 year). The parameter is updated every 10 days, whereas the VaR forecasting is done every day with the updated log return data. The total forecasting number N is 4,270. The last estimated parameters of P-GARCH(r=3) are $\widehat{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B})) = 10^{-3} \times$ (54.33, 0.00, 0.00, 0.00, 79.23, 143.68, 0.00, 0.00, 0.00, 20.81, 0.00, 8.61, 500.66, 705.18, 14.39, 360.23, 88.21, 0.48, 0.26, 773.81, 16.58). Figure 5 depicts one sample path of the estimated one-step ahead 1%-level VaR under t-distribution with portfolio size 5. It shows predicted VaR of the P-GARCH model can capture the dynamics of extreme loss.

To test whether the predicted VaR is correct, we used the unconditional coverage test (Kupiec, 1995) and the conditional coverage test (Christoffersen, 1998). We defined the hit rate as the number of hitting N_α divided by the number of total samples N. Then the unconditional coverage test is a likelihood ratio test whether the hit rate is sufficiently close to α. Under the null hypothesis, the hit rate should be an unbiased estimator of the significance level α (Jorion, 2000). Then the test statistic is defined as follows:

$$
\begin{aligned}
LR_{uc} = & -2\left\{(N - N_\alpha)\log(1-\alpha) + N_\alpha \log \alpha\right\} \\
& + 2\left\{(N - N_\alpha)\log\left(1 - \frac{N_\alpha}{N}\right) + N \log \frac{N_\alpha}{N}\right\}.
\end{aligned}
$$

The conditional coverage test checks whether the hit is consecutively occurring or not, and

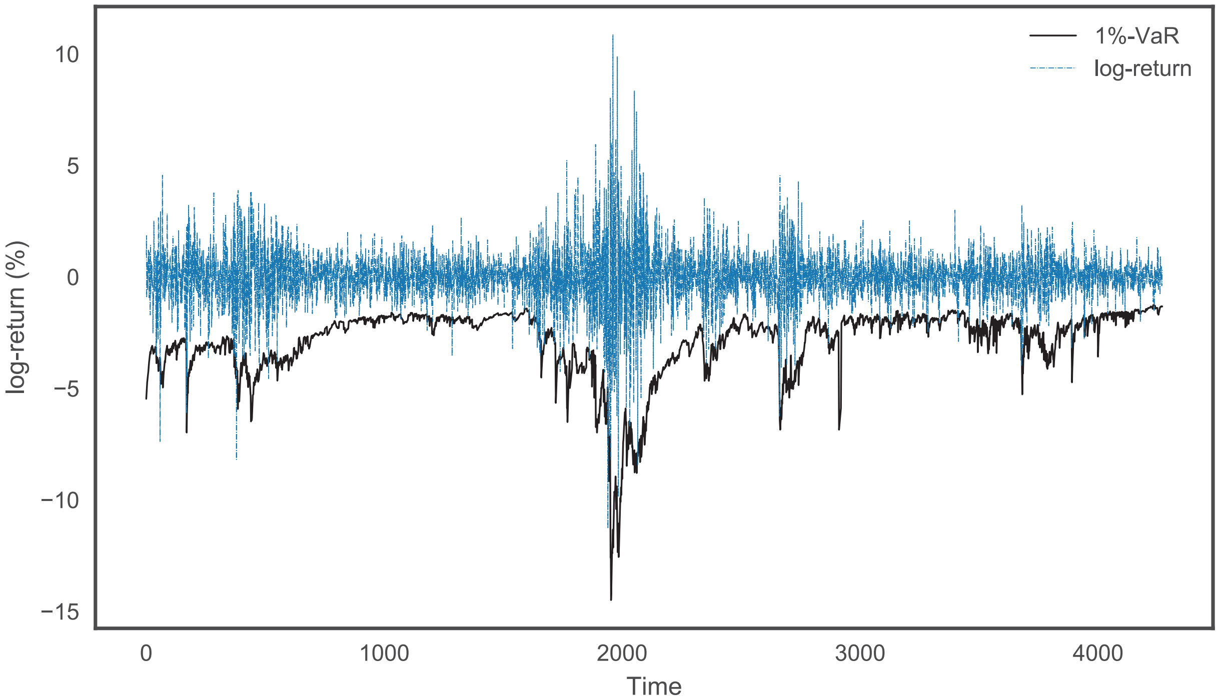

Figure 5: The plots of daily log return and predicted one-step ahead VaR of P-GARCH model.

its test statistics is defined as follows:

$$
\begin{aligned}
LR_{cc} &= -2\left\{(N - N_\alpha)\log(1-\alpha) + N_\alpha \log \alpha\right\} \\
&\quad + 2\log\left\{\left(\frac{N_{00}}{N_0}\right)^{N_{00}}\left(\frac{N_{10}}{N_0}\right)^{N_{10}}\left(\frac{N_{01}}{N_1}\right)^{N_{01}}\left(\frac{N_{11}}{N_1}\right)^{N_{11}}\right\},
\end{aligned}
$$

where $N_{00}, N_{01}, N_{10}, N_{11}$ are conditional frequencies defined in Table 2. Specifically, N_{00} is the number of hits when there was a hit right before, whereas N_{01} is the number of hits when there was no hit right before. N_{10} and N_{11} are the number of opposite cases. Hence, the summation $N_{00} + N_{01} + N_{10} + N_{11}$ is equal to N. Kupiec (1995) and Christoffersen (1998) showed that under the null, LR_{uc} and LR_{cc} follow the chi-square distribution with degrees of freedom 1 and 2, respectively. To obtain more robust results, we also considered the dynamic quantile test (Kuester et al., 2006). It generalizes the conditional coverage test by considering the relation of the current hit and the multiple lagged information. For example, we denote the dummy variable H_t an indicator of the current hit. Then we can have a regression with lag L as follows:

$$H_t = \beta_0 + \sum_{i=1}^{L} \beta_i H_{t-i} + \varepsilon_t,$$

and with the vector notation, we have

$$\mathbf{H} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{1}$ is a vector of ones and $\mathbf{X}$ is a matrix of lagged variables. Under the null hypothesis, β_0 should be equal to α, and $\boldsymbol{\beta}$ should be zero. Moreover, Engle and Manganelli (2004) showed that the test statistic (DQ) converges in distribution to the chi-square as follows:

$$DQ = \frac{\widehat{\boldsymbol{\beta}}^{\top} \mathbf{X}^{\top} \mathbf{X} \widehat{\boldsymbol{\beta}}}{\alpha(1-\alpha)} \xrightarrow{d} \chi^2_{L+2},$$

where χ^2_{L+2} is the chi-square with the degrees of freedom $L+2$. For the dynamic quantile test, when $\mathbf{X}$ is $H_{t-1}, \ldots, H_{t-4}$, we denote the test by $\mathrm{DQ}_{\mathrm{Hit}}$. When $\mathbf{X}$ is $H_{t-1}, \ldots, H_{t-4}, \widehat{\mathrm{VaR}}_{\alpha,t}$, we denote the test by $\mathrm{DQ}_{\mathrm{VaR}}$.

Table 2: The conditional frequency table of the VaR hit events.

t \ $t-1$	$r_{t-1} < \widehat{\mathrm{VaR}}_{\alpha,t-1}$	$r_{t-1} \geq \widehat{\mathrm{VaR}}_{\alpha,t-1}$	
$r_t < \widehat{\mathrm{VaR}}_{\alpha,t}$	N_{00}	N_{01}	
$r_t \geq \widehat{\mathrm{VaR}}_{\alpha,t}$	N_{10}	N_{11}	
Total	N_0	N_1	N

We compared the out-of-sample VaR forecast performances with the small portfolio risk analysis methods considered in Section 5. For example, we used the P-GARCH, CCC, BEKK, portfolio GARCH, Hist-Vol, and POET for the volatility prediction and the standard normal quantile, t-quantile with degrees of freedom 6, and sample quantile for the α-quantile. Table 3 reports the averaged hit rates (N_α/N) and the p-values of the LR_{uc}, LR_{cc}, $\mathrm{DQ}_{\mathrm{Hit}}$, and $\mathrm{DQ}_{\mathrm{VaR}}$ for the portfolio sizes 1, 5, 20 and $\alpha = 1\%$ with the normal distribution, student-t distribution, and sample quantile. Figure 6 draws box-plots of LR_{cc} p-values for the portfolio sizes 1, 5, and 20 and $\alpha = 1\%$ with the normal distribution, t-distribution, and sample quantile. From Table 3 and Figure 6, we find that when comparing the dynamic models and static models, the dynamic models such as the P-GARCH, BEKK, CCC, and portfolio GARCH show better performance. This indicates that the financial market is not static and the GARCH-type models can account for the market dynamics. When comparing the dynamic models, the P-GARCH has the highest p-values for most of the VaR tests. From this empirical result, we can conjecture that the P-GARCH accounts for the market dynamics by incorporating financial big data, and this leads to better performance in VaR forecasting for the relatively small portfolio. Finally, when comparing the quantile methods, the t-distribution shows the best performance. Moreover, the normal distribution has the smallest p-values for all the portfolio sizes. This pattern coincides with the stylized fact that the log return follows the conditional heavy-tailed distribution (Cont, 2001).

Table 3: Average hit rate (N_α/N) and the p-values of the 1%-level VaR test statistics under the normal, t-distribution $\nu = 6$ and sample quantile for the portfolio sizes 1, 5, and 20.

