Estimation of Large Dynamic Covariance Matrices: A Selective Review

Degui Li*

Department of Mathematics, University of York, UK

Version: April 27, 2021

Abstract

A personal review of some recent developments on estimating large dynamic covariance matrices whose entries are allowed to change over time is provided. The underlying covariance matrices are assumed to satisfy structural assumptions such as GARCH, approximate sparsity and conditional sparsity. Initially the review considers extensions of the classic GARCH model to multivariate and high-dimensional time series settings, and then focuses on some data-driven non- and semi-parametric models and estimation approaches for large covariance matrices which evolve smoothly over time or with some conditioning variables. Detection of multiple structural breaks in large covariance structures is also reviewed. Finally some relevant future directions are discussed.

Keywords: Covariance matrix, CUSUM statistic, Factor model, GARCH, Generalised shrinkage, Kernel estimation, Semi-parametric estimation, Sparsity, Structural breaks

*Email: degui.li@york.ac.uk.

1 Introduction

Estimation of covariance matrices is one of the most important problems in modern statistics and econometrics, with applications in various areas including biology, economics, finance, health science and social science. Suppose that we collect $\mathbf{X}_1, \cdots, \mathbf{X}_n$ from a sequence of p-dimensional random vectors which may be independent or stationary weakly dependent, $\mathsf{E}(\mathbf{X}_t) = \mathbf{0}$ and $\mathsf{Cov}(\mathbf{X}_t) = \mathsf{E}(\mathbf{X}_t\mathbf{X}_t^\intercal) = \mathbf{\Sigma}_X$, where $\mathbf{0}$ is a null vector and $\mathbf{\Sigma}_X$ is a $p \times p$ positive definite covariance matrix. The main focus of this review article is on estimating the covariance structure of $\mathbf{X}_t$. In the classic setting, the dimension of $\mathbf{X}_t$ is either fixed or much smaller than the sample size n. The population covariance matrix is often estimated by the conventional sample covariance matrix (e.g., Anderson, 2003) defined by

$$\widetilde{\mathbf{\Sigma}}_X = \frac{1}{n-1}\sum_{t=1}^{n}\left(\mathbf{X}_t - \overline{\mathbf{X}}\right)\left(\mathbf{X}_t - \overline{\mathbf{X}}\right)^\intercal = (\widetilde{\sigma}_{ij})_{p\times p} \quad \text{with } \overline{\mathbf{X}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t. \tag{1.1}$$

Under some moment conditions on $\mathbf{X}_t$, by the Law of Large Numbers, we may show that $\widetilde{\mathbf{\Sigma}}_X$ converges to $\mathbf{\Sigma}_X$ in probability when $p \ll n$, and consequently the smallest eigenvalue of $\widetilde{\mathbf{\Sigma}}_X$ would be positive and bounded away from zero with a high probability.

However, in big data analytics, a typical feature of the collected data is that the dimension p diverges to infinity at a fast rate of the sample size n. It is well known that the sample covariance matrix $\widetilde{\mathbf{\Sigma}}_X$ becomes singular if p is larger than n, making it challenging to obtain its inverse. To address this problem, some structural assumptions are imposed on $\mathbf{\Sigma}_X$ and various regularisation techniques have been introduced in the estimation procedure. A comprehensive review of recent developments on this topic is available at Bai and Shi (2011), Pourahmadi (2013), Cai, Ren and Zhou (2016), Fan, Liao and Liu (2016) and Bose and Bhattacharjee (2018). We review two commonly-used large covariance structures and discuss the corresponding estimation approaches in the next two subsections.

1.1 *Sparse covariance matrices and generalised shrinkage*

In order to construct reliable covariance matrix estimation when p exceeds n, we often need to assume that $\mathbf{\Sigma}_X$ satisfies the following sparsity condition (e.g., Bickel and Levina, 2008a).

Assumption 1. *The population covariance matrix $\mathbf{\Sigma}_X$ is approximately sparse, i.e., $\mathbf{\Sigma}_X \in \mathcal{S}(q, \varpi_1(p), M_1)$,*

$$\mathcal{S}(q, \varpi_1(p), M_1) = \left\{\mathbf{\Sigma} = [\sigma_{ij}]_{p\times p} \;\middle|\; \forall i,\ \sigma_{ii} \leqslant M_1,\ \sum_{j=1}^{p}|\sigma_{ij}|^q \leqslant \varpi_1(p)\right\},$$

where $0 \leqslant q < 1$ and M_1 is a positive constant.

The rate $\varpi_1(p)$ in Assumption 1 is allowed to depend on the dimension p and would affect the estimation convergence rate, see (1.3) below. For the special case of $q = 0$, Assumption 1 is reduced to the exact sparsity assumption. The sparsity assumption indicates that most of the off-diagonal entries in $\mathbf{\Sigma}_X$ are either zero or close to zero, so the corresponding random elements in $\mathbf{X}_t$ are either uncorrelated or very weakly correlated.

As recommended by the existing literature (e.g., Bickel and Levina, 2008a,b; Rothman, Levina and Zhu, 2009; Cai and Liu, 2011), we may apply the following generalised shrinkage technique to the off-diagonal entries of the sample covariance matrix and estimate the sparse $\mathbf{\Sigma}_X$ by

$$\overline{\mathbf{\Sigma}}_X = [\overline{\sigma}_{ij}]_{p\times p}, \quad \overline{\sigma}_{ij} = \begin{cases} \widetilde{\sigma}_{ii}, & i = j, \\ s_\rho\left(\widetilde{\sigma}_{ij}\right), & i \neq j, \end{cases} \tag{1.2}$$

where $s_\rho(\cdot)$ is a shrinkage function and ρ is a tuning parameter. The function $s_\rho(\cdot)$ satisfies the following three conditions: (i) $|s_\rho(u)| \leqslant |u|$ for $u \in \mathcal{R}$; (ii) $s_\rho(u) = 0$ if $|u| \leqslant \rho$; (iii) $|s_\rho(u) - u| \leqslant \rho$. Some commonly-used shrinkage functions satisfying these restrictions include

- hard thresholding: $s_\rho(u) = u \cdot I(|u| \geqslant \rho)$ with $I(\cdot)$ being an indicator function;
- soft thresholding: $s_\rho(u) = \mathsf{sign}(u) \cdot (|u| - \rho)_+$ with $u_+ = \max\{u, 0\}$;
- SCAD (Fan and Li, 2001): $s_\rho(u)$ equals to the soft thresholding for $|u| \leqslant 2\rho$, $[(a-1)u - \mathsf{sign}(u)a\rho]/(a-2)$ for $2\rho < |u| \leqslant a\rho$, and u for $|u| > a\rho$, where a is often chosen as 3.7.

The tuning parameter ρ in the generalised shrinkage may be determined by the cross-validation rule (e.g., Bickel and Levina, 2008a). Theorem 1 in Rothman, Levina and Zhu (2009) shows that, choosing $\rho = M\sqrt{\log p/n}$ with $M > 0$ being sufficiently large, under some regularity conditions,

$$\left\|\overline{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X\right\|_O = O_P\left(\varpi_1(p)(\log p/n)^{(1-q)/2}\right), \tag{1.3}$$

where $\|\cdot\|_O$ denotes the matrix operator norm. The dimension p affects the convergence rate via $\varpi_1(p)$ and $\log p$. Assuming $\varpi_1(p)$ is bounded and $q = 0$, the rate in (1.3) would be simplified to $(\log p/n)^{1/2}$, indicating that estimation consistency may be achieved even when p diverges at an exponential rate of n. Optimality of the above convergence rate is discussed in Cai and Zhou (2012) and Cai, Ren and Zhou (2016).

1.2 *Estimation of conditionally sparse covariance matrices*

The sparsity assumption discussed in Section 1.1 may be too restrictive when there exists high correlation among the random components of $\mathbf{X}_t$. The latter is a typical feature for large data collected in economics and finance. For instance, high-dimensional stock return time series data

collected in the stock market are often affected by market and industry factors, leading to high correlation among individual stock return series. To allow non-sparse covariance structure, we may employ the following approximate factor model (e.g., Chamberlain and Rothschild, 1983):

$$\mathbf{X}_t = \mathbf{\Gamma}\mathbf{F}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\chi}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \cdots, n, \tag{1.4}$$

where $\mathbf{\Gamma}$ is a $p \times k$ factor loading matrix, $\mathbf{F}_t$ is a k-dimensional factor vector, and $\boldsymbol{\varepsilon}_t$ is a p-dimensional idiosyncratic error vector. Through the factor model (1.4), $\mathbf{X}_t$ is decomposed into the common component $\boldsymbol{\chi}_t$ and the error component $\boldsymbol{\varepsilon}_t$. Assuming that $\mathbf{F}_t$ and $\boldsymbol{\varepsilon}_t$ are uncorrelated with mean zero, the covariance matrix of $\mathbf{X}_t$ can be written as

$$\mathbf{\Sigma}_X = \mathbf{\Gamma}\mathbf{\Sigma}_F\mathbf{\Gamma}^\intercal + \mathbf{\Sigma}_\varepsilon \text{ with } \mathbf{\Sigma}_F = \mathsf{Cov}(\mathbf{F}_t) \text{ and } \mathbf{\Sigma}_\varepsilon = \mathsf{Cov}(\boldsymbol{\varepsilon}_t). \tag{1.5}$$

The approximate sparsity condition may be imposed on $\mathbf{\Sigma}_\varepsilon$, leading to the so-called conditionally sparse structure on $\mathbf{\Sigma}_X$. Assuming that $\mathbf{F}_t$ is observable (say, Fama-French three factors as in Fama and French, 1992) and $\mathbf{\Sigma}_\varepsilon \in \mathcal{S}(q, \varpi_1(p), M_1)$ defined in Assumption 1, Fan, Liao and Mincheva (2011) estimate $\mathbf{\Sigma}_X$ by applying the generalised shrinkage to the estimate of $\mathbf{\Sigma}_\varepsilon$. Fan, Liao and Mincheva (2013) further introduce a novel estimation technique of Principal Orthogonal complEment Thresholding (POET) to deal with the case of latent factors. We next briefly describe the POET method.

