A New Semiparametric Estimation of Large Dynamic Covariance Matrix with Multiple Conditioning Variables

Article *in* SSRN Electronic Journal · September 2017

DOI: 10.2139/ssrn.3210726

Jia Chen
The University of York

Degui Li
The University of York

Oliver Linton
University of Cambridge

A New Semiparametric Estimation of Large Dynamic Covariance Matrix with Multiple Conditioning Variables

Jia Chen* Degui Li† Oliver Linton‡

Version: September 21, 2017

Abstract

This paper studies estimation of dynamic covariance matrix with multiple conditioning variables, where the matrix size can be ultra large (divergent at an exponential rate of the sample size). We introduce an easy-to-implement semiparametric method to estimate each entry of the covariance matrix via model averaging marginal regression, and then apply a shrinkage technique to obtain an estimate of the large dynamic covariance matrix estimation. Under some regularity conditions, we derive the asymptotic properties for the proposed estimators including the uniform consistency with general convergence rates. We also consider extending our methodology to deal with the scenario where the number of conditioning variables is diverging. Simulation studies are conducted to illustrate the finite-sample performance of the developed methodology.

Keywords: dynamic covariance matrix, MAMAR, semiparametric estimation, sparsity, uniform consistency.

*Department of Economics and Related Studies, University of York, YO10 5DD, UK. E-mail: jia.chen@york.ac.uk
†Department of Mathematics, University of York, YO10 5DD, UK. E-mail: degui.li@york.ac.uk.
‡Faculty of Economics, University of Cambridge, CB3 9DD, UK. E-mail: obl20@cam.ac.uk.

1 Introduction

The classical theory of mean/variance portfolio choice was developed by Markowitz (1952), see Merton (1969) and Fama (1970) for some other important developments. More recently this topic has been at the centre of a lot of research, see Back (2010) and Brandt (2010) for some recent surveys. In practice, it is not uncommon that the dynamic portfolio choice depends on many conditioning (or forecasting) variables, which reflects the varying investment opportunities over the time. Generally speaking, there are two ways to describe the dependence of portfolio choice on the conditioning variables. One is to assume a parametric model that relates the returns of risky assets to the conditioning variables and then solve for an investor's portfolio choice using some traditional econometric approaches to estimate the unknown parameters. However, the assumed parametric models might be misspecified, which would lead to inconsistent estimation of the optimal portfolio and invalid inference. One way to avoid the possible model misspecification issue is to use some nonparametric methods such as the kernel estimation method to describe the dependence of the portfolio choice on conditioning variables. The latter method is introduced in Brandt (1999) in the case of a univariate conditioning variable. Aït-Sahalia and Brandt (2001) further develop a single-index strategy to handle multiple conditioning variables. This literature has worked with the case where the number of assets is fixed and relatively small. However, another literature has considered the case where there are no covariates but there are a large number of assets (c.f., Ledoit and Wolf, 2003, 2004, 2014; Kan and Zhou, 2007; Fan, Fan and Lv, 2008; DeMiguel *et al*, 2009; DeMiguel, Garlappi and Uppal, 2009; Pesaran and Zaffaroni, 2009; Frahm and Memmel, 2010; Tu and Zhou, 2011).

As seen from the aforementioned literature, accurate covariance matrix estimation plays a crucial role in portfolio choice problem. In this paper suppose that the observations $X_t = (X_{t1}, \ldots, X_{td})^\intercal$, $t = 1, \ldots, n$, are collected from a d-dimensional stationary process with covariance matrix $\mathsf{E}[(X_t - EX_t)(X_t - EX_t)^\intercal] = \mathbf{\Sigma}$, where the matrix $\mathbf{\Sigma}$ is invariant over time. There have been extensive studies on estimating such a *static* covariance matrix. For instance, when the dimension d is fixed or significantly smaller than the sample size n, $\mathbf{\Sigma}$ can be consistently estimated by the sample covariance matrix (c.f. Anderson, 2003):

$$\overline{\mathbf{\Sigma}} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})(X_t - \overline{X})^\intercal, \quad \overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{1.1}$$

However, the above conventional sample covariance matrix would fail when the dimension d is large and exceeds the sample size n. In the latter case, the matrix $\overline{\mathbf{\Sigma}}$ becomes singular. In order to obtain a proper estimation of $\mathbf{\Sigma}$ when $d > n$, some structural assumptions such as sparsity and factor modelling are usually imposed in the literature, and then various regularisation techniques are used to produce consistent and reliable estimates (c.f., Wu and Pourahmadi, 2003; Bickel and Levina, 2008a,b; Lam and Fan, 2009; Rothman, Levina and Zhu, 2009; Cai and Liu, 2011; Fan, Liao and Mincheva, 2013).

The aforementioned literature on large covariance matrix estimation assumes that the underlying covariance matrix is constant over time. Such an assumption is very restrictive and may be violated in many practical applications such as in dynamic optimal portfolio allocation (Guo, Box and Zhang, 2017). This motivates us to consider a dynamic large covariance matrix, whose entries may evolve over time. In recent years, there have been increasing interests in estimating dynamic covariance or correlation matrices and exploring their applications. For example, Engle (2002) uses the parametric multivariate GARCH modelling method to estimate dynamic conditional correlation; Guo, Box and Zhang (2017) combine semiparametric adaptive functional-coefficient and GARCH modelling approaches to estimate dynamic covariance structure with the dimension d diverging at a polynomial rate of n; Chen, Xu and Wu (2013) and Chen and Leng (2016) use the kernel smoothing method to nonparametrically estimate each entry in the dynamic covariance matrix and then apply the thresholding or generalised shrinkage technique when the dimension d can be divergent at an exponential rate but the conditioning variable is univariate; Engle, Ledoit and Wolf (2016) extends Engle (2002)'s dynamic conditional correlation models to large dimensional case using a nonlinear shrinkage technique derived from the random matrix theory.

Let $U_t = (U_{t1}, \ldots, U_{tp})^\intercal$ be a p-dimensional vector of conditioning variables which are stationary over time. We consider the conditional covariance matrix of X_{t+1} given U_t:

$$\mathbf{\Sigma}_0(u) = \mathsf{E}\left(X_{t+1}X_{t+1}^\intercal | U_t = u\right) - \left[\mathsf{E}(X_{t+1}|U_t = u)\right]\left[\mathsf{E}(X_{t+1}|U_t = u)\right]^\intercal,$$

where $u = (u_1, \ldots, u_p)^\intercal$ is a vector of fixed constants. To simplify notation, we let

$$\mathcal{C}_0(u) = \mathsf{E}\left(X_{t+1}X_{t+1}^\intercal | U_t = u\right) \quad \text{and} \quad \mathcal{M}_0(u) = \mathsf{E}(X_{t+1}|U_t = u),$$

and rewrite the conditional covariance matrix as

$$\mathbf{\Sigma}_0(u) = \mathcal{C}_0(u) - \mathcal{M}_0(u)\mathcal{M}_0^\intercal(u). \tag{1.2}$$

In order to estimate $\mathbf{\Sigma}_0(u)$, one only needs to estimate $\mathcal{C}_0(u)$ and $\mathcal{M}_0(u)$. A natural way to estimate $\mathcal{C}_0(u)$ and $\mathcal{M}_0(u)$ is via nonparametric smoothing. However, although the nonparametric estimation is robust to model misspecification, its finite-sample performance is often poor when the dimension of conditioning variables U_t, p, is moderately large (or even as small as three), owing to the "curse of dimensionality". Therefore, when U_t is a multivariate vector, a direct use of the nonparametric kernel approach as in Chen, Xu and Wu (2013) or Chen and Leng (2016) would be inappropriate, and an alternative technique is needed. In practice, many variables including past lags, momentum measures, seasonal dummy variables, past earnings, transaction volume, have been used to predict the mean and variance of stock returns.

Letting $\sigma_{ij}^0(u)$ and $c_{ij}^0(u)$ be the (i,j)-entry of the matrices $\mathbf{\Sigma}_0(u)$ and $\mathcal{C}_0(u)$, respectively, and $m_i^0(u)$ be the i-th element of $\mathcal{M}_0(u)$, it follows from (1.2) that

$$\sigma_{ij}^0(u) = c_{ij}^0(u) - m_i^0(u)m_j^0(u), \quad 1 \leq i, j \leq d. \tag{1.3}$$

Instead of estimating $m_i^0(u)$ and $c_{ij}^0(u)$ directly via nonparametric smoothing, we approximate them using the Model Averaging MArginal Regression (MAMAR) approximation (Li, Linton and Lu, 2015), i.e.,

$$m_i^0(u) \approx b_{i,0} + \sum_{k=1}^{p} b_{i,k}\mathsf{E}(X_{t+1,i}|U_{tk} = u_k) =: b_{i,0} + \sum_{k=1}^{p} b_{i,k}m_{i,k}(u_k), \quad 1 \leq i \leq d, \tag{1.4}$$

where $b_{i,k}$ are unknown parameters which may be regarded as "weights" for marginal mean regression models; and similarly for $c_{ij}^0(u)$

$$c_{ij}^0(u) \approx a_{ij,0} + \sum_{k=1}^{p} a_{ij,k}\mathsf{E}(X_{t+1,i}X_{t+1,j}|U_{tk} = u_k) =: a_{ij,0} + \sum_{k=1}^{p} a_{ij,k}c_{ij,k}(u_k), \quad 1 \leq i, j \leq d, \tag{1.5}$$

where $a_{ij,k}$ are unknown weighting parameters. In (1.4) and (1.5), both $m_{i,k}(u_k)$ and $c_{ij,k}(u_k)$ are univariate nonparametric functions and can be well estimated by commonly-used nonparametric methods without incurring the curse of dimensionality. The MAMAR method provides an alternative way to estimate nonparametric joint mean regression with multiple regressors. The MAMAR approximation is introduced by Li, Linton and Lu (2015) in a semiparametric setting, and is applied to semiparametric dynamic portfolio choice by Chen *et al* (2016) and further generalised to the ultra-high dimensional time series setting by Chen *et al* (2017). A similar idea is also used by Fan *et al* (2016) in high-dimensional classification.

The accuracy of the MAMAR approximation to the joint regression functions relies on the choice of the weight parameters, e.g., $b_{i,k}$ and $a_{ij,k}$ in (1.4) and (1.5), respectively. Section 2.1 below derives the theoretically optimal weights and consequently obtains a proxy, $\mathbf{\Sigma}_A^\star(u)$, of the true dynamic covariance matrix $\mathbf{\Sigma}_0(u)$. A two-stage semiparametric method is proposed to estimate each entry of $\mathbf{\Sigma}_A^\star(u)$: in stage 1, the kernel smoothing method is used to estimate the marginal regression functions $m_{i,k}(u_k)$ and $c_{ij,k}(u_k)$; in stage 2, the least squares method is used to estimate the optimal weights in the MAMAR approximation by replacing the marginal regression functions with their estimates obtained from stage 1 and then treating them as "regressors" in approximate linear models associated with $m_i^0(u)$ and $c_{ij}^0(u)$. Based on the above, an estimate of the optimal MAMAR approximation of $m_i^0(u)$ and $c_{ij}^0(u)$ can be constructed via (1.4) and (1.5), and subsequently the optimal MAMAR approximation of $\sigma_{ij}^0(u)$ can be estimated via (1.3). Finally, a generalised shrinkage technique is applied to the obtained covariance matrix to produce a non-degenerate estimate that has its small entries forced to zero. Under some mild conditions and the assumption that $\mathbf{\Sigma}_A^\star(u)$ is approximately sparse, we derive the uniform consistency results for estimators of $\mathbf{\Sigma}_A^\star(u)$ and its inverse. These results also hold for the true covariance matrix $\mathbf{\Sigma}_0(u)$ as long as $\mathbf{\Sigma}_A^\star(u)$ and $\mathbf{\Sigma}_0(u)$ are sufficiently "close". The sparsity result for the semiparametric shrinkage estimator is also established.

The rest of the paper is organised as follows. Section 2 derives the optimal weights in the MAMAR approximation (1.4) and (1.5), and introduces the semiparametric shrinkage method to estimate the dynamic covariance matrix. Section 3 gives the limit theorems of the developed estimators. Section 4 introduces a modification technique to guarantee the positive definiteness of the dynamic covariance matrix estimation, and discusses the choice of tuning parameter in the generalised shrinkage method. Section 5 reports finite-sample simulation studies of our methodology. Section 6 concludes the paper and discusses some possible extensions. The proofs of the main results and some technical lemmas are given in the appendix. Throughout the paper, we use $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ to denote the minimum and maximum eigenvalues of a matrix; $\|\cdot\|_O$ to denote the operator (or spectral) norm defined as $\|\mathbf{\Delta}\|_O = \sup_{\mathbf{x}} \{\|\mathbf{\Delta x}\| : \|\mathbf{x}\| = 1\}$ for a $q \times q$ matrix $\mathbf{\Delta} = (\delta_{ij})_{q\times q}$, where $\|\mathbf{x}\| = (\sum_{i=1}^q x_i^2)^{1/2}$ is the Euclidean norm; and $\|\cdot\|_F$ to denote the Frobenius norm defined as$\|\mathbf{\Delta}\|_F = \left(\sum_{i=1}^q \sum_{j=1}^q \delta_{ij}^2\right)^{1/2} = \mathsf{Tr}^{1/2}(\mathbf{\Delta\Delta}^\intercal)$, where $\mathsf{Tr}(\cdot)$ denotes the trace of a matrix.

2 Estimation methodology

In this section we introduce an estimation method for the dynamic covariance matrix via the MAMAR approximation. It combines a semiparametric least squares method and the generalised shrinkage technique to produce reliable large covariance matrix estimation. We start with an introduction of the MAMAR approximation in our context and then derive the theoretically optimal weights for the approximation.