Models	N_α/N	LR_{uc}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	N_α/N	LR_{uc}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	N_α/N	LR_{uc}	LR_{cc}	DQ_{Hit}	DQ_{VaR}
	normal-distribution (size = 1)					normal-distribution (size = 5)					normal-distribution (size = 20)				
P-GARCH(r=1)	0.016	0.043	0.040	0.024	0.021	0.018	0.002	0.003	0.001	0.001	0.019	0.000	0.000	0.000	0.000
P-GARCH(r=2)	0.015	0.049	0.047	0.032	0.029	0.017	0.004	0.004	0.001	0.001	0.019	0.000	0.000	0.000	0.000
P-GARCH(r=3)	0.015	0.045	0.042	0.028	0.025	0.017	0.002	0.002	0.000	0.001	0.019	0.000	0.000	0.000	0.000
CCC	0.018	0.002	0.004	0.003	0.001	0.018	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000
BEKK	0.017	0.006	0.009	0.007	0.004	0.020	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000
port-GARCH	0.018	0.003	0.004	0.003	0.001	0.019	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000
POET(r=1)	0.018	0.008	0.007	0.001	0.001	0.021	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000
POET(r=2)	0.018	0.008	0.007	0.001	0.001	0.020	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000
POET(r=3)	0.018	0.008	0.007	0.001	0.001	0.020	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000
Hist-Vol	0.018	0.008	0.007	0.001	0.001	0.020	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000
	t-distribution (size = 1)					t-distribution (size = 5)					t-distribution (size = 20)				
P-GARCH(r=1)	0.011	0.367	0.243	0.110	0.105	0.012	0.339	0.279	0.039	0.044	0.012	0.208	0.206	0.005	0.005
P-GARCH(r=2)	0.011	0.405	0.262	0.121	0.117	0.012	0.377	0.309	0.060	0.061	0.012	0.190	0.202	0.014	0.012
P-GARCH(r=3)	0.011	0.390	0.246	0.120	0.118	0.012	0.353	0.272	0.053	0.053	0.013	0.150	0.154	0.012	0.012
CCC	0.013	0.202	0.195	0.093	0.060	0.013	0.212	0.208	0.012	0.016	0.013	0.125	0.153	0.001	0.001
BEKK	0.012	0.264	0.250	0.090	0.078	0.014	0.065	0.066	0.000	0.001	0.016	0.004	0.004	0.000	0.000
port-GARCH	0.013	0.202	0.195	0.093	0.061	0.013	0.114	0.128	0.012	0.008	0.014	0.048	0.072	0.001	0.001
POET(r=1)	0.013	0.201	0.107	0.027	0.021	0.015	0.031	0.023	0.000	0.000	0.016	0.003	0.002	0.000	0.000
POET(r=2)	0.013	0.201	0.107	0.027	0.021	0.015	0.036	0.027	0.000	0.000	0.016	0.004	0.003	0.000	0.000
POET(r=3)	0.013	0.201	0.107	0.027	0.021	0.015	0.036	0.028	0.000	0.000	0.016	0.004	0.003	0.000	0.000
Hist-Vol	0.013	0.201	0.107	0.027	0.021	0.015	0.032	0.024	0.000	0.000	0.016	0.003	0.002	0.000	0.000
	sample quantile (size = 1)					sample quantile (size = 5)					sample quantile (size = 20)				
P-GARCH(r=1)	0.013	0.172	0.130	0.050	0.032	0.013	0.104	0.120	0.029	0.028	0.014	0.024	0.031	0.004	0.005
P-GARCH(r=2)	0.013	0.183	0.144	0.066	0.041	0.013	0.097	0.112	0.033	0.031	0.014	0.021	0.034	0.009	0.007
P-GARCH(r=3)	0.013	0.152	0.125	0.050	0.031	0.013	0.075	0.088	0.025	0.027	0.015	0.012	0.021	0.005	0.005
CCC	0.014	0.042	0.059	0.031	0.013	0.014	0.043	0.058	0.003	0.004	0.015	0.014	0.023	0.000	0.000
BEKK	0.013	0.107	0.125	0.047	0.027	0.014	0.054	0.063	0.001	0.001	0.014	0.017	0.010	0.000	0.000
port-GARCH	0.014	0.042	0.059	0.031	0.013	0.014	0.024	0.039	0.005	0.003	0.015	0.007	0.012	0.000	0.000
POET(r=1)	0.013	0.107	0.071	0.015	0.006	0.014	0.043	0.029	0.000	0.000	0.015	0.016	0.008	0.000	0.000
POET(r=2)	0.013	0.107	0.071	0.015	0.006	0.014	0.043	0.029	0.000	0.000	0.015	0.016	0.008	0.000	0.000
POET(r=3)	0.013	0.107	0.071	0.015	0.006	0.014	0.043	0.029	0.000	0.000	0.015	0.016	0.008	0.000	0.000
Hist-Vol	0.013	0.107	0.071	0.015	0.006	0.014	0.043	0.029	0.000	0.000	0.015	0.016	0.008	0.000	0.000

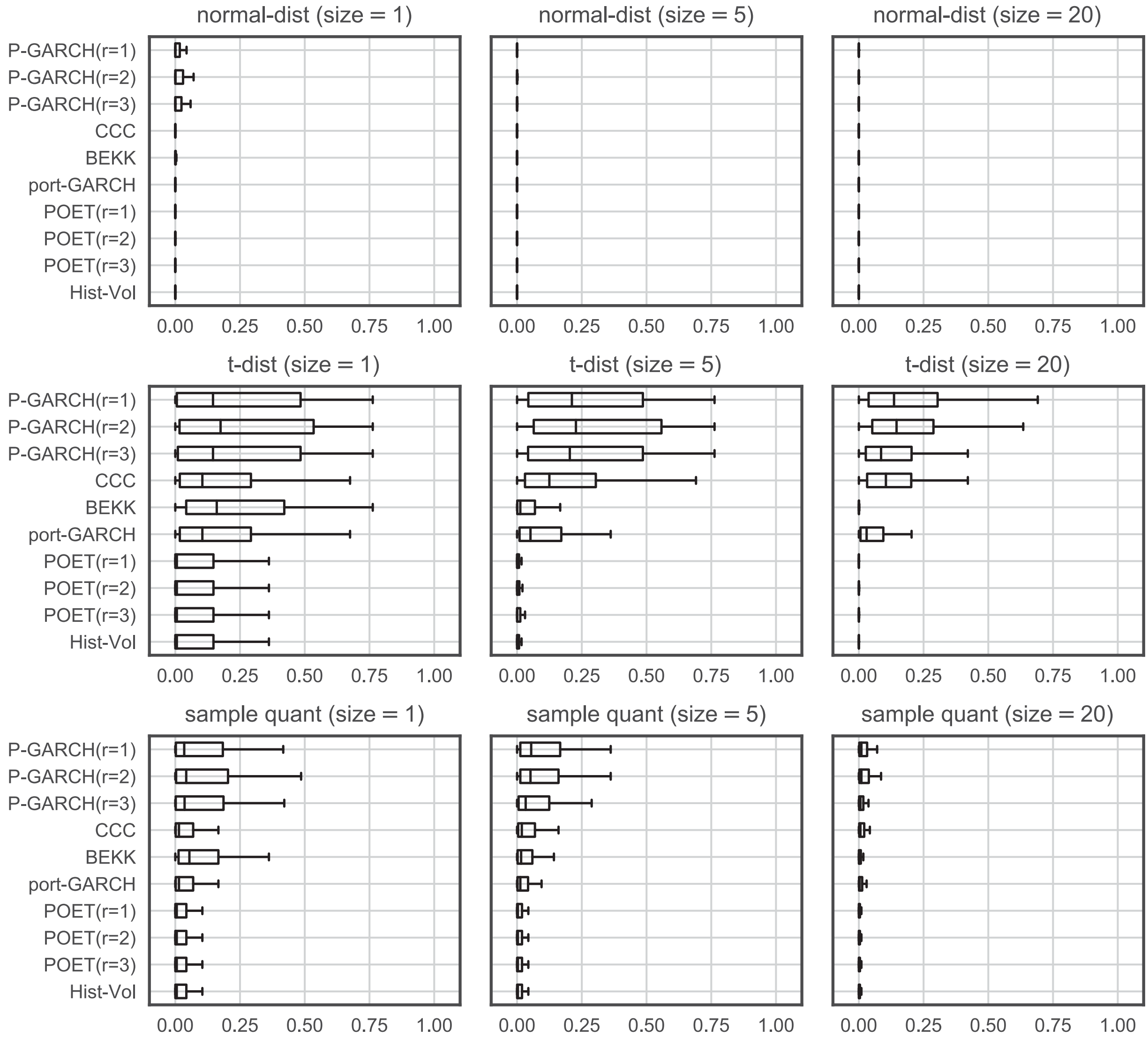

Figure 6: The box plots of p-values for LR_{cc} with $\alpha = 1\%$. Each column shows three different VaR estimation results with the normal, t-distribution ($\nu = 6$), and sample quantile. Each row shows four different VaR estimates for the portfolio sizes 1, 5, 20.

7 Conclusion

In this paper, we studied the blessing of dimensionality in risk analysis by incorporating financial big data and understanding how to overcome the curse of dimensionality. Specifically, under the latent factor model, to account for the common market dynamics, we proposed the dynamic factor model, which is the famous GARCH form. For the factor volatility, we showed that a large number of assets help to estimate the factor volatility more accurately. The intuition behind is that every asset contains common market information, so as the number of assets increases, we have more information for the latent factor. That is, we can enjoy the blessing of dimensionality in estimating the latent common factor. This fact indicates that even if we analyze small portfolio risk, incorporating financial big data can be helpful. However, when handling big data, we must inevitably face the curse of dimensionality. We overcame the curse of dimensionality by adopting the POET method (Fan et al., 2013). Then we applied the proposed P-GARCH model to measuring VaR for the relatively small portfolio. The numerical studies showed that P-GARCH can capture the market dynamics better than the small portfolio dynamic risk models which consider only the assets in the portfolio. From these results, we can conjecture that by incorporating financial big data, we can better account for the market dynamics.

In this paper, we only consider the factor dynamics, but there are several studies that the idiosyncratic volatilities are also affected by the market common factors (Barigozzi and Hallin, 2016; Connor et al., 2006; Herskovic et al., 2016; Rangel and Engle, 2012). Thus, it is interesting but difficult to develop idiosyncratic dynamic models. We leave this for the future study.

References

Acerbi, C. and Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic notes*, 31(2):379–388.

Ahmadi-Javid, A. (2012). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123.

Ait-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Bali, T. G., Cakici, N., Yan, X., and Zhang, Z. (2005). Does idiosyncratic risk really matter? *The Journal of Finance*, 60(2):905–929.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, 19(1).

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bensaid, B., Lesne, J.-P., Pages, H., and Scheinkman, J. (1992). Derivative asset pricing with transaction costs 1. *Mathematical Finance*, 2(2):63–86.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

Billio, M. and Pelizzon, L. (2000). Value-at-risk: a multivariate switching regime approach. *Journal of Empirical Finance*, 7(5):531–554.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Review of Economics and Statistics*, 72(3):498–505.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172.

Brooks, C. and Persand, G. (2003). Volatility forecasting for risk management. *Journal of Forecasting*, 22(1):1–22.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684.

Cajueiro, D. O. and Tabak, B. M. (2004). The hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Physica A: Statistical Mechanics and its Applications*, 336(3-4):521–537.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chan, L. K., Karceski, J., and Lakonishok, J. (1998). The risk and return from factors. *Journal of Financial and Quantitative Analysis*, 33(2):159–188.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, 3(2):227–255.

Christoffersen, P. and Langlois, H. (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis*, 48(5):1371–1404.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, pages 841–862.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate garch processes. *Journal of Multivariate Analysis*, 84(1):61–84.

Connor, G., Korajczyk, R. A., and Linton, O. (2006). The common and specific components of dynamic volatility. *Journal of Econometrics*, 132(1):231–255.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000):32.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

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

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.

Engle, R. F., Ledoit, O., and Wolf, M. (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, pages 1–13.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Engle, R. F., Shephard, N., and Sheppard, K. (2008). Fitting vast dimensional time-varying covariance models. Technical Report FIN-08-009.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fan, J., Furger, A., and Xiu, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics*, 34(4):489–503.

Fan, J. and Gu, J. (2003). Semiparametric estimation of value at risk. *The Econometrics Journal*, 6(2):261–290.

Fan, J. and Kim, D. (2018). Robust high-dimensional volatility matrix estimation for high-frequency factor model. *Journal of the American Statistical Association*, 113(523):1268–1283.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680.

Fan, J., Liu, H., and Wang, W. (2018). Large covariance estimation through elliptical factor models. *Annals of statistics*, 46(4):1383.

Fan, J., Wang, W., and Zhong, Y. (2017). An eigenvector perturbation bound and its application to robust covariance estimation. *The Journal of Machine Learning Research*, 18(1):7608–7649.

Fan, J., Wang, W., and Zhong, Y. (2019). Robust covariance estimation for approximate factor models. *Journal of Econometrics*, 208(1):5–22.

Fantazzini, D. (2008). Dynamic copula modelling for value at risk. *Frontiers in Finance and Economics*, 5(2):72–108.

Giot, P. and Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance*, 11(3):379–398.

Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269.

Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic Policy Review*, 2(1).

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283.

Huisman, R., Koedijk, K. G., and Pownall, R. A. (1998). Var-x: Fat tails in financial risk management. *Journal of Risk*, 1(1):47–61.

Hull, J. and White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1):5–19.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6):47–56.

Jorion, P. (2000). *Value at risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill Professional Publishing.

Kalnina, I. and Tewou, K. (2015). Cross-sectional dependence in idiosyncratic volatility.

Kawakatsu, H. (2006). Matrix exponential garch. *Journal of Econometrics*, 134(1):95–128.

Kim, D. and Fan, J. (2019). Factor garch-itô models for high-frequency data with application to large volatility matrix prediction. *Journal of Econometrics*, 208(2):395–417.

Kim, D., Liu, Y., and Wang, Y. (2018). Large volatility matrix estimation with factor-based diffusion model for high-frequency financial data. *Bernoulli*, 24(4B):3657–3682.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2).