Let $\widetilde{\mathbf{\Sigma}}_X$ be the sample covariance matrix defined in (1.1), and $\left(\widetilde{\lambda}_j, \widetilde{\eta}_j\right)$, $j = 1, \cdots, p$, be pairs of eigenvalues and eigenvectors of $\widetilde{\mathbf{\Sigma}}_X$, where $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \cdots \geqslant \widetilde{\lambda}_p$. By the spectral decomposition,

$$\widetilde{\mathbf{\Sigma}}_X = \sum_{j=1}^{p} \widetilde{\lambda}_j \widetilde{\eta}_j \widetilde{\eta}_j^\intercal = \sum_{j=1}^{k} \widetilde{\lambda}_j \widetilde{\eta}_j \widetilde{\eta}_j^\intercal + \widetilde{\mathbf{\Sigma}}_\varepsilon \text{ with } \widetilde{\mathbf{\Sigma}}_\varepsilon = \sum_{j=k+1}^{p} \widetilde{\lambda}_j \widetilde{\eta}_j \widetilde{\eta}_j^\intercal = \left(\widetilde{\sigma}_{\varepsilon,ij}\right)_{p\times p}. \tag{1.6}$$

As $\mathbf{\Sigma}_\varepsilon \in \mathcal{S}(q, \varpi_1(p), M_1)$, we may apply the generalised shrinkage technique to $\widetilde{\mathbf{\Sigma}}_\varepsilon$, i.e.,

$$\widehat{\mathbf{\Sigma}}_\varepsilon = \left(\widehat{\sigma}_{\varepsilon,ij}\right)_{p\times p} \text{ with } \widehat{\sigma}_{\varepsilon,ij} = \begin{cases} \widetilde{\sigma}_{\varepsilon,ii}, & i = j, \\ s_\rho\left(\widetilde{\sigma}_{\varepsilon,ij}\right), & i \neq j, \end{cases} \tag{1.7}$$

where $s_\rho(\cdot)$ is defined as in Section 1.1. Theorem 3.1 in Fan, Liao and Mincheva (2013) proves that

$$\left\|\widehat{\mathbf{\Sigma}}_\varepsilon - \mathbf{\Sigma}_\varepsilon\right\|_O = O_P\left(\varpi_1(p)\left((\log p/n)^{1/2} + p^{-1/2}\right)^{1-q}\right), \tag{1.8}$$

where the additional rate $p^{-1/2}$ is due to the cost of estimating the latent factors and can be removed when $p \gg n$. Consequently, we estimate $\mathbf{\Sigma}_X$ by

$$\widehat{\mathbf{\Sigma}}_X = \sum_{j=1}^{k} \widetilde{\lambda}_j \widetilde{\eta}_j \widetilde{\eta}_j^\intercal + \widehat{\mathbf{\Sigma}}_\varepsilon. \tag{1.9}$$

The number of latent factors, k, is assumed to be known a priori in the above POET estimation. In practice, this number can be estimated by the ratio or information criterion (e.g., Bai and Ng, 2002; Lam and Yao, 2012; Ahn and Horenstein, 2013; Fan, Liao and Mincheva, 2013).

1.3 *Large dynamic covariance matrices*

The large covariance matrices discussed in Sections 1.1 and 1.2 are static ones with constant entries over time. This condition is somehow restrictive and may be violated when large data are collected over a long time span. We next give two motivating examples to further interpret this.

Example 1. Consider a U.S. macroeconomic dataset used by Stock and Watson (2009), containing 108 variables and covering the time period between 01/1959 and 12/2006. Let $\mathbf{X}_t$ be the first-order difference of the original monthly macroeconomic data, removing possible nonstationarity involved. Note that the dimension of $\mathbf{X}_t$ is $p = 108$ and the sample size is $n = 575$ (after taking the first-order difference). As co-movement of many macroeconomic time series variables is often driven by some latent factors, it seems sensible to apply the approximate factor model (1.4) to $\mathbf{X}_t$ and then use Fan, Liao and Mincheva (2013)'s POET to estimate the covariance structure $\mathbf{\Sigma}_X$. However, as revealed by some recent literature such as Stock and Watson (2009), Chen, Dolado and Gonzalo (2014) and Ma and Su (2018), it is inappropriate to apply a stable factor model with constant factor loadings to this macroeconomic dataset. For example, Ma and Su (2018) identify five structural breaks in the factor model. Consequently, a direct application of the POET method would lead to biased covariance matrix estimation. It thus becomes imperative to detect possible structural breaks in large covariance matrices. Some recent developments in this direction will be reviewed in Section 5.

Example 2. Consider an empirical dataset which consists of daily returns (in %) on stocks listed on New York Stock Exchange over the period between 11/12/2008 and 10/04/2018 (available at Datastream). We may select 500 stocks which have largest capitalisation among those stocks that have complete trading history over this period. Let $\mathbf{X}_t$ be the vector of returns for day t with $\mathbf{\Sigma}_X$ as the covariance matrix. Note that the dimension of $\mathbf{X}_t$ is $p = 500$. The main interest lies in construction of global minimum variance portfolios, which is obtained by solving

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}^\tau \mathbf{\Sigma}_X \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^\tau \mathbf{1} = 1,$$

where $\mathbf{1}$ is a p-dimensional column vector with each element being one. Allowing for short selling, i.e., elements of the optimal weight vector can be negative, and assuming no transaction cost, the analytical solution to the above problem can be written as $\boldsymbol{w}^\star = \mathbf{\Sigma}_X^{-1}\mathbf{1}/\mathbf{1}^\tau\mathbf{\Sigma}_X^{-1}\mathbf{1}$. Note that $\boldsymbol{w}^\star$ is a vector of time-invariant constants, relying on accurate estimation of the static covariance matrix $\mathbf{\Sigma}_X$ and its inverse. In practical applications, however, the dynamic optimal portfolio allocation which allows $\boldsymbol{w}^\star$ to vary over time is often preferred, and outperforms the static one (e.g., Guo,

Box and Zhang, 2017; Chen, Li and Linton, 2019; Wang *et al*, 2021). For example, Chen, Li and Linton (2019) estimate the dynamic covariance matrix which is allowed to change smoothly with Fama-French three factors (to be reviewed in Section 3.2) and then use it to construct the global minimum variance portfolio.

There have been extensive studies in the literature on dynamic volatility and covariance since the seminal works by Engle (1982) and Bollerslev (1986) on ARCH and GARCH models, respectively. In recent decades, there has been increasing interest on extending the ARCH/GARCH modelling approach to the multivariate and high-dimensional setting. Bollerslev, Engle and Wooldridge (1988) introduce the VEC-GARCH model, a straightforward multivariate extension of the univariate GARCH model; Engle, Ng and Rothschild (1990) and Hafner and Preminger (2009) consider the factor ARCH and GARCH models; Engle (2002) models dynamic conditional correlation through multivariate GARCH; van der Weide (2002) proposes a generalised orthogonal GARCH model; Guo, Box and Zhang (2017) introduce a semi-parametric method by combining GARCH and adaptive functional-coefficient models to estimate dynamic covariance matrix with size divergent to infinity at a polynomial rate of n; Engle, Ledoit and Wolf (2019) further combine Engle (2002)'s dynamic conditional correlation with a nonlinear shrinkage technique based on the random matrix theory in the high-dimensional case. Section 2 below will provide a selective review of the aforementioned GARCH-based models in multivariate and high-dimensional time series setting and discuss their estimation methodology.

However, the ARCH/GARCH modelling approaches are essentially parametric and may be misspecified in practice. Hence, it becomes imperative to develop data-driven non- and semi-parametric methods to estimate large dynamic covariance matrices. It is not uncommon to observe smooth structural changes or structural breaks (over time) in covariance structures of large economic and financial data. Section 3 considers large sparse dynamic covariance matrices which evolve smoothly over time or change with a single or multiple conditioning variables. We mainly review two estimation methods: a non-parametric kernel-based method combined with the generalised shrinkage to estimate large dynamic covariance matrices when the conditioning variable is univariate or scaled time; and a semi-parametric method combining the Model Averaging MArginal Regression (MAMAR) and the generalised shrinkage when the conditioning variables are multiple. Section 4 first extends the approximate factor model (1.4) by assuming factor loadings as functions of an index variable or scaled time, and introduces a localised version of Principal Component Analysis (PCA) to estimate both the factor loadings and common factors (subject to appropriate rotation). Then, we extend the POET method to estimate large dynamic covariance matrices with conditional sparsity. Section 5 reviews recent developments on multiple structural break detection in large covariance matrices which are either sparse or conditionally sparse. The classic binary segmentation with the CUSUM statistic is extended to estimate breaks in covariance of large data. Section 6 finally discusses some relevant future directions.

Throughout this review article, we assume the underlying covariance matrices to satisfy certain structural assumptions such as GARCH, approximate sparsity and conditional sparsity, and review the relevant parametric, nonparametric and semiparametric frequentist estimation techniques. Other large covariance matrix estimation methods which have been extensively studied in the literature but are not covered in this review include the robust estimation (e.g., Fan, Liu and Wang, 2018), nonlinear shrinkage method (e.g., Ledoit and Wolf, 2012, 2015) and the Bayesian method (e.g., Yang and Berger, 1994; Daniels and Pourahmadi, 2002; Rajaratnam, Massam and Corvallo, 2008).

2 GARCH-based dynamic covariance matrices

Recall that the univariate GARCH(m, q) model (Bollerslev, 1986) is defined by

$$x_t = \sigma_t \eta_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2.1}$$

where η_t are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $i = 1, \cdots, m$, $\beta_j \geqslant 0$, $j = 1, \cdots, q$, and $\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. With the above constraints on the parameters α_i and β_j, we may show that the unconditional variance of x_t is time-invariant, whereas its conditional variance σ_t^2 changes over time. Since introduction of the GARCH model in 1986, there have been extensive studies on generalising it to multivariate time series data, see Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) for a comprehensive review. In recent years, there have been some further extensions of the GARCH model to high-dimensional time series. Throughout this section, we let $\mathbf{X}_t = (x_{t1}, \cdots, x_{tp})^\tau$ and assume that

$$\mathbf{X}_t = \mathbf{\Sigma}_t^{1/2} \boldsymbol{\eta}_t, \tag{2.2}$$

where $\mathbf{\Sigma}_t$ are $p \times p$ positive definite conditional covariance matrices, $\boldsymbol{\eta}_t = (\eta_{t1}, \cdots, \eta_{tp})^\tau$ are i.i.d. random vectors with $\mathsf{E}(\boldsymbol{\eta}_t) = \mathbf{0}$ and $\mathsf{Cov}(\boldsymbol{\eta}_t) = \mathbf{I}_p$, a $p \times p$ identity matrix. We next briefly review some commonly-used model specifications for $\mathbf{\Sigma}_t$.

2.1 *VEC-GARCH models*

The VEC-GARCH model introduced by Bollerslev, Engle and Wooldridge (1988) is a straightforward extension of model (2.1) to the multivariate setting. It is defined as

$$\mathsf{vech}(\mathbf{\Sigma}_t) = \mathbf{A}_0 + \sum_{i=1}^{m} \mathbf{A}_i \mathsf{vech}\left(\mathbf{X}_{t-i}\mathbf{X}_{t-i}^\tau\right) + \sum_{j=1}^{q} \mathbf{B}_j \mathsf{vech}(\mathbf{\Sigma}_{t-j}), \tag{2.3}$$

where $\mathsf{vech}(\cdot)$ denotes half vectorisation of a symmetric matrix obtained by vectorising the lower triangular elements, $\boldsymbol{A}_0$ is a column vector with dimension $p(p+1)/2$, both $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are $p(p+1)/2 \times p(p+1)/2$ matrices. The model specification (2.3) shows that each entry in $\boldsymbol{\Sigma}_t$ is a linear function of all lagged conditional variances and covariances, as well as all lagged components of $\mathsf{vech}\left(\boldsymbol{X}_t\boldsymbol{X}_t^\intercal\right)$. Let $\boldsymbol{\Lambda} = \left(\boldsymbol{A}_0^\intercal, \mathsf{vec}(\boldsymbol{A}_1), \cdots, \mathsf{vec}(\boldsymbol{A}_m), \mathsf{vec}(\boldsymbol{B}_1), \cdots, \mathsf{vec}(\boldsymbol{B}_q)\right)^\intercal$ with $\mathsf{vec}(\cdot)$ denoting the conventional vectorisation of a matrix. As suggested by Bollerslev, Engle and Wooldridge (1988), we may estimate $\boldsymbol{\Lambda}$ by the maximum likelihood if $\boldsymbol{\eta}_t$ follows a p-variate normal distribution. Specifically, define the log-likelihood function $L(\boldsymbol{\Lambda}) = \sum_{t=1}^{n} L_t(\boldsymbol{\Lambda})$ with $L_t(\boldsymbol{\Lambda}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_t(\boldsymbol{\Lambda})| - \frac{1}{2}\boldsymbol{X}_t^\intercal\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\Lambda})\boldsymbol{X}_t + c_0$, where c_0 is a constant independent of the unknown parameters and $\boldsymbol{\Sigma}_t(\boldsymbol{\Lambda})$ satisfies (2.3). The maximum likelihood estimate is obtained by maximising $L(\boldsymbol{\Lambda})$ with respect to $\boldsymbol{\Lambda}$. Unfortunately, this maximum likelihood problem may be difficult to solve when the dimension p is moderately large, partly due to involvement of the inverse $\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\Lambda})$. Engle and Kroner (1995) thus propose the so-called BEKK model, which is a restricted version of VEC-GARCH but has a desirable property that the conditional covariance matrices are guaranteed to be positive definite.