2.1 Optimal weights in the MAMAR approximation

For each $k = 0, 1, \ldots, p$, let $\mathcal{A}_k = (a_{ij,k})_{d\times d}$ be a matrix consisting of the weights in (1.5) and $\mathcal{C}_k(u_k) = [c_{ij,k}(u_k)]_{d\times d}$ be a matrix consisting of the conditional means of $X_{t+1,i}X_{t+1,j}$ (for given $U_{tk} = u_k$) in (1.5). Then, the MAMAR approximation for $\mathcal{C}_0(u)$ can be written in matrix form as

$$\mathcal{C}_0(u) \approx \mathcal{A}_0 + \mathcal{A}_1 \odot \mathcal{C}_1(u_1) + \cdots + \mathcal{A}_p \odot \mathcal{C}_p(u_p) =: \mathcal{C}_A(u), \tag{2.1}$$

where $\odot$ denotes the Hadamard product. Similarly, we have the following MAMAR approximation for $\mathcal{M}_0(u)$

$$\mathcal{M}_0(u) \approx \mathcal{B}_0 + \mathcal{B}_1 \odot \mathcal{M}_1(u_1) + \cdots + \mathcal{B}_p \odot \mathcal{M}_p(u_p) =: \mathcal{M}_A(u), \tag{2.2}$$

where for $k = 0, 1, \ldots, p$, $\mathcal{B}_k = (b_{1,k}, b_{2,k}, \ldots, b_{d,k})^\intercal$ is the vector consisting of the weights in (1.4) and $\mathcal{M}_k(u_k) = [m_{1,k}(u_k), m_{2,k}(u_k), \ldots, m_{d,k}(u_k)]^\intercal$ is the vector consisting of the conditional means of $X_{t+1,i}$ (for given $U_{tk} = u_k$) in (1.4). Combining (2.1) and (2.2), we have the following MAMAR approximation for $\mathbf{\Sigma}_0(u)$

$$\begin{aligned}\mathbf{\Sigma}_0(u) &\approx \left[\mathcal{A}_0 + \sum_{k=1}^{p} \mathcal{A}_k \odot \mathcal{C}_k(u_k)\right] - \left[\mathcal{B}_0 + \sum_{k=1}^{p} \mathcal{B}_k \odot \mathcal{M}_k(u_k)\right]\left[\mathcal{B}_0 + \sum_{k=1}^{p} \mathcal{B}_k \odot \mathcal{M}_k(u_k)\right]^\intercal \\ &= \mathcal{C}_A(u) - \mathcal{M}_A(u)\mathcal{M}_A^\intercal(u) =: \mathbf{\Sigma}_A(u).\end{aligned} \tag{2.3}$$

The matrix $\mathbf{\Sigma}_A(u)$ on the right hand side of (2.3) can be viewed as a semiparametric approximation of $\mathbf{\Sigma}_0(u)$, in which the weights $a_{ij,k}$ and $b_{i,k}$ play an important role. These weights have to be appropriately chosen in order to achieve optimal MAMAR approximation. We next derive the theoretically optimal weights. For $1 \le i, j \le d$, we may choose the optimal weights $a^\star_{ij,k}$, $k = 0, 1, \ldots, p$, so that they minimise

$$\mathsf{E}\left[X_{t+1,i}X_{t+1,j} - a_{ij,0} - \sum_{k=1}^{p} a_{ij,k}\mathsf{E}(X_{t+1,i}X_{t+1,j}|U_{tk})\right]^2.$$

Following standard calculations (c.f., Li, Linton and Lu, 2015), we have the following solution for the theoretically optimal weights

$$\left(a^\star_{ij,1}, \ldots, a^\star_{ij,p}\right)^\intercal = \mathbf{\Omega}^{-1}_{XX,ij}\mathbf{V}_{XX,ij}, \quad a^\star_{ij,0} = \left(1 - \sum_{k=1}^{p} a^\star_{ij,k}\right)\mathsf{E}(X_{ti}X_{tj}), \tag{2.4}$$

where $\mathbf{\Omega}_{XX,ij}$ is a $p \times p$ matrix with the (k, l)-entry being

$$\omega_{ij,kl} = \mathsf{Cov}\left[\mathsf{E}(X_{t+1,i}X_{t+1,j}|U_{tk}), \mathsf{E}(X_{t+1,i}X_{t+1,j}|U_{tl})\right] = \mathsf{Cov}\left[c_{ij,k}(U_{tk}), c_{ij,l}(U_{tl})\right],$$

and $\mathbf{V}_{XX,ij}$ is a p-dimensional column vector with the k-th element being

$$v_{ij,k} = \mathsf{Cov}\left[\mathsf{E}(X_{t+1,i}X_{t+1,j}|U_{tk}), X_{t+1,i}X_{t+1,j}\right] = \mathsf{Cov}\left[c_{ij,k}(U_{tk}), X_{t+1,i}X_{t+1,j}\right] = \mathsf{Var}\left[c_{ij,k}(U_{tk})\right].$$

We thus can obtain the optimal weight matrix $\mathcal{A}^\star_k$ from $a^\star_{ij,k}$, $k = 0, 1, \ldots, p$, and subsequently the theoretically optimal MAMAR approximation to $\mathcal{C}_0(u)$:

$$\mathcal{C}^\star_A(u) = \mathcal{A}^\star_0 + \sum_{k=1}^{p} \mathcal{A}^\star_k \odot \mathcal{C}_k(u_k). \tag{2.5}$$

Similarly, we can derive the optimal weights $b^\star_{i,k}$ in the MAMAR approximation (1.4):

$$\left(b^\star_{i,1}, \ldots, b^\star_{i,p}\right)^\intercal = \mathbf{\Omega}_{X,i}^{-1}\mathbf{V}_{X,i}, \quad b^\star_{i,0} = \left(1 - \sum_{k=1}^{p} b^\star_{i,k}\right) \mathsf{E}(X_{ti}), \tag{2.6}$$

where $\mathbf{\Omega}_{X,i}$ is a $p \times p$ matrix with the (k, l)-entry being

$$\omega_{i,kl} = \mathsf{Cov}\left[\mathsf{E}(X_{t+1,i}|U_{tk}), \mathsf{E}(X_{t+1,i}|U_{tl})\right] = \mathsf{Cov}\left[m_{i,k}(U_{tk}), m_{i,l}(U_{tl})\right],$$

and $\mathbf{V}_{X,i}$ is a p-dimensional column vector with the k-th element being

$$v_{i,k} = \mathsf{Cov}\left[\mathsf{E}(X_{t+1,i}|U_{tk}), X_{t+1,i}\right] = \mathsf{Cov}\left[m_{i,k}(U_{tk}), X_{t+1,i}\right] = \mathsf{Var}\left[m_{i,k}(U_{tk})\right].$$

We can then obtain the optimal weight vector $\mathcal{B}^\star_k$ from $b^\star_{i,k}$, $k = 0, 1, \ldots, p$, and consequently the optimal MAMAR approximation to $\mathcal{M}_0(u)$:

$$\mathcal{M}^\star_A(u) = \mathcal{B}^\star_0 + \sum_{k=1}^{p} \mathcal{B}^\star_k \odot \mathcal{M}_k(u_k). \tag{2.7}$$

Combining (2.3), (2.5) and (2.7), we obtain the optimal MAMAR approximation to $\mathbf{\Sigma}_0(u)$:

$$\mathbf{\Sigma}^\star_A(u) = \mathcal{C}^\star_A(u) - \mathcal{M}^\star_A(u)\left[\mathcal{M}^\star_A(u)\right]^\intercal \tag{2.8}$$

The matrix $\mathbf{\Sigma}^\star_A(u)$ serves as a proxy for $\mathbf{\Sigma}_0(u)$. Our aim is to consistently estimate $\mathbf{\Sigma}^\star_A(u)$. This will be done by a semiparametric shrinkage method.

2.2 Semiparametric shrinkage estimation

We next introduce a two-stage semiparametric method to estimate $m^0_i(u)$ and $c^0_{ij}(u)$, respectively.

Stage 1. As both $m_{i,k}(u_k)$ and $c_{ij,k}(u_k)$ are univariate functions, they can be well estimated by the kernel method, i.e.,

$$\widehat{m}_{i,k}(u_k) = \left[\sum_{t=1}^{n-1} K\left(\frac{U_{tk} - u_k}{h_1}\right) X_{t+1,i}\right] \Big/ \left[\sum_{t=1}^{n-1} K\left(\frac{U_{tk} - u_k}{h_1}\right)\right], \quad 1 \le k \le p, \ 1 \le i \le d,$$

and

$$\widehat{c}_{ij,k}(u_k) = \left[\sum_{t=1}^{n-1} K\left(\frac{U_{tk} - u_k}{h_2}\right) X_{t+1,i}X_{t+1,j}\right] \Big/ \left[\sum_{t=1}^{n-1} K\left(\frac{U_{tk} - u_k}{h_2}\right)\right], \quad 1 \le k \le p, \ 1 \le i, j \le d,$$

where $K(\cdot)$ is a kernel function, h_1 and h_2 are two bandwidths. Other nonparametric estimation methods such as the local polynomial method (Fan and Gijbels, 1996) and the sieve method (Chen, 2007) are equally applicable here.

Stage 2. With the kernel estimates obtained in stage 1, obtain the following approximate linear regression models:

$$X_{t+1,i} \approx b_{i,0} + \sum_{k=1}^{p} b_{i,k}\widehat{m}_{i,k}(U_{tk}), \quad 1 \le i \le d, \tag{2.9}$$

and

$$X_{t+1,i}X_{t+1,j} \approx a_{ij,0} + \sum_{k=1}^{p} a_{ij,k}\widehat{c}_{ij,k}(U_{tk}), \quad 1 \le i,j \le d. \tag{2.10}$$

Treating $\widehat{m}_{i,k}(U_{tk})$ and $\widehat{c}_{ij,k}(U_{tk})$ as "regressors" and using the ordinary least squares, we may obtain an estimate of the optimal weights defined in (2.4) and (2.6), i.e.,

$$\left(\widehat{b}_{i,1},\ldots,\widehat{b}_{i,p}\right)^{\intercal} = \widehat{\mathbf{\Omega}}_{X,i}^{-1}\widehat{\mathbf{V}}_{X,i}, \quad \widehat{b}_{i,0} = \frac{1}{n-1}\sum_{t=1}^{n-1} X_{t+1,i} - \sum_{k=1}^{p}\widehat{b}_{i,k}\left(\frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{m}_{i,k}(U_{tk})\right), \tag{2.11}$$

where $\widehat{\mathbf{\Omega}}_{X,i}$ is a $p \times p$ matrix with the (k,l)-entry being

$$\widehat{\omega}_{i,kl} = \frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{m}_{i,k}^{c}(U_{tk})\widehat{m}_{i,l}^{c}(U_{tl}), \quad \widehat{m}_{i,k}^{c}(U_{tk}) = \widehat{m}_{i,k}(U_{tk}) - \frac{1}{n-1}\sum_{s=1}^{n-1}\widehat{m}_{i,k}(U_{sk}),$$

and $\widehat{\mathbf{V}}_{X,i}$ is a p-dimensional column vector with the k-th element being

$$\widehat{v}_{i,k} = \frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{m}_{i,k}^{c}(U_{tk})X_{t+1,i}^{c}, \quad X_{t+1,i}^{c} = X_{t+1,i} - \frac{1}{n-1}\sum_{s=1}^{n-1}X_{s+1,i};$$

and

$$\left(\widehat{a}_{ij,1},\ldots,\widehat{a}_{ij,p}\right)^{\intercal} = \widehat{\mathbf{\Omega}}_{XX,ij}^{-1}\widehat{\mathbf{V}}_{XX,ij}, \quad \widehat{a}_{ij,0} = \frac{1}{n-1}\sum_{t=1}^{n-1} X_{t+1,i}X_{t+1,j} - \sum_{k=1}^{p}\widehat{a}_{ij,k}\left(\frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{c}_{ij,k}(U_{tk})\right), \tag{2.12}$$

where $\widehat{\mathbf{\Omega}}_{XX,ij}$ is a $p \times p$ matrix with the (k,l)-entry being

$$\widehat{\omega}_{ij,kl} = \frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{c}_{ij,k}^{c}(U_{tk})\widehat{c}_{ij,l}^{c}(U_{tl}), \quad \widehat{c}_{ij,k}^{c}(U_{tk}) = \widehat{c}_{ij,k}(U_{tk}) - \frac{1}{n-1}\sum_{s=1}^{n-1}\widehat{c}_{ij,k}(U_{sk}),$$

and $\widehat{\mathbf{V}}_{XX,ij}$ is a p-dimensional column vector with the k-th element being

$$\widehat{v}_{ij,k} = \frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{c}^{c}_{ij,k}(U_{tk})X^{c}_{t+1,(i,j)}, \quad X^{c}_{t+1,(i,j)} = X_{t+1,i}X_{t+1,j} - \frac{1}{n-1}\sum_{s=1}^{n-1}X_{s+1,i}X_{s+1,j}.$$

As a result, an estimate of $\sigma^{\star}_{ij}(u)$, the (i,j)-entry in $\mathbf{\Sigma}^{\star}_{A}(u)$, can be obtained as

$$\widehat{\sigma}_{ij}(u) = \widehat{c}_{ij}(u) - \widehat{m}_i(u)\widehat{m}_j(u), \tag{2.13}$$

where

$$\widehat{c}_{ij}(u) = \widehat{a}_{ij,0} + \sum_{k=1}^{p}\widehat{a}_{ij,k}\widehat{c}_{ij,k}(u_k), \quad \widehat{m}_i(u) = \widehat{b}_{i,0} + \sum_{k=1}^{p}\widehat{b}_{i,k}\widehat{m}_{i,k}(u_k).$$

A naive estimate, $\widehat{\mathbf{\Sigma}}(u)$, of $\mathbf{\Sigma}^{\star}_{A}(u)$ uses $\widehat{\sigma}_{ij}(u)$ directly as its entries, i.e.,

$$\widehat{\mathbf{\Sigma}}(u) = \left[\widehat{\sigma}_{ij}(u)\right]_{d\times d}.$$

Unfortunately, this matrix gives a poor estimation of $\mathbf{\Sigma}_0(u)$ when the dimension d is ultra large. In the latter case, a commonly-used approach is to use a shrinkage method on $\widehat{\mathbf{\Sigma}}(u)$ so that very small values of $\widehat{\sigma}_{ij}(u)$ are forced to zero. We follow the same approach and denote $s_\rho(\cdot)$ a shrinkage function that satisfies the following three conditions: (i) $|s_\rho(z)| \le \|z\|$ for $z \in \mathcal{R}$ (the real line); (ii) $s_\rho(z) = 0$ if $|z| \le \rho$; (iii) $|s_\rho(z) - z| \le \rho$, where ρ is a tuning parameter. It is easy to show that some commonly-used shrinkage methods including the hard thresholding, soft thresholding and SCAD satisfy the above three conditions. Then define

$$\widetilde{\sigma}_{ij}(u) = s_{\rho(u)}\left(\widehat{\sigma}_{ij}(u)\right), \quad 1 \le i, j \le d, \tag{2.14}$$

where $\rho(u)$ is a variable tuning parameter which may depend on the value of conditioning variables. Then we construct

$$\widetilde{\mathbf{\Sigma}}(u) = \left[\widetilde{\sigma}_{ij}(u)\right]_{d\times d}, \tag{2.15}$$

as the final estimate of $\mathbf{\Sigma}^{\star}_{A}(u)$. The asymptotic properties of $\widetilde{\mathbf{\Sigma}}(u)$ will be explored in Section 3 below. Section 4.1 will introduce a modified version of $\widetilde{\mathbf{\Sigma}}(u)$ to guarantee the positive definiteness of the estimated covariance matrix.