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52.

Li, Q., Cheng, G., Fan, J., and Wang, Y. (2018). Embracing the blessing of dimensionality in factor models. *Journal of the American Statistical Association*, 113(521):380–389.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82.

Malkiel, B. G. and Xu, Y. (1997). Risk and return revisited. *Journal of Portfolio Management*, 23(3):9.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.

Narayan, P. K. and Smyth, R. (2004). Is south korea's stock market efficient? *Applied Economics Letters*, 11(11):707–710.

Natarajan, K., Pachamanova, D., and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, 54(3):573–585.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2017). Fitting vast dimensional time-varying covariance models.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*.

Pedersen, R. S. and Rahbek, A. (2014). Multivariate variance targeting in the bekk–garch model. *The Econometrics Journal*, 17(1):24–55.

Rangel, J. G. and Engle, R. F. (2012). The factor–spline–garch model for high and low frequency correlations. *Journal of Business & Economic Statistics*, 30(1):109–124.

Sadorsky, P. (2005). Stochastic volatility forecasting and risk management. *Applied Financial Economics*, 15(2):121–135.

Shiller, R. J. (1995). Aggregate income risks and hedging mechanisms. *The Quarterly Review of Economics and Finance*, 35(2):119–152.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Van der Weide, R. (2002). Go-garch: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics*, 17(5):549–564.

Wu, P.-T. and Shieh, S.-J. (2007). Value-at-risk analysis for long-term interest rate futures: Fat-tail and long memory in return innovations. *Journal of Empirical Finance*, 14(2):248–259.

Zoia, M. G., Biffi, P., and Nicolussi, F. (2018). Value at risk and expected shortfall based on gram-charlier-like expansions. *Journal of Banking & Finance*, 93:92–104.

A Proofs

We first introduce some notations. For a symmetric matrix $\mathbf{A}$, $\lambda_k(\mathbf{A})$ is the k^{th} largest eigenvalue of $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue, and $\lambda_{\min}(\mathbf{A})$ is the smallest eigenvalue. $\|\cdot\|_\infty$ is the matrix ℓ_∞ norm. Denote $\mathbf{A}\circ\mathbf{B}$ the Hadamard (or element-wise) product and $\mathbf{x}^2 = \mathbf{x}\circ\mathbf{x}$ the element-wise square of a vector $\mathbf{x}$. $\mathrm{tr}(\mathbf{A})$ and $\mathrm{vec}(\mathbf{A})$ are the trace and vectorization of $\mathbf{A}$, respectively. Denote $\mathbb{1}_E$ the indicator function of the event E and $\mathbf{e}_i$ the i^{th} standard basis vector. To simplify the notations, we define ∂ as a partial derivatives: $\partial_x y(\mathbf{x}) = \partial y(\mathbf{x})/\partial x$; $\partial_x \mathbf{y}(\mathbf{x}) = \partial \mathbf{y}(\mathbf{x})/\partial x = [\partial y_1/\partial x, \ldots, \partial y_{\dim(\mathbf{y})}/\partial x]^\top$; $\partial_{\mathbf{x}} y(\mathbf{x}) = [\partial y/\partial x_1, \ldots, \partial y/\partial x_{\dim(\mathbf{x})}]$; $\partial_{\mathbf{x}}\mathbf{y}(\mathbf{x}) = [\partial \mathbf{y}/\partial x_1, \ldots, \partial \mathbf{y}/\partial x_{\dim(\mathbf{x})}]$, where the functions $y(\mathbf{x})$ and $\mathbf{y}(\mathbf{x})$ are differentiable. C is a generic constant that may differ from each equations.

A.1 Proof of Theorem 3.1

Lemma A.1. *Under the assumptions of Theorem 3.1, we have*

$$\mathbb{E}\left\|\widehat{\overline{\mathbf{\Sigma}}} - \overline{\mathbf{\Sigma}}\right\|_F^4 = O\left(\frac{p^4}{T^2}\right).$$

Proof of Lemma A.1. We have

$$\widehat{\overline{\mathbf{\Sigma}}} - \overline{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\left((\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})^\top - \mathbf{\Sigma}_t\right)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{V}\left(\mathbf{f}_t\mathbf{f}_t^\top - \mathrm{diag}(\mathbf{h}_{0,t})\right)\mathbf{V}^\top + \mathbf{V}\mathbf{f}_t\mathbf{u}_t^\top + \mathbf{u}_t\mathbf{f}_t^\top\mathbf{V}^\top + \mathbf{u}_t\mathbf{u}_t^\top - \boldsymbol{\Sigma}_u\right)$$
$$- (\bar{\mathbf{y}} - \boldsymbol{\mu})(\bar{\mathbf{y}} - \boldsymbol{\mu})^\top,$$

where $\bar{\mathbf{y}} = (1/T)\sum_{t=1}^{T}\mathbf{y}_t$ and $\mathbf{h}_{0,t} = \mathbf{h}_t(\boldsymbol{\theta}_0)$. Then each element is a martingale difference, and so the Burkholder-Davis-Gundy inequality shows the statement. □

Proof of Theorem 3.1. First consider (3.2). By the Davis-Kahan's sine theorem, we have

$$\begin{aligned}\mathbb{E}\left\|\widehat{\mathbf{V}} - \mathbf{VO}\right\|^4 &= p^2\mathbb{E}\left\|\frac{1}{\sqrt{p}}\widehat{\mathbf{V}} - \frac{1}{\sqrt{p}}\mathbf{VO}\right\|^4 \\ &\le p^2\mathbb{E}\left[\frac{\left\|\widehat{\overline{\boldsymbol{\Sigma}}} - \overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_u\right\|^4}{p^4\delta_r^4}\right] \\ &\le C\left(\frac{p^2}{T^2} + \frac{s_p^4}{p^2}\right), \end{aligned} \tag{A.1}$$

where the last equality is due to Lemma A.1 and the fact that $\|\boldsymbol{\Sigma}_u\| \le \|\boldsymbol{\Sigma}_u\|_\infty \le Cs_p$.

Consider (3.3). Without of loss of generality, we assume that $\mathbf{O}$ is the identity matrix. Algebraic manipulations show

$$\begin{aligned}&\mathbb{E}\left\|\widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2\right\|^2 \\ &= \mathbb{E}\left\|\left(\frac{1}{p}\widehat{\mathbf{V}}^\top(\mathbf{y}_t - \bar{\mathbf{y}})\right)^2 - \left(\frac{1}{p}\mathbf{V}^\top(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{u}_t)\right)^2\right\|^2 \\ &= \frac{1}{p^4}\mathbb{E}\left\|(\widehat{\mathbf{V}}^\top\mathbf{y}_t)^2 - 2(\widehat{\mathbf{V}}^\top\mathbf{y}_t \circ \widehat{\mathbf{V}}^\top\bar{\mathbf{y}}) + (\widehat{\mathbf{V}}^\top\bar{\mathbf{y}})^2 - (\mathbf{V}^\top\mathbf{y}_t)^2 + 2p\mathbf{f}_t \circ \mathbf{V}^\top(\boldsymbol{\mu} + \mathbf{u}_t) + (\mathbf{V}^\top(\boldsymbol{\mu} + \mathbf{u}_t))^2\right\|^2 \\ &= \frac{1}{p^4}\mathbb{E}\Big\|(\widehat{\mathbf{V}} - \mathbf{V})^\top\mathbf{y}_t \circ (\widehat{\mathbf{V}} + \mathbf{V})^\top\mathbf{y}_t + 2p\mathbf{f}_t \circ \mathbf{V}^\top\mathbf{u}_t + (\mathbf{V}^\top\mathbf{u}_t)^2 \\ &\qquad - 2(\widehat{\mathbf{V}}^\top\mathbf{y}_t \circ \widehat{\mathbf{V}}^\top\bar{\mathbf{y}}) + (\widehat{\mathbf{V}}^\top\bar{\mathbf{y}})^2 + 2p\mathbf{f}_t \circ \mathbf{V}^\top\boldsymbol{\mu} + (\mathbf{V}^\top\boldsymbol{\mu})^2 + 2(\mathbf{V}^\top\mathbf{u}_t \circ \mathbf{V}^\top\boldsymbol{\mu})\Big\|^2 \\ &\le \frac{C}{p^4}\Bigg(\mathbb{E}\left\|(\widehat{\mathbf{V}} - \mathbf{V})^\top\mathbf{y}_t \circ (\widehat{\mathbf{V}} + \mathbf{V})^\top\mathbf{y}_t\right\|^2 + p^2\mathbb{E}\left\|\mathbf{f}_t \circ \mathbf{V}^\top\mathbf{u}_t\right\|^2 + \mathbb{E}\left\|(\mathbf{V}^\top\mathbf{u}_t)^2\right\|^2 \\ &\qquad + \mathbb{E}\left\| - 2(\widehat{\mathbf{V}}^\top\mathbf{y}_t \circ \widehat{\mathbf{V}}^\top\bar{\mathbf{y}}) + (\widehat{\mathbf{V}}^\top\bar{\mathbf{y}})^2 + 2(\mathbf{V}^\top\mathbf{y}_t \circ \mathbf{V}^\top\boldsymbol{\mu}) - (\mathbf{V}^\top\boldsymbol{\mu})^2\right\|^2\Bigg)\end{aligned}$$

$$= \frac{C}{p^4}\Big((\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV})\Big).$$

For (I), we have

$$\begin{aligned}
(\mathrm{I}) &= \mathbb{E}\left\|(\widehat{\mathbf{V}} - \mathbf{V})^\top \mathbf{y}_t \circ (\widehat{\mathbf{V}} + \mathbf{V})^\top \mathbf{y}_t\right\|^2 \\
&\le \mathbb{E}\left[\left\|(\widehat{\mathbf{V}} - \mathbf{V})^\top \mathbf{y}_t\right\|^2 \left\|(\widehat{\mathbf{V}} + \mathbf{V})^\top \mathbf{y}_t\right\|^2\right] \\
&\le \mathbb{E}\left[\left\|\widehat{\mathbf{V}} - \mathbf{V}\right\|^2 \left\|\widehat{\mathbf{V}} + \mathbf{V}\right\|^2 \|\mathbf{y}_t\|^4\right] \\
&\le 4p\,\mathbb{E}\left[\left\|\widehat{\mathbf{V}} - \mathbf{V}\right\|^2 \|\mathbf{y}_t\|^4\right] \\
&\le 4p\sqrt{\mathbb{E}\left\|\widehat{\mathbf{V}} - \mathbf{V}\right\|^4}\sqrt{\mathbb{E}\|\mathbf{y}_t\|^8} \\
&\le C\left(\frac{p^4}{T} + p^2 s_p^2\right),
\end{aligned}$$

where the fourth inequality is due to the Hölder's inequality and the last inequality is due to (A.1). For (II), we have

$$\begin{aligned}
(\mathrm{II}) &\le \mathbb{E}\|\mathbf{f}_t\|^2 \mathbb{E}\left\|\mathbf{V}^\top \mathbf{u}_t\right\|^2 \\
&\le C\left(p s_p\right),
\end{aligned}$$

where the last inequality is by the fact that $\mathbb{E}\|\mathbf{V}^\top \mathbf{u}_t\|^2 = \mathrm{tr}(\mathbf{V}^\top \boldsymbol{\Sigma}_u \mathbf{V}) = O(p\, s_p)$. For (III), we have

$$\begin{aligned}
(\mathrm{III}) &= \mathbb{E}\left[\sum_{i=1}^{r} (\mathbf{v}_i^\top \mathbf{u}_t)^4\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{r} \frac{(\mathbf{v}_i^\top \mathbf{u}_t)^4}{(\mathbf{v}_i^\top \boldsymbol{\Sigma}_u \mathbf{v}_i)^2}\left(\mathbf{v}_i^\top \boldsymbol{\Sigma}_u \mathbf{v}_i\right)^2\right] \\
&\le Cp^2 \|\boldsymbol{\Sigma}_u\|^2 \mathbb{E}\left[\sum_{i=1}^{r} \frac{(\mathbf{v}_i^\top \mathbf{u}_t)^4}{(\mathbf{v}_i^\top \boldsymbol{\Sigma}_u \mathbf{v}_i)^2}\right] \\
&\le C s_p^2 p^2,
\end{aligned}$$

where $\mathbf{v}_i$ is the i^{th} column of $\mathbf{V}$. Similarly, we can show

$$
\begin{aligned}
\text{(IV)} &\leq 2\mathbb{E}\Big\| - \widehat{\mathbf{V}}^\top \mathbf{y}_t \circ \widehat{\mathbf{V}}\bar{\mathbf{y}} + \mathbf{V}^\top \boldsymbol{\mu} \circ \mathbf{V}^\top \mathbf{y}_t \Big\|^2 + \mathbb{E}\Big\| (\widehat{\mathbf{V}}^\top \bar{\mathbf{y}})^2 - (\mathbf{V}^\top \boldsymbol{\mu})^2 \Big\|^2 \\
&\leq C\left(\frac{p^4}{T} + p^2 s_p^2\right).
\end{aligned}
$$