2.2 *Factor-GARCH models*

Engle, Ng and Rothschild (1990) propose the following decomposition for $\boldsymbol{\Sigma}_t$ in (2.2):

$$\boldsymbol{\Sigma}_t = \sum_{i=1}^{k} \boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^\intercal f_{ti} + \boldsymbol{\Sigma}_\diamond, \tag{2.4}$$

where $\boldsymbol{\Sigma}_\diamond$ is a $p \times p$ positive semi-definite matrix, $\boldsymbol{\gamma}_i$, $i = 1, \cdots, k$, are linearly independent p-dimensional column vectors and f_{ti}, $i = 1, \cdots, k$, are positive random variables. The formulation (2.4) shows that the dynamic variance and covariance are mainly driven by the underlying factors whose number k is often considerably smaller than p. We may assume that each factor process f_{ti} follows a univariate GARCH model defined as in (2.1). The dynamic factors in (2.4) are allowed to be correlated. If the number k is small, we may fit $(f_{t1}, \cdots, f_{tk})^\intercal$ by the VEC-GARCH or BEKK model mentioned in Section 2.1.

Vrontos, Dellaportas and Politis (2003) and Hafner and Preminger (2009) consider a full factor-GARCH model defined by

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t, \quad \boldsymbol{F}_t = (\boldsymbol{\Sigma}_t^*)^{1/2}\,\boldsymbol{\xi}_t, \tag{2.5}$$

where $\boldsymbol{R}$ is a $p \times p$ constant loading matrix which is of full rank, $\boldsymbol{\Sigma}_t^* = \mathsf{diag}\left\{\sigma_{t1,*}^2, \cdots, \sigma_{tp,*}^2\right\}$, and $\boldsymbol{\xi}_t$ are i.i.d. random vectors with $\mathsf{E}(\boldsymbol{\xi}_t) = \boldsymbol{0}$ and $\mathsf{Cov}(\boldsymbol{\xi}_t) = \boldsymbol{I}_p$. It follows from (2.5) that the conditional covariance matrix of $\boldsymbol{X}_t$ is $\boldsymbol{\Sigma}_t = \boldsymbol{R}\boldsymbol{\Sigma}_t^*\boldsymbol{R}^\intercal$. Note that the factor space has the same dimension as p and there is no idiosyncratic error, making the full factor-GARCH model substantially different from the conventional factor model discussed in Section 1.2. Model (2.5) covers some special cases such as the generalised orthogonal GARCH model (van der Weide, 2002). Each diagonal component

of $\mathbf{\Sigma}_t^*$ is assumed to follow the univariate GARCH process (2.1). Hafner and Preminger (2009) use a quasi-maximum likelihood method to estimate the unknown parameters involved in the GARCH model formulation for $\sigma_{ti,*}^2$, $i = 1, \cdots, p$. They show that the estimate is consistent and asymptotically normal distributed under some regularity conditions.

2.3 *Dynamic conditional correlation models*

Engle (2002) proposes a dynamic conditional correlation (DCC) model:

$$\mathbf{\Sigma}_t = \mathbf{D}_t \mathbf{C}_t \mathbf{D}_t, \tag{2.6}$$

where $\mathbf{D}_t = \mathsf{diag}\{\sigma_{t1}, \cdots, \sigma_{tp}\}$ with σ_{ti}^2 being the i-th diagonal element of $\mathbf{\Sigma}_t$, and

$$\mathbf{C}_t = \mathsf{Corr}(\mathbf{X}_t | \mathcal{F}_{t-1}) = \mathbf{D}_t^{-1} \mathbf{\Sigma}_t \mathbf{D}_t^{-1}.$$

Model (2.6) is a natural generalisation of the constant conditional correlation model introduced by Bollerslev (1990). The univariate GARCH defined in (2.1) can be used to model σ_{ti}^2, $i = 1, \cdots, p$. Letting $\mathbf{X}_t^* = \mathbf{D}_t^{-1} \mathbf{X}_t$, Engle (2002) introduces the following expression:

$$\mathbf{Q}_t = (1 - \alpha - \beta)\mathbf{C} + \alpha \mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*\intercal} + \beta \mathbf{Q}_{t-1},$$

where $\mathbf{Q}_t$ is called as a conditional pesudo-correlation matrix, α and β are the DCC parameters, and $\mathbf{C}$ is the unconditional correlation matrix of $\mathbf{X}_t$. Consequently, $\mathbf{C}_t$ is constructed via

$$\mathbf{C}_t = [\mathsf{diag}(\mathbf{Q}_t)]^{-1/2} \, \mathbf{Q}_t \, [\mathsf{diag}(\mathbf{Q}_t)]^{-1/2}.$$

In general, estimation of the DCC model includes three steps: (i) a low-order univariate GARCH model is used to fit σ_{ti}^2 and obtain the fitted matrices $\widehat{\mathbf{D}}_t$ and $\widehat{\mathbf{X}}_t^*$; (ii) estimate the large unconditional correlation matrix; and (iii) finally estimate the conditional correlation matrix. Note that the estimation problems in Steps (ii) and (iii) are challenging when the dimension p is large. Engle, Ledoit and Wolf (2019) recommend a nonlinear shrinkage method (e.g., Ledoit and Wolf, 2012, 2015) to estimate $\mathbf{C}$ in Step (ii) when p and n are of the same order, using the random matrix theory. The composite likelihood method introduced by Pakel *et al.* (2020) may be used to estimate the large dynamic correlation matrix in Step (iii). A recent paper by Morana (2019) proposes a regularised semiparametric method to estimate high-dimensional DCC matrices.

3 Non- and semi-parametric estimation of sparse dynamic covariance matrices

The GARCH-based conditional covariance matrix models introduced in Section 2 heavily rely on pre-specified parametric model assumptions, which might be misspecified in practice. The potential model misspecification issue often leads to biased covariance matrix estimates and misleading conclusions in statistical inference. Non- and semi-parametric models and estimation methods can avoid such a problem by allowing data to "speak for themselves". In this section, we review two non- and semi-parametric methods to estimate large sparse dynamic covariance matrices, extending the methodology and theory in Section 1.1.

3.1 *Non-parametric kernel estimation and generalised shrinkage*

Define the conditional mean and covariance matrix of $\mathbf{X}_t$ given $z_t = z$ as

$$\mathfrak{m}_0(z) = \mathsf{E}(\mathbf{X}_t|z_t = z) \text{ and } \mathbf{\Sigma}_X(z) = \mathsf{Cov}(\mathbf{X}_t|z_t = z) = \mathbf{\Lambda}_0(z) - \mathfrak{m}_0(z)\mathfrak{m}_0^\intercal(z), \tag{3.1}$$

where $\mathbf{\Lambda}_0(z) = \mathsf{E}\left(\mathbf{X}_t\mathbf{X}_t^\intercal|z_t = z\right)$ and z_t is a univariate random variable. In practice, z_t may also be chosen as scaled time point t/n with n as the sample size (e.g., Chen, Xu and Wu, 2013). Without loss of generality, we assume that z_t is defined on a compact set $\mathcal{C}$. In contrast to the static covariance matrix considered in Section 1, the covariance matrix defined in (3.1) is allowed to evolve smoothly over time or with a stationary process z_t, providing a more flexible way to capture the dynamic covariance structure. When $\mathbf{X}_t$ is univariate ($p = 1$), there have been extensive studies on non-parametric estimation of the conditional variance function (e.g., Fan and Yao, 1998; Yu and Jones, 2004; Chen, Cheng and Peng, 2009). When the dimension p diverges to infinity or even exceeds n, Chen, Xu and Wu (2013) combine the non-parametric kernel smoothing method with hard thresholding to estimate $\mathbf{\Sigma}_X(\cdot)$ for the fixed-design case; whereas Chen and Leng (2016) use the generalised shrinkage in the estimation for the random-design case. Before describing their methodology, we extend the sparsity condition in Assumption 1 to a uniform version.

Assumption 2. *The population conditional covariance matrix $\mathbf{\Sigma}_X(\cdot)$ is approximately sparse uniformly over $\mathcal{C}$, i.e., $\mathbf{\Sigma}_X(\cdot) \in \mathcal{S}_{\mathcal{C}}(q, \varpi_2(p), M_2)$,*

$$\mathcal{S}_{\mathcal{C}}(q, \varpi_2(p), M_2) = \left\{ \mathbf{\Sigma}(\cdot) = [\sigma_{ij}(\cdot)]_{p\times p} \;\middle|\; \forall i, \sup_{z\in\mathcal{C}} \sigma_{ii}(z) \leqslant M_2, \sup_{z\in\mathcal{C}} \sum_{j=1}^{p} |\sigma_{ij}(z)|^q \leqslant \varpi_2(p) \right\},$$

where $0 \leqslant q < 1$ and M_2 is a positive constant.