3 Large sample theory

In this section we first state the regularity conditions required for establishing the limit theorems of the large dynamic covariance matrix estimators developed in Section 2, and then present these theorems in Section 3.2.

3.1 Technical assumptions

Some of the assumptions presented below may not be the weakest possible, but they are imposed to facilitate proofs of our limit theorems and can be relaxed at the cost of more lengthy proofs.

ASSUMPTION 1. (i) The process $\{(X_t, U_t)\}_{t\geq 1}$ is stationary and α-mixing dependent with the mixing coefficient decaying to zero at a geometric rate, i.e., $\alpha_k \sim c_\alpha \gamma^k$ with $0 < \gamma < 1$ and c_α being a positive constant.

(ii) The variables X_{ti}, $1 \leq i \leq d$, satisfy the following moment condition:

$$\max_{1\leq i\leq d} \mathsf{E}\left[\exp\{sX_{ti}^2\}\right] \leq c_X, \quad 0 < s \leq s_0, \tag{3.1}$$

where c_X and s_0 are two positive constants.

(iii) The conditioning variables U_t have a compact support denoted by $\mathcal{U} = \prod_{k=1}^p \mathcal{U}_k$, where $\mathcal{U}_k = [a_k, b_k]$ is the support of the k-th conditional variable U_{tk}. The marginal density functions, $f_k(\cdot)$, of U_{tk}, $1 \leq k \leq p$, are continuous and uniformly bounded away from zero on $\mathcal{U}_k$, i.e.,

$$\min_{1\leq k\leq p} \inf_{a_k \leq u_k \leq b_k} f_k(u_k) \geq c_f > 0.$$

In addition, the marginal density functions $f_k(\cdot)$, $1 \leq k \leq p$, have continuous derivatives up to the second order.

ASSUMPTION 2. (i) The regression functions $c_{ij,k}(\cdot)$ and $m_{i,k}(\cdot)$ are continuous and uniformly bounded over $1 \leq i, j \leq d$ and $1 \leq k \leq p$. Furthermore, they have continuous and uniformly bounded derivatives up to the second order.

(ii) For each $i = 1, \ldots, d$ and $j = 1, \ldots, d$, the $p \times p$ matrix $\mathbf{\Omega}_{XX,ij}$ defined in (2.4) is positive definite and satisfies

$$0 < \underline{c}_{\mathbf{\Omega}_{XX}} \leq \min_{1\leq i,j\leq d} \lambda_{\min}(\mathbf{\Omega}_{XX,ij}) \leq \max_{1\leq i,j\leq d} \lambda_{\max}(\mathbf{\Omega}_{XX,ij}) \leq \bar{c}_{\mathbf{\Omega}_{XX}} < \infty. \tag{3.2}$$

The analogous condition also holds for the matrix $\mathbf{\Omega}_{X,i}$ defined in (2.6).

ASSUMPTION 3. (i) The kernel function $K(\cdot)$ is symmetric and Lipschitz continuous and has a compact support, say $[-1, 1]$.

(ii) The bandwidths h_1 and h_2 satisfy $h_1 \to 0$ and $h_2 \to 0$, and there exists $0 < \iota < 1/2$ so that

$$\frac{n^{1-\iota}h_1}{\log^2(d \vee n)} \to \infty, \quad \frac{n^{1-2\iota}h_2}{\log^2(d \vee n)} \to \infty, \tag{3.3}$$

where $x \vee y$ denotes the maximum of x and y.

(iii) The dimension, d, of X satisfies $(dn)\exp\{-sn^\iota\} = o(1)$ for some $0 < s < s_0$, where ι is defined as in Assumption 3(ii).

Assumption 4. The variable tuning parameter can be written as $\rho(u) = M_0(u)\tau_{n,d}$, where $M_0(u)$ is positive and can be sufficiently large at each $u \in \mathcal{U}$ with $\sup_{u\in\mathcal{U}} M_0(u) < \infty$, and

$$\tau_{n,d} = \sqrt{\log(d \vee n)/(nh_1)} + \sqrt{\log(d \vee n)/(nh_2)} + h_1^2 + h_2^2.$$

Most of the above assumptions are commonly used and can be found in some existing literature. The stationarity and α-mixing dependence condition in Assumption 1(i) relaxes the restriction of independent observations usually imposed in the literature on high-dimensional covariance matrix estimation (c.f. Bickel and Levina, 2008a,b). For some classic vector time series processes such as vector auto-regressive processes, it is easy to verify Assumption 1(i) under some mild conditions. It is possible to allow the even more general setting of local stationarity, Vogt (2012), which includes deterministic local trends, but for simplicity we have chosen not to go there. The moment condition (3.1) is similar to those in Bickel and Levina (2008a,b) and Chen and Leng (2016), and can be replaced by the weaker condition of $\mathsf{E}(|X_{ti}|^\kappa)$ for $\kappa > 2$ sufficiently large if the dimension d diverges at a polynomial rate of n. The restriction of the conditioning variables U_t having a compact support in Assumption 1(iii) is imposed mainly in order to facilitate the proofs of our asymptotic results and can be removed by using an appropriate truncation technique on U_t (c.f., Remark 1 in Chen *et al*, 2017). The smoothness condition on $c_{ij,k}(\cdot)$ and $m_{i,k}(\cdot)$ in Assumption 2(i) is commonly used in kernel smoothing, and it is relevant to asymptotic bias of the kernel estimators (c.f., Wand and Jones, 1995). Assumption 2(ii) is crucial to the unique existence of optimal weights in the MAMAR approximation of $c_{ij}^0(\cdot)$ and $m_i^0(\cdot)$. Many commonly-used kernel functions, such as the uniform kernel and the Epanechnikov kernel, all satisfy the conditions in Assumption 3(i). The conditions in Assumptions 3(ii) and (iii) indicate that the dimension d can be divergent at an exponential rate of n. For example, when h_1 and h_2 have the well-known optimal rate of $n^{-1/5}$, we may show that d can be divergent at a rate of $\exp\{n^\zeta\}$ with $0 < \zeta < 1/5$ while Assumptions 3(ii) and (iii) hold. Assumption 4 is critical to ensure the validity of the shrinkage method, and Section 4.2 below will discuss how to select $\rho(u)$ in numerical studies.

3.2 Asymptotic properties

In order to derive some sensible asymptotic results for the dynamic covariance matrix estimators defined in Section 2.2, we extend the sparsity assumption in Bickel and Levina (2008a), Rothman, Levina and Zhu (2009) and Cai and Liu (2011) and assume that $\mathbf{\Sigma}_A^\star(u)$ is *approximately sparse* uniformly over $u \in \mathcal{U}$. Specifically, this means that $\mathbf{\Sigma}_A^\star(u) \in \mathcal{S}_A(q, c_d, M_\star, \mathcal{U})$ uniformly over $u \in \mathcal{U}$, where

$$\mathcal{S}_A(q, c_d, M_\star, \mathcal{U}) = \left\{ \mathbf{\Sigma}(u), u \in \mathcal{U} \;\middle|\; \sup_{u \in \mathcal{U}} \sigma_{ii}(u) \le M_\star < \infty,\ \sup_{u \in \mathcal{U}} \sum_{j=1}^d |\sigma_{ij}(u)|^q \le c_d\ \forall\, 1 \le i \le d \right\} \quad (3.4)$$

with $0 \le q < 1$. In particular, if $q = 0$, $\mathcal{S}_A(q, c_d, M_\star, \mathcal{U})$ becomes

$$\mathcal{S}_A(0, c_d, M_\star, \mathcal{U}) = \left\{ \mathbf{\Sigma}(u), u \in \mathcal{U} \;\middle|\; \sup_{u \in \mathcal{U}} \sigma_{ii}(u) \le M_\star < \infty,\ \sup_{u \in \mathcal{U}} \sum_{j=1}^d I\left(|\sigma_{ij}(u)| \neq 0\right) \le c_d\ \forall\, 1 \le i \le d \right\},$$

and we have $\mathbf{\Sigma}_A^\star(u) \in \mathcal{S}_A(0, c_d, M_\star, \mathcal{U})$, the *exact sparsity* assumption, uniformly over $u \in \mathcal{U}$. The above assumption is natural for nonparametric estimation of large covariance matrices (c.f., Chen, Xu and Wu, 2013; Chen and Leng, 2016). Define $\mathcal{U}_{h_\star} = \prod_{k=1}^p \mathcal{U}_{k,h_\star}$ with $\mathcal{U}_{k,h_\star} = [a_k + h_\star, b_k - h_\star]$ and $h_\star = h_1 \vee h_2$. Without loss of generality, we assume that, for each $1 \le k \le p$, all of the observations U_{tk}, $1 \le t \le n$, are located in the intervals $[a_k + h_\star, b_k - h_\star]$ (otherwise a truncation technique can be applied when constructing the semiparametric estimators defined in Section 2.2). Theorem 1 below gives the uniform consistency for the semiparametric shrinkage estimator of the matrix $\mathbf{\Sigma}_A^\star(u)$ and its inverse.

Theorem 1. *Suppose that Assumptions 1–4 are satisfied, p is fixed, and $\mathbf{\Sigma}_A^\star(u) \in \mathcal{S}_A(q, c_d, M_\star, \mathcal{U})$.*

(i) For $\widetilde{\mathbf{\Sigma}}(u)$, we have

$$\sup_{u \in \mathcal{U}_{h_\star}} \left\| \widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}_A^\star(u) \right\|_O = O_P\left(c_d \cdot \tau_{n,d}^{1-q}\right), \quad 0 \le q < 1, \quad (3.5)$$

where $\tau_{n,d}$ was defined in Assumption 4 and $\|\cdot\|_O$ denotes the operator norm.

(ii) If, in addition, $c_d \tau_{n,d}^{1-q} = o(1)$ and

$$\inf_{u \in \mathcal{U}} \lambda_{\min}\left(\mathbf{\Sigma}_A^\star(u)\right) \ge c_{\mathbf{\Sigma}} > 0, \quad (3.6)$$

we have

$$\sup_{u \in \mathcal{U}_{h_\star}} \left\| \widetilde{\mathbf{\Sigma}}^{-1}(u) - \mathbf{\Sigma}_A^{\star-1}(u) \right\|_O = O_P\left(c_d \cdot \tau_{n,d}^{1-q}\right), \quad 0 \le q < 1. \quad (3.7)$$

The main reason for considering the uniform consistency only over the set $\mathcal{U}_{h_\star}$ rather than the whole support $\mathcal{U}$ of the conditioning variables is to avoid the boundary effect in kernel estimation (c.f., Fan and Gijbels, 1996). The uniform convergence rate in the above theorem is quite general. Its dependence on the sparsity structure of the matrix $\mathbf{\Sigma}_A^\star(u)$ is shown through c_d, which controls the sparsity level in the covariance matrix and may be divergent to infinity. If we assume that $h_1 = h_2 = h$ and $h^2 = O\left(\sqrt{\log(d \vee n)/(nh)}\right)$, $\tau_{n,d}$ can be simplified to $\sqrt{\log(d \vee n)/(nh)}$. Then we may find that our uniform convergence rate is comparable to the rate derived by Bickel and Levina (2008a) and Rothman, Levina and Zhu (2009) when we treat nh as the "effective" sample size in nonparametric kernel-based estimation. In the special case of $q = 0$ and fixed d, $\log(d \vee n) = \log n$ and it would be reasonable to assume that c_d is fixed. Consequently, the rate in (3.5) and (3.7) reduces to

$$O_P(\tau_{n,d}) = O_P\left(\sqrt{\log n/(nh_1)} + \sqrt{\log n/(nh_2)} + h_1^2 + h_2^2\right),$$

the same as the uniform convergence rate for nonparametric kernel-based estimators (c.f. Bosq, 1998).