Therefore, we have

$$
\mathbb{E}\big\|\widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2\big\|^2 = O\left(1/T + s_p^2/p^2 + s_p/p\right).
$$

□

A.2 Proof of Theorem 3.2

Define

$$
\begin{aligned}
L_{0,T}(\boldsymbol{\theta}) &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\left(\log h_{it}(\boldsymbol{\theta}) + h_{it}^{-1}(\boldsymbol{\theta}) h_{0,it}\right) = \frac{1}{T}\sum_{t=1}^{T} l_{0,t}(\boldsymbol{\theta}), \\
L_T(\boldsymbol{\theta}) &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\left(\log h_{it}(\boldsymbol{\theta}) + h_{it}^{-1}(\boldsymbol{\theta}) f_{it}^2\right) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \\
\widehat{L}_T(\boldsymbol{\theta}) &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\left(\log \widehat{h}_{it}(\boldsymbol{\theta}) + \widehat{h}_{it}^{-1}(\boldsymbol{\theta}) \widehat{f}_{it}^2\right) = \frac{1}{T}\sum_{t=1}^{T} \widehat{l}_t(\boldsymbol{\theta}),
\end{aligned}
$$

where $\mathbf{h}_{0,t} = \mathbf{h}_t(\boldsymbol{\theta}_0)$ and $\widehat{\mathbf{h}}_t(\boldsymbol{\theta}) = \boldsymbol{\omega} + \mathbf{A}\widehat{\mathbf{f}}_{t-1}^2 + \mathbf{B}\widehat{\mathbf{h}}_{t-1}(\boldsymbol{\theta})$.

Lemma A.2. *Under the assumptions of Theorem 3.2, let $\mathbf{h}_t$ to be a function of $\boldsymbol{\theta}$ and $\mathbf{h}_1$,*

$$
\mathbf{h}_t(\boldsymbol{\theta}, \mathbf{h}_1) = \mathbf{B}^{t-1}\mathbf{h}_1 + (\mathbf{I}_r - \mathbf{B})^{-1}\left(\mathbf{I}_r - \mathbf{B}^{t-1}\right)\boldsymbol{\omega} + \sum_{k=0}^{t-2}\mathbf{B}^k \mathbf{A}\mathbf{f}_{t-1-k}^2.
$$

Then for some true initial value $\mathbf{h}_{0,1}$,

$$
\begin{aligned}
\|\mathbf{h}_t(\boldsymbol{\theta}, \mathbf{h}_1) - \mathbf{h}_t(\boldsymbol{\theta}, \mathbf{h}_{0,1})\| &= O_p\left(\|\mathbf{B}\|^{t-1}\right), \\
|L_T(\boldsymbol{\theta}, \mathbf{h}_1) - L_T(\boldsymbol{\theta}, \mathbf{h}_{0,1})| &= O_p\left(\frac{1}{T}\right).
\end{aligned}
$$

Proof of Lemma A.2. By the sub-multiplicativity, we have

$$\begin{aligned}\|\mathbf{h}_t(\boldsymbol{\theta},\mathbf{h}_1)-\mathbf{h}_t(\boldsymbol{\theta},\mathbf{h}_{0,1})\| &= \left\|\mathbf{B}^{t-1}\mathbf{h}_1-\mathbf{B}^{t-1}\mathbf{h}_{0,1}\right\| \\ &\le \|\mathbf{B}\|^{t-1}\|\mathbf{h}_1-\mathbf{h}_{0,1}\| \\ &= C\|\mathbf{B}\|^{t-1}.\end{aligned}$$

This implies that

$$|L_T(\boldsymbol{\theta},\mathbf{h}_1)-L_T(\boldsymbol{\theta},\mathbf{h}_{0,1})| = O_p\left(\frac{1}{T}\right).$$

□

Lemma A.2 shows that the effect of the initial value is negligible. Thus, for the mathematical convenience, we assume that $\mathbf{h}_1(\boldsymbol{\theta})=(\mathbf{I}_r-\mathbf{B})^{-1}\boldsymbol{\omega}+\sum_{k=0}^{\infty}\mathbf{B}^k\mathbf{A}\mathbf{f}_{-k}^2$. Then we have $\mathbf{h}_t(\boldsymbol{\theta}) = (\mathbf{I}_r-\mathbf{B})^{-1}\boldsymbol{\omega}+\sum_{k=0}^{\infty}\mathbf{B}^k\mathbf{A}\mathbf{f}_{t-1-k}^2$. To obtain the consistent $\widehat{\boldsymbol{\theta}}$, it is sufficient to show that $\widehat{L}_T(\boldsymbol{\theta}) \xrightarrow{p} L_{0,T}$ uniformly in $\boldsymbol{\theta}$ (Lemma A.3) and $L_{0,T}(\boldsymbol{\theta})$ attains a unique global maximum at $\boldsymbol{\theta}_0$ (Amemiya, 1985).

Lemma A.3 (Uniform Convergence)**.** *Under the assumptions of Theorem 3.2, we have*

$$\widehat{L}_T(\boldsymbol{\theta}) \xrightarrow{p} L_{0,T}(\boldsymbol{\theta}) \text{ uniformly in } \boldsymbol{\theta}.$$

Proof of Lemma A.3. We have

$$|\widehat{L}_T(\boldsymbol{\theta})-L_{0,T}(\boldsymbol{\theta})| \le |\widehat{L}_T(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})|+|L_T(\boldsymbol{\theta})-L_{0,T}(\boldsymbol{\theta})|.$$

For the simplicity, we omit the parameter $\boldsymbol{\theta}$. Consider $|\widehat{L}_T(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})|$. We have

$$\begin{aligned}\left|\widehat{L}_T(\boldsymbol{\theta})-L_T(\boldsymbol{\theta})\right| &\le \frac{1}{T}\sum_{t=1}^{T}\left(\left|\sum_{i=1}^{r}\log\widehat{h}_{it}-\log h_{it}\right|+\left|\widehat{\mathbf{f}}_t^2-\mathbf{f}_t^2\right|^\top\widehat{\mathbf{h}}_t^{-1}+\mathbf{f}_t^{2\top}\left|\widehat{\mathbf{h}}_t^{-1}-\mathbf{h}_t^{-1}\right|\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\Big((\mathrm{I})+(\mathrm{II})+(\mathrm{III})\Big),\end{aligned}$$

where $\widehat{\mathbf{h}}_t^{-1}$ is the element-wise inverse of $\widehat{\mathbf{h}}_t$. Recall $\mathbf{h}_t(\boldsymbol{\theta}) = (\mathbf{I}_r - \mathbf{B})^{-1}\boldsymbol{\omega} + \sum_{k=0}^{\infty} \mathbf{B}^k \mathbf{A} \mathbf{f}_{t-1-k}^2$. Then by the fact that $\log(1+x) \le x$ for all $x > -1$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}} (\mathrm{I}) &= \sup_{\boldsymbol{\theta}} \left| \sum_{i=1}^{r} \log \left(1 + \frac{\Delta \widehat{h}_{it}}{h_{it}} \right) \right| \\
&\le \sup_{\boldsymbol{\theta}} \sum_{i=1}^{r} \left| \frac{\Delta \widehat{h}_{it}}{h_{it}} \right| \\
&\le \frac{1}{\min_{i \le r} \omega_{\min,i}} \mathbf{1}^\top \sup_{\boldsymbol{\theta}} \left| \sum_{k=0}^{\infty} \mathbf{B}^k \mathbf{A} \big(\Delta \widehat{\mathbf{f}}_{t-1-k}^2 \big) \right| \\
&= o_p(1),
\end{aligned}$$

where $\Delta \widehat{\mathbf{h}}_t = \widehat{\mathbf{h}}_t - \mathbf{h}_t$, $\Delta \widehat{\mathbf{f}}_t^2 = \widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2$, and the last equality is due to Theorem 3.1. Similarly, we can show that (II) and (III) uniformly converges to zero. Therefore, we have

$$\sup_{\boldsymbol{\theta}} |\widehat{L}_T(\boldsymbol{\theta}) - L_T(\boldsymbol{\theta})| = o_p(1).$$

Now we consider $L_T(\boldsymbol{\theta}) - L_{0,T}(\boldsymbol{\theta})$. By Theorem 2.1 in Newey (1991), uniform convergence is equivalent to the pointwise convergence and stochastic equicontinuity, where the stochastic equicontinuity is satisfied by the Lipschitz condition $\big|\partial_\theta \big(L_T(\boldsymbol{\theta}) - L_{0,T}(\boldsymbol{\theta}) \big)\big| \le O_p(1)$. Thus, it is enough to show

1. $L_T(\boldsymbol{\theta}) \xrightarrow{p} L_{0,T}(\boldsymbol{\theta})$ pointwise in $\boldsymbol{\theta}$;

2. $\mathbb{E}\big[\sup_{\boldsymbol{\theta}} \big|\partial_\theta \big(L_T(\boldsymbol{\theta}) - L_{0,T}(\boldsymbol{\theta}) \big)\big| \big] < \infty$ for all $\theta \in \boldsymbol{\theta}$.