With kernel smoothing (Wand and Jones, 1995), we estimate $\mathbf{m}_0(z)$ by

$$\widetilde{\mathbf{m}}(z) = \frac{\sum_{t=1}^{n} K_h(z_t - z)\mathbf{X}_t}{\sum_{t=1}^{n} K_h(z_t - z)} = [\widetilde{m}_1(z), \cdots, \widetilde{m}_p(z)]^\intercal, \tag{3.2}$$

where $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$, $K(\cdot)$ is a kernel function and h is a bandwidth. Letting λ_{ij} be the (i,j)-entry of $\mathbf{\Lambda}_0(z) = \mathsf{E}\left(\mathbf{X}_t\mathbf{X}_t^\intercal | z_t = z\right)$, we estimate it by

$$\widetilde{\lambda}_{ij}(z) = \frac{\sum_{t=1}^{n} K_h(z_t - z)x_{ti}x_{tj}}{\sum_{t=1}^{n} K_h(z_t - z)}, \ \ 1 \leqslant i, j \leqslant p, \tag{3.3}$$

where x_{ti} is the i-th component of $\mathbf{X}_t$. Combining (3.2) and (3.3), we estimate $\sigma_{ij}(z)$, the (i,j)-entry of $\mathbf{\Sigma}_X(z)$, by $\widetilde{\sigma}_{ij}(z) = \widetilde{\lambda}_{ij}(z) - \widetilde{m}_i(z)\widetilde{m}_j(z)$. Finally, the generalised shrinkage is applied to $\widetilde{\sigma}_{ij}(z)$ when $i \neq j$, and estimate $\mathbf{\Sigma}_X(z)$ by

$$\overline{\mathbf{\Sigma}}_X(z) = [\overline{\sigma}_{ij}(z)]_{p\times p}, \ \ \overline{\sigma}_{ij}(z) = \begin{cases} \widetilde{\sigma}_{ii}(z), & i = j, \\ s_{\rho(z)}\left(\widetilde{\sigma}_{ij}(z)\right), & i \neq j, \end{cases} \tag{3.4}$$

where $s_{\rho(z)}(\cdot)$ is a shrinkage function with a variable tuning parameter $\rho(z)$ (Chen and Leng, 2016). When z_t is univariate random and letting $\mathcal{C} = [C_1, C_2]$ with $C_1 < C_2$, Theorem 1 in Chen and Leng (2016) shows that, under some technical conditions,

$$\sup_{C_1+\epsilon\leqslant z\leqslant C_2-\epsilon} \left\|\overline{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z)\right\|_O = O_P\left(\varpi_2(p)\left([\log p/(nh)]^{1/2} + h^2(\log p)^{1/2}\right)^{1-q}\right), \tag{3.5}$$

where $\varpi_2(p)$ is defined in Assumption 2 and ϵ is a small positive number. The rate $h^2(\log p)^{1/2}$ in the uniform consistency result (3.5) is caused by the kernel-based estimation bias. Choosing h to have the optimal rate $n^{-1/5}$ and treating (nh) as the effective sample size in the non-parametric local estimation, the uniform convergence rate in (3.5) is comparable to that in (1.3). The convergence result of $\overline{\mathbf{\Sigma}}_X(z)$ when $z_t = t/n$ is available in Theorem 4.1 of Chen, Xu and Wu (2013).

A potential issue of the above estimate $\overline{\mathbf{\Sigma}}_X(z)$ is that it is not necessarily positive definite in finite samples. To address this problem, Chen and Leng (2016) recommend the following modification. Let $\overline{\lambda}_1(z) \geqslant \overline{\lambda}_2(z) \geqslant \cdots \geqslant \overline{\lambda}_p(z)$ be the eigenvalues of $\overline{\mathbf{\Sigma}}_X(z)$ which are arranged in a non-increasing order. When the smallest eigenvalue $\overline{\lambda}_p(z)$ is negative, the matrix $\overline{\mathbf{\Sigma}}_X(z)$ is not positive definite and a modification is thus required. Define

$$\overline{\mathbf{\Sigma}}_X^C(z) = \overline{\mathbf{\Sigma}}_X(z) + \left[\delta_n - \overline{\lambda}_p(z)\right] \cdot \mathbf{I}_p,$$

where δ_n is a tuning parameter which tends to zero as $n \to \infty$. The above modification guarantees that $\overline{\mathbf{\Sigma}}_X^C(z)$ is uniformly positive definite, and retains the approximately sparse covariance structure. Hence, it is sensible to use $\overline{\mathbf{\Sigma}}_X^C(z)$ as an alternative estimate of $\mathbf{\Sigma}_X(z)$ when $\overline{\lambda}_p(z)$ is negative. Then,

we may define the following modified version of $\overline{\mathbf{\Sigma}}_X(z)$:

$$\begin{aligned}\overline{\mathbf{\Sigma}}_X^M(z) &= \overline{\mathbf{\Sigma}}_X(z)\cdot \mathrm{I}\left(\overline{\lambda}_p(z)>0\right)+\overline{\mathbf{\Sigma}}_X^C(z)\cdot \mathrm{I}\left(\overline{\lambda}_p(z)\leqslant 0\right)\\ &= \overline{\mathbf{\Sigma}}_X(z)+\left[\delta_n-\overline{\lambda}_p(z)\right]\mathbf{I}_p\cdot \mathrm{I}\left(\overline{\lambda}_p(z)\leqslant 0\right).\end{aligned}$$

It is easy to verify that $\overline{\mathbf{\Sigma}}_X^M(z)$ is alway positive definite (uniformly over z).

3.2 *Semiparametric MAMAR and generalised shrinkage*

In practical data analysis, the conditioning variables in (3.1) are often multivariate. For example, Chen, Li and Linton (2019) choose them as the Fama-French three factors in construction of optimal dynamic portfolios. In this case, a direct use of non-parametric kernel estimation as in Section 3.1 often results in poor finite-sample performance due to the "curse of dimensionality" problem. To address this problem, Chen, Li and Linton (2019) consider MAMAR approximation for both the multivariate conditional mean and covariance matrix, and subsequently introduce a semi-parametric shrinkage method to estimate the large dynamic covariance matrix. More detailed introduction of the MAMAR-based semi-parametric estimation together with its applications in high-dimensional classification and portfolio selection can be found in Li, Linton and Lu (2015), Chen *et al* (2016), Fan *et al* (2016) and Chen *et al* (2018).

Replacing z_t by $\mathbf{Z}_t=(z_{t1},\cdots,z_{tr})^\tau$ in (3.1), we have

$$\sigma_{ij}(\boldsymbol{z})=\lambda_{ij}(\boldsymbol{z})-m_i(\boldsymbol{z})m_j(\boldsymbol{z}),\ \ 1\leqslant i,j\leqslant p,$$

where $\boldsymbol{z}=(z_1,\cdots,z_r)^\tau$. Letting $\lambda_{ij,k}=\mathsf{E}\left(x_{ti}x_{tj}|z_{tk}=z_k\right)$, we consider the following MAMAR approximation to $\lambda_{ij}(\boldsymbol{z})$:

$$\lambda_{ij}(\boldsymbol{z})\approx a_{ij,0}^*+\sum_{k=1}^r a_{ij,k}^*\lambda_{ij,k}(z_k)=:\lambda_{ij}^*(\boldsymbol{z}), \tag{3.6}$$

where the constant weights $a_{ij,0}^*, a_{ij,1}^*,\cdots,a_{ij,r}^*$ are determined by minimising

$$\mathsf{E}\left[x_{ti}x_{tj}-a_{ij,0}-\sum_{k=1}^r a_{ij,k}\lambda_{ij,k}(z_{tk})\right]^2$$

with respect to $a_{ij,0}, a_{ij,1},\cdots,a_{ij,r}$. Similarly, we also have the MAMAR approximation to $m_i(\boldsymbol{z})$:

$$m_i(\boldsymbol{z})\approx b_{i,0}^*+\sum_{k=1}^r b_{i,k}^* m_{i,k}(z_k)=:m_i^*(\boldsymbol{z}), \tag{3.7}$$

where $m_{i,k}(z_k)=\mathsf{E}(x_{ti}|z_{tk}=z_k)$ and $b_{i,0}^*, b_{i,1}^*,\cdots,b_{i,r}^*$ are the constant optimal weights. Combin-

ing (3.6) and (3.7), we have

$$\mathbf{\Sigma}_X(z) \approx \mathbf{\Sigma}_X^*(z) = \left[\sigma_{ij}^*(z)\right]_{p\times p} \quad \text{with} \quad \sigma_{ij}^*(z) = \lambda_{ij}^*(z) - m_i^*(z)m_j^*(z). \tag{3.8}$$

Instead of estimating $\mathbf{\Sigma}_X(z)$ directly, we may estimate its MAMAR approximation $\mathbf{\Sigma}_X^*(z)$ which is assumed to satisfy Assumption 2 in Section 3.1 (with slight modification). We next briefly review a two-stage semi-parametric estimation method introduced by Chen, Li and Linton (2019).

- Stage 1: Estimate $m_{i,k}(\cdot)$ and $\lambda_{ij,k}(\cdot)$ non-parametrically and denote these estimates by $\widetilde{m}_{i,k}(\cdot)$ and $\widetilde{\lambda}_{ij,k}(\cdot)$, respectively. The non-parametric method can be the kernel (as in Section 3.1), local polynomial (Fan and Gijbels, 1996) or sieve method Chen (2007).

- Stage 2: Consider the approximate linear regression:

$$x_{ti}x_{tj} \approx a_{ij,0}^* + \sum_{k=1}^{r} a_{ij,k}^* \widetilde{\lambda}_{ij,k}(z_{tk}), \quad t = 1, \cdots, n,$$

  and use ordinary least squares (OLS) to obtain the optimal weight estimates denoted by $\widetilde{a}_{ij,0}, \widetilde{a}_{ij,1}, \cdots, \widetilde{a}_{ij,r}$. Analogously, we can also obtain $\widetilde{b}_{i,0}, \widetilde{b}_{i,1}, \cdots, \widetilde{b}_{i,r}$, as the OLS estimates of $b_{i,0}^*, b_{i,1}^*, \cdots, b_{i,r}^*$.

With the estimates obtained in the above two stages, we have

$$\widetilde{\lambda}_{ij}^*(z) = \widetilde{a}_{ij,0} + \sum_{k=1}^{r} \widetilde{a}_{ij,k}\widetilde{\lambda}_{ij,k}(z_k), \quad \widetilde{m}_i^*(z) = \widetilde{b}_{i,0} + \sum_{k=1}^{r} \widetilde{b}_{i,k}\widetilde{m}_{i,k}(z_k),$$

and consequently,

$$\widetilde{\sigma}_{ij}^*(z) = \widetilde{\lambda}_{ij}^*(z) - \widetilde{m}_i^*(z)\widetilde{m}_j^*(z).$$

Finally, applying the generalised shrinkage to $\widetilde{\sigma}_{ij}^*(z)$ as in (3.4), we obtain the covariance matrix estimate $\overline{\mathbf{\Sigma}}_X^*(z)$. Suppose that z_{tk} has a compact support $[C_{1k}, C_{2k}]$, $k = 1, \cdots, r$, and construct $\widetilde{m}_{i,k}(\cdot)$ and $\widetilde{\lambda}_{ij,k}(\cdot)$ in Stage 1 via the kernel smoothing method with equal bandwidth h. When p is larger than n, Theorem 1(i) in Chen, Li and Linton (2019) shows that, under some conditions,

$$\sup_{z=(z_1,\cdots,z_r)^\intercal : C_{1k}+\epsilon \leqslant z_k \leqslant C_{2k}-\epsilon} \left\| \overline{\mathbf{\Sigma}}_X^*(z) - \mathbf{\Sigma}_X^*(z) \right\|_O = O_P\left(\varpi_2(p) \left([\log p/(nh)]^{1/2} + h^2 \right)^{1-q} \right),$$

which is similar to the uniform consistency in (3.5). When the MAMAR approximation $\mathbf{\Sigma}_X^*(z)$ is sufficiently close to $\mathbf{\Sigma}_X(z)$, i.e., there exists $b_{n,p} \to 0$ such that

$$\|\mathbf{\Sigma}_X^*(z) - \mathbf{\Sigma}_X(z)\|_O = O(b_{n,p})$$

uniformly over z, and additionally

$$\max_{1\leqslant i,j\leqslant p}\left|\sigma_{ij}^{*}(z)-\sigma_{ij}(z)\right|=O\left([\log p/(nh)]^{1/2}+h^{2}\right)$$

uniformly over z, we may further show that

$$\sup_{z=(z_1,\cdots,z_r)^{\intercal}:C_{1k}+\epsilon\leqslant z_k\leqslant C_{2k}-\epsilon}\left\|\overline{\boldsymbol{\Sigma}}_{X}^{*}(z)-\boldsymbol{\Sigma}_{X}(z)\right\|_{O}=O_{P}\left(\varpi_{2}(p)\left([\log p/(nh)]^{1/2}+h^{2}\right)^{1-q}+b_{n,p}\right).$$

A modification technique as in Section 3.1 may be applied to $\overline{\boldsymbol{\Sigma}}_{X}^{*}(z)$ so that the resulting covariance matrix estimate is positive definite in finite samples.