If we assume that the true dynamic covariance matrix $\mathbf{\Sigma}_0(u)$ belongs to $\mathcal{S}_A(q, c_d, M_\star, \mathcal{U})$, and $\mathbf{\Sigma}_A^\star(u)$ is sufficiently close to $\mathbf{\Sigma}_0(u)$ in the sense that $\sup_{u\in\mathcal{U}} \left\|\mathbf{\Sigma}_A^\star(u) - \mathbf{\Sigma}_0(u)\right\|_O = O(b_n)$ with $b_n \to 0$ and $\max_{1\le i,j\le d} \sup_{u\in\mathcal{U}} \left|\sigma_{ij}^\star(u) - \sigma_{ij}(u)\right| = O(\tau_{n,d})$, by Theorem 1 and its proof in Appendix A, we may show that

$$\begin{aligned}\sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}_0(u)\right\|_O &\le \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}_A^\star(u)\right\|_O + \sup_{u\in\mathcal{U}_{h_\star}} \left\|\mathbf{\Sigma}_A^\star(u) - \mathbf{\Sigma}_0(u)\right\|_O \\ &= O_P\left(c_d \cdot \tau_{n,d}^{1-q}\right) + O\left(b_n\right) = O_P\left(c_d \cdot \tau_{n,d}^{1-q} + b_n\right). \end{aligned} \tag{3.8}$$

The following theorem shows the sparsity property of the semiparametric shrinkage estimator defined in Section 2.2.

THEOREM 2. *Suppose that Assumptions 1–4 are satisfied and p is fixed. For any $u \in \mathcal{U}_{h_\star}$ and $1 \le i, j \le d$, if $\sigma_{ij}^\star(u) = 0$, we must have $\widetilde{\sigma}_{ij}(u) = 0$ with probability approaching one.*

4 Modified dynamic covariance matrix estimation and variable tuning parameter selection

In this section we first modify the semiparametric covariance matrix estimator developed in Section 2.2 to ensure the positive definiteness of the estimated matrix in finite samples, and then discuss the choice of the variable tuning parameter $\rho(u)$ in the generalised shrinkage method.

4.1 Modified dynamic covariance matrix estimation

In practical application, the estimated covariance matrix $\widetilde{\mathbf{\Sigma}}(u)$ constructed in Section 2.2 is not necessarily positive definite uniformly on $\mathcal{U}$. To fix this problem, we next introduce a simple modification of our estimation method. Let $\widetilde{\lambda}_{\min}(u)$ be the smallest eigenvalue of $\widetilde{\mathbf{\Sigma}}(u)$ and m_n be a tuning parameter which tends to zero as the sample size n goes to infinity. As in Chen and Leng (2016), a corrected version of $\widetilde{\mathbf{\Sigma}}(u)$ is defined by

$$\widetilde{\mathbf{\Sigma}}_C(u) = \widetilde{\mathbf{\Sigma}}(u) + \left[m_n - \widetilde{\lambda}_{\min}(u)\right] \mathbf{I}_{d\times d}, \tag{4.1}$$

where $\mathbf{I}_{d\times d}$ is the $d \times d$ identity matrix. The above correction guarantees that the smallest eigenvalue of $\widetilde{\mathbf{\Sigma}}_C(u)$ is uniformly larger than zero, indicating that $\widetilde{\mathbf{\Sigma}}_C(u)$ is uniformly positive definite. Hence, we may use $\widetilde{\mathbf{\Sigma}}_C(u)$ as an alternative estimate of $\mathbf{\Sigma}^\star_A(u)$ when $\widetilde{\lambda}_{\min}(u)$ is negative. We thus define the following modified version of $\widetilde{\mathbf{\Sigma}}(u)$:

$$\begin{aligned} \widetilde{\mathbf{\Sigma}}_M(u) &= \widetilde{\mathbf{\Sigma}}(u) \cdot I\left(\widetilde{\lambda}_{\min}(u) > 0\right) + \widetilde{\mathbf{\Sigma}}_C(u) \cdot I\left(\widetilde{\lambda}_{\min}(u) \leq 0\right) \\ &= \widetilde{\mathbf{\Sigma}}(u) + \left[m_n - \widetilde{\lambda}_{\min}(u)\right] \mathbf{I}_{d\times d} \cdot I\left(\widetilde{\lambda}_{\min}(u) \leq 0\right), \end{aligned} \tag{4.2}$$

where $I(\cdot)$ is an indicator function. Note that when $\widetilde{\lambda}_{\min}(u) \leq 0$ for $u \in \mathcal{U}_{h_\star}$, by Weyl's inequality, we have

$$\left|\widetilde{\lambda}_{\min}(u)\right| \leq \left|\widetilde{\lambda}_{\min}(u) - \lambda_{\min}(\mathbf{\Sigma}^\star_A(u))\right| \leq \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}^\star_A(u)\right\|_O = O_P\left(c_d \cdot \tau_{n,d}^{1-q}\right).$$

Hence,

$$\sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}_M(u) - \mathbf{\Sigma}^\star_A(u)\right\|_O \leq \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}^\star_A(u)\right\|_O + \sup_{u\in\mathcal{U}_{h_\star}} \left|\widetilde{\lambda}_{\min}(u)\right| + m_n = O_P\left(c_d \cdot \tau_{n,d}^{1-q} + m_n\right). \tag{4.3}$$

By choosing $m_n = O(c_d \tau_{n,d}^{1-q})$, we obtain the same uniform convergence rate for $\widetilde{\mathbf{\Sigma}}_M(u)$ as that for $\widetilde{\mathbf{\Sigma}}(u)$ in Theorem 1. Glad, Hjort and Ushakov (2003) consider a similar modification for density estimators that are not bona fide densities; indeed they show that the correction improves the performance according to integrated mean squared error.

4.2 Choice of the variable tuning parameter

For any shrinkage method for covariance matrix estimation, it is essential to choose an appropriate tuning parameter. Since the variables (X_t, U_t) are allowed to be serially correlated over time, the tuning parameter selection criteria proposed in Bickel and Levina (2008b) or Chen and Leng (2016) for independent data may no longer work well in the numerical studies. We hence need to modify their method for our own setting, which is described as follows.

STEP 1: For given $u \in \mathcal{U}$, use a rolling window of size $\lfloor n/2 \rfloor + 10$ and split data within each window into two samples of sizes $n_1 = \left\lfloor \frac{n}{2}\left(1 - \frac{1}{\log(n/2)}\right) \right\rfloor$ and $n_2 = \lfloor n/2 \rfloor - n_1$ by leaving out 10 observations in-between them, where n is the sample size and $\lfloor \cdot \rfloor$ denotes the floor function.

STEP 2: Obtain $\widetilde{\mathbf{\Sigma}}_{\rho(u),1,k}(u)$ (the semiparametric shrinkage estimate of the dynamic covariance matrix from the first sample of the k-th rolling window) constructed as in (2.15), and $\widehat{\mathbf{\Sigma}}_{2,k}(u)$ (the naive estimate without the general shrinkage technique from the second sample of the k-th rolling window), $k = 1, \ldots, N$ with $N = \lfloor n/20 \rfloor$.

STEP 3: Choose the tuning parameter $\rho(u)$ so that it minimises

$$\sum_{k=1}^{N} \left\| \widetilde{\mathbf{\Sigma}}_{\rho(u),1,k}(u) - \widehat{\mathbf{\Sigma}}_{2,k}(u) \right\|_F^2. \tag{4.4}$$

Note that, by leaving out 10 observations inbetween, the correlation between the two samples within each rolling window is expected to be negligible for weekly dependent time series data. Our simulation studies in Section 5 show that the above selection method has reasonably good numerical performance.

5 Simulation studies

In this section, we conduct some simulation experiments to examine the finite sample performance of the proposed large dynamic covariance matrix estimation methods. In order to provide a full performance study, we consider three different sparsity patterns of the underlying covariance matrix, i.e., the dynamic banded structure, the dynamic AR(1) structure, and the varying-sparsity structure. These are the multivariate conditioning variables extension of the covariance models considered in Examples 1-3 of Chen and Leng (2016). To measure estimation accuracy, we consider both the operator and Frobenius losses, i.e., $\|\mathbf{\Sigma}_0(u) - \breve{\mathbf{\Sigma}}(u)\|_O$ and $\|\mathbf{\Sigma}_0(u) - \breve{\mathbf{\Sigma}}(u)\|_F$, for an estimate $\breve{\mathbf{\Sigma}}(u)$ at a point u. We compare the accuracy of our semiparametric shrinkage estimation defined in (2.15) with that of the generalised thresholding of the sample covariance matrix (which treats the covariance matrix as static). Note that the proposed method sometimes produces covariance matrices that are not positive definite, in which case the modification in (4.2) can be used. Hence, we also report the accuracy of the modified dynamic covariance matrix estimation. Four commonly used shrinkage methods – the hard thresholding, soft thresholding, adaptive LASSO (A. LASSO) and Smoothly Clipped Absolute Deviation (SCAD) – are considered in the simulation. Throughout this section, the dimension of X_t takes one of the values of 100, 200, and 300, and the dimension of the conditioning vector U_t is set to $p = 3$. The sample size is fixed at $n = 200$.

EXAMPLE 5.1. (Dynamic banded covariance matrix) The conditioning variables $U_t = (U_{t1}, U_{t2}, U_{t3})^\intercal$ are drawn from a VAR(1) process:

$$U_t = 0.5U_{t-1} + v_t, \qquad t = 1, \dots, n, \tag{5.1}$$

where v_t are i.i.d. three-dimensional random vectors following the $\mathsf{N}(\mathbf{0}, \mathbf{I}_{3\times 3})$ distribution. For each $t = 1, \dots, n$, the d-dimensional vector X_t is generated from the multivariate Gaussian distribution $\mathsf{N}(\mathbf{0},\ \mathbf{\Sigma}_0(U_t))$, where

$$\mathbf{\Sigma}_0(U_t) = \left\{\sigma_{ij}^0(U_t)\right\}_{d\times d} \text{ with } \sigma_{ij}^0(U_t) = 0.08\varsigma_{ij}(U_{t1}) + 0.04\varsigma_{ij}(U_{t2}) + 0.04\varsigma_{ij}(U_{t3}) \tag{5.2}$$

and

$$\varsigma_{ij}(v) = \exp(v/2)\left\{I(i=j) + [\phi(v) + 0.1]I(|i-j| = 1) + \phi(v)I(|i-j| = 2)\right\}$$

for any $v \in \mathcal{R}$, in which $\phi(v)$ is the probability density function of the standard normal distribution.

The dynamic covariance matrix is estimated at the following three points: $(-0.5, -0.5, -0.5)$, $(0, 0, 0)$, and $(0.5, 0.5, 0.5)$. The average operator and Frobenius losses over 30 replications and their standard errors (in parenthesises) are summarised in Tables 5.1(a)–5.1(c). In these tables, "Static" refers to the estimation by treating the underlying covariance matrix as static, "Dynamic" refers to the estimation by using our semiparametric shrinkage method detailed in Section 2, and "Modified Dynamic" refers to the modified dynamic covariance matrix estimation defined in (4.2). In addition, "Hard", "Soft", "A. LASSO" and "SCAD" in the tables represent the hard thresholding, soft thresholding, adaptive LASSO and the smoothly clipped absolute deviation, respectively.

The results in Tables 5.1(a)–5.1(c) reveal that at the point $u = (-0.5, -0.5, -0.5)$, the three methods are comparable in their estimation accuracy. However, at the points $u = (0, 0, 0)$ and $u = (0.5, 0.5, 0.5)$, our semiparametric dynamic covariance matrix estimation via the proposed MAMAR approximation and its modified version outperform the static covariance matrix estimation in almost all thresholding methods. The modified dynamic estimator has similar performance to its non-modified estimator.

EXAMPLE 5.2. (Dynamic non-sparse covariance matrix) The specifications of the data generating process are the same as those in Example 5.1, except that the dynamic covariance matrix $\mathbf{\Sigma}_0(U_t)$ is non-sparse. Specifically, the function $\varsigma_{ij}(\cdot)$ in (5.2) is assumed to follow the covariance pattern of an AR(1) process:

$$\varsigma_{ij}(v) = \exp(v/2)\left[\phi(v)\right]^{|i-j|}$$

Table 5.1(a): Average (standard error) losses at point $(-0.5, -0.5, -0.5)$ for Example 5.1

Method		operator loss			Frobenius loss		
		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.221(0.017)	0.227(0.014)	0.225(0.012)	1.195(0.049)	1.711(0.066)	2.074(0.063)
	Soft	0.313(0.059)	0.250(0.062)	0.241(0.007)	1.082(0.095)	1.504(0.081)	1.845(0.031)
	A. LASSO	0.185(0.009)	0.193(0.010)	0.200(0.009)	1.062(0.030)	1.502(0.037)	1.820(0.027)
	SCAD	0.250(0.084)	0.231(0.010)	0.239(0.008)	1.089(0.096)	1.499(0.011)	1.851(0.024)
Dynamic	Hard	0.201(0.041)	0.219(0.052)	0.213(0.047)	0.970(0.152)	1.482(0.329)	1.786(0.350)
	Soft	0.255(0.008)	0.256(0.007)	0.260(0.014)	1.088(0.025)	1.545(0.032)	1.944(0.162)
	A. LASSO	0.225(0.014)	0.232(0.022)	0.232(0.020)	0.978(0.022)	1.429(0.148)	1.763(0.193)
	SCAD	0.250(0.010)	0.252(0.009)	0.257(0.015)	1.066(0.027)	1.515(0.035)	1.912(0.169)
Modified Dynamic	Hard	0.251(0.021)	0.252(0.020)	0.259(0.030)	1.414(0.134)	1.972(0.253)	2.451(0.408)
	Soft	0.187(0.017)	0.183(0.007)	0.187(0.010)	1.032(0.100)	1.466(0.075)	1.792(0.057)
	A. LASSO	0.189(0.036)	0.187(0.023)	0.187(0.017)	1.112(0.227)	1.561(0.189)	1.896(0.168)
	SCAD	0.190(0.021)	0.188(0.012)	0.190(0.009)	1.084(0.118)	1.547(0.085)	1.869(0.064)

Table 5.1(b): Average (standard error) losses at point $(0, 0, 0)$ for Example 5.1