Since $\mathbf{f}_t^2 - \mathbf{h}_{0,t}$ is a martingale difference, we can show

$$\mathbb{E}\left[\left(L_T(\boldsymbol{\theta}) - L_{0,T}(\boldsymbol{\theta}) \right)^2 \right] = o(1).$$

Now, we show the Lipschitz condition. We have

$$\mathbb{E}\big[\sup_{\boldsymbol{\theta}} \big|\partial_\theta \big(L_T(\boldsymbol{\theta}) - L_{0,T}(\boldsymbol{\theta}) \big)\big| \big] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}} \left| \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{r} -h_{it}^{-2} (f_{it}^2 - h_{0,it}) \partial_\theta h_{it} \right| \right]$$

$$
\begin{aligned}
&\leq C\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\mathbb{E}\left[\sup_{\boldsymbol{\theta}}(\partial_\theta h_{it})\left(f_{it}^2+h_{0,it}\right)\right] \\
&= \frac{C}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\mathbb{E}\left[2\sup_{\boldsymbol{\theta}}(\partial_\theta h_{it})\left(h_{0,it}\right)\right] \\
&\leq \frac{C}{T}\sum_{t=1}^{T}\sum_{i=1}^{r}\mathbb{E}\left[h_{0,it}^2\right]^{\frac{1}{2}}\mathbb{E}\left[\sup_{\boldsymbol{\theta}}(\partial_\theta h_{it})^2\right]^{\frac{1}{2}} \\
&< \infty,
\end{aligned}
$$

where the first inequality is due to $\partial_\theta h_{it} > 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and the second inequality is due to Hölder's inequality. □

Lemma A.4 (Uniqueness of $\boldsymbol{\theta}$)**.** *Under the assumptions of Theorem 3.2,*

$$
\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} L_{0,T}(\boldsymbol{\theta})
$$

is unique almost surely, and $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$. *Moreover,*

$$
\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0. \tag{A.2}
$$

Proof of Lemma A.4. Since $\log x + t/x$ has a unique minimizer at $x = t$, $\mathbf{h}_{it}(\boldsymbol{\theta}_0)$ is a unique maximizer of $l_{0,t}(\mathbf{h}_{it})$ for all t. If $\boldsymbol{\theta}_0$ is not a unique parameter to have $\mathbf{h}_{it}(\boldsymbol{\theta}_0)$, then there exists $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$ such that $\mathbf{h}_t(\boldsymbol{\theta}^*) = \mathbf{h}_t(\boldsymbol{\theta}_0)$. Then $\{\mathbf{h}_t(\boldsymbol{\theta}^*) - \mathbf{h}_t(\boldsymbol{\theta}_0)\}_{t\leq T} = \{\mathbf{0}\}_{t\leq T}$, and

$$
\begin{aligned}
&\Big(\mathbf{h}_2(\boldsymbol{\theta}^*) - \mathbf{h}_2(\boldsymbol{\theta}_0), \ldots, \mathbf{h}_T(\boldsymbol{\theta}^*) - \mathbf{h}_T(\boldsymbol{\theta}_0)\Big) \\
&= \Big(\boldsymbol{\omega}^* - \boldsymbol{\omega}_0 \quad \mathbf{A}^* - \mathbf{A}_0 \quad \mathbf{B}^* - \mathbf{B}_0\Big)\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{f}_1^2 & \mathbf{f}_2^2 & \cdots & \mathbf{f}_{T-1}^2 \\ \mathbf{h}_1(\boldsymbol{\theta}_0) & \mathbf{h}_2(\boldsymbol{\theta}_0) & \cdots & \mathbf{h}_{T-1}(\boldsymbol{\theta}_0) \end{pmatrix} \\
&= \Big(\boldsymbol{\omega}^* - \boldsymbol{\omega}_0 \quad \mathbf{A}^* - \mathbf{A}_0 \quad \mathbf{B}^* - \mathbf{B}_0\Big)\mathbf{M} = \mathbf{0},
\end{aligned}
$$

where $\mathbf{0}$ is a zero matrix. By Assumption 2(b), $\mathbf{M}\mathbf{M}^\top$ is invertible a.s., which implies

$$\begin{pmatrix}\boldsymbol{\omega}^* - \boldsymbol{\omega}_0 & \mathbf{A}^* - \mathbf{A}_0 & \mathbf{B}^* - \mathbf{B}_0\end{pmatrix} = \mathbf{0}.$$

Therefore, $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ a.s. which is a contradiction.

With the uniqueness solution result, by Lemma A.3 and Theorem 4.1.2 in Amemiya (1985), we can show (A.2). □

Lemma A.5. *Under the assumptions of Theorem 3.2, for all fixed $c \geq 1$, $l \leq r$ and $\theta \in \boldsymbol{\theta}$,*

$$\mathbb{E}\left[\left|\sup_{\boldsymbol{\theta}} \frac{\partial_\theta \widehat{h}_{lt}}{\widehat{h}_{lt}}\right|^c\right] < \infty,\ \mathbb{E}\left[\left|\sup_{\boldsymbol{\theta}} \frac{\partial_\theta^2 \widehat{h}_{lt}}{\widehat{h}_{lt}}\right|^c\right] < \infty,\ \mathbb{E}\left[\left|\sup_{\boldsymbol{\theta}} \frac{\partial_\theta^3 \widehat{h}_{lt}}{\widehat{h}_{lt}}\right|^c\right] < \infty.$$

Proof of Lemma A.5. Notice that

$$\partial_\theta \widehat{\mathbf{h}}_t = \partial_\theta \boldsymbol{\varpi} + \sum_{k=0}^{\infty} \mathbf{B}^k (\partial_\theta \mathbf{A}) \widehat{\mathbf{f}}_{t-1-k}^2 + \sum_{k=1}^{\infty} \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi (\partial_\theta \mathbf{B}) \mathbf{B}^{k-1-\xi} \right) \mathbf{A} \widehat{\mathbf{f}}_{t-1-k}^2 \geq \mathbf{0}.$$

where $\boldsymbol{\varpi} = (\mathbf{I}_r - \mathbf{B})^{-1}\boldsymbol{\omega}$. For the case of $\theta = \varpi_i$, $\mathbb{E}\big[\big|\sup_{\boldsymbol{\theta}} \partial_\theta \widehat{h}_{lt}/\widehat{h}_{lt}\big|^c\big] < \infty$ is trivial. With the case $\theta = A_{ij}$,

$$\begin{aligned}
\frac{\partial_\theta \widehat{h}_{lt}}{\widehat{h}_{lt}} &\leq \frac{\mathbf{e}_l^\top \sum_{k=0}^\infty \mathbf{B}^k (\partial_\theta \mathbf{A}) \widehat{\mathbf{f}}_{t-1-k}^2}{\mathbf{e}_l^\top \sum_{k=0}^\infty \mathbf{B}^k \mathbf{A} \widehat{\mathbf{f}}_{t-1-k}^2} \\
&\leq \frac{\mathbf{e}_l^\top \sum_{k=0}^\infty \mathbf{B}^k \mathbf{e}_i \mathbf{e}_j^\top \widehat{\mathbf{f}}_{t-1-k}^2}{\mathbf{e}_l^\top \sum_{k=0}^\infty \mathbf{B}^k \mathbf{e}_i A_{ij} \mathbf{e}_j^\top \widehat{\mathbf{f}}_{t-1-k}^2} \\
&= \frac{1}{A_{ij}} \leq \frac{1}{A_{\min,ij}},
\end{aligned}$$

where $\mathbf{e}_i$ is the i^{th} standard basis vector. Thus, $\mathbb{E}\big[\big|\sup_{\boldsymbol{\theta}} \partial_\theta \widehat{h}_{lt}/\widehat{h}_{lt}\big|^c\big] < \infty$. For the case of $\theta = B_{ij}$,

$$\frac{\partial_\theta \widehat{h}_{lt}}{\widehat{h}_{lt}} = \frac{\mathbf{e}_l^\top \sum_{k=1}^\infty \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi (\partial_\theta \mathbf{B}) \mathbf{B}^{k-1-\xi} \right) \mathbf{A} \widehat{\mathbf{f}}_{t-1-k}^2}{\varpi_l + \mathbf{e}_l^\top \sum_{k=0}^\infty \mathbf{B}^k \mathbf{A} \widehat{\mathbf{f}}_{t-1-k}^2}$$

$$
\begin{aligned}
&\le \sum_{k=1}^{\infty} \frac{\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2}{\varpi_l + \mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i B_{ij} \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2} \\
&= \sum_{k=0}^{\infty} \frac{\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2}{\varpi_l + B_{ij}\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2} \\
&= \frac{1}{B_{ij}} \sum_{k=0}^{\infty} \frac{(B_{ij}/\varpi_l)\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2}{1 + (B_{ij}/\varpi_l)\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2} \\
&\le \frac{1}{B_{ij}} \frac{B_{ij}^\alpha}{\varpi_l^\alpha} \sum_{k=0}^{\infty} \left[\mathbf{e}_l^\top \left(\sum_{\xi=0}^{k-1} \mathbf{B}^\xi \mathbf{e}_i \mathbf{e}_j^\top \mathbf{B}^{k-1-\xi}\right) \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2\right]^\alpha \\
&\le \frac{1}{B_{ij}} \frac{1}{\varpi_l^\alpha} \sum_{k=0}^{\infty} \left[\mathbf{e}_l^\top \mathbf{B}^k \mathbf{A}\widehat{\mathbf{f}}_{t-1-k}^2\right]^\alpha,
\end{aligned}
$$

where the second inequality is due to $x/(1+x) \le x^\alpha$ for $\forall \alpha \in [0,1]$. By choosing $\alpha = 1/c$,

$$
\begin{aligned}
\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}} \left|\frac{\partial_\theta \widehat{h}_{lt}}{\widehat{h}_{lt}}\right|^c\right]\right)^{\frac{1}{c}} &\le C \sum_{k=0}^{\infty} \|\mathbf{B}_{\max}\|^k \mathbb{E}\left[\|\widehat{\mathbf{f}}_{t-1-k}^2\|\right] \\
&< \infty,
\end{aligned}
$$

where first inequality is due to the Minkowski's inequality, and the last inequality follows from $\mathbb{E}\left[|f_{it}|^8\right] \le C$ and $\mathbb{E}\left\|\widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2\right\|^2 = o(1)$. Similarly, we can show the higher order derivatives bounds. □

Lemma A.6. *Under the assumptions of Theorem 3.2, let* $\mathbb{E}\left\|\Delta \mathbf{f}_t^2\right\|^2 = O(\beta_T^2)$. *Then we have*

$$
\left|\partial_\theta \widehat{L}_T(\boldsymbol{\theta}_0) - \partial_\theta L_{0,T}(\boldsymbol{\theta}_0)\right| = O_p\left(\beta_T + \frac{1}{\sqrt{T}}\right).
$$

Proof of Lemma A.6. Denote $\widehat{\mathbf{h}}_{0,t} = \widehat{\mathbf{h}}_t(\boldsymbol{\theta}_0)$ and $\partial_\theta \widehat{\mathbf{h}}_{0,t} = \partial_\theta \widehat{\mathbf{h}}_t(\boldsymbol{\theta}_0)$. Since $\mathbb{E}\left\|\Delta \mathbf{f}_t^2\right\|^2 = O(\beta_T^2)$, we have

1. $\mathbb{E}\left\|\widehat{\mathbf{h}}_{0,t} - \mathbf{h}_{0,t}\right\|^2 = O(\beta_T^2)$;
2. $\mathbb{E}\left\|\widehat{\mathbf{h}}_{0,t}^{-1} - \mathbf{h}_{0,t}^{-1}\right\|^2 = O(\beta_T^2)$;

3. $\mathbb{E}\left\|\partial_\theta \widehat{\mathbf{h}}_{0,t} - \partial_\theta \mathbf{h}_{0,t}\right\|^2 = O(\beta_T^2)$;