4 Non-parametric estimation of dynamic covariance matrices with conditional sparsity

In this section, we review a non-parametric estimation method for the conditionally sparse dynamic covariance matrix, which is a natural extension of Fan, Liao and Mincheva (2013)'s POET method. As this covariance structure relies on the factor model framework for $\mathbf{X}_t$, we first generalise model (1.4) by allowing factor loadings to change over time, and review a localised version of PCA.

4.1 *Factor models with varying loadings and local PCA*

Consider the time-dependent approximate factor model:

$$\mathbf{X}_t=\boldsymbol{\Gamma}_t\mathbf{F}_t+\boldsymbol{\varepsilon}_t,\quad t=1,\cdots,n, \tag{4.1}$$

where $\boldsymbol{\Gamma}_t$ is a $p\times k$ factor loading matrix whose entries are allowed to change over time, and the remaining components are the same as those in (1.4). Assume that each factor loading is defined as a smooth function of a univariate index variable or scaled time point, i.e., $\boldsymbol{\Gamma}_t=\boldsymbol{\Gamma}(z_t)=[\gamma_{ij}(z_t)]_{p\times k}$. When z_t is a stationary random state process, Pelger and Xiong (2019) consider estimation and testing problems for model (4.1); when $z_t=t/n$, Motta, Hafner and von Sachs (2011) and Su and Wang (2017) estimate the time-varying factor model (4.1) via a localised version of PCA, which we will describe shortly. Other estimation methods include the maximum likelihood method (Mikkelsen, Hillebrand and Urga, 2019).

Let $K(\cdot)$ be the kernel function defined as in Section 3.1 and h be the bandwidth. Multiplying

both sides of (4.1) by

$$K_{t,h}(z) = K_h(z_t - z) \Big/ \left[\frac{1}{n}\sum_{s=1}^{n} K_h(z_s - z)\right] \quad \text{with} \quad K_h(\cdot) = \frac{1}{h}K(\cdot/h),$$

we obtain the following local approximation:

$$\mathbf{X}_t(z) = \mathbf{X}_t K_{t,h}^{1/2}(z) \approx \mathbf{\Gamma}(z)\mathbf{F}_t K_{t,h}^{1/2}(z) + \varepsilon_t K_{t,h}^{1/2}(z) = \mathbf{\Gamma}(z)\mathbf{F}_t(z) + \varepsilon_t(z), \quad t = 1, \cdots, n, \tag{4.2}$$

when $z_t \approx z$. Let $\mathbb{X}(z) = [\mathbf{X}_1(z), \cdots, \mathbf{X}_n(z)]$, $\mathbb{F}(z) = [\mathbf{F}_1(z), \cdots, \mathbf{F}_n(z)]^\intercal$, $\mathbf{\Gamma}(z) = [\mathbf{\Gamma}_1(z), \cdots, \mathbf{\Gamma}_p(z)]^\intercal$. Suppose that the number of common factors is pre-specified, and consider the following local identification condition (Su and Wang, 2017):

$$\frac{1}{n}\mathbb{F}^\intercal(z)\mathbb{F}(z) = \mathbf{I}_k, \quad \mathbf{\Gamma}^\intercal(z)\mathbf{\Gamma}(z) \text{ is diagonal.} \tag{4.3}$$

From (4.2), we consider the local least squares objective function defined by $\left\|\mathbb{X}(z) - \mathbf{\Gamma}\mathbb{F}^\intercal\right\|_F^2$, where $\|\cdot\|_F$ denotes the matrix Frobenius norm, and $\mathbf{\Gamma}$ and $\mathbb{F}$ are $p \times k$ and $n \times k$ matrices, respectively. With the identification condition (4.3), given $\mathbb{F}$, we estimate $\mathbf{\Gamma}(z)$ by $\frac{1}{n}\mathbb{X}(z)\mathbb{F}$. Consequently, the local least squares objective function becomes

$$\left\|\mathbb{X}(z) - \frac{1}{T}\mathbb{X}(z)\mathbb{F}\mathbb{F}^\intercal\right\|_F^2 = \text{trace}\left\{\mathbb{X}^\intercal(z)\mathbb{X}(z)\right\} - \frac{1}{n}\text{trace}\left\{\mathbb{F}^\intercal\mathbb{X}^\intercal(z)\mathbb{X}(z)\mathbb{F}\right\}, \tag{4.4}$$

which indicates that minimisation of $\left\|\mathbb{X}(z) - \mathbf{\Gamma}\mathbb{F}^\intercal\right\|_F^2$ subject to the restriction (4.3) is equivalent to maximisation of the second term on the right side of (4.4) subject to $\frac{1}{T}\mathbb{F}^\intercal\mathbb{F} = \mathbf{I}_k$. Let $\widetilde{\mathbb{F}}(z) = \left[\widetilde{\mathbf{F}}_1(z), \cdots, \widetilde{\mathbf{F}}_n(z)\right]^\intercal$ with each column vector being the eigenvector (multiplied by $\sqrt{n}$) of the $n \times n$ matrix $\mathbb{X}^\intercal(z)\mathbb{X}(z)$ corresponding to the first k largest eigenvalues, and estimate the factor loading matrix by

$$\widetilde{\mathbf{\Gamma}}(z) = \frac{1}{n}\mathbb{X}(z)\widetilde{\mathbb{F}}(z) = \left[\widetilde{\mathbf{\Gamma}}_1(z), \cdots, \widetilde{\mathbf{\Gamma}}_p(z)\right]^\intercal.$$

Without loss of generality, we assume that z_t has the compact support $[0, 1]$ if it is random, and p diverges at a polynomial rate of n. Under some regularity conditions, Wang *et al* (2021) give the uniform consistency results for $\widetilde{\mathbf{\Gamma}}_i(z)$ and $\widetilde{\mathbf{F}}_t(z)$, i.e.,

$$\max_{1\leqslant i\leqslant p}\sup_{\epsilon\leqslant z\leqslant 1-\epsilon}\left\|\widetilde{\mathbf{\Gamma}}_i(z) - \left[\mathbf{H}^{-1}(z)\right]^\intercal\mathbf{\Gamma}_i(z)\right\| = O_P\left(\left(\frac{\log p}{nh} + h^2\right)^{1/2}\right),$$

and

$$\sup_{\epsilon\leqslant z\leqslant 1-\epsilon}\frac{1}{n}\sum_{t=1}^{n}\left\|\widetilde{\mathbf{F}}_t(z) - \mathbf{H}(z)\mathbf{F}_t(z)\right\|^2 = O_P\left(\frac{\log p}{nh} + h^2\right),$$

where $\mathbf{H}(z)$ is a $k \times k$ time-dependent rotation matrix and $\|\cdot\|$ denotes the Euclidean norm of a vector. A similar uniform consistency result is also derived by Su and Wang (2017) when z_t is fixed. More asymptotic properties for the local PCA estimates (such as the limit distribution theory) are available at Motta, Hafner and von Sachs (2011), Su and Wang (2017), Mikkelsen, Hillebrand and Urga (2019), Pelger and Xiong (2019) and Wang *et al* (2021).

4.2 *Local POET*

A recent paper by Wang *et al* (2021) further uses the local PCA method introduced in Section 4.1 to estimate the dynamic covariance matrix of $\mathbf{X}_t$ satisfying (4.1). Assuming that $\mathbf{F}_t$ and $\boldsymbol{\varepsilon}_t$ are conditionally uncorrelated given z_t, we obtain

$$\boldsymbol{\Sigma}_X(z) = \boldsymbol{\Gamma}(z)\boldsymbol{\Sigma}_F\boldsymbol{\Gamma}^\intercal(z) + \boldsymbol{\Sigma}_\varepsilon(z),$$

where $\boldsymbol{\Sigma}_\varepsilon(z)$ denotes the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ given $z_t = z$ (when z_t is random) and $\boldsymbol{\Sigma}_F = \mathsf{Cov}(\mathbf{F}_t)$ is assumed to be time-invariant, ensuring that the index variable z_t and the factor $\mathbf{F}_t$ are separable (Pelger and Xiong, 2019).

We next combine the local PCA estimates in Section 4.1 with the generalised shrinkage to estimate the large dynamic covariance matrix, and then introduce a localised version of POET, extending the POET methodology in Section 1.2. With the local PCA estimates $\widetilde{\boldsymbol{\Gamma}}_i(z)$ and $\widetilde{\mathbf{F}}_t(z)$, we approximate the local residuals $\boldsymbol{\varepsilon}_t(z) = \boldsymbol{\varepsilon}_t K_{t,h}^{1/2}(z)$ in (4.2) by

$$\widetilde{\boldsymbol{\varepsilon}}_t(z) = [\widetilde{\varepsilon}_{1t}(z), \cdots, \widetilde{\varepsilon}_{pt}(z)]^\intercal \quad \text{with} \quad \widetilde{\varepsilon}_{it}(z) = x_{it}(z) - \widetilde{\boldsymbol{\Gamma}}_i^\intercal(z)\widetilde{\mathbf{F}}_t(z),$$

where $x_{it}(z)$ is the i-th element in $\mathbf{X}_t(z)$, and then obtain the sample covariance matrix of $\widetilde{\boldsymbol{\varepsilon}}_t(z)$:

$$\widetilde{\boldsymbol{\Sigma}}_\varepsilon(z) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{\varepsilon}}_t(z)\widetilde{\boldsymbol{\varepsilon}}_t^\intercal(z) = [\widetilde{\sigma}_{\varepsilon,ij}(z)]_{p\times p}\,.$$

However, the above estimate is unstable when the dimension p exceeds n. Hence, we may apply the generalised shrinkage to the off-diagonal entries of $\widetilde{\boldsymbol{\Sigma}}_\varepsilon(z)$, and denote the resulting estimate by $\check{\boldsymbol{\Sigma}}_\varepsilon(z)$. Finally, we estimate $\boldsymbol{\Sigma}_X(z)$ by

$$\check{\boldsymbol{\Sigma}}_X(z) = \widetilde{\boldsymbol{\Gamma}}(u)\widetilde{\boldsymbol{\Gamma}}^\intercal(z) + \check{\boldsymbol{\Sigma}}_\varepsilon(z),$$

where $\widetilde{\boldsymbol{\Gamma}}(z)$ is the local PCA estimate of the factor loading matrix defined in Section 4.1.