Method		operator loss			Frobenius loss		
		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.262(0.010)	0.266(0.008)	0.270(0.007)	1.431(0.016)	2.067(0.025)	2.537(0.019)
	Soft	0.295(0.042)	0.358(0.018)	0.361(0.007)	1.173(0.182)	2.180(0.112)	2.752(0.048)
	A. LASSO	0.302(0.012)	0.312(0.010)	0.319(0.009)	1.444(0.011)	2.067(0.011)	2.549.(0.019)
	SCAD	0.340(0.035)	0.350(0.011)	0.359(0.008)	1.457(0.122)	2.187(0.044)	2.730(0.053)
Dynamic	Hard	0.231(0.019)	0.244(0.012)	0.250(0.008)	1.131(0.039)	1.617(0.039)	2.049(0.040)
	Soft	0.321(0.007)	0.325(0.006)	0.329(0.005)	1.396(0.020)	1.986(0.029)	2.465(0.023)
	A. LASSO	0.277(0.009)	0.283(0.007)	0.286(0.005)	1.300(0.021)	1.853(0.030)	2.304(0.022)
	SCAD	0.291(0.010)	0.297(0.009)	0.304(0.007)	1.376(0.014)	1.958(0.018)	2.422(0.015)
Modified Dynamic	Hard	0.264(0.018)	0.274(0.012)	0.290(0.016)	1.416(0.060)	2.072(0.072)	2.634(0.095)
	Soft	0.295(0.009)	0.295(0.009)	0.299(0.008)	1.358(0.015)	1.930(0.022)	2.388(0.017)
	A. LASSO	0.271(0.013)	0.273(0.014)	0.277(0.009)	1.309(0.031)	1.871(0.043)	2.318(0.019)
	SCAD	0.278(0.013)	0.276(0.013)	0.280(0.012)	1.388(0.024)	1.993(0.035)	2.459(0.039)

Table 5.1(c): Average (standard error) losses at point $(0.5, 0.5, 0.5)$ for Example 5.1

Method		operator loss			Frobenius loss		
		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.349(0.010)	0.354(0.008)	0.358(0.007)	1.613(0.016)	2.333(0.024)	2.880(0.025)
	Soft	0.326(0.042)	0.437(0.021)	0.449(0.007)	1.348(0.182)	2.718(0.209)	3.459(0.066)
	A. LASSO	0.390(0.012)	0.399(0.010)	0.407(0.009)	1.735(0.011)	2.497(0.046)	3.103(0.053)
	SCAD	0.408(0.035)	0.438(0.011)	0.447(0.008)	1.763(0.122)	2.721(0.077)	3.415(0.081)
Dynamic	Hard	0.318(0.019)	0.323(0.011)	0.337(0.040)	1.691(0.039)	2.400(0.018)	2.959(0.033)
	Soft	0.358(0.007)	0.366(0.009)	0.389(0.067)	1.523(0.020)	2.172(0.035)	2.748(0.159)
	A. LASSO	0.317(0.009)	0.330(0.025)	0.357(0.062)	1.449(0.021)	2.053(0.088)	2.607(0.181)
	SCAD	0.316(0.010)	0.326(0.016)	0.361(0.092)	1.522(0.014)	2.171(0.031)	2.753(0.193)
Modified Dynamic	Hard	0.325(0.018)	0.326(0.024)	0.350(0.064)	1.726(0.060)	2.456(0.085)	3.098(0.294)
	Soft	0.348(0.009)	0.351(0.009)	0.371(0.082)	1.506(0.015)	2.137(0.036)	2.676(0.092)
	A. LASSO	0.314(0.013)	0.333(0.042)	0.359(0.093)	1.458(0.031)	2.114(0.110)	2.719(0.244)
	SCAD	0.333(0.013)	0.334(0.019)	0.360(0.090)	1.513(0.024)	2.160(0.038)	2.737(0.110)

for any $v \in \mathcal{R}$. The dynamic covariance matrix is again estimated at the points $(-0.5, -0.5, -0.5)$, $(0, 0, 0)$, $(0.5, 0.5, 0.5)$, and the average operator and Frobenius losses are summarised in Tables 5.2(a)–5.2(c). The same conclusion as that from Example 5.1 can be observed. The proposed semiparametric shrinkage method can estimate non-sparse dynamic covariance matrices with satisfactory accuracy.

Example 5.3. (Dynamic covariance matrix with varying sparsity) Data on U_t and X_t are generated in the same way as in Example 5.1 except that the dynamic covariance matrix $\mathbf{\Sigma}_0(U_t)$ has varying sparsity patterns. Specifically, the function $\varsigma_{ij}(\cdot)$ in (5.2) is defined as

$$\varsigma_{ij}(v) = \exp(v/2)\big\{I(i=j) + 0.5\exp\big[-\frac{(v-0.25)^2}{0.75^2-(v-0.25)^2}\big]I(-0.49 \le v \le 0.99)I(|i-j|=1) + 0.4\exp\big[-\frac{(v-0.65)^2}{0.35^2-(v-0.65)^2}I(0.31 \le v \le 0.99)I(|i-j|=2)\big]\big\}$$

for any $v \in \mathcal{R}$. The dynamic covariance matrix is estimated at the following three points: $(-0.6, -0.6, -0.6)$, $(0, 0, 0)$, $(0.6, 0.6, 0.6)$, where the corresponding covariance matrices have different sparsity structures. The average operator and Frobenius losses are presented in Tables 5.3(a)–5.3(c), from which the same conclusion as that from Example 5.1 can be observed.

6 Conclusion and extension

In this paper we estimate the ultra large dynamic covariance matrix for high-dimensional time series data where the conditioning random variables are multivariate. Through the semiparametric MAMAR approximation to each entry in the underlying dynamic covariance matrix, we successfully circumvent

Table 5.2(a): Average (standard error) losses at point $(-0.5, -0.5, -0.5)$ for Example 5.2

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.165(0.019)	0.173(0.017)	0.174(0.019)	0.932(0.055)	1.330(0.082)	1.640(0.109)
	Soft	0.235(0.077)	0.171(0.008)	0.174(0.008)	0.891(0.195)	1.041(0.034)	1.291(0.045)
	A. LASSO	0.135(0.011)	0.140(0.008)	0.142(0.009)	0.753(0.035)	1.054(0.047)	1.285(0.063)
	SCAD	0.159(0.010)	0.168(0.010)	0.172(0.010)	0.736(0.015)	1.056(0.020)	1.307(0.034)
Dynamic	Hard	0.141(0.033)	0.149(0.032)	0.171(0.073)	0.721(0.143)	1.015(0.092)	1.340(0.333)
	Soft	0.191(0.006)	0.193(0.006)	0.206(0.049)	0.831(0.034)	1.165(0.036)	1.464(0.049)
	A. LASSO	0.162(0.009)	0.164(0.010)	0.192(0.096)	0.704(0.022)	0.992(0.024)	1.272(0.172)
	SCAD	0.186(0.008)	0.188(0.008)	0.204(0.050)	0.804(0.039)	1.124(0.043)	1.421(0.056)
Modified Dynamic	Hard	0.174(0.037)	0.197(0.044)	0.216(0.103)	0.978(0.252)	1.520(0.357)	2.065(1.258)
	Soft	0.127(0.010)	0.131(0.012)	0.146(0.072)	0.717(0.039)	1.037(0.080)	1.514(0.988)
	A. LASSO	0.135(0.020)	0.140(0.026)	0.180(0.130)	0.785(0.091)	1.152(0.182)	1.873(1.792)
	SCAD	0.146(0.021)	0.154(0.024)	0.177(0.077)	0.784(0.051)	1.142(0.098)	1.628(1.000)

Table 5.2(b): Average (standard error) losses at point $(0, 0, 0)$ for Example 5.2

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.191(0.007)	0.193(0.008)	0.194(0.010)	1.035(0.022)	1.474(0.033)	1.813(0.050)
	Soft	0.271(0.023)	0.280(0.008)	0.283(0.008)	1.088(0.066)	1.650(0.057)	2.046(0.073)
	A. LASSO	0.230(0.009)	0.238(0.011)	0.240(0.010)	1.014(0.010)	1.450(0.017)	1.785(0.024)
	SCAD	0.264(0.010)	0.276(0.011)	0.280(0.010)	1.104(0.038)	1.618(0.063)	2.016(0.790)
Dynamic	Hard	0.189(0.006)	0.193(0.008)	0.196(0.006)	0.968(0.017)	1.372(0.020)	1.712(0.027)
	Soft	0.253(0.005)	0.255(0.005)	0.258(0.004)	1.048(0.019)	1.475(0.017)	1.833(0.020)
	A. LASSO	0.213(0.005)	0.215(0.005)	0.218(0.004)	0.980(0.008)	1.387(0.009)	1.713(0.006)
	SCAD	0.224(0.008)	0.226(0.008)	0.233(0.007)	1.016(0.010)	1.438(0.009)	1.771(0.008)
Modified Dynamic	Hard	0.184(0.013)	0.194(0.018)	0.203(0.025)	1.026(0.044)	1.517(0.060)	1.884(0.144)
	Soft	0.230(0.007)	0.230(0.006)	0.227(0.007)	0.993(0.008)	1.401(0.009)	1.723(0.009)
	A. LASSO	0.223(0.008)	0.223(0.007)	0.219(0.009)	0.982(0.010)	1.387(0.010)	1.719(0.017)
	SCAD	0.209(0.011)	0.210(0.008)	0.208(0.008)	1.030(0.023)	1.467(0.028)	1.824(0.046)

Table 5.2(c): Average (standard error) losses at point $(0.5, 0.5, 0.5)$ for Example 5.2

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.248(0.007)	0.250(0.008)	0.249(0.008)	1.119(0.013)	1.588(0.018)	1.948(0.018)
	Soft	0.293(0.038)	0.337(0.008)	0.340(0.008)	1.296(0.188)	2.159(0.078)	2.676(0.101)
	A. LASSO	0.286(0.009)	0.294(0.011)	0.297(0.011)	1.239(0.040)	1.788(0.064)	2.212(0.082)
	SCAD	0.321(0.010)	0.333(0.010)	0.337(0.010)	1.416(0.067)	2.092(0.101)	2.614(0.126)
Dynamic	Hard	0.219(0.026)	0.233(0.025)	0.241(0.037)	1.168(0.023)	1.664(0.025)	2.037(0.030)
	Soft	0.269(0.006)	0.275(0.007)	0.280(0.015)	1.147(0.024)	1.623(0.027)	2.022(0.020)
	A. LASSO	0.241(0.041)	0.270(0.054)	0.299(0.080)	1.106(0.048)	1.595(0.078)	1.992(0.090)
	SCAD	0.233(0.029)	0.248(0.028)	0.264(0.040)	1.138(0.027)	1.619(0.036)	2.008(0.038)
Modified Dynamic	Hard	0.257(0.020)	0.262(0.020)	0.262(0.023)	1.168(0.027)	1.652(0.045)	2.017(0.043)
	Soft	0.266(0.010)	0.265(0.013)	0.269(0.024)	1.137(0.028)	1.594(0.032)	1.963(0.031)
	A. LASSO	0.259(0.038)	0.272(0.057)	0.310(0.098)	1.118(0.043)	1.603(0.111)	2.034(0.190)
	SCAD	0.246(0.022)	0.254(0.025)	0.260(0.036)	1.130(0.030)	1.622(0.068)	2.005(0.036)

Table 5.3(a): Average (standard error) losses at point $(-0.6, -0.6, -0.6)$ for Example 5.3

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.123(0.018)	0.132(0.017)	0.134(0.016)	0.721(0.097)	1.021(0.119)	1.234(0.140)
	Soft	0.132(0.119)	0.070(0.008)	0.075(0.009)	0.478(0.394)	0.399(0.061)	0.532(0.096)
	A. LASSO	0.096(0.020)	0.100(0.015)	0.102(0.018)	0.412(0.101)	0.535(0.117)	0.608(0.117)
	SCAD	0.090(0.027)	0.091(0.016)	0.094(0.017)	0.332(0.064)	0.467(0.037)	0.582(0.066)
Dynamic	Hard	0.107(0.030)	0.115(0.013)	0.105(0.022)	0.503(0.360)	0.759(0.506)	0.935(0.731)
	Soft	0.103(0.015)	0.115(0.025)	0.111(0.019)	0.547(0.053)	0.841(0.222)	1.033(0.268)
	A. LASSO	0.094(0.026)	0.112(0.042)	0.100(0.022)	0.398(0.237)	0.554(0.292)	0.796(0.530)
	SCAD	0.103(0.015)	0.115(0.025)	0.111(0.019)	0.519(0.064)	0.808(0.234)	0.996(0.281)
Modified Dynamic	Hard	0.133(0.076)	0.147(0.064)	0.114(0.067)	0.613(0.406)	1.003(0.448)	0.868(0.562)
	Soft	0.089(0.024)	0.100(0.033)	0.099(0.029)	0.403(0.127)	0.681(0.332)	0.759(0.302)
	A. LASSO	0.115(0.040)	0.140(0.055)	0.121(0.033)	0.563(0.242)	1.054(0.584)	0.956(0.411)
	SCAD	0.125(0.030)	0.136(0.041)	0.135(0.038)	0.510(0.128)	0.816(0.339)	0.903(0.314)

Table 5.3(b): Average (standard error) losses at point $(0, 0, 0)$ for Example 5.3