4. $\mathbb{E}\left[\mathbf{f}_t^{2\top}(\widehat{\mathbf{h}}_{0,t}^{-2} - \mathbf{h}_{0,t}^{-2}) \circ \partial_\theta \widehat{\mathbf{h}}_{0,t}\right] = O(\beta_T)$,

where the last equation is by the fact that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{f}_t^{2\top}(\widehat{\mathbf{h}}_{0,t}^{-2} - \mathbf{h}_{0,t}^{-2}) \circ \partial_\theta \widehat{\mathbf{h}}_{0,t}\right] \\
&= \mathbb{E}\left[\sum_{l=1}^{r}\left(\frac{h_{0,lt}^2 - \widehat{h}_{0,lt}^2}{h_{0,lt}^2 \widehat{h}_{0,lt}^2}\right) f_{lt}^2 \partial_\theta \widehat{h}_{0,lt}\right] \\
&= \mathbb{E}\left[\sum_{l=1}^{r}\left(\frac{h_{0,lt}^2 - \widehat{h}_{0,lt}^2}{h_{0,lt} \widehat{h}_{0,lt}}\right) \epsilon_{lt}^2 \frac{\partial_\theta \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}}\right] \\
&= \mathbb{E}\left[\sum_{l=1}^{r}\left(\frac{h_{0,lt} - \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}}\right) \epsilon_{lt}^2 \frac{\partial_\theta \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}} + \left(\frac{h_{0,lt} - \widehat{h}_{0,lt}}{h_{0,lt}}\right) \epsilon_{lt}^2 \frac{\partial_\theta \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}}\right] \\
&\leq C\mathbb{E}\left[\sum_{l=1}^{r}\left|h_{0,lt} - \widehat{h}_{0,lt}\right| \epsilon_{lt}^2 \frac{\partial_\theta \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}}\right] \\
&\leq C\sum_{l=1}^{r}\left(\mathbb{E}\left[\left(h_{0,lt} - \widehat{h}_{0,lt}\right)^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\left(\frac{\partial_\theta \widehat{h}_{0,lt}}{\widehat{h}_{0,lt}}\right)^2\right]\right)^{\frac{1}{2}} \\
&= O(\beta_T),
\end{aligned}
$$

where the last equality is due to Lemma A.5. Thus, we have

$$
\begin{aligned}
&\partial_\theta \widehat{L}_T(\boldsymbol{\theta}_0) - \partial_\theta L_T(\boldsymbol{\theta}_0) \\
&= -\frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{h}}_{0,t} - \widehat{\mathbf{f}}_t^2)^\top(\widehat{\mathbf{h}}_{0,t}^{-2} \circ \partial_\theta \widehat{\mathbf{h}}_{0,t}) - (\mathbf{h}_{0,t} - \mathbf{f}_t^2)^\top(\mathbf{h}_{0,t}^{-2} \circ \partial_\theta \mathbf{h}_{0,t}) \\
&= \frac{1}{T}\sum_{t=1}^{T}\Big\{(\widehat{\mathbf{f}}_t^2 - \mathbf{f}_t^2)^\top(\widehat{\mathbf{h}}_{0,t}^{-2} \circ \partial_\theta \widehat{\mathbf{h}}_{0,t}) + \mathbf{f}_t^{2\top}(\widehat{\mathbf{h}}_{0,t}^{-2} - \mathbf{h}_{0,t}^{-2}) \circ \partial_\theta \widehat{\mathbf{h}}_{0,t} \\
&\qquad + \mathbf{f}_t^{2\top}\mathbf{h}_{0,t}^{-2} \circ (\partial_\theta \widehat{\mathbf{h}}_{0,t} - \partial_\theta \mathbf{h}_{0,t}) - (\widehat{\mathbf{h}}_{0,t}^{-1} - \mathbf{h}_{0,t}^{-1})^\top \partial_\theta \widehat{\mathbf{h}}_{0,t} \\
&\qquad - \mathbf{h}_{0,t}^{-1\top}(\partial_\theta \widehat{\mathbf{h}}_{0,t} - \partial_\theta \mathbf{h}_{0,t})\Big\}
\end{aligned}
$$

and

$$\mathbb{E}\left|\partial_\theta \widehat{L}_T(\boldsymbol{\theta}_0) - \partial_\theta L_T(\boldsymbol{\theta}_0)\right| \leq \frac{C}{T}\sum_{t=1}^{T}\{\beta_T + 4\beta_T + \beta_T + \beta_T + \beta_T\}$$
$$= O(\beta_T). \tag{A.3}$$

Moreover, we have

$$\begin{aligned}\partial_\theta L_T(\boldsymbol{\theta}_0) - \partial_\theta L_{0,T}(\boldsymbol{\theta}_0) &= \frac{1}{T}\sum_{t=1}^{T}(\mathbf{f}_t^2 - \mathbf{h}_{0,t})^\top(\mathbf{h}_{0,t}^{-2} \circ \partial_\theta \mathbf{h}_{0,t}) \\ &= \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t^2 - \mathbf{1})^\top(\mathbf{h}_{0,t}^{-1} \circ \partial_\theta \mathbf{h}_{0,t})\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[(\partial_\theta L_T(\boldsymbol{\theta}_0) - \partial_\theta L_{0,T}(\boldsymbol{\theta}_0))^2] \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\operatorname{tr}\left((\mathbb{E}[\boldsymbol{\varepsilon}_t^2\boldsymbol{\varepsilon}_t^{2\top}] - \mathbf{1}\mathbf{1}^\top)\mathbb{E}\left[(\mathbf{h}_{0,t}^{-1} \circ \partial_\theta \mathbf{h}_{0,t})(\mathbf{h}_{0,t}^{-1} \circ \partial_\theta \mathbf{h}_{0,t})^\top\right]\right) \\ &= O\left(\frac{1}{T}\right),\end{aligned} \tag{A.4}$$

where the last equality is due to Lemma A.5. By combining (A.3) and (A.4), we complete the proof. □

Lemma A.7. *Under the assumptions of Theorem 3.2, we have*

$$\partial_{\boldsymbol{\theta}}^2 \widehat{L}_T(\boldsymbol{\theta}^*) \xrightarrow{p} \partial_{\boldsymbol{\theta}}^2 L_{0,T}(\boldsymbol{\theta}_0),$$

and $-\partial_{\boldsymbol{\theta}}^2 L_{0,T}(\boldsymbol{\theta}_0)$ *is a positive definite matrix.*

Proof of Lemma A.7. We have

$$\partial_{\boldsymbol{\theta}}^2 \widehat{L}_T(\boldsymbol{\theta}^*) - \partial_{\boldsymbol{\theta}}^2 L_{0,T}(\boldsymbol{\theta}_0) = \left(\partial_{\boldsymbol{\theta}}^2 \widehat{L}_T(\boldsymbol{\theta}^*) - \partial_{\boldsymbol{\theta}}^2 \widehat{L}_T(\boldsymbol{\theta}_0)\right) + \left(\partial_{\boldsymbol{\theta}}^2 \widehat{L}_T(\boldsymbol{\theta}_0) - \partial_{\boldsymbol{\theta}}^2 L_{0,T}(\boldsymbol{\theta}_0)\right)$$

$$= \text{(I)} + \text{(II)}.$$

By the mean value theorem and (A.2), we have (I) $= \partial^3_{\boldsymbol\theta} \widehat{L}_T(\boldsymbol\theta^{**})(\boldsymbol\theta^* - \boldsymbol\theta_0)$ and $\boldsymbol\theta^* - \boldsymbol\theta_0 \le \left|\widehat{\boldsymbol\theta} - \boldsymbol\theta_0\right| \xrightarrow{p} 0$, respectively. Thus, it is enough to show $\partial^3_{\boldsymbol\theta} \widehat{L}_T(\boldsymbol\theta^{**}) = O_p(1)$. Denote $\partial^3_{ijk} = \partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_j}$, then for all $i, j, k \le \dim(\boldsymbol\theta)$,

$$\begin{aligned}
&\partial^3_{ijk}\widehat{L}_T(\boldsymbol\theta) \\
&= \frac{1}{T}\sum_{t=1}^{T}\Bigg\{ \left(\widehat{\mathbf{h}}_t - \widehat{\mathbf{f}}_t^2\right)^\top \widehat{\mathbf{h}}_t^{\circ -2} \circ \partial^3_{ijk}\widehat{\mathbf{h}}_t \\
&\qquad + \left(2\widehat{\mathbf{f}}_t^2 - \widehat{\mathbf{h}}_t\right)^\top \widehat{\mathbf{h}}_t^{-3} \circ \left(\partial_i\widehat{\mathbf{h}}_t \circ \partial_{jk}\widehat{\mathbf{h}}_t + \partial_j\widehat{\mathbf{h}}_t \circ \partial_{ki}\widehat{\mathbf{h}}_t + \partial_k\widehat{\mathbf{h}}_t \circ \partial_{ij}\widehat{\mathbf{h}}_t\right) \\
&\qquad + \left(2\widehat{\mathbf{h}}_t - 6\widehat{\mathbf{f}}_t^2\right)^\top \widehat{\mathbf{h}}_t^{-4} \circ \partial_i\widehat{\mathbf{h}}_t \circ \partial_j\widehat{\mathbf{h}}_t \circ \partial_k\widehat{\mathbf{h}}_t \Bigg\}.
\end{aligned}$$

By $\mathbb{E}\left[\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t^\top\right] < \infty$ and Lemma A.5, we can show $\mathbb{E}\left|\partial^3_{ijk}\widehat{L}_T(\theta^{**})\right| < \infty$. Thus, (I) $\xrightarrow{p} 0$. Similar to the proof of Lemma A.6, we can show

$$\begin{aligned}
\text{(II)} &= \partial^2_{ij}\widehat{L}_T(\boldsymbol\theta_0) - \partial^2_{ij}L_{0,T}(\boldsymbol\theta_0) \\
&= \partial^2_{ij}\widehat{L}_T(\boldsymbol\theta_0) - \partial^2_{ij}L_T(\boldsymbol\theta_0) + \partial^2_{ij}L_T(\boldsymbol\theta_0) - \partial^2_{ij}L_{0,T}(\boldsymbol\theta_0) \\
&= O_p(\beta_T) + O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Moreover, since $\mathbf{h}_t^{-2} > \mathbf{0}$ and $\mathbf{f}_t^2$'s are non-degenerate, we can show

$$-\partial^2_{\boldsymbol\theta}L_{0,T}(\boldsymbol\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{r} h_{0,it}^{-2}\left(\partial_{\boldsymbol\theta}h_{0,it}\;\partial_{\boldsymbol\theta}h_{0,it}{}^\top\right) \succ 0.$$

□

Proof of Theorem 3.2. By the mean value theorem, there is $\boldsymbol\theta^*$ between $\widehat{\boldsymbol\theta}$ and $\boldsymbol\theta_0$ such that

$$\widehat{\boldsymbol\theta} - \boldsymbol\theta_0 = \partial^2_{\boldsymbol\theta}\widehat{L}_T(\boldsymbol\theta^*)^{-1}\left(\partial_{\boldsymbol\theta}\widehat{L}_T(\widehat{\boldsymbol\theta}) - \partial_{\boldsymbol\theta}\widehat{L}_T(\boldsymbol\theta_0)\right)$$

$$= \partial^2_{\boldsymbol{\theta}} \widehat{L}_T(\boldsymbol{\theta}^*)^{-1} \left(\partial_{\boldsymbol{\theta}} L_{0,T}(\boldsymbol{\theta}_0) - \partial_{\boldsymbol{\theta}} \widehat{L}_T(\boldsymbol{\theta}_0) \right),$$

where the last equality is by the fact that $\partial_{\boldsymbol{\theta}} \widehat{L}_T(\widehat{\boldsymbol{\theta}}) = \partial_{\boldsymbol{\theta}} L_{0,T}(\boldsymbol{\theta}_0) = 0$. Then, by Lemmas A.6 and A.7, we have

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p\left(\beta_T + \frac{1}{\sqrt{T}} \right),$$

and thus the result of Theorem 3.1 completes the proof. □

A.3 Proof of Theorem 4.1

Lemma A.8. *Under the assumptions of Theorem 4.1, we have*

$$\left\| \widehat{\mathbf{V}} \widehat{\boldsymbol{\Sigma}}_{f,t+1} \widehat{\mathbf{V}}^\top - \mathbf{V} \boldsymbol{\Sigma}_{f,t+1} \mathbf{V}^\top \right\|_{\boldsymbol{\Sigma}_{t+1}} = O_p\left(\frac{1}{\sqrt{T}} + \frac{\sqrt{p}}{T} + \frac{s_p^2}{p\sqrt{p}} + \frac{s_p}{p} \right). \tag{A.5}$$