On the other hand, consider the kernel-weighted sample covariance matrix:

$$\widetilde{\mathbf{\Sigma}}_X(z) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t(z)\mathbf{X}_t^\intercal(z) = \left[\frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t\mathbf{X}_t^\intercal K_h\left(z_t - z\right)\right] / \left[\frac{1}{n}\sum_{t=1}^{n} K_h\left(z_t - z\right)\right].$$

We next introduce the localised version of POET using the eigen-analysis of $\widetilde{\mathbf{\Sigma}}_X(z)$. Let $\left(\widetilde{\lambda}_{i,z}, \widetilde{\boldsymbol{\eta}}_{i,z}\right)$, $i = 1, \cdots, p$, be pairs of eigenvalues and normalised eigenvectors of $\widetilde{\mathbf{\Sigma}}_X(z)$, where $\widetilde{\lambda}_{1,z} \geqslant \widetilde{\lambda}_{2,z} \geqslant \cdots \geqslant \widetilde{\lambda}_{p,z}$. By the spectral decomposition, we have

$$\widetilde{\mathbf{\Sigma}}_X(z) = \sum_{i=1}^{k} \widetilde{\lambda}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}^\intercal + \overline{\mathbf{\Sigma}}_\varepsilon(z), \;\; \overline{\mathbf{\Sigma}}_\varepsilon(z) = \sum_{i=k+1}^{p} \widetilde{\lambda}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}^\intercal.$$

Apply the generalised shrinkage (3.4) to the off-diagonal entries of $\overline{\mathbf{\Sigma}}_\varepsilon(z)$ and denote the resulting estimate by $\widehat{\mathbf{\Sigma}}_\varepsilon(z)$. Consequently, $\mathbf{\Sigma}_X(z)$ is estimated by

$$\widehat{\mathbf{\Sigma}}_X(z) = \sum_{i=1}^{k} \widetilde{\lambda}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}\widetilde{\boldsymbol{\eta}}_{i,z}^\intercal + \widehat{\mathbf{\Sigma}}_\varepsilon(z).$$

Proposition 1 in Wang *et al* (2021) shows the equivalence of the local POET estimate and the local PCA one, i.e., $\widehat{\mathbf{\Sigma}}_\varepsilon(z) = \check{\mathbf{\Sigma}}_\varepsilon(z)$ and $\widehat{\mathbf{\Sigma}}_X(z) = \check{\mathbf{\Sigma}}_X(z)$. Theorems 1 and 2 in that paper further provide uniform convergence results for the local POET estimate.

5 Detection of structural breaks in large covariance matrices

Structural breaks are common in many fields and occur for various reasons. For example, the environmental Kuznets curve may shift (thus leading to breaks) due to international cooperation on environmental protection and growing public awareness of environmental problems (e.g., Li, Qian and Su, 2016). It is important to take possible structural breaks into account in the covariance matrix estimation in order to conduct valid inference in the subsequent statistical analysis. The binary segmentation introduced by Vostrikova (1981) is probably the most commonly-used technique in detecting and estimating structural breaks and has been extended in recent years to handle high-dimensional time series and large panel data with instabilities (Cho and Fryzlewicz, 2015; Jirak, 2015; Cho, 2016; Aston and Kirch, 2018; Wang and Samworth, 2018). In this section, we review recent developments on structural break detection and estimation for large covariance matrices.

5.1 *Sparsified binary segmentation for covariance break detection*

Throughout this section, we let $\mathbf{\Sigma}_{X,t} = \mathsf{E}\left(\mathbf{X}_t\mathbf{X}_t^\intercal\right)$ be the time-dependent second-order moment of $\mathbf{X}_t$, and assume the following multiple structural breaks:

$$\mathbf{\Sigma}_{X,t} = \begin{cases} \mathbf{\Sigma}_X(1), & 1 \leqslant t \leqslant \eta_1, \\ \mathbf{\Sigma}_X(2), & \eta_1 + 1 \leqslant t \leqslant \eta_2, \\ \vdots & \vdots \\ \mathbf{\Sigma}_X(K+1), & \eta_K + 1 \leqslant t \leqslant n, \end{cases}$$

where $\mathbf{\Sigma}_X(j) \neq \mathbf{\Sigma}_X(j+1)$, η_j, $j = 1, \cdots, K$, are unobservable break times, and the break number K is unobservable either. To estimate the break times and break number over the interval $[l, u]$, we may use the classic cumulative sum (CUSUM) statistic in the matrix form (e.g., Wang, Yu and Rinaldo, 2021):

$$\mathbf{C}_{l,u}^X(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}}\left(\frac{1}{s-l+1}\sum_{t=l}^{s}\mathbf{X}_t\mathbf{X}_t^\intercal - \frac{1}{u-s}\sum_{t=s+1}^{u}\mathbf{X}_t\mathbf{X}_t^\intercal\right), \tag{5.1}$$

for $1 \leqslant l \leqslant s < u \leqslant n$, or the CUSUM statistic in the vector form (e.g., Aue *et al.*, 2009; Li, Li and Fryzlewicz, 2020):

$$C_{l,u}^X(s) = \sqrt{\frac{(s-l+1)(u-s)}{u-l+1}}\left[\frac{1}{s-l+1}\sum_{t=l}^{s}\mathsf{vech}\left(\mathbf{X}_t\mathbf{X}_t^\intercal\right) - \frac{1}{u-s}\sum_{t=s+1}^{u}\mathsf{vech}\left(\mathbf{X}_t\mathbf{X}_t^\intercal\right)\right]. \tag{5.2}$$

Aue *et al.* (2009) test the hypothesis of a single break in the covariance matrix with a fixed size. Wang, Yu and Rinaldo (2021) consider the matrix operator norm for $\mathbf{C}_{l,u}^X(s)$ and then use the binary segmentation (and its extended version) to estimate the unknown breaks. Li, Li and Fryzlewicz (2020) compare numerical performance between various norms on the CUSUM statistics and find that the one defined in (5.2) with the L_2-norm has reliable finite-sample performance.

Let $C_{l,u}^X(s;i,j)$ be the (i,j)-entry of $\mathbf{C}_{l,u}^X(s)$. As it is often the case that breaks in large covariance matrices are sparse, Cho and Fryzlewicz (2015) introduce a sparsified binary segmentation for break detection in mean of high-dimensional time series. It is easy to generalise such an idea to the large covariance structure. Define the L_2-norm based and sparsified CUSUM statistic:

$$\widetilde{C}_{l,u}^X(s) = \sum_{i=1}^{p}\sum_{j=i}^{p}\left|C_{l,u}^X(s;i,j)\right|^2 I\left(\max_{l\leqslant t<u}\left|C_{l,u}^X(s;i,j)\right| > \nu_{n,p}\right),$$

where $\nu_{n,p}$ is a thresholding parameter. By comparing $C_{l,u}^X(s;i,j)$ with $\nu_{n,p}$, we delete the index pair (i,j) when $\max_{l\leqslant t<u}\left|C_{l,u}^X(s;i,j)\right|$ does not exceed $\nu_{n,p}$. The latter indicates that there may be

no break for the index pair (i, j). By maximising $\widetilde{C}^{X}_{l,u}(s)$ with respect to s, we obtain

$$\widetilde{\eta}_1 = \arg \max_{l \leqslant s < u} \widetilde{C}^{X}_{l,u}(s),$$

the first candidate break point (which is not necessarily estimate of the first break point η_1), and split the closed interval $[l, u]$ into two sub-intervals: $[l, \widetilde{\eta}_1]$ and $[\widetilde{\eta}_1+1, u]$. Then, letting $(l, u) = (l, \widetilde{\eta}_1)$ or $(\widetilde{\eta}_1 + 1, u)$, we compute $\widetilde{C}^{X}_{l,u}(s)$ with s in either of the two sub-intervals, maximise it with respect to s and estimate the next candidate break points. The procedure is terminated when no more split is possible or $\widetilde{C}^{X}_{l,u}(s) \equiv 0$. Li, Li and Fryzlewicz (2020) further combine the above technique with the wild binary segmentation introduced by Fryzlewicz (2014) and find that it performs well in practice when breaks in the large covariance structure are indeed sparse.

5.2 *Detection of breaks in conditionally sparse covariance matrices*

In Section 4 above, we review the estimation method for large dynamic covariance matrices with conditional sparsity. The main interest of this section lies in estimating sudden changes in the conditionally sparse covariance structure. As the high-dimensional vector $\mathbf{X}_t$ satisfies the factor model assumption, we are particularly interested on identifying whether the breaks originate in the common or idiosyncratic error components.

Existing literature on break estimation and testing for factor models mainly focuses on breaks in the factor loadings. Breitung and Eickmeier (2011) study the hypothesis of a structural break by using the LR, LM and Wald tests; Chen, Dolado and Gonzalo (2014) test the break by regressing one of the PCA estimated factors on the remaining ones; Cheng, Liao and Schorfheide (2016) propose a shrinkage method to estimate the break in the factor loadings; Han and Inoue (2015) and Baltagi, Kao and Wang (2017) observe that breaks in the factor loadings can be transformed to breaks in the factor covariance structure and then test and estimate breaks using the PCA estimates of the transformed factors. We next review this transformation mechanism and then introduce the break detection technique. Consider

$$\mathbf{X}_t = \mathbf{\Gamma}_j \mathbf{F}_{t,j} + \boldsymbol{\varepsilon}_t, \quad \eta^1_{j-1} + 1 \leqslant t \leqslant \eta^1_j, \tag{5.3}$$

where $\mathbf{\Gamma}_j$ is a $p \times k_j$ factor loading matrix, $\mathbf{F}_{t,j}$ is the corresponding k_j-dimensional factor vector when $\eta^1_{j-1} + 1 \leqslant t \leqslant \eta^1_j$, $\eta^1_1, \cdots, \eta^1_{k_1}$ are break times, $\eta^1_0 = 0$ and $\eta^1_{k_1+1} = n$. The factor model formulation (5.3) indicates that structural breaks in covariance of the common components may be caused by either or a combination of the following three scenarios: (i) breaks in the factor loadings; (ii) breaks in the covariance for common factors; and (iii) changes in the factor number. Proposition 2.1 in Li, Li and Fryzlewicz (2020) proves that model (5.3) can be equivalently written as

$$\mathbf{X}_t = \mathbf{\Gamma}^\star \mathbf{F}^\star_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \cdots, n, \tag{5.4}$$

where $\Gamma^\star$ is the transformed factor loading matrix which is time invariant, and $F_t^\star$ denotes the transformed factors with dimension k_0. In addition, the number of factors in (5.3) and that in (5.4) satisfy $\max_{1\leqslant i\leqslant K_1+1} k_i \leqslant k_0 \leqslant \sum_{i=1}^{K_1+1} k_i$. With the transformed factor model (5.4), we may use the classic PCA to estimate the transformed factors, and construct the CUSUM statistic as in (5.1) and (5.2) but with X_t replaced by the PCA estimates of $F_t^\star$. Then, the binary segmentation could be used to estimate the break times $\eta_1^1, \cdots, \eta_{K_1}^1$ as well as the break number K_1.