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.184(0.013)	0.188(0.010)	0.189(0.010)	1.054(0.031)	1.494(0.037)	1.827(0.037)
	Soft	0.230(0.041)	0.218(0.009)	0.223(0.008)	1.154(0.048)	1.670(0.056)	2.084(0.074)
	A. LASSO	0.171(0.008)	0.180(0.010)	0.184(0.010)	1.030(0.008)	1.471(0.015)	1.814(0.035)
	SCAD	0.203(0.011)	0.216(0.010)	0.222(0.009)	1.117(0.038)	1.639(0.061)	2.055(0.082)
Dynamic	Hard	0.188(0.015)	0.198(0.015)	0.206(0.012)	1.042(0.017)	1.474(0.017)	1.818(0.019)
	Soft	0.190(0.004)	0.195(0.005)	0.196(0.005)	1.070(0.013)	1.525(0.020)	1.875(0.021)
	A. LASSO	0.164(0.007)	0.168(0.006)	0.170(0.007)	1.019(0.007)	1.443(0.007)	1.772(0.007)
	SCAD	0.175(0.008)	0.181(0.006)	0.183(0.006)	1.043(0.009)	1.480(0.010)	1.817(0.013)
Modified Dynamic	Hard	0.176(0.018)	0.193(0.023)	0.200(0.023)	1.026(0.029)	1.470(0.054)	1.816(0.065)
	Soft	0.170(0.006)	0.169(0.006)	0.169(0.007)	1.025(0.009)	1.445(0.008)	1.768(0.008)
	A. LASSO	0.167(0.005)	0.168(0.006)	0.172(0.010)	1.022(0.008)	1.444(0.006)	1.776(0.012)
	SCAD	0.182(0.012)	0.193(0.010)	0.204(0.014)	1.051(0.017)	1.505(0.028)	1.872(0.048)

Table 5.3(c): Average (standard error) losses at point $(0.6, 0.6, 0.6)$ for Example 5.3

		operator loss			Frobenius loss		
Method		$p = 100$	$p = 200$	$p = 300$	$p = 100$	$p = 200$	$p = 300$
Static	Hard	0.365(0.009)	0.369(0.009)	0.370(0.008)	1.689(0.013)	2.398(0.017)	2.942(0.025)
	Soft	0.430(0.028)	0.457(0.009)	0.461(0.008)	1.918(0.129)	2.875(0.066)	3.565(0.082)
	A. LASSO	0.403(0.011)	0.413(0.011)	0.419(0.011)	1.789(0.037)	2.569(0.052)	3.180(0.072)
	SCAD	0.437(0.013)	0.452(0.012)	0.459(0.010)	1.931(0.065)	2.818(0.085)	3.516(0.100)
Dynamic	Hard	0.336(0.071)	0.338(0.021)	0.362(0.104)	1.720(0.029)	2.439(0.028)	2.996(0.047)
	Soft	0.386(0.023)	0.395(0.018)	0.405(0.045)	1.727(0.017)	2.492(0.074)	3.079(0.091)
	A. LASSO	0.362(0.027)	0.382(0.027)	0.391(0.026)	1.726(0.029)	2.465(0.037)	3.046(0.041)
	SCAD	0.342(0.053)	0.360(0.034)	0.384(0.076)	1.722(0.024)	2.475(0.075)	3.072(0.125)
Modified Dynamic	Hard	0.388(0.063)	0.373(0.015)	0.410(0.147)	1.733(0.033)	2.459(0.116)	3.124(0.599)
	Soft	0.385(0.024)	0.379(0.009)	0.398(0.060)	1.727(0.022)	2.441(0.017)	3.022(0.044)
	A. LASSO	0.373(0.020)	0.369(0.034)	0.369(0.039)	1.729(0.019)	2.462(0.035)	3.018(0.058)
	SCAD	0.363(0.046)	0.363(0.020)	0.383(0.073)	1.716(0.021)	2.454(0.042)	3.033(0.061)

the curse of dimensionality problem in multivariate nonparametric estimation. The subsequent two-stage semiparametric estimation method, combined with the general shrinkage technique commonly used in high-dimensional data analysis, produces reliable dynamic covariance matrix estimation. Under some mild conditions such as the approximate sparsity assumption, the developed covariance matrix estimation is proved to be uniformly consistent with convergence rates comparable to those obtained in the literature. In addition, a modified version of the semiparametric dynamic covariance matrix estimation is introduced to ensure that the estimated covariance matrix is positive definite. Furthermore, a new selection criterion to determine the optimal local tuning parameter is provided to implement the proposed semiparametric large covariance matrix estimation for high-dimensional weakly dependent time series data. Extensive simulation studies conducted in Section 5 show that the proposed approaches have reliable numerical performance.

In the present paper, we limit attention to the case where the number of conditioning variables is fixed. However, it is often not uncommon to have a very large number of conditioning variables in practice. In this latter case, a direct application of the MAMAR approximation and the semiparametric method proposed in Section 2.2 may result in poor and unstable matrix estimation results. Motivated by a recent paper by Chen *et al* (2017) on high-dimensional MAMAR method, we can circumvent this problem by assuming that the number of conditioning variables which make "significant" contribution to estimating joint regression functions, $m_i^0(u)$ and $c_{ij}^0(u)$, in (1.4) and (1.5) is relatively small, i.e., for each i and j, when p is divergent, the number of nonzero weights $b_{i,k}$ and $a_{ij,k}$, $1 \le k \le p$, is relatively small. This makes equations (1.4) and (1.5) fall into the classic sparsity framework commonly used in high-dimensional variable or feature selection literature. To remove the insignificant conditioning variables, we combine the penalisation and MAMAR techniques when estimating $m_i^0(u)$ and $c_{ij}^0(u)$. Specifically, for each $1 \le i \le d$, to estimate $\left(b_{i,1}^\star, \dots, b_{i,p}^\star\right)^\intercal$, we define the penalised objective function:

$$\mathcal{Q}_i(b_{i,1}, \dots, b_{i,p}) = \sum_{t=1}^{n-1} \left[X_{t+1,i}^c - \sum_{k=1}^{p} b_{i,k} \widehat{m}_{i,k}^c(U_{tk}) \right]^2 + n \sum_{k=1}^{p} p_{\lambda_1}(|b_{i,k}|), \tag{6.1}$$

where $X_{t+1,i}^c = X_{t+1,i} - \frac{1}{n-1}\sum_{s=1}^{n-1} X_{s+1,i}$, $\widehat{m}_{i,k}^c(U_{tk}) = \widehat{m}_{i,k}(U_{tk}) - \frac{1}{n-1}\sum_{s=1}^{n-1} \widehat{m}_{i,k}(U_{sk})$, and $p_{\lambda_1}(\cdot)$ is a penalty function with a tuning parameter λ_1. The solution to the minimisation of $\mathcal{Q}_i(b_{i,1}, \dots, b_{i,p})$ is the penalised estimator of the optimal weights and is denoted by $(\bar{b}_{i,1}, \dots, \bar{b}_{i,p})^\intercal$. The subsequent intercept estimate, denoted by $\bar{b}_{i,0}$, can be calculated in the same way as $\widehat{b}_{i,0}$ in (2.11), but with $\widehat{b}_{i,k}$ replaced by $\bar{b}_{i,k}$, $k = 1, \dots, p$. Similarly, for each $1 \le i, j \le d$, to estimate $\left(a_{ij,1}^\star, \dots, a_{ij,p}^\star\right)^\intercal$, we define the penalised objective function:

$$\mathcal{Q}_{ij}(a_{ij,1}, \dots, a_{ij,p}) = \sum_{t=1}^{n-1} \left[X_{t+1,(i,j)}^c - \sum_{k=1}^{p} a_{ij,k} \widehat{c}_{ij,k}^c(U_{tk}) \right]^2 + n \sum_{k=1}^{p} p_{\lambda_2}(|a_{ij,k}|), \tag{6.2}$$

where $X^c_{t+1,(i,j)} = X_{t+1,i}X_{t+1,j} - \frac{1}{n-1}\sum_{s=1}^{n-1} X_{s+1,i}X_{s+1,j}$, $\widehat{c}^c_{ij,k}(U_{tk}) = \widehat{c}_{ij,k}(U_{tk}) - \frac{1}{n-1}\sum_{s=1}^{n-1}\widehat{c}_{ij,k}(U_{sk})$, and $p_{\lambda_2}(\cdot)$ is a penalty function with a tuning parameter λ_2. The solution to the minimisation of $\mathcal{Q}_{ij}(a_{ij,1},\ldots,a_{ij,p})$ is denoted by $(\overline{a}_{ij,1},\ldots,\overline{a}_{ij,p})^\intercal$, and the intercept estimate, $\overline{a}_{ij,0}$, can be obtained accordingly by replacing $\widehat{a}_{ij,k}$ with $\overline{a}_{ij,k}$, $k=1,\ldots,p$, on the right hand side of the equation for $\widehat{a}_{ij,0}$ in (2.12). By Theorem 2(ii) in Chen *et al* (2017), under the sparsity assumption and some technical conditions, the zero optimal weights can be estimated exactly as zeros with probability approaching one. After obtaining $\overline{b}_{i,k}$ and $\overline{a}_{ij,k}$, $0 \le k \le p$, we can calculate the penalised estimates of the optimal MAMAR approximation to $c^0_{ij}(u)$ and $m^0_i(u)$ as

$$\overline{c}_{ij}(u) = \overline{a}_{ij,0} + \sum_{k=1}^{p}\overline{a}_{ij,k}\widehat{c}_{ij,k}(u_k), \quad \overline{m}_i(u) = \overline{b}_{i,0} + \sum_{k=1}^{p}\overline{b}_{i,k}\widehat{m}_{i,k}(u_k),$$

and subsequently the penalised estimate of $\sigma^0_{ij}(u)$ as

$$\overline{\sigma}_{ij}(u) = \overline{c}_{ij}(u) - \overline{m}_i(u)\overline{m}_j(u). \tag{6.3}$$

Finally, we apply the shrinkage technique detailed in Section 2.2 to $\overline{\sigma}_{ij}(u)$ to obtain the estimate of the dynamic covariance matrix. Their asymptotic property and numerical performance will be explored in a separate project.

Another feasible way to deal with high-dimensional conditioning variables is to impose the so-called approximate factor modelling structure on U_t (Bai and Ng, 2002). Instead of directly using U_t whose size can be very large, we may use the relatively low-dimensional latent common factors F_t (c.f., Stock and Watson, 2002), which can be estimated (up to a possible rotation) by some classic approaches like the principal component analysis and maximum likelihood method. As a result, our semiparametric dynamic covariance matrix estimation method may be still applicable after replacing U_t by the estimates of F_t.

Appendix A: Proofs of the main limit theorems

In this appendix, we provide the detailed proofs of the main asymptotic theorems. We start with some technical lemmas whose proofs will be given in Appendix B.

LEMMA 1. *Suppose that Assumptions 1, 2(i) and 3 in Section 3.1 are satisfied. Then we have*

$$\max_{1\le i\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\widehat{m}_{i,k}(u_k) - m_{i,k}(u_k)\right| = O_P\left(\sqrt{\log(d\vee n)/(nh_1)} + h_1^2\right), \tag{A.1}$$

and

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\widehat{c}_{ij,k}(u_k)-c_{ij,k}(u_k)\right| = O_P\left(\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right), \tag{A.2}$$

where $h_\star = h_1 \vee h_2$.

Lemma 2. *Suppose that Assumptions 1–3 in Section 3.1 are satisfied. Then we have*

$$\max_{1\le i,j\le d}\sum_{k=0}^{p}\left|\widehat{a}_{ij,k}-a_{ij,k}^\star\right| = O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right), \tag{A.3}$$

and

$$\max_{1\le i\le d}\sum_{k=1}^{p}\left|\widehat{b}_{i,k}-b_{i,k}^\star\right| = O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_1)}+h_1^2\right). \tag{A.4}$$

The following proposition gives an uniform consistency (with convergence rates) for the nonparametric conditional covariance matrix estimation via the MAMAR approximation.

Proposition 1. *Suppose that Assumptions 1–3 in Section 3.1 are satisfied. Then we have*

$$\max_{1\le i,j\le d}\sup_{u\in\mathcal{U}}\left|\widehat{\sigma}_{ij}(u)-\sigma_{ij}^\star(u)\right| = O_P(\tau_{n,d}), \tag{A.5}$$

where $\tau_{n,d}$ *is defined in Assumption 4, and* $\sigma_{ij}^\star(u) = c_{ij}^\star(u) - m_i^\star(u)m_j^\star(u)$, $c_{ij}^\star(u)$ *is the* (i,j)*-entry of* $\mathcal{C}_{\mathcal{A}}^\star(u)$ *and* $m_i^\star(u)$ *is the* i*-th element of* $\mathcal{M}_{\mathcal{B}}^\star(u)$, $\mathcal{C}_{\mathcal{A}}^\star(u)$ *and* $\mathcal{M}_{\mathcal{B}}^\star(u)$ *are defined in Section 2.1.*

Proof of Proposition 1. By (A.2) and (A.3), we have

$$\begin{aligned}
\widehat{c}_{ij}(u)-c_{ij}^\star(u) &= \left[\widehat{a}_{ij,0}+\sum_{k=1}^{p}\widehat{a}_{ij,k}\widehat{c}_{ij,k}(u_k)\right]-\left[a_{ij,0}^\star+\sum_{k=1}^{p}a_{ij,k}^\star c_{ij,k}(u_k)\right]\\
&= \left(\widehat{a}_{ij,0}-a_{ij,0}^\star\right)+\sum_{k=1}^{p}\left(\widehat{a}_{ij,k}-a_{ij,k}^\star\right)c_{ij,k}(u_k)+\sum_{k=1}^{p}a_{ij,k}^\star\left[\widehat{c}_{ij,k}(u_k)-c_{ij,k}(u_k)\right]+\\
&\quad\sum_{k=1}^{p}\left(\widehat{a}_{ij,k}-a_{ij,k}^\star\right)\left[\widehat{c}_{ij,k}(u_k)-c_{ij,k}(u_k)\right]\\
&= O_P\left(\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right)
\end{aligned} \tag{A.6}$$

uniformly for $1\le i,j\le d$ and $u\in\mathcal{U}_{h_\star}$. On the other hand, note that

$$\begin{aligned}
\widehat{m}_i(u)\widehat{m}_j(u)-m_i^\star(u)m_j^\star(u) &= \left[\widehat{m}_i(u)-m_i^\star(u)\right]m_j^\star(u)-m_i^\star(u)\left[\widehat{m}_j(u)-m_j^\star(u)\right]+\\
&\quad\left[\widehat{m}_i(u)-m_i^\star(u)\right]\left[\widehat{m}_j(u)-m_j^\star(u)\right]
\end{aligned} \tag{A.7}$$

with

$$\widehat{m}_i(u) - m_i^\star(u) = \left(\widehat{b}_{i,0} - b_{i,0}^\star\right) + \sum_{k=1}^{p} \left(\widehat{b}_{i,k} - b_{ij,k}^\star\right) m_{i,k}(u_k) + \sum_{k=1}^{p} b_{i,k}^\star \left[\widehat{m}_{i,k}(u_k) - m_{i,k}(u_k)\right] +$$

$$\sum_{k=1}^{p} \left(\widehat{b}_{i,k} - b_{i,k}^\star\right) \left[\widehat{m}_{i,k}(u_k) - m_{i,k}(u_k)\right]$$

$$= O_P\left(\sqrt{\log(d \vee n)/(nh_1)} + h_1^2\right) \tag{A.8}$$

uniformly for $1 \leq i \leq d$ and $u \in \mathcal{U}_{h_\star}$, where (A.1) and (A.4) have been used.