Proof of Lemma A.8. Without loss of generality, we assume that $\mathbf{O}$ is the identity matrix. Similar to the proofs of Theorem 4.2 in Kim et al. (2018), we have

$$\begin{aligned}
&\left\| \widehat{\mathbf{V}} \widehat{\boldsymbol{\Sigma}}_{f,t+1} \widehat{\mathbf{V}}^\top - \mathbf{V} \boldsymbol{\Sigma}_{f,t+1} \mathbf{V}^\top \right\|_{\boldsymbol{\Sigma}_{t+1}} \\
&\le \left\| \mathbf{V} \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}} \left(\boldsymbol{\Sigma}_{f,t+1}^{-\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}_{f,t+1} \boldsymbol{\Sigma}_{f,t+1}^{-\frac{1}{2}} - \mathbf{I}_r \right) \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \right\|_{\boldsymbol{\Sigma}_{t+1}} \\
&\quad + \left\| (\widehat{\mathbf{V}} - \mathbf{V}) \widehat{\boldsymbol{\Sigma}}_{f,t+1} (\widehat{\mathbf{V}} - \mathbf{V})^\top \right\|_{\boldsymbol{\Sigma}_{t+1}} + 2 \left\| \mathbf{V} \widehat{\boldsymbol{\Sigma}}_{f,t+1} (\widehat{\mathbf{V}} - \mathbf{V})^\top \right\|_{\boldsymbol{\Sigma}_{t+1}} \\
&= \text{(I)} + \text{(II)} + \text{(III)}.
\end{aligned}$$

First consider (I). By the matrix inversion lemma, we have

$$(\mathbf{V} \boldsymbol{\Sigma}_{f,t+1} \mathbf{V}^\top + \boldsymbol{\Sigma}_u)^{-1} = \boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1} \mathbf{V} \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}} (\mathbf{I}_r + \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \boldsymbol{\Sigma}_u^{-1} \mathbf{V} \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}})^{-1} \boldsymbol{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \boldsymbol{\Sigma}_u^{-1}$$

and by denoting $\mathbf{X} = \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \mathbf{\Sigma}_u^{-1} \mathbf{V} \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}}$,

$$
\begin{aligned}
\left\| \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \mathbf{\Sigma}_{t+1}^{-1} \mathbf{V} \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \right\| &= \left\| \mathbf{X} - \mathbf{X}(\mathbf{I}_r + \mathbf{X})^{-1} \mathbf{X} \right\| \\
&= \left\| \mathbf{X}(\mathbf{I}_r + \mathbf{X})^{-1} \right\| \\
&= \left\| \mathbf{I}_r - (\mathbf{I}_r + \mathbf{X})^{-1} \right\| \\
&\leq 2.
\end{aligned}
$$

Then, since $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\| \|\mathbf{B}\|_F$ and $\left\| \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \right\| \leq \sqrt{2}$, we have

$$
\begin{aligned}
(\mathrm{I}) &= \frac{1}{\sqrt{p}} \left\| \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \mathbf{V} \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \left(\mathbf{\Sigma}_{f,t+1}^{-\frac{1}{2}} \widehat{\mathbf{\Sigma}}_{f,t+1} \mathbf{\Sigma}_{f,t+1}^{-\frac{1}{2}} - \mathbf{I}_r \right) \mathbf{\Sigma}_{f,t+1}^{\frac{1}{2}} \mathbf{V}^\top \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \right\|_F \\
&\leq \frac{2}{\sqrt{p}} \left\| \mathbf{\Sigma}_{f,t+1}^{-\frac{1}{2}} \widehat{\mathbf{\Sigma}}_{f,t+1} \mathbf{\Sigma}_{f,t+1}^{-\frac{1}{2}} - \mathbf{I}_r \right\|_F \\
&= \frac{2}{\sqrt{p}} \left(\sum_{i=1} \frac{\left| \widehat{h}_{it+1}(\widehat{\boldsymbol{\theta}}) - h_{0,it+1} \right|^2}{h_{0,it+1}^2} \right)^{\frac{1}{2}} \\
&= O_p \left(\frac{1}{\sqrt{pT}} + \frac{\sqrt{s_p}}{p} \right),
\end{aligned}
$$

where the last equality is by the fact that

$$
\begin{aligned}
\widehat{h}_{it+1}(\widehat{\boldsymbol{\theta}}) - h_{0,it+1} &= \widehat{h}_{it+1}(\widehat{\boldsymbol{\theta}}) - \widehat{h}_{it+1}(\boldsymbol{\theta}_0) + \widehat{h}_{it+1}(\boldsymbol{\theta}_0) - h_{0,it+1} \\
&= O_p \left(\frac{1}{\sqrt{T}} + \sqrt{\frac{s_p}{p}} \right).
\end{aligned}
$$

For (II) and (III), similarly, by Theorem 3.1, we can show

$$
\begin{aligned}
(\mathrm{II}) &= \frac{1}{\sqrt{p}} \left\| \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} (\widehat{\mathbf{V}} - \mathbf{V}) \widehat{\mathbf{\Sigma}}_{f,t+1} (\widehat{\mathbf{V}} - \mathbf{V})^\top \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \right\|_F \\
&\leq \frac{1}{\sqrt{p}} \left\| \mathbf{\Sigma}_{t+1}^{-1} \right\| \left\| \widehat{\mathbf{\Sigma}}_{f,t+1} \right\|_F \left\| \widehat{\mathbf{V}} - \mathbf{V} \right\|^2 \\
&= O_p \left(\frac{\sqrt{p}}{T} + \frac{s_p^2}{p\sqrt{p}} \right)
\end{aligned}
$$

and

$$
\begin{aligned}
\text{(III)} &= \frac{1}{\sqrt{p}} \left\| \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \mathbf{V} \widehat{\mathbf{\Sigma}}_{f,t+1} (\widehat{\mathbf{V}} - \mathbf{V})^\top \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \right\|_F \\
&= \frac{\sqrt{2}}{\sqrt{p}} \left\| \widehat{\mathbf{\Sigma}}_{f,t+1}^{\frac{1}{2}} (\widehat{\mathbf{V}} - \mathbf{V})^\top \mathbf{\Sigma}_{t+1}^{-\frac{1}{2}} \right\|_F \\
&\leq \frac{C}{\sqrt{p}} \left\| \widehat{\mathbf{\Sigma}}_{f,t+1}^{\frac{1}{2}} \right\|_F \left\| \widehat{\mathbf{V}} - \mathbf{V} \right\| \\
&= O_p \left(\frac{1}{\sqrt{T}} + \frac{s_p}{p} \right).
\end{aligned}
$$

□

Lemma A.9. *Under the assumptions of Theorem 4.1, suppose that* $\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|_{\max} = O_p(\gamma_T)$, *and the thresholding level satisfies the condition* $\tau_T = C\gamma_T$ *such that* $|\widehat{\Sigma}_{u,ij} - \Sigma_{u,ij}| < \tau_T(\widehat{\Sigma}_{u,ii}\widehat{\Sigma}_{u,jj})^{1/2}/2$. *Then we have*

$$
\left\| \mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u \right\| = O_p \left(s_p \gamma_T^{1-q} \right).
$$

Proof of Lemma A.9. Let $\tau_{ij} = \tau_T(\widehat{\Sigma}_{u,ii}\widehat{\Sigma}_{u,jj})^{1/2}$. Similar to the proofs of Section 3 in Fan and Kim (2018), under the event $|\widehat{\Sigma}_{u,ij} - \Sigma_{u,ij}| < \tau_{ij}/2$,

$$
\begin{aligned}
\left\| \mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u \right\| &\leq \left\| \mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u \right\|_\infty \\
&= \max_i \sum_{j=1}^p \left| s_{ij}(\widehat{\Sigma}_{u,ij}) \mathbb{1}_{|\widehat{\Sigma}_{u,ij}| \geq \tau_{ij}} - \Sigma_{u,ij} \mathbb{1}_{|\Sigma_{u,ij}| \geq \tau_{ij}} - \Sigma_{u,ij} \mathbb{1}_{|\Sigma_{u,ij}| < \tau_{ij}} \right| \\
&\leq \max_i \sum_{j=1}^p \Bigg\{ \left| s_{ij}(\widehat{\Sigma}_{u,ij}) - \Sigma_{u,ij} \right| \mathbb{1}_{|\widehat{\Sigma}_{u,ij}| \geq \tau_{ij}} + |\Sigma_{u,ij}| \left| \mathbb{1}_{|\widehat{\Sigma}_{u,ij}| \geq \tau_{ij}} - \mathbb{1}_{|\Sigma_{u,ij}| \geq \tau_{ij}} \right| \\
&\qquad + |\Sigma_{u,ij}| \, \mathbb{1}_{|\Sigma_{u,ij}| < \tau_{ij}} \Bigg\} \\
&\leq \max_i \sum_{j=1}^p \left\{ \frac{3}{2} \tau_{ij} \mathbb{1}_{|\Sigma_{u,ij}| \geq \frac{1}{2}\tau_{ij}} + |\Sigma_{u,ij}| \, \mathbb{1}_{|\Sigma_{u,ij}| \leq \frac{3}{2}\tau_{ij}} + |\Sigma_{u,ij}|^q \, \tau_{ij}^{1-q} \right\} \\
&\leq C \max_i \sum_{j=1}^p |\Sigma_{u,ij}|^q \tau_{ij}^{1-q}
\end{aligned}
$$

$$= O\left(s_p \gamma_T^{1-q}\right),$$

where the last equality is due to the sparse condition. □

Proof of Theorem 4.1. By Assumption 3(a), we have

$$\left\| \widehat{\overline{\boldsymbol{\Sigma}}} - \overline{\boldsymbol{\Sigma}} \right\|_{\max} = O_p\left(\sqrt{\frac{\log p}{T}}\right). \tag{A.6}$$

We denote $\widehat{\mathbf{v}}_i$ and $\mathbf{v}_i$ the i^{th} column vectors of $\widehat{\mathbf{V}}$ and $\mathbf{V}$, respectively. Then by Theorem 2.1 in Fan et al. (2017),

$$\begin{aligned}
\sum_{i=1}^{r} \|\widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^\top - \mathbf{v}_i \mathbf{v}_i^\top\|_{\max} &\le \sum_{i=1}^{r} \left\{ \|(\widehat{\mathbf{v}}_i - \mathbf{v}_i)\widehat{\mathbf{v}}_i^\top\|_{\max} + \|\mathbf{v}_i(\widehat{\mathbf{v}}_i - \mathbf{v}_i)^\top\|_{\max} \right\} \\
&\le \sum_{i=1}^{r} \left\{ \|\widehat{\mathbf{v}}_i - \mathbf{v}_i\|_\infty \|\widehat{\mathbf{v}}_i - \mathbf{v}_i + \mathbf{v}_i\|_\infty + \|\widehat{\mathbf{v}}_i - \mathbf{v}_i\|_\infty \|\mathbf{v}_i\|_\infty \right\} \\
&\le \sum_{i=1}^{r} \left\{ \|\widehat{\mathbf{v}}_i - \mathbf{v}_i\|_\infty^2 + 2\|\widehat{\mathbf{v}}_i - \mathbf{v}_i\|_\infty \|\mathbf{v}_i\|_\infty \right\} \\
&\le C\left(\frac{\left\|\widehat{\overline{\boldsymbol{\Sigma}}} - \overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_u\right\|_\infty^2}{p^2\delta_r^2} + \frac{\left\|\widehat{\overline{\boldsymbol{\Sigma}}} - \overline{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}_u\right\|_\infty}{p\delta_r} \right) \\
&= O_p\left(\sqrt{\frac{\log p}{T}} + \frac{s_p}{p}\right),
\end{aligned}$$

where the last equality is due to $\|\mathbf{A}\|_\infty \le p\|\mathbf{A}\|_{\max}$. Thus,

$$\begin{aligned}
&\left\|\widehat{\mathbf{V}}\widehat{\boldsymbol{\Sigma}}_{f,t+1}\widehat{\mathbf{V}}^\top - \mathbf{V}\boldsymbol{\Sigma}_{f,t+1}\mathbf{V}^\top\right\|_{\max} \\
&\le \sum_{i=1}^{r} \left\{ \left\|\widehat{h}_{it+1}\left(\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^\top - \mathbf{v}_i\mathbf{v}_i^\top\right)\right\|_{\max} + \left\|(\widehat{h}_{it+1} - h_{0,it+1})\mathbf{v}_i\mathbf{v}_i^\top\right\|_{\max} \right\} \\
&\le \left\|\widehat{\mathbf{h}}_{t+1} - \mathbf{h}_{0,t+1} + \mathbf{h}_{0,t+1}\right\|_\infty \sum_{i=1}^{r} \|\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^\top - \mathbf{v}_i\mathbf{v}_i^\top\|_{\max} + \left\|\widehat{\mathbf{h}}_{t+1} - \mathbf{h}_{0,t+1}\right\|_\infty \sum_{i=1}^{r} \|\mathbf{v}_i\mathbf{v}_i^\top\|_{\max} \\
&\le O_p\left(\sqrt{\frac{\log p}{T}} + \frac{s_p}{p}\right) + O_p\left(\frac{1}{\sqrt{T}} + \sqrt{\frac{s_p}{p}}\right).
\end{aligned} \tag{A.7}$$