We next turn to break detection in the covariance structure of X_t defined via (5.3). Observe that the time-invariant matrix decomposition (1.5) is now replaced by

$$\Sigma_{X,t} = \Sigma_{\chi,t} + \Sigma_{\varepsilon,t}, \tag{5.5}$$

where $\Sigma_{\chi,t}$ is the covariance matrix of the common components χ_t with breaks:

$$\Sigma_{\chi,t} = \begin{cases} \Sigma_\chi(1), & 1 \leqslant t \leqslant \eta_1^1, \\ \Sigma_\chi(2), & \eta_1^1 + 1 \leqslant t \leqslant \eta_2^1, \\ \vdots & \vdots \\ \Sigma_\chi(K_1+1), & \eta_{K_1}^1 + 1 \leqslant t \leqslant n, \end{cases} \tag{5.6}$$

and $\Sigma_{\varepsilon,t}$ is the covariance matrix of the idiosyncratic components ε_t with breaks:

$$\Sigma_{\varepsilon,t} = \begin{cases} \Sigma_\varepsilon(1), & 1 \leqslant t \leqslant \eta_1^2, \\ \Sigma_\varepsilon(2), & \eta_1^2 + 1 \leqslant t \leqslant \eta_2^2, \\ \vdots & \vdots \\ \Sigma_\varepsilon(K_2+1), & \eta_{K_2}^2 + 1 \leqslant t \leqslant n. \end{cases} \tag{5.7}$$

From the formulation (5.6) and (5.7), it is clear that there are K_1 breaks in covariance of the common components and K_2 breaks in the idiosyncratic error components. The main challenge of the break detection is that neither the common components χ_t nor the error components ε_t are observable. However, the transformation mechanism given in (5.4) shows that the PCA method can be used to estimate both χ_t and ε_t. Barigozzi, Cho and Fryzlewicz (2018) combine the wavelet transformation and double CUSUM (Cho, 2016) to estimate the breaks. Li, Li and Fryzlewicz (2020) provide an alternative detection method, where breaks in the common components are detected by using the wild binary segmentation for the PCA estimates of $F_t^\star$ and breaks in the error components are detected by the wild sparsified binary segmentation for the estimates of ε_t (similar to that described in Section 5.1).

6 Discussions on future topics

In this section, we discuss some possible future topics which are relevant to the methodologies reviewed in the previous sections.

6.1 *Large dynamic precision matrices*

The precision matrix, the inverse of the covariance matrix, plays an important role in various areas such as discriminant analysis, Gaussian graphical models and optimal portfolio allocation. There has been a rich literature in recent years on estimating large static precision matrices (e.g., Lam and Fan, 2009; Yuan, 2010; Cai, Liu and Luo, 2011; Cai *et al*, 2012; Leng and Tang, 2012). With the non- and semi-parametric estimates of large dynamic covariance matrices introduced in Sections 3 and 4, we may directly take their inverse, obtain the precision matrix estimates and then derive sensible asymptotic theory. For example, Proposition 3 in Chen and Leng (2016) shows that

$$\sup_{C_1+\epsilon\leqslant z\leqslant C_2-\epsilon}\left\|\overline{\mathbf{\Sigma}}_X^{-1}(z)-\mathbf{\Sigma}_X^{-1}(z)\right\|_O=O_P\left(\varpi_2(p)\left([\log p/(nh)]^{1/2}+h^2(\log p)^{1/2}\right)^{1-q}\right),$$

where $\overline{\mathbf{\Sigma}}_X^{-1}(z)$ is the inverse of $\overline{\mathbf{\Sigma}}_X(z)$ defined in (3.4), and Theorem 3 in Wang *et al* (2021) derives the uniform convergence rate for the inverse of $\check{\mathbf{\Sigma}}_X(z)$ defined in Section 4.2.

When the primary interest is to estimate large precision matrices, it is often common to impose a sparsity assumption that is similar to Assumption 1 in Section 1.1. Specifically, we may assume that the large static precision matrix $\mathbf{\Omega}_X:=\mathbf{\Sigma}_X^{-1}$ satisfies $\mathbf{\Omega}_X\in\mathcal{S}_I(q,\varpi_3(p),M_3)$, where

$$\mathcal{S}_I(q,\varpi_3(p),M_3)=\left\{\mathbf{\Omega}=[\omega_{ij}]_{p\times p}\ \middle|\ \mathbf{\Omega}>0,\ \max_{1\leqslant j\leqslant p}\sum_{i=1}^{p}|\omega_{ij}|\leqslant M_3,\ \max_{1\leqslant i\leqslant p}\sum_{j=1}^{p}|\omega_{ij}|^q\leqslant\varpi_3(p)\right\},$$

where "$\mathbf{\Omega}>0$" denotes that $\mathbf{\Omega}$ is positive definite. Commonly-used estimation methods for large static precision matrices include penalised likelihood (Lam and Fan, 2009), graphical Danzig selector (Yuan, 2010) and CLIME (Cai, Liu and Luo, 2011). It would be an interesting future topic to apply these methods (with modifications) to estimate large dynamic precision matrices which evolve smoothly over time. The methodologies reviewed in Sections 3 and 4 may be potentially useful to this future research.

6.2 *Large correlation matrices and their inverse for copula models*

In recent years, there has been increasing interest in the statistical literature on high-dimensional semiparametric Gaussian copula models (e.g., Liu, Lafferty and Wasserman, 2009; Xue and Zou, 2012, 2014; Liu *et al*., 2012). Let $f=\{f_1,\cdots,f_p\}$ be a set of univariate monotone transformation

functions and $\mathbf{R}$ be a $p \times p$ positive definite correlation matrix. Assume that

$$(f_1(x_1), \cdots, f_p(x_p)) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R}).$$

Then $\mathbf{X} = (x_1, \cdots, x_p)^\tau$ follows a p-dimensional nonparanormal model (Liu, Lafferty and Wasserman, 2009). The main interest often lies in estimating the high-dimensional correlation matrix $\mathbf{R}$ and its inverse. Liu, Lafferty and Wasserman (2009) propose a two-step normal-score estimation method. To achieve optimal convergence rates, Liu *et al.* (2012) and Xue and Zou (2012) suggest estimating $\mathbf{R}$ via Spearman's rho or Kendall's tau which is invariant under monotone transformation, and obtaining $\mathbf{R}^{-1}$ via the techniques such as the graphical Danzig selector and CLIME mentioned in Section 6.1.

Some recent econometric literature further introduces dynamic copula models for multivariate and high-dimensional time series (e.g., Creal and Tsay, 2015; Oh and Patton, 2017, 2020; Opschoor *et al.*, 2020). For example, Oh and Patton (2020) propose a dynamic factor copula model by assuming factor loadings to satisfy the generalised autoregressive score model and estimate the large dynamic correlation matrix. It may be also interesting to develop a data-driven approach to estimate large correlation matrices and their inverse under the copula model framework.

6.3 *Large spot volatility matrices for high-frequency data*

Modelling high-frequency financial data, as one of the most important problems in financial economics, has received increasing attention in the past two decades. Continuous-time models like the Brownian semi-martingale and the Itô semi-martingale are often used to capture dynamics of asset prices, where the main component is the volatility function or matrix. In the classic setting with a fixed number of assets, there have been extensive studies on using the realised volatility to estimate the so-called integrated volatility over a fixed time interval (e.g., Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002). In recent years, due to wide availability of large data sets, financial practitioners often need to deal with a large amount of high-frequency financial data collected for a large number of assets. So it becomes imperative to estimate large volatility structures for high-frequency data. Motivated by recent developments in large covariance matrix estimation for low-frequency data (see the review in Sections 1.1 and 1.2), Wang and Zou (2010) and Tao, Wang and Zhou (2013) estimate large volatility matrices satisfying the sparsity assumption; whereas Aït-Sahalia and Xiu (2017) and Dai, Lu and Xiu (2019) assume the factor model for high-frequency data and then estimate conditionally sparse volatility matrices.

The estimation methods in the aforementioned literature cannot capture the large instantaneous/spot volatility structure and its dynamic change over time. Existing literature on spot volatility estimation usually considers the low-dimensional high-frequency data setting with a single or a fixed number of assets (e.g., Kristensen, 2010; Zu and Boswijk, 2014; Kanaya and

Kristensen, 2016). It would be an interesting future topic to combine the spot volatility estimation technique with the large dynamic covariance matrix estimation methods reviewed in Sections 3 and 4 to estimate large spot volatility matrices. However, such an extension is non-trivial as there often exist market microstructure noises and non-synchronisation for large high-frequency financial data. Consequently, additional techniques such as the pre-averaging (Jacod *et al.*, 2009; Christensen, Kinnebrock and Podolskij, 2010) and the generalised sampling time (Aït-Sahalia, Fan and Xiu, 2010) need to be incorporated in the estimation procedure.

6.4 *Large spectral density matrix estimation*

The main focus of this review article is to estimate the contemporaneous second-order moment structure with possible time-variation. In practice, it is often interesting to further explore the entire second-order moment structure including the autocovariance. The spectral density which relies on the Fourier transform of autocovariance functions is a commonly-used tool to address this problem. It is essentially the frequency domain analogue of autocovariances. Letting $\boldsymbol{\Delta}(k) = \mathsf{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})$, the spectral density function/matrix is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \boldsymbol{\Delta}(k) e^{-ik\omega}, \quad \omega \in [-\pi, \pi].$$

For the low-dimensional time series, as in Brockwell and Davis (2009), we may estimate $f(\omega)$ by

$$\widehat{f}_m(\omega) = \frac{1}{2\pi(2m+1)} \sum_{|j| \leqslant m} I(\omega + \omega_j),$$

where $\omega_j = 2\pi j/n$, m is a user-specified truncation parameter and $I(\omega)$ is the periodogram defined as

$$I(\omega) = \sum_{k=-(n-1)}^{n-1} \widehat{\Delta}(k) e^{-ik\omega}$$

with $\widehat{\boldsymbol{\Delta}}(k)$ being the sample version of the k-th autocovariance. For the high-dimensional time series, as recommended by Sun *et al.* (2018) and Fiecas *et al.* (2019), we may apply the generalised shrinkage technique to $\widehat{f}_m(\omega)$. Specifically, let $\widehat{f}_{m,ij}(\omega)$ be the (i,j)-entry of $\widehat{f}_m(\omega)$ and construct

$$\overline{f}_m(\omega) = \left[\overline{f}_{m,ij}(\omega)\right], \quad \overline{f}_{m,ij}(\omega) = s_\rho\left(\widehat{f}_{m,ij}(\omega)\right),$$

where $s_\rho(\cdot)$ is defined as in Section 2.1. Sun *et al.* (2018) and Fiecas *et al.* (2019) derive the consistency results for $\overline{f}_m(\omega)$ under different dependence conditions on large time series. It seems worthwhile to further study the time-varying spectral density matrix for large time series with possible structural changes or breaks, and extend the methodology and theory which have been developed

by Dahlhaus (1997, 2009), Preuss, Puchstein and Dette (2015) and Kawka (2020) for low-dimensional time series.

Acknowledgments

The author thanks the Co-Editor, an Associate Editor and two reviewers for the valuable comments, which improve a former version of the article. The author is partly supported by the BA/Leverhulme Small Research Grant funded by the Leverhulme Trust (No. SRG1920/100603)

References

AHN, S. AND HORENSTEIN, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

AÏT-SAHALIA, Y. FAN, J. AND XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* **105** 1504-1517.

AÏT-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**, 384–399.