Therefore, by (A.6)–(A.8), we have

$$\max_{1\leq i,j\leq d} \sup_{u\in\mathcal{U}_{h_\star}} \left|\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right|$$

$$= \max_{1\leq i,j\leq d} \sup_{u\in\mathcal{U}_{h_\star}} \left|\widehat{c}_{ij}(u) - c_{ij}^\star(u)\right| + \max_{1\leq i,j\leq d} \sup_{u\in\mathcal{U}_{h_\star}} \left|\widehat{m}_i(u)\widehat{m}_j(u) - m_i^\star(u)m_j^\star(u)\right|$$

$$= O_P\left(\sqrt{\log(d \vee n)/(nh_1)} + \sqrt{\log(d \vee n)/(nh_2)} + h_1^2 + h_2^2\right), \tag{A.9}$$

completing the proof of Proposition 1. □

PROOF OF THEOREM 1. From the definition of $\widetilde{\mathbf{\Sigma}}(u)$ and $\widetilde{\sigma}_{ij}(u)$, we have

$$\sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}_A^\star(u)\right\|_O \leq \sup_{u\in\mathcal{U}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \left|\widetilde{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right|$$

$$= \sup_{u\in\mathcal{U}_{h_\star}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \left|s_{\rho(u)}\left(\widehat{\sigma}_{ij}(u)\right) I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) - \sigma_{ij}^\star(u)\right|$$

$$= \sup_{u\in\mathcal{U}_{h_\star}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \Big|s_{\rho(u)}\left(\widehat{\sigma}_{ij}(u)\right) I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) -$$

$$\sigma_{ij}^\star(u) I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) - \sigma_{ij}^\star(u) I\left(|\widehat{\sigma}_{ij}(u)| \leq \rho(u)\right)\Big|$$

$$\leq \sup_{u\in\mathcal{U}_{h_\star}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \left|s_{\rho(u)}\left(\widehat{\sigma}_{ij}(u)\right) - \widehat{\sigma}_{ij}(u)\right| I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) +$$

$$\sup_{u\in\mathcal{U}_{h_\star}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \left|\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right| I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) +$$

$$\sup_{u\in\mathcal{U}_{h_\star}} \max_{1\leq i\leq d} \sum_{j=1}^{d} \left|\sigma_{ij}^\star(u)\right| I\left(|\widehat{\sigma}_{ij}(u)| \leq \rho(u)\right)$$

$$=: I_1 + I_2 + I_3. \tag{A.10}$$

From Proposition 1, we define an event

$$\mathcal{E} = \left\{ \max_{1\le i,j\le d} \sup_{u\in\mathcal{U}_{h_\star}} \left|\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right| \le M_1\tau_{n,d} \right\},$$

where M_1 is a positive constant such that $\mathsf{P}(\mathcal{E}) \ge 1-\epsilon$ with $\epsilon > 0$ being arbitrarily small. By Property (iii) of the shrinkage function and Proposition 1, we readily have

$$I_1 \le \sup_{u\in\mathcal{U}_{h_\star}} \rho(u) \left[\max_{1\le i\le d} \sum_{j=1}^d I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) \right] \tag{A.11}$$

and

$$I_2 \le M_1\tau_{n,d} \sup_{u\in\mathcal{U}_{h_\star}} \max_{1\le i\le d} \sum_{j=1}^d I\left(|\widehat{\sigma}_{ij}(u)| > \rho(u)\right) \tag{A.12}$$

conditional on the event $\mathcal{E}$. Note that on $\mathcal{E}$,

$$|\widehat{\sigma}_{ij}(u)| \le |\sigma_{ij}^\star(u)| + |\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)| \le |\sigma_{ij}^\star(u)| + M_1\tau_n.$$

Recall that $\rho(u) = M_0(u)\tau_{n,h}$ in Assumption 4 and choose $M_0(u)$ such that $\inf_{u\in\mathcal{U}} M_0(u) = 2M_1$. Then, it is easy to see the event $\{|\widehat{\sigma}_{ij}(u)| > \rho(u)\}$ indicates that $\{|\sigma_{ij}^\star(u)| > M_1\tau_{n,d}\}$ holds. As $\mathbf{\Sigma}_A^\star(\cdot) \in \mathcal{S}(q, c_d, M_\star, \mathcal{U})$ defined in (3.4), we may show that

$$\begin{aligned}
I_1 + I_2 &\le \tau_{n,d}\left[\sup_{u\in\mathcal{U}} M_0(u) + M_1\right]\left[\sup_{u\in\mathcal{U}_{h_\star}} \max_{1\le i\le d} \sum_{j=1}^d I\left(|\widehat{\sigma}_{ij}(u)| > M_1\tau_{n,d}\right)\right] \\
&\le \tau_{n,d}\left[\sup_{u\in\mathcal{U}} M_0(u) + M_1\right]\left[\sup_{u\in\mathcal{U}} \max_{1\le i\le d} \sum_{j=1}^d \frac{\left|\sigma_{ij}^\star(u)\right|^q}{M_1^q\tau_{n,d}^q}\right] \\
&= O\left(c_d \cdot \tau_{n,d}^{1-q}\right)
\end{aligned} \tag{A.13}$$

on the event $\mathcal{E}$.

On the other hand, by the triangle inequality, we have for any $u \in \mathcal{U}_{h_\star}$,

$$|\widehat{\sigma}_{ij}(u)| \ge |\sigma_{ij}^\star(u)| - |\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)| \ge |\sigma_{ij}^\star(u)| - M_1\tau_n$$

on the event $\mathcal{E}$. Hence, we readily show that $\{|\widehat{\sigma}_{ij}(u)| \le \rho(u)\}$ indicates

$$\left\{ |\sigma_{ij}^\star(u)| \le \left(\sup_{u\in\mathcal{U}} M_0(u) + M_1\right)\tau_{n,d} \right\}.$$

Then, for I_3, by Assumption 4 and the definition of $\mathcal{S}(q, c_d, M_\star, \mathcal{U})$, we have

$$\begin{aligned} I_3 &\le \sup_{u\in\mathcal{U}_{h_\star}} \max_{1\le i\le d} \sum_{j=1}^d \left|\sigma_{ij}^\star(u)\right| I\left(\left|\sigma_{ij}^\star(u)\right| \le (\sup_{u\in\mathcal{U}} M_0(u) + M_1)\tau_{n,d}\right) \\ &\le (\sup_{u\in\mathcal{U}} M_0(u) + M_1)^{1-q}\tau_{n,d}^{1-q} \sup_{u\in\mathcal{U}} \max_{1\le i\le d} \sum_{j=1}^d \left|\sigma_{ij}^\star(u)\right|^q \\ &= O_P\left(c_d \cdot \tau_{n,d}^{1-q}\right). \end{aligned} \tag{A.14}$$

The proof of (3.5) in Theorem 1(i) can be completed by (A.10), (A.13) and (A.14).

Note that

$$\begin{aligned} \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}^{-1}(u) - \mathbf{\Sigma}_A^{\star-1}(u)\right\|_O &= \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}^{-1}(u)\mathbf{\Sigma}_A^\star(u)\mathbf{\Sigma}_A^{\star-1}(u) - \widetilde{\mathbf{\Sigma}}^{-1}(u)\widetilde{\mathbf{\Sigma}}(u)\mathbf{\Sigma}_A^{\star-1}(u)\right\|_O \\ &\le \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}^{-1}(u)\right\|_O \sup_{u\in\mathcal{U}_{h_\star}} \left\|\widetilde{\mathbf{\Sigma}}(u) - \mathbf{\Sigma}_A^\star(u)\right\|_O \sup_{u\in\mathcal{U}_{h_\star}} \left\|\mathbf{\Sigma}_A^{\star-1}(u)\right\|_O. \end{aligned}$$

It is easy to prove (3.7) in Theorem 1(ii) from (3.6) and (3.5) in Theorem 1(i). □

Proof of Theorem 2. By the definition of $s_{\rho(u)}(\cdot)$, it is easy to show that $\left\{\widetilde{\sigma}_{ij}(u) = s_{\rho(u)}(\widehat{\sigma}_{ij}(u)) \neq 0\right\}$ is equivalent to $\{|\widehat{\sigma}_{ij}(u)| > \rho(u)\}$ for any $u \in \mathcal{U}_{h_\star}$ and $1 \le i, j \le d$. Hence, $\left\{\widetilde{\sigma}_{ij}(u) \neq 0 \text{ and } \sigma_{ij}^\star(u) = 0\right\}$ indicates that

$$\left|\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right| > \rho(u). \tag{A.15}$$

Note that $\rho(u) = M_0(u)\tau_{n,d}$ with $\inf_{u\in\mathcal{U}} M_0(u) \ge c_M > 0$. From (A.15) and Proposition 1 above, taking $c_M > 0$ sufficiently large, we have

$$\begin{aligned} &\mathsf{P}\left(\widetilde{\sigma}_{ij}(u)) \neq 0 \ \text{ and } \ \sigma_{ij}^\star(u) = 0 \ \text{ for } u \in \mathcal{U}_{h_\star} \text{ and } 1 \le i, j \le d\right) \\ &\le \mathsf{P}\left(\max_{1\le i,j\le d} \sup_{u\in\mathcal{U}_{h_\star}} \left|\widehat{\sigma}_{ij}(u) - \sigma_{ij}^\star(u)\right| > c_M\tau_{n,d}\right) \\ &\to 0, \end{aligned}$$

completing the proof of Theorem 2. □

Appendix B: Proofs of the technical lemmas

We next give the detailed proofs of the lemmas used in Appendix A to prove the main results.

Proof of Lemma 1. We next only give a detailed proof of (A.2) as the proof of (A.1) is similar. By the definitions of $\widehat{c}_{ij,k}(u_k)$ and $c_{ij,k}(u_k)$, we have

$$\begin{aligned}\widehat{c}_{ij,k}(u_k) - c_{ij,k}(u_k) &= \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)[X_{t+1,i}X_{t+1,j} - c_{ij,k}(u_k)]\right\} / \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)\right\} \\ &= \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi_{t+1,ij,k}\right\} / \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)\right\} + \\ &\quad \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)[c_{ij,k}(U_{tk}) - c_{ij,k}(u_k)]\right\} / \left\{\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)\right\} \\ &=: I_{ij,k}^{(1)}(u_k) + I_{ij,k}^{(2)}(u_k), \end{aligned} \tag{B.1}$$

where $\xi_{t+1,ij,k} = X_{t+1,i}X_{t+1,j} - c_{ij,k}(U_{tk})$.

We first consider $I_{ij,k}^{(1)}(u_k)$ and prove that

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi_{t+1,ij,k}\right| = O_P\left(\sqrt{\log(d\vee n)/(nh_2)}\right), \tag{B.2}$$

and

$$\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1} K\left(\frac{U_{tk}-u_k}{h_2}\right) - f_k(u_k)\right| = O_P\left(h_2^2 + \sqrt{\log n/(nh_2)}\right). \tag{B.3}$$

In fact, by (B.2) and (B.3) and noting that $f_k(\cdot)$ is positive and uniformly bounded away from zero in Assumption 1(iii), we readily have

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|I_{ij,k}^{(1)}(u_k)\right| = O_P\left(\sqrt{\log(d\vee n)/(nh_2)}\right). \tag{B.4}$$

We next only prove (B.2) as (B.3) can be proved in a similar (and simpler) way. Define

$$\xi_{t+1,ij,k}^{*} = \xi_{t+1,ij,k} I\left(|\xi_{t+1,ij,k}| \le n^{\iota}\right), \quad \xi_{t+1,ij,k}^{\diamond} = \xi_{t+1,ij,k} - \xi_{t+1,ij,k}^{*}, \tag{B.5}$$

where $I(\cdot)$ is an indicator function and ι is defined as in Assumption 3(ii). Observe that

$$\begin{aligned}\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi_{t+1,ij,k} &= \frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi^*_{t+1,ij,k}+\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi^\diamond_{t+1,ij,k}\\ &= \frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]+\\ &\quad \frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^\diamond_{t+1,ij,k}-\mathsf{E}(\xi^\diamond_{t+1,ij,k})\right] \qquad \text{(B.6)}\end{aligned}$$

as $\mathsf{E}(\xi_{t+1,ij,k})=\mathsf{E}(\xi^*_{t+1,ij,k})+\mathsf{E}(\xi^\diamond_{t+1,ij,k})=0$.