By combining (A.6) and (A.7), we have

$$\left\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\right\|_{\max} \leq \left\|\widehat{\overline{\mathbf{\Sigma}}} - \overline{\mathbf{\Sigma}}\right\|_{\max} + \left\|\widehat{\mathbf{V}}\widehat{\mathbf{\Sigma}}_{f,t+1}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{\Sigma}_{f,t+1}\mathbf{V}^\top\right\|_{\max}$$
$$= O_p\left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}}\right), \tag{A.8}$$

which leads to $\gamma_T = C\big(\sqrt{\log p/T} + \sqrt{s_p/p}\big)$ and $\|\widehat{\mathbf{\Sigma}}_{t+1} - \mathbf{\Sigma}_{t+1}\|_{\max} = O_p(\gamma_T)$. Combining Lemmas A.8 and A.9, and (A.8), we have

$$\left\|\widehat{\mathbf{\Sigma}}_{t+1} - \mathbf{\Sigma}_{t+1}\right\|_{\mathbf{\Sigma}_{t+1}} \leq \left\|\widehat{\mathbf{V}}\widehat{\mathbf{\Sigma}}_{f,t+1}\widehat{\mathbf{V}}^\top - \mathbf{V}\mathbf{\Sigma}_{f,t+1}\mathbf{V}^\top\right\|_{\mathbf{\Sigma}_{t+1}} + \left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u\right\|_{\mathbf{\Sigma}_{t+1}}$$
$$= O_p\left(\frac{\sqrt{p}}{T} + s_p\gamma_T^{1-q}\right),$$

where the last equality is by the fact

$$\left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u\right\|_{\mathbf{\Sigma}_{t+1}} \leq \frac{1}{\sqrt{p}}\left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u\right\| \left\|\mathbf{\Sigma}_{t+1}^{-1}\right\|_F$$
$$\leq \left\|\mathcal{T}(\widehat{\mathbf{\Sigma}}_u) - \mathbf{\Sigma}_u\right\| \left\|\mathbf{\Sigma}_{t+1}^{-1}\right\|$$
$$= O_p\left(s_p\gamma_T^{1-q}\right).$$

□

A.4 Proof of Theorem 4.2

Proof of Theorem 4.2. Consider the parametric VaR estimator case. By the definition of VaR in (2.2), we have

$$\left|\widehat{\mathrm{VaR}}_{\alpha,t+1} - \mathrm{VaR}_{\alpha,t+1}\right| \leq C\left|c_\alpha\sqrt{\mathbf{w}^\top\widehat{\mathbf{\Sigma}}_{t+1}\mathbf{w}} - c_\alpha\sqrt{\mathbf{w}^\top\mathbf{\Sigma}_{t+1}\mathbf{w}}\right| + C\left|\mathbf{w}^\top(\bar{\mathbf{y}} - \boldsymbol{\mu})\right|.$$

Then, since $\|\bar{\mathbf{y}} - \boldsymbol{\mu}\|_{\max}^2 = O_p(\log p/T)$ and $\|\bar{\mathbf{y}} - \boldsymbol{\mu}\|^2 = O_p(p/T)$, we have

$$\left(\mathbf{w}^\top(\bar{\mathbf{y}} - \boldsymbol{\mu})\right)^2 = O_p\left(\min\left\{\frac{\log p}{T}, \|\mathbf{w}\|^2\frac{p}{T}\right\}\right). \tag{A.9}$$

Consider $\left|c_\alpha\sqrt{\mathbf{w}^\top\widehat{\boldsymbol{\Sigma}}_{t+1}\mathbf{w}} - c_\alpha\sqrt{\mathbf{w}^\top\boldsymbol{\Sigma}_{t+1}\mathbf{w}}\right|$. We have

$$\begin{aligned}\left|c_\alpha\sqrt{\mathbf{w}^\top\widehat{\boldsymbol{\Sigma}}_{t+1}\mathbf{w}} - c_\alpha\sqrt{\mathbf{w}^\top\boldsymbol{\Sigma}_{t+1}\mathbf{w}}\right| &\le C\left|\mathbf{w}^\top\left(\widehat{\boldsymbol{\Sigma}}_{t+1} - \boldsymbol{\Sigma}_{t+1}\right)\mathbf{w}\right| \\ &\le C\sum_{i,j=1}^{p}|w_iw_j|\left\|\widehat{\boldsymbol{\Sigma}}_{t+1} - \boldsymbol{\Sigma}_{t+1}\right\|_{\max} \\ &= O_p\left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}}\right), \end{aligned} \tag{A.10}$$

where the last is due to Theorem 4.1 and $\|\mathbf{w}\|_1 \le C$. Combining (A.9) and (A.10), we have

$$\left|\widehat{\mathrm{VaR}}_{\alpha,t+1} - \mathrm{VaR}_{\alpha,t+1}\right| = O_p\left(\sqrt{\frac{\log p}{T}} + \sqrt{\frac{s_p}{p}}\right).$$

Consider the non-parametric σ-based VaR estimator case. Define

$$\widehat{F}_T(x) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\{\widehat{x}_t \le x\}}, \qquad F_T(x) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\{x_t \le x\}},$$

where x is an α-quantile value, and $F(x)$ as a cumulative distribution function of x_t. Then the expected value of the absolute difference between $\widehat{F}_T(x)$ and $F_T(x)$ is

$$\mathbb{E}\left|\widehat{F}_T(x) - F_T(x)\right| \le \frac{1}{T}\sum_{t=1}^{T}\big(\mathbb{P}\{\widehat{x}_t \le x, x_t > x\} + \mathbb{P}\{\widehat{x}_t > x, x_t \le x\}\big).$$

Let $\vartheta_T = \sqrt{\log p/T} + \sqrt{s_p/p}$. Then, since Assumption 4(a) implies $|\widehat{x}_t - x_t| = O_p(\vartheta_T\sqrt{\log T})$, we have, for large enough C,

$$\mathbb{P}\{\widehat{x}_t \le x, x_t > x\} \le \mathbb{P}\{x_t \le x + |\widehat{x}_t - x_t|, x_t > x\}$$

$$
\begin{aligned}
&= \mathbb{P}\{x < x_t \le x + |\widehat{x}_t - x_t|\} \\
&\le \mathbb{P}\{x < x_t \le x + C\vartheta_T\sqrt{\log T}\} + \mathbb{P}\{|\widehat{x}_t - x_t| > C\vartheta_T\sqrt{\log T}\} \\
&\le F(x + C\vartheta_T\sqrt{\log T}) - F(x) + \frac{1}{\sqrt{T}} \\
&= O\left(\vartheta_T\sqrt{\log T} + \frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the last equality is due to Assumption 4(b). Similarly, the bound for $\mathbb{P}\{\widehat{x}_t > x, x_t \le x\}$ can be found. Then $\big|\widehat{F}_T(x) - F_T(x)\big| = O_p\big(\vartheta_T\sqrt{\log T} + 1/\sqrt{T}\big)$. By the Dvoretzky-Kiefer-Wolfowitz inequality, we have $\big|F_T(x) - F(x)\big| = O_p\big(1/\sqrt{T}\big)$. Thus,

$$
\begin{aligned}
\big|\widehat{F}_T(x) - F(x)\big| &\le \big|\widehat{F}_T(x) - F_T(x)\big| + \big|F_T(x) - F(x)\big| \\
&= O_p\big(\vartheta_T\sqrt{\log T} + 1/\sqrt{T}\big).
\end{aligned}
$$

Define

$$
F^{-1}(y) = \inf\big\{x : F(x) > y\big\}.
$$

Then the $\lceil \alpha T \rceil$-th smallest value of $\{x_t\}_{t=1}^T$ is bounded as follows:

$$
\begin{aligned}
\widehat{F}_T^{-1}(\alpha) &= \inf\big\{x : \widehat{F}_T(x) > \alpha\big\} \\
&\le \inf\big\{x : F(x) > \alpha + \big|\widehat{F}_T(x) - F(x)\big|\big\} \\
&\le F^{-1}\left(\alpha + \big|\widehat{F}_T(x) - F(x)\big|\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{F}_T^{-1}(\alpha) &\ge \inf\big\{x : F(x) > \alpha - \big|\widehat{F}_T(x) - F(x)\big|\big\} \\
&\ge F^{-1}\left(\alpha - \big|\widehat{F}_T(x) - F(x)\big|\right).
\end{aligned}
$$

Thus, we have

$$\left|\widehat{F}_T^{-1}(\alpha) - F^{-1}(\alpha)\right| \leq C\left|\widehat{F}_T(x) - F(x)\right|. \tag{A.11}$$

We also have

$$\begin{aligned}
\mathbf{w}^\top \widehat{\mathbf{\Sigma}}_{t+1}\mathbf{w} &= \mathbf{w}^\top \mathbf{\Sigma}_{t+1}\mathbf{w} + \mathbf{w}^\top(\widehat{\mathbf{\Sigma}}_{t+1} - \mathbf{\Sigma}_{t+1})\mathbf{w} \\
&= \mathbf{w}^\top \mathbf{V}\mathbf{\Sigma}_{f,t+1}\mathbf{V}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{\Sigma}_u \mathbf{w} + o_p(1) \\
&\leq \max_{i\leq r} h_{it+1}\left\|\mathbf{V}^\top \mathbf{w}\right\|^2 + C + o_p(1) \\
&= O_p(1), \qquad \text{(A.12)}
\end{aligned}$$

where the inequality is due to the gross exposure condition and $|\Sigma_{u,ij}| \leq C$, and the last equality is due to $\|\mathbf{V}^\top \mathbf{w}\|^2 = \|\sum_{i=1}^p w_i \widetilde{\mathbf{v}}_i\|^2 \leq \sum_{i=1}^p |w_i| \|\widetilde{\mathbf{v}}_i\|^2 \leq C$, where $\widetilde{\mathbf{v}}_i$ is the i^{th} row vector of $\mathbf{V}$. Combining results from (A.9) to (A.12), we have

$$\begin{aligned}
\left|\widehat{\mathrm{VaR}}_{\alpha,t+1} - \mathrm{VaR}_{\alpha,t+1}\right| &\leq C\sqrt{\mathbf{w}^\top \widehat{\mathbf{\Sigma}}_{t+1}\mathbf{w}}\left|\widehat{F}_T^{-1}(\alpha) - F^{-1}(\alpha)\right| + C\left|\mathbf{w}^\top(\bar{\mathbf{y}} - \boldsymbol{\mu})\right| \\
&\quad + CF^{-1}(\alpha)\left|\sqrt{\mathbf{w}^\top \widehat{\mathbf{\Sigma}}_{t+1}\mathbf{w}} - \sqrt{\mathbf{w}^\top \mathbf{\Sigma}_{t+1}\mathbf{w}}\right| \\
&= O_p\left(\vartheta_T\sqrt{\log T}\right).
\end{aligned}$$

□