ANDERSEN, T. G. AND BOLLERSLEV, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**, 885–905.

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd Edition)*. Wiley Series in Probability and Statistics.

ASTON, J. A. AND KIRCH, C. (2018). High dimensional efficiency with applications to change point tests. *Electronic Journal of Statistics* **12**, 1901–1947.

AUE, A., HÖRMANN, HORVÁTH, L. AND REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**, 4046–4087.

BAI, J. AND NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **90**, 191–221.

BAI, J. AND SHI, S. (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance* **12**, 199–215.

BALTAGI, B., KAO, C. AND WANG, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics* **197**, 87–100.

BARIGOZZI, M., CHO, H. AND FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics* **206**, 187–225.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B* **64**, 253–280.

BAUWENS, L., LAURENT, S. AND ROMBOUTS, J. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics* **21**, 79–109.

BICKEL, P. AND LEVINA, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

BICKEL, P. AND LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

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

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

BOLLERSLEV, T., ENGLE, R. F. AND WOOLDRIDGE, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy* **96**, 116–131.

BOSE, A. AND BHATTACHARJEE, M. (2018). *Large Covariance and Autocovariance Matrices*. Chapman and Hall/CRC.

BREITUNG, J. AND EICKMEIER, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* **163**, 71–84.

BROCKWELL, P. J. AND DAVIS, R. A. (2009). *Time Series: Theory and Methods* (2nd Edition). Springer.

CAI, T. T., LI, H., LIU, W. AND XIE, J. (2012). Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika* **100**, 139–156.

CAI, T. T. AND LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

CAI, T. T., LIU, W. AND LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106**, 594–607.

CAI, T. T., REN, Z. AND ZHOU, H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10**, 1–59.

CAI, T. T. AND ZHOU, H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics* **40**, 2389–2420.

CHAMBERLAIN, G. AND ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305–1324.

CHEN, J., LI, D. AND LINTON, O. (2019). A new semiparametric estimation approach of large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics* **212**, 155–176.

CHEN, J., LI, D., LINTON, O. AND LU, Z. (2016). Semiparametric dynamic portfolio choice with multiple conditioning variables. *Journal of Econometrics* **194**, 309–318.

CHEN, J., LI, D., LINTON, O. AND LU, Z. (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association* **113**, 919–932 .

CHEN, L., DOLADO, J. J. AND GONZALO, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics* **180**, 30–48.

CHEN, L. H., CHENG, M. Y. AND PENG, L. (2009). Conditional variance estimation in heteroscedastic regression models. *Journal of Statistical Planning and Inference* **139**, 236–245.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*,**76**, North Holland, Amsterdam.

CHEN, X., XU, M. AND WU, W. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics* **41**, 2994–3021.

CHEN, Z. AND LENG, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association* **111**, 1196–1207.

CHENG, X., LIAO, Z. AND SCHORFHEIDE, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies* **83**, 1511–1543.

CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics* **10**, 2000–2038.

CHO, H. AND FRYZLEWICZ P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society Series B* **77**, 475–507.

CHRISTENSEN, K., KINNEBROCK, S. AND PODOLSKIJ, M.. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* **159**, 116–133.

CREAL, D. D. AND TSAY, R. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics* **189**, 335–345.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

DAHLHAUS, R. (2009). Local inference of locally stationary time series based on the empirical spectral measure. *Journal of Econometrics* **151**, 101–112.

DAI, C., LU, K. AND XIU, D. (2019). Knowing factors or factor loadings, or neither? Evaluating estimators for large covariance matrices with noisy and asynchronous data. *Journal of Econometrics* **208**, 43–79.

DANIELS, M. AND POURAHMADI, M. (2002). Dynamic models and Bayesian analysis of covariance matrices in longitudinal data. *Biometrika* **89**, 553–566.

ENGLE, R. F. (1982). Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

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

ENGLE, R. F. AND KRONER K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.

ENGLE, R. F., LEDOIT, O. AND WOLF, M. (2019). Large dynamic covariance matrices. *Journal of Business and Economic Statistics* **37**, 363–375.

ENGLE, R. F., NG, V. AND ROTHSCHILD M. (1990). Asset pricing with a factor-ARCH structure: Empirical estimates for Treasury bills. *Journal of Econometrics* **45**, 213–237.

FAMA, E. AND FRENCH, K. (1992). The cross-section of expected stock returns. *Journal of Finance* **47**, 427–465.

FAN, J., FENG, Y., JIANG, J. AND TONG, X. (2016). Feature augmentation via nonparametrics and selection (FANS) in high dimensional classification. *Journal of American Statistical Association* **111**, 275–287.

FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

FAN, J. AND LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

FAN, J., LIAO, Y. AND LIU, H. (2016). An overview on the estimation of large covariance and precision matrices. *Econometrics Journal* **19**, 1–32.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics* **39**, 3320–3356.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* **75**, 603–680.

FAN, J., LIU, H. AND WANG, W. (2018). Large covariance estimation through elliptical factor models. *The Annals of Statistics* **46**,1383–1414.

FAN, J. AND YAO, Q. (1998). Efficient estimation of conditional variance function in stochastic regression. *Biometrika* **85**, 645–660.

FIECAS, M., LENG, C., LIU, W. AND YU, Y. (2019). Spectral analysis of high-dimensional time series. *Electronic Journal of Statistics* **13**, 4079–4101.

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics* **42**, 2243–2281.

GUO, S., BOX, J. AND ZHANG, W. (2017). A dynamic structure for high dimensional covariance matrices and its application in portfolio allocation. *Journal of the American Statistical Association* **112**, 235–253.

HAFNER, C. AND PREMINGER, A. (2009). Asymptotic theory for a factor GARCH model. *Econometric Theory* **25**, 336–363.

HAN X. AND INOUE, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory* **31**, 1117–1152.

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. AND VETTER, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications* **119**, 2249–2276.

JIRAK, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics* **43**, 2451–2483.

KANAYA, S. AND KRISTENSEN, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* **32**, 861–916.

KAWKA, R. (2020). Convergence of spectral density estimators in the locally stationary framework. Forthcoming in *Econometrics and Statistics*.

KRISTENSEN, D. (2010). Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory* **26**, 60–93.

LAM, C. AND FAN, J. (2009). Sparsity and rates of convergence in large covariance matrix estimation. *The Annals of Statistics* **37**, 4254–4278.

LAM, C. AND YAO, Q. (2012). Factor modelling for high-dimensional time series: Inference for the number of factor. *The Annals of Statistics* **40**, 694–726.

LEDOIT, O.. AND WOLF, M. (2012). Nonlinear shrinkage and estimation of large-dimensional covariance matrices. *The Annals of Statistics* **40**, 1024–1060.

LEDOIT, O.. AND WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis* **139**, 360–384.

LENG, C. AND TANG, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association* **107**, 1187–1200.

LIU, H., LAFFERTY, J. AND WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research* **10**, 2295–2328.

LIU, H., HAN, F., YUAN, M., LAFFERTY, J. AND WASSERMAN, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40**, 2293–2326.

LI, D., LINTON, O. AND LU, Z. (2015). A flexible semiparametric forecasting model for time series. *Journal of Econometrics* **187**, 345–357.

LI, D., QIAN, J. AND SU, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* **111**, 1804–1819.

LI, Y., LI, D. AND FRYZLEWICZ, P. (2020). Detection of multiple structural breaks in large covariance matrices. Working paper available at http://stats.lse.ac.uk/fryzlewicz/wbscov/wbscov.pdf.

MA, S. AND SU, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics* **207**, 1–29.

MIKKELSEN, J. G., HILLEBRAND, E. AND URGA, G. (2019). Consistent estimation of time-varying loadings in high-dimensional factor models. *Journal of Econometrics* **208**, 535–562.

MORANA, C. (2019). Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices. *Econometrics and Statistics* **12**, 42–65.

MOTTA, G., HAFNER, C. AND VON SACHS, R. (2011). Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* **27**, 1279–1319.

OH, D. H. AND PATTON A. J. (2017). Modelling dependence in high dimensions with factor copulas. *Journal of Business and Economic Statistics* **35**, 139–154.

OH, D. H. AND PATTON A. J. (2020). Dynamic factor copula models with estimated cluster assignments. *Working paper*.

OPSCHOOR, A., LUCAS, A., BARRA I. AND VAN DIJK, D. (2020). Closed-form multi-factor copula models with observation-driven dynamic factor loadings. Forthcoming in *Journal of Business and Economic Statistics*.

PAKEL, C., SHEPHARD, N., SHEPPARD, K. AND ENGLE, R. F. (2020). Fitting vast dimensional time-varying covariance models. Forthcoming in *Journal of Business and Economic Statistics*.

PELGER, M. AND XIONG, R. (2019). State-varying factor models of large dimensions. Working paper available at https://arxiv.org/pdf/1807.02248.pdf.

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Wiley.

PREUSS, P., PUCHSTEIN, R. AND DETTE, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* **110**, 654–668.

RAJARATNAM, B., MASSAM, H. AND CORVALLO, C. (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of Statistics* **36**, 2818–2849.

ROTHMAN, A. J., LEVINA, E. AND ZHU, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104**, 177–186.

SILVENNOINEN, A. AND TERÄSVIRTA, T. (2009). Multivariate GARCH models. *Handbook of Financial Time Series* (ed: T. G. Anderson, R. A. Davis, J. P. Kreiß, T. Mikosch), 201–229.

STOCK, J. H. AND WATSON, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry* (ed: J. Castle and N. Shephard). Oxford University Press, pp. 173–205.

SU, L. AND WANG, X. (2017). On time-varying factor models: estimation and testing. *Journal of Econometrics* **198**, 84–101.

SUN, Y., LI, Y., KUCEYESKI, A. AND BASU, S. (2018). Large spectral density matrix estimation by thresholding. Working paper available at https://arxiv.org/abs/1812.00532.

TAO, M., WANG, Y. AND ZHOU, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *The Annals of Statistics* **41**, 1816–1864.

VAN DER WEIDE, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* **17**, 549–564.

VOSTRIKOVA L. YU. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady* **24**, 55–59.

VRONTOS, I. D., DELLAPORTAS, P. AND POLITIS, D. N. (2003). A full-factor multivariate GARCH model. *Econometrics Journal* **6**, 311–333.

WAND, M. P. AND JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall.

WANG, D., YU, Y. AND RINALDO, A. (2021). Optimal covariance change point localization in high dimensions. *Bernoulli* **27**, 554–575.

WANG, H., PENG, B., LI, D. AND LENG, C. (2021). Nonparametric estimation of large covariance matrices with conditional sparsity. Forthcoming in *Journal of Econometrics*.

WANG, T. AND SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B* **80**, 57–83.

WANG, Y. AND ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics* **38**, 943–978.

XUE, L. AND ZOU, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics* **40**, 2541–2571.

XUE, L. AND ZOU, H. (2014). Optimal estimation of sparse correlation matrices of semiparametric Gaussian copulas. *Statistics and Its Interface* **7**, 201–209.

YANG, R. AND BERGER, J. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics* **22**, 1195–1121.

YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research* **11**, 2261–2286.

YU, K. AND JONES, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association* **99**, 139–144.

ZU, Y. AND BOSWIJK, H. P. (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics* **181**, 117–135.