By the moment condition (3.1) in Assumption 1(ii), we have

$$\mathsf{E}\left(\left|\xi^\diamond_{t+1,ij,k}\right|\right)=\mathsf{E}\left[|\xi_{t+1,ij,k}|I\left(|\xi_{t+1,ij,k}|>n^\iota\right)\right]=O\left(n^{-\iota M_1}\right), \qquad \text{(B.7)}$$

where M_1 can be arbitrarily large. Then, by (B.7), Assumptions 1(ii), 2(i) and 3(iii) and the definition

$$\begin{aligned}&\mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^\diamond_{t+1,ij,k}-\mathsf{E}(\xi^\diamond_{t+1,ij,k})\right]\right|>M_0\sqrt{\log(d\vee n)/(nh_2)}\right)\\ &\le \mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\xi^\diamond_{t+1,ij,k}\right|>\frac{1}{2}M_0\sqrt{\log(d\vee n)/(nh_2)}\right)\\ &\le \mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\max_{1\le t\le n-1}\left|\xi^\diamond_{t+1,ij,k}\right|>0\right)\le \mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\max_{1\le t\le n-1}|\xi_{t+1,ij,k}|>n^\iota\right)\\ &\le \mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le t\le n-1}|X_{t+1,i}X_{t+1,j}|>n^\iota-\bar c\right)\le \mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le t\le n-1}\left(X^2_{t+1,i}+X^2_{t+1,j}\right)>2(n^\iota-\bar c)\right)\\ &\le 2\mathsf{P}\left(\max_{1\le i\le d}\max_{1\le t\le n-1}X^2_{t+1,i}>n^\iota-\bar c\right)\le 2\sum_{i=1}^{d}\sum_{t=1}^{n-1}\mathsf{P}\left(X^2_{t+1,i}>n^\iota-\bar c\right)\\ &= O_P\left(dn\exp\{-sn^\iota\}\max_{1\le i\le d}\mathsf{E}\left[\exp\left\{sX^2_{ti}\right\}\right]\right)=o(1) \qquad \text{(B.8)}\end{aligned}$$

for $0<s<s_0$, where $\bar c=\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k\le u_k\le b_k}|c_{ij,k}(u_k)|$ is bounded by Assumption 2(i).

We next consider covering the set $\mathcal{U}_k$ by some disjoint intervals $\mathcal{U}_{k,l}$, $l=1,\ldots,q$ with the center $u_{k,l}$ and length $h_2^2n^{-\iota}\sqrt{\log(d\vee n)/(nh_2)}$. It is easy to find that q is of order $n^\iota h_2^{-2}\sqrt{(nh_2)/\log(d\vee n)}$.

Note that

$$
\begin{aligned}
&\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|\\
&\le \max_{1\le i,j\le d}\max_{1\le k\le p}\max_{1\le l\le q}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|+\\
&\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{u_k\in\mathcal{U}_{k,l}}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}\left[K\left(\frac{U_{tk}-u_k}{h_2}\right)-K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\right]\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|\\
&\le \max_{1\le i,j\le d}\max_{1\le k\le p}\max_{1\le l\le q}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|+\\
&\max_{1\le k\le p}\sup_{u_k\in\mathcal{U}_{k,l}}\frac{2n^\iota}{nh_2}\sum_{t=1}^{n-1}\left|K\left(\frac{U_{tk}-u_k}{h_2}\right)-K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\right|\\
&\le \max_{1\le i,j\le d}\max_{1\le k\le p}\max_{1\le l\le q}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|+O_P\left(\sqrt{\log(d\vee n)/(nh_2)}\right),
\end{aligned}
$$

where Assumption 3(i) and the definition of $\xi^*_{t+1,ij,k}$ in (B.5) are used.

By the exponential inequality for the α-mixing dependent sequence such as Theorem 1.3 in Bosq (1998), we may show that

$$
\begin{aligned}
&\mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|>M_0\sqrt{\log(d\vee n)/(nh_2)}\right)\\
&\le\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{p}\sum_{l=1}^{q}\mathsf{P}\left(\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_{k,l}}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|>M_0\sqrt{\log(d\vee n)/(nh_2)}\right)\\
&=O\left(d^2pq\exp\left\{-M_*\log(d\vee n)\right\}\right)+O\left(d^2pq\left[n^{5+2\iota}/(h_2\log^3(d\vee n))\right]^{1/4}\gamma^{\sqrt{M_\diamond\log(d\vee n)}}\right)=o_P(1).
\end{aligned}
$$

where M_* and $M_\diamond$ are two positive constants which could be sufficiently large, and $0<\gamma<1$ is defined in Assumption 1(i). Therefore, we have

$$
\mathsf{P}\left(\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}K\left(\frac{U_{tk}-u_k}{h_2}\right)\left[\xi^*_{t+1,ij,k}-\mathsf{E}(\xi^*_{t+1,ij,k})\right]\right|>M_0\sqrt{\log(d\vee n)/(nh_2)}\right)=o(1). \tag{B.9}
$$

By (B.8) and (B.9), we can complete the proof of (B.2).

Similarly, we can also show that

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{nh_2}\sum_{t=1}^{n-1}\left\{K\left(\frac{U_{tk}-u_k}{h_2}\right)c_{ij,k}(U_{tk})-\mathsf{E}\left[K\left(\frac{U_{tk}-u_k}{h_2}\right)c_{ij,k}(U_{tk})\right]\right\}\right|$$
$$=O_P\left(\sqrt{\log(d\vee n)/(nh_2)}\right), \tag{B.10}$$

and by Taylor's expansion for $c_{ij,k}(\cdot)$ and $f_k(\cdot)$

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|\frac{1}{h_2}\mathsf{E}\left[K\left(\frac{U_{tk}-u_k}{h_2}\right)c_{ij,k}(U_{tk})\right]-c_{ij,k}(u_k)f_k(u_k)\right|=O_P\left(h_2^2\right). \tag{B.11}$$

By (B.3), (B.10) and (B.11), we have

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\sup_{a_k+h_\star\le u_k\le b_k-h_\star}\left|I_{ij,k}^{(2)}(u_k)\right|=O_P\left(\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right). \tag{B.12}$$

Then the proof of (A.2) is completed in view of (B.1), (B.4) and (B.12). □

Proof of Lemma 2. From the definition of $(\widehat{a}_{ij,1},\ldots,\widehat{a}_{ij,p})^\intercal$ in (2.12), we have

$$(\widehat{a}_{ij,1},\ldots,\widehat{a}_{ij,p})^\intercal=\widehat{\mathbf{\Omega}}_{ij}^{-1}\widehat{\mathbf{V}}_{ij}=\left[\widetilde{\mathbf{\Omega}}_{ij}+\left(\widehat{\mathbf{\Omega}}_{ij}-\widetilde{\mathbf{\Omega}}_{ij}\right)\right]^{-1}\left[\widetilde{\mathbf{V}}_{ij}+\left(\widehat{\mathbf{V}}_{ij}-\widetilde{\mathbf{V}}_{ij}\right)\right], \tag{B.13}$$

where $\widetilde{\mathbf{\Omega}}_{ij}$ is a $p\times p$ matrix with the (k,l)-entry being

$$\widetilde{\omega}_{ij,kl}=\frac{1}{n-1}\sum_{t=1}^{n-1}c_{ij,k}^c(U_{tk})c_{ij,l}^c(U_{tl}),\quad c_{ij,k}^c(U_{tk})=c_{ij,k}(U_{tk})-\mathsf{E}\left[c_{ij,k}(U_{tk})\right],$$

and $\widetilde{\mathbf{V}}_{ij}$ is a p-dimensional column vector with the k-th element being

$$\widetilde{v}_{ij,k}=\frac{1}{n-1}\sum_{t=1}^{n-1}c_{ij,k}^c(U_{tk})X_{t+1,(i,j)}^*,\quad X_{t+1,(i,j)}^*=X_{t+1,i}X_{t+1,j}-\mathsf{E}\left[X_{t+1,i}X_{t+1,j}\right].$$

Following the proof of (B.2) above, we may show that

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\left|\frac{1}{n-1}\sum_{t=1}^{n-1}c_{ij,k}(U_{tk})-\mathsf{E}\left[c_{ij,k}(U_{tk})\right]\right|=O_P\left(\sqrt{\log(d\vee n)/n}\right) \tag{B.14}$$

and

$$\max_{1\le i,j\le d}\max_{1\le k\le p}\left|\frac{1}{n-1}\sum_{t=1}^{n-1}X_{t+1,i}X_{t+1,j}-\mathsf{E}\left[X_{t+1,i}X_{t+1,j}\right]\right|=O_P\left(\sqrt{\log(d\vee n)/n}\right). \tag{B.15}$$

By (B.14), (B.15) and Lemma 1, we readily have

$$\max_{1\le i,j\le d}\left\|\widehat{\mathbf{\Omega}}_{ij}-\widetilde{\mathbf{\Omega}}_{ij}\right\|_F=O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right) \tag{B.16}$$

and

$$\max_{1\le i,j\le d}\left\|\widehat{\mathbf{V}}_{ij}-\widetilde{\mathbf{V}}_{ij}\right\|=O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right). \tag{B.17}$$

By (B.13), (B.16) and (B.17), we have

$$(\widehat{a}_{ij,1},\dots,\widehat{a}_{ij,p})^{\intercal}=\widetilde{\mathbf{\Omega}}_{ij}^{-1}\widetilde{\mathbf{V}}_{ij}+O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right). \tag{B.18}$$

Again, following the proof of (B.2), we can easily show that

$$\max_{1\le i,j\le d}\left\|\widetilde{\mathbf{\Omega}}_{ij}-\mathbf{\Omega}_{ij}\right\|_F=O_P\left(\sqrt{\log(d\vee n)/n}\right) \tag{B.19}$$

and

$$\max_{1\le i,j\le d}\left\|\widetilde{\mathbf{V}}_{ij}-\mathbf{V}_{ij}\right\|=O_P\left(\sqrt{\log(d\vee n)/n}\right), \tag{B.20}$$

which together with (B.18), indicates that

$$\max_{1\le i,j\le d}\sum_{k=1}^{p}\left|\widehat{a}_{ij,k}-a_{ij,k}^{\star}\right|=O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right). \tag{B.21}$$

We finally consider $\widehat{a}_{ij,0}$. Note that uniformly for $1\le i,j\le d$,

$$\begin{aligned}
\widehat{a}_{ij,0}&=\frac{1}{n-1}\sum_{t=1}^{n-1}X_{t+1,i}X_{t+1,j}-\sum_{k=1}^{p}\widehat{a}_{ij,k}\left[\frac{1}{n-1}\sum_{t=1}^{n-1}\widehat{c}_{ij,k}(U_{tk})\right]\\
&=\frac{1}{n-1}\sum_{t=1}^{n-1}X_{t+1,i}X_{t+1,j}-\sum_{k=1}^{p}\widehat{a}_{ij,k}\left[\frac{1}{n-1}\sum_{t=1}^{n-1}c_{ij,k}(U_{tk})+O_P\left(\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right)\right]\\
&=\mathsf{E}\left(X_{ti}X_{tj}\right)+O_P\left(\sqrt{\log(d\vee n)/n}\right)-\sum_{k=1}^{p}\widehat{a}_{ij,k}\left[\mathsf{E}\left(X_{ti}X_{tj}\right)+O_P\left(\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right)\right]\\
&=\left(1-\sum_{k=1}^{p}a_{ij,k}^{\star}\right)\mathsf{E}\left(X_{ti}X_{tj}\right)+O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right)\\
&=a_{ij,0}^{\star}+O_P\left(\sqrt{\log(d\vee n)/n}+\sqrt{\log(d\vee n)/(nh_2)}+h_2^2\right),
\end{aligned} \tag{B.22}$$

where (B.14), (B.15) and (B.21) have been used.

From (B.21) and (B.22), we can complete the proof of (B.3). The proof of (B.4) is similar, so details are omitted here. The proof of Lemma 2 has been completed. □

References

Aït-Sahalia, Y. and Brandt, M. W. (2001). Variable selection for portfolio choice. *Journal of Finance* **56**, 1297–1351.

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

Back, K. E. (2010). *Asset Pricing and Portfolio Choice Theory*. Oxford University Press.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

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.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: a conditional Euler equations approach. *Journal of Finance* **54**, 1609–1645.

Brandt, M. W. (2010). Portfolio choice problems. *Handbook of Financial Econometrics, Volume 1: Tools and Techniques* (Editors: Aït-Sahalia Y. and Hansen L. P.), 269-336.

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

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. (2017). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. Forthcoming in *Journal of the American Statistical Association*.

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. *Annals of Statistics* **41**, 2994–3021.

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

DEMIGUEL, V., GARLAPPI, L., NOGALES, F. J. AND UPPAL, R. (2009). A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Management Science* **55**(5), 798-812.

DEMIGUEL, V., GARLAPPI, L. AND UPPAL, R. (2009). Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Review of Financial Studies* **22**, 1915-1953.

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., LEDOIT, O. AND WOLF, M. (2016). Large dynamic covariance matrices. *Working Paper*, Department of Economics, University of Zurich.

FAMA, E. F. (1970). Multiperiod consumption-investment decisions. *American Economic Review* **60**, 163–174.

FAN, J., FAN, Y. AND LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147**, 186-197.

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., 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.

FRAHM, G. AND MEMMEL, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics* **159**(2), 289-302.58.

GLAD, I., HJORT, N. L. AND USHAKOV, N. G. (2003). Correction of Density Estimators that are not Densities. *Scandinavian Journal of Statistics* **30** 415–427.

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.

KAN, J. AND ZHOU, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* **42**, 621-656.

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

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* **10**(5), 603-621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**(2), 365-411.

Ledoit, O. and Wolf, M. (2014). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Working Paper.

Li, D., Linton, O. and Lu, Z. (2015). A flexible semiparametric forecasting model for time series. *Journal of Econometrics* **187**, 345–357.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance* **7**, 77–91.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics* **51**, 247–257.

Pesaran, M. H. and Zaffaroni, P. (2009). Optimality and diversifiability of mean variance and arbitrage pricing portfolios. CESifo Working Paper Series No. 2857.

Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104**, 177–186.

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** 1167–1179.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* **99**, 204-215.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics* **40**(5), 2601–2633.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall.

Wu, W. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831–